Electronic Journal of Statistics
Vol. 0 (2006) 1–8
ISSN: 1935-7524

Estimating transformation function

Yunyi Zhang, Jiazheng Liu, Zexin Pan and Dimitris N. Politis

Department of Mathematics
University of California, San Diego
La Jolla, CA 92093
e-mail:
yuz334@ucsd.edu; liujiazzheng@gmail.com; zep002@stanford.edu; dpolitis@ucsd.edu

Abstract: In this paper, we propose an estimator for $g(x)$ under the model $Y_i = g(Z_i),\ i = 1, 2, ..., n$ where $Z_i,\ i = 1, 2, ...$ are random variables with known distribution but unknown observed values, $Y_i,\ i = 1, 2, ...$ are observed data and $g(x)$ is an unknown strictly monotonically increasing function(we call $g(x)$ transformation function). We prove the almost sure convergence of the estimator and construct confidence intervals and bands when $Z_i, i = 1, 2, ...$ are i.i.d data based on their asymptotic distribution. Corresponding case when Z_i being linear process is handled by resampling method. We also design the hypothesis test regarding whether $g(x)$ equals an expected transformation function or not. The finite sample performance is evaluated by applying the method to simulated data and an urban waste water treatment plant's dataset.

MSC 2010 subject classifications: 62G05, 62G09, 62G10.
Keywords and phrases: Transformation function, quantile process, resampling method.

Contents

1. Introduction and assumptions

1.1. Introduction

In this article, we focus on model

$$Y_i = g(Z_i), i = 1, 2, ..., n \tag{1.1}$$

Here Y_i are observed data and Z_i are random variables with known distribution but unknown observed values. We are interested in estimating strictly increasing function $g(x)$ (we call it transformation function) for given x under this situation. We first provide some examples to clarify the motivation to estimate transformation function $g(x)$.

Example 1. *Suppose there is a production line and we want to control the quality of products and minimize the cost of materials at the same time. It is reasonable to assume the quality Y of products as a decreasing function of property of materials, $|Z - z_0|$ with z_0 being the design point. Moreover, the distribution of quality of materials can assume to be known. (For example, tensile strength of materials satisfies Weibull distribution [17].) However, it is difficult to use regression model since testing materials' quality is of great cost and always brings damage to materials. Instead, if the distribution of quality of materials is known, then distribution of $|Z - z_0|$ can be calculated and model (1.1) can be applied to understand relationship between quality of products and property of materials. After estimating g, we know how sensitive the quality of products is affected by quality of materials.*

Example 2. *Consider the model in figure 1. Suppose the probability distribution of input signal is known and the output signal data can be acquired. Then, two things are worth considering. The first one is to understand how the amplifier enlarges the input signal, that is, to estimate the transformation function $g(x)$. The second thing is to test whether the transformation function g coincides with the expected transformation function h, which comes from physical laws or experience. For example, according to [14], observed concentration $Y_i,\ i = 1, 2, ...$ from an experiment can be modelled as*

$$Y_i = \mu \exp(a + b \times Z_i),\ i = 1, 2, ... \tag{1.2}$$

With μ being the true concentration and a, b unknown constant. Z_i is assumed to be a standard normal random variable(but its value cannot be observed in measurement). Researchers having concentration data may hope to justify the correctness of model (1.2), especially whether $\log(Y_i)$ is a linear function of Z_i or not.

Example 3. *In the third example, we consider a type of time series data*

$$Y_n = g(Z_n),\ Z_n = \Sigma_{k=1}^{m} a_k \epsilon_{n-k},\ \epsilon_i \sim i.i.d\ N(0, 1) \tag{1.3}$$

Since we suppose that $\epsilon_i,\ i = ..., -1, 0, 1, ...$ are standard normal random variables, Z_n is also of normal distribution. We want to estimate $g(x)$ for some x

FIG 1. *A standard amplifier system*

in this situation. For example, in [9], daily number of respiratory symptoms per child is recorded and is related to daily SO_2 *and* NO_2*. In that paper, transformation function* $g(x) = v_0 \log(x) + \epsilon$ *with* ϵ *being an ARIMA series are considered. If instead, we ignore the error* ϵ *and want to estimate transformation function in a non-parametric way, then model* (1.1) *can be applied to this problem.*

To summarize, example 1 and 3 involves estimating transformation function $g(x)$ in i.i.d data and dependent data, and example 2 involves testing equivalence of a transformation function. All of these three topics will be covered in this paper.

According to [15], suppose Z is a random variable with continuous cumulative distribution F_Z, then $F_Z(Z)$ is of uniform distribution. Thus, random variables Z with strictly increasing cumulative distribution function (which is invertible) can be naturally related to a random variable U with uniform distribution by choosing $g(x) = F_Z^{-1}(x)$. There are discussions on estimating $F_Z(x)$ and $F_Z^{-1}(x)$, related results can be found in [6] and [18]. There are also some researches related to estimating monotone functions. For example, Zhao and Woodroofe [19] considered model $Y_k = \mu_k + Z_k$ and used isotonic method to estimate monotone trend, Dietz and Killeen [8] proposed a test on whether time series data have an increasing order, Mukerjee [11] considered monotone regression problem, etc.

The aforementioned models mainly consider estimating trends of data, but model (1.1) composites function g on data $Z_k, k = ..., -1, 0, 1, ...$ Worse still, we do not know exact observed values of Z_k, so regression methods (like [11]) cannot be applied to this problem. However, the methods we propose in this paper can be applied to estimate $g(x)$ and perform tests under model (1.1).

In this paper, we provide an estimator for strictly increasing transformation function $g(x)$ and discuss its asymptotic properties when random variable Z_i are i.i.d or short range dependent. In section 2, we demonstrate how to estimate transformation function and construct confidence intervals and bands for i.i.d data. We also provide a test similar to Kolmogorov-Smironv test[10] on testing whether $g(x) = h(x)$, the expected transformation function. In section 3, we discuss how to estimate transformation function and how to construct confidence interval through sub-sampling methods for linear processes. In section 4, several numerical examples are provided and conclusion is made in section 5. Proofs of main theorems will be given in the appendix.

1.2. Frequently used notations and assumptions

In this part, we introduce frequently used notations for this paper, other symbols will be defined when being used. Besides, we will list basic assumptions and constraints on random variables and transformation function below.

Suppose that $Z_i,\ i = 1, 2, ..., n$ are random variables with known cumulative distribution function $F_Z(x)$ and density $f_Z(x)$, Y_i being unknown random variables satisfying $Y_i = g(Z_i)\ \forall i$. We define empirical distribution function as

$$\widehat{F}_K(x) = \frac{1}{n}\Sigma_{i=1}^{n}\mathbf{1}_{K_i \leq x} \tag{1.4}$$

Here subscript K can be chosen as Y or Z. Similarly, quantile and sample quantile function are respectively defined as

$$\xi_K(p) = \inf\left\{x|F_K(x) \geq p\right\},\ \widehat{\xi}_K(p) = \inf\left\{x|\widehat{F}_K(x) \geq p\right\} \tag{1.5}$$

Assumption A1: $Z_i,\ i = 1, 2, ...$ are i.i.d with strictly increasing cumulative distribution function(but we do not assume continuity).

Assumption A2: $Z_i,\ i = 1, 2, ...$ are causal stationary linear short range dependent processes (details can be seen in [18]). That is, $Z_k = \Sigma_{i=0}^{\infty} a_i \epsilon_{k-i}$ with $a_0 = 1$, $\epsilon_i,\ i = 1, 2, ...$ being i.i.d. random variables and satisfy

$$\sup_{x\in\mathbf{R}} f_\epsilon(x) + |f_\epsilon^{'}(x)| + |f_\epsilon^{''}(x)| < \infty \tag{1.6}$$

Here f_ϵ is density of innovation ϵ. Moreover, suppose $\exists\alpha > 0$, $q \geq 2$ such that $\mathbf{E}|\epsilon_k|^\alpha < \infty$ and

$$\Sigma_{i=1}^{\infty}|a_i|^{\min(\alpha/q,1)} < \infty \tag{1.7}$$

Assumption A3: $Z_i,\ i = 1, 2, ...$ satisfy $\alpha-$ mixing condition (Details can be seen at [12] and [1])

Assumption B1: g is strictly monotonically increasing (for decreasing g, $h = -g$ is increasing)

Assumption B2: g is differentiable

Assumption B3: g is twice continuously differentiable, f_Z is continuously differentiable on (a, b) defined in table 1. Moreover, we assume that $\exists\gamma > 0$ such that

$$\sup_{a<x<b} \frac{F_Z(x)(1 - F_Z(x))}{f_Z^2(x)}|f_Z^{'}(x) - f_Z(x)\frac{g^{''}(x)}{g^{'}(x)}| \leq \gamma \tag{1.8}$$

Notice that this equation implies that $g^{'}(x), f_Z(x) > 0$ on (a, b), and correspondingly $f_Y(x) > 0,\ x \in (g(a), g(b))$.

While other conditions are natural and frequently used in density and quantile estimation, condition A2 and B3 seems complex and needs explanation. For condition A2, uniform bound of density f_ϵ and its derivative is used to make sure that Bahadur representation [2] of density f_Z exists and uniform convergence in theorem 5 holds. For point-wise estimation or construction of point-wise

TABLE 1
Frequently used notations

Notation	Meaning
$F_Z(x),\ f_Z(x)$	Cumulative distribution and density of known random variable Z
$F_Y(x),\ f_Y(x)$	Cumulative distribution and density of unknown random variable Y
$\widehat{f}_Y(x)$	Estimated density of unknown random variable
$\xi_K(p)$	pth quantile function of distribution of random variable K
$\widehat{\xi}_K(p)$	pth sample quantile function of random variable K
$g(x)$	Transformation function satisfying $Y_i = g(Z_i),\ i = 1, 2, ...n$
$\widehat{g}(x)$	Estimated transformation function at x
$\widehat{F}_K(x)$	Empirical distribution function of random variable K
a, b	Here, $-\infty \le a = \sup\{x \| F_Z(x) = 0\}$, $\infty \ge b = \inf\{x \| F_Z(x) = 1\}$
$\mathbf{1}_{K\in A}$	If $K \in A$, then function is equal to 1 and the function is equal to 0 otherwise

confidence intervals, (1.6) can be weakened by introducing a stronger mixing condition[16]. In section 4, we will construct a counter example to see what happens when condition A2 is violated. Necessity of (1.7) can be illustrated by an example when $\alpha = 2$. Suppose $\alpha = 2$, covariance of Z_0 and $Z_k, k > 0$ is given by

$$\begin{aligned} Cov(Z_k, Z_0) &= \Sigma_{i=0}^{\infty}\Sigma_{s=0}^{\infty} a_i a_s \mathbf{E}\epsilon_{k-i}\epsilon_{-s} = \mathbf{E}\epsilon_0^2 \Sigma_{s=0}^{\infty} a_{k+s} a_s \\ &\Rightarrow \Sigma_{k=0}^{\infty} |Cov(Z_k, Z_0)| \le \Sigma_{s=0}^{\infty} |a_s| (\Sigma_{k=s}^{\infty} |a_k|) \end{aligned} \tag{1.9}$$

If (1.7) holds in this example, then summation of covariance is finite, which implies that dependency of data is not strong.

According to lemma 1.4.1 in [5], if $F_Y(x)(1 - F_Y(x))\frac{|f_Y'(x)|}{f_Y^2(x)}$ is uniformly bounded, then deviation of composite function $f_Y(F_Y^{-1}(y))$ with $y \ne y_0$, a fixed point, can be controlled by a simple function of $y, y_0 \in (0, 1)$ globally. This implies uniform convergence of quantile processes. Property of quantile function is relatively hard to study because it grows fast when x is close to a and b near which density is always small and we need an easily-controlled upper bound to perform analysis. Besides, in confidence band estimation and testing, we need uniform convergence of quantile process $\widehat{\xi}_Y$, so this condition is a must. Combine this condition with (1.1) and lemma 1 and 2, we get (1.8). Example 4, figure 2(c) and 2(d) shows that, when assumption B3 is violated, confidence bands will be wide even when sample size is relatively large.

2. Estimation of transformation function with i.i.d data

In this section, we discuss estimation and test of transformation function on i.i.d data, including estimation, construction of confidence intervals and confidence bands. Based on Kolmogorov-Smirnov test, we provide a test on whether the transformation function is equal to the desired one and discuss performance of test under an alternative. First we provide two lemmas.

Lemma 1. *Assume random variable Y, Z satisfy $Y = g(Z)$ and g satisfies B1, with the notation in table 1, then we have*

$$F_Y(g(x)) = F_Z(x), \forall x \in [a, b] \tag{2.1}$$

Proof. Because g is strictly increasing, we have

$$F_Y(g(x)) = P(Y \leq g(x)) = P(g(Z) \leq g(x)) = P(Z \leq x) = F_Z(x) \tag{2.2}$$

and the lemma is proved □

Lemma 2. *Assume B1, random variable $Y = g(Z)$, then we have, $\forall p \in (0, 1)$*

$$\xi_Y(p) = g(\xi_Z(p)),\ \widehat{\xi}_Y(p) = g(\widehat{\xi}_Z(p)) \tag{2.3}$$

Proof. From definition, on one hand, $F_Y(g(\xi_Z(p))) = F_Z(\xi_Z(p)) \geq p$, this is because F_Z is right continuous. Therefore, $\xi_Y(p) \leq g(\xi_Z(p))$. On the other hand, since g is strictly increasing, its inverse function $g^{-1}(y)$ is strictly increasing. Therefore we have $\xi_Z(p) \leq g^{-1}(\xi_Y(p)) \Rightarrow g(\xi_Z(p)) \leq \xi_Y(p)$, and the first part is proved. For the second part, we notice that $\widehat{F}_K(x), K = Y, Z$ are also a right continuous cumulative distribution functions, thus the discussion above can be directly applied to $\widehat{\xi}_Y(p), \widehat{\xi}_Z(p)$, and the second part is proved. □

2.1. Estimation of transformation function

This section aims at providing an estimator and constructing confidence intervals and confidence bands for transformation function. Combine with lemma 1 and 2, the estimator is not difficult to provide.

Theorem 1. *Suppose A1 and B1, and for $\forall x \in (a, b)$ being given, define $\widehat{g}(x) = \widehat{\xi}_Y(F_Z(x))$. Then we have*

$$\widehat{g}(x) \rightarrow_{a.s.} g(x),\ n \rightarrow \infty \tag{2.4}$$

Moreover, for $\alpha \in (0, 1/2)$ being given, suppose $\zeta(y)$ being quantile function of standard normal distribution, then

$$\liminf_{n\rightarrow\infty} P(\widehat{\xi}_Y(c_1) \leq g(x) \leq \widehat{\xi}_Y(c_2)) \geq 1 - \alpha \tag{2.5}$$

Here, $c_1 = F_Z(x) + \frac{\zeta(\alpha/2)\sqrt{F_Z(x)(1-F_Z(x))}}{\sqrt{n}}$, $c_2 = F_Z(x) + \frac{\zeta(1-\alpha/2)\sqrt{F_Z(x)(1-F_Z(x))}}{\sqrt{n}}$

Notice that $|c_2 - c_1| = O\left(\frac{1}{\sqrt{n}}\right)$, at $x \in \mathbf{R}$ such that $\xi_Y(F_Z(x))$ is continuous, according to convergence of sample quantile (proposition 5.7 in [4] and Glivenko−Cantelli theorem), we have $c_2 - c_1 \rightarrow 0 \Rightarrow \widehat{\xi}_Y(c_2) - \widehat{\xi}_Y(c_1) \rightarrow 0$ in probability as sample size increases. Thus, we can use the result in theorem 1 to construct point-wise confidence intervals. If in addition we assume density of F_Z exists, then we can apply uniform convergence theorem in [5] to construct confidence bands.

Theorem 2. *Suppose A1, B1, B3, and suppose* $\delta_n = (25 \log\log n)/n$*, define* $\phi(x)$ *as a kernel function satisfying the following condition:*

1) ϕ *is of finite support, i.e. there exists a compact interval* $[d_1, d_2]$ *such that supp* $\phi \subseteq [d_1, d_2]$.

2) ϕ *is continuously differentiable on* $[d_1, d_2]$.

3) $\int_{d_1}^{d_2} \phi(x) = 1$. *We define the estimated density* $\widehat{f}_Y(x)$ *as*

$$\widehat{f}_Y(x) = \frac{1}{nh}\Sigma_{i=1}^{n}\phi(\frac{x - Y_i}{h}) \tag{2.6}$$

Here, $h = h(n)$ *is a bandwidth satisfying* $(\log\log n)^{1/2}h \to 0$ *and* $\frac{\sqrt{n}h^2}{\log\log n} \to \infty$. *Also suppose that* $[c, d]$ *is a closed interval in* $\mathbf{R}$ *such that* $a < c < d < b$. *Then we can find a Kiefer process* $K(y, n), 0 \le y \le 1$*[6] such that*

$$\sup_{c\le x\le d} |\sqrt{n}(\widehat{g}(x) - g(x))\widehat{f}_Y(\widehat{g}(x)) - \frac{K(F_Z(x), n)}{\sqrt{n}}| \to 0 \ a.s. \tag{2.7}$$

Remark 1. ϕ *and* $h(n)$ *satisfying requirements in theorem 2 exist. For example, we can choose* ϕ *as*

$$\phi(x) = \begin{cases} \frac{1}{2\pi}(1 + \cos(x)) \ \textit{if } x \in [-\pi, \pi] \\ 0 \ \textit{otherwise} \end{cases} \tag{2.8}$$

and $h(n) = (1/n)^{1/6}$.

Remark 2 (Estimating derivative of g)**.** *If we assume A1, B1 and B3, according to lemma 1,*

$$f_Y(g(x))g^{'}(x) = f_Z(x) \Rightarrow g^{'}(x) = f_Z(x)/f_Y(g(x)) \tag{2.9}$$

This implies that we can use estimator $\frac{f_Z(x)}{\widehat{f}_Y(\widehat{g}(x))}$ *to estimate the derivative of* g. *Here we prove*

$$\frac{f_Z(x)}{\widehat{f}_Y(\widehat{g}(x))} \to_{a.s.} g^{'}(x) \tag{2.10}$$

with given $x \in [c, d]$ *and bandwidth* $h(n)$ *is chosen similar as in theorem 2.*

Proof. From (6.8), (6.9) and (6.14) and assumption B3,

$$\begin{aligned} |\widehat{f}_Y(\widehat{g}(x)) - f_Y(g(x))| &\le |\widehat{f}_Y(\widehat{g}(x)) - \widehat{f}_Y(g(x))| + |\widehat{f}_Y(g(x)) - f_Y(g(x))| \\ &= O_{a.s.}\left(\frac{(\log\log n)^{1/2}}{h^2\sqrt{n}}\right) + O_{a.s.}\left(h + \frac{(\log\log n)^{1/2}}{h\sqrt{n}}\right) \end{aligned} \tag{2.11}$$

This implies that $\widehat{f}_Y(\widehat{g}(x)) \to_{a.s.} f_Y(g(x))$. For $f_Y(g(x)) \neq 0$ and $f_Z(x)$ is continuous at x, the result is proved. ☐

TABLE 2

Bisection method for finding c in constructing confidence band(discussion can be seen in corollary 2)

Input: Confidence level $1-\alpha,\ 0<\alpha<1$, error tolerance ϵ, start point $0<a<b$
while $s(a)\times s(b)>0$, **do**
$\quad a=a/2,\ b=2\times b$
while $\vert b-a\vert>\epsilon$, **do**
$\quad$ **If** $s(a)\times s(\frac{a+b}{2})\le 0$
$\qquad a=a,\ b=\frac{a+b}{2}$
$\quad$ **Else**
$\qquad a=(a+b)/2,\ b=b$
Output $(a+b)/2$

By applying theorem 2, we are able to construct confidence band for transformation function. Compare with point-wise confidence intervals, confidence band is more reliable since we do not have to assign x a priori and we can monitor different x in once observation. For example, in example 2, acceptable design points of a product can be a closed interval instead of a fix point. If this happens, we want to control the estimation error uniformly among the acceptable design points and we need a uniformly confidence band.

Corollary 1 (Confidence band within an interval)**.** *Suppose the same conditions as in theorem 2, and suppose $c>0$ is a positive number, then we have*

$$\begin{aligned}\lim_{n\to\infty}\sup P(\sup_{c\le x\le d}|\sqrt{n}(\widehat{g}(x)-g(x))\widehat{f}_Y(\widehat{g}(x))|>c) &\le P(\sup_{0\le y\le 1}|B(y)|>c)\\ &=\Sigma_{k\neq 0}(-1)^{k+1}\exp(-2k^2c^2)\end{aligned} \tag{2.12}$$

In real situation, we always construct confidence bands with given confidence level $1-\alpha$. Thus, in corollary 2, we use bisection method in [13] to derive constant c in (2.12) such that $P(\sup_{0\le y\le 1}|B(y)|>c)\to\alpha$ when tolerance ϵ in table 2 tends to 0.

Corollary 2. *Suppose $0<\alpha<1$ is a given constant and c is derived from table 2 with $s(x)=P(\sup_{0\le y\le 1}|B(y)|>x)-\alpha$ and tolerance ϵ. Then we have $P(\sup_{0\le y\le 1}|B(y)|>c)\to\alpha$ as $\epsilon\to 0$.*

Proof. From (6.20), we know that $s(x)$ is continuous on $(0,\infty)$. Also, from definition of s, we know that $s(x)\to 1-\alpha>0$ as $x\to 0$ and $s(x)\to-\alpha<0$ as $x\to\infty$ and $s(x)$ is decreasing. Therefore, $s(x)=0$ has a solution c^* in $(0,\infty)$ and for arbitrary start point a,b, after iterations we have $a\le c^*\le b$. From bisection method, we have $|c-c^*|\le\epsilon$ and since $s(x)$ is continuous at c^*, we know that $P(\sup_{0\le y\le 1}|B(y)|>c)\to\alpha$ as $\epsilon\to 0$. □

2.2. Testing

In this section, we mainly consider testing $H_0 : g = h$ versuses $H_1 : g \neq h$ under uniform norm. Here h is a known or desired transformation function and g is the underlying one. We consider the test that reject H_0 when

$$\sup_{\delta_n \leq F_Z(x) \leq 1-\delta_n} \sqrt{n}\frac{f_Z(x)}{h^{'}(x)}|\widehat{g}(x) - h(x)| > c \tag{2.13}$$

Here c is a positive constant and δ_n is the same as in theorem 2. Similar as confidence band estimation, we need to quantify influence of randomness uniformly. However, in the test setting, asymptotically we want to know value of $|g(x) - h(x)|$ in $[a, b] - \{\pm\infty\}$ in order to calculate the infinity norm, so we choose compact set $\delta_n \leq F_Z \leq 1 - \delta_n$ with $\delta_n \to 0$ as $n \to \infty$, which asymptotically equals $[a, b] - \{\pm\infty\}$ instead of a fixed interval $[c, d]$. One of the purposes for testing is illustrated in example 2, another purpose is to detect abnormal status of a device. Transformation function h for a normal operated device is fixed, and if estimated transformation function $\widehat{g} \neq h$ with high probability, it is possible that something goes wrong with the device. We will discuss asymptotic behavior of test statistics (2.13) under the null in theorem 3 and one alternative in theorem 4.

Theorem 3. *Suppose A1, B1, B3. Consider testing $H_0 : g(x) = h(x)\ \forall\ x \in (a, b)$ versus $H_1 : \exists a < x < b$ such that $g(x) \neq h(x)$. Suppose δ_n is defined the same as in theorem 2. Then under the null hypothesis, we have, given $c > 0$,*

$$\begin{aligned} P(\sup_{\delta_n \leq F_Z(x) \leq 1-\delta_n} \sqrt{n}\frac{f_Z(x)}{h^{'}(x)}|\widehat{g}(x) - h(x)| > c) \to P(\sup_{0\leq y\leq 1}|B(y)| > c) \\ = \Sigma_{k\neq 0}(-1)^{k+1}\exp(-2k^2c^2) \end{aligned} \tag{2.14}$$

Here $B(y)$ is a Brownian bridge.

Like theorem 14.2.2 in [10], we also consider power of test (2.13) under non-asymptotic alternatives. Theorem 4 shows that power of test (2.13) will decrease if deviation of h and g in uniform norm is of order $O\left(\frac{1}{\sqrt{n}}\right)$. Theorem 4 also provides a term $\sup_{a<x<b} \frac{\sqrt{n}f_Z(x)|h(x)-g(x)|}{g^{'}(x)}$ to quantify influence of closeness of h and g on power. In the abnormality detection problem, this term can be used to evaluate whether the test result is trustful or not.

Theorem 4. *Consider the test (2.13) and same condition as in theorem 3, h is continuously differentiable on $[a, b]$ and has positive derivative on (a, b). Define events $M_n = \sup_{\delta_n \leq F_Z(x) \leq 1-\delta_n} \sqrt{n}\frac{f_Z(x)}{h^{'}(x)}|\widehat{g}(x) - h(x)| > c$.*

1) If $\exists x_0 \in (a, b)$ such that $g(x_0) \neq h(x_0)$, then $P(M_n) \to 1$ as $n \to \infty$.

2) We suppose alternative $H_1^{'} : h(x) = g(x) + \frac{1}{\sqrt{n}}s(x)$, here $s(x) \in C^1$ on $[a, b]$ and $s^{'}(x) \geq 0$ on $[a, b]$. Suppose $B(y)$ is a standard Brownian bridge, then *the power of test satisfies*

$$\lim_{n\to\infty} \sup P(M_n) \leq P(\sup_{0\leq y\leq 1} |B(y)| \geq c - \sup_{a<x<b} \frac{f_Z(x)|s(x)|}{g^{'}(x)}) \tag{2.15}$$

Theorem 4 shows that $\sup_{a<x<b} \frac{\sqrt{n} f_Z(x)|h(x)-g(x)|}{g^{'}(x)}$ influences power of test. If it is bigger than c, asymptotically power of test gets close to 1. On the contrary, if this term is less than c, then the power of test will be less than 1 even when sample size is large. From another perspective, if $\sup_{a<x<b} \frac{\sqrt{n} f_Z(x)|h(x)-g(x)|}{g^{'}(x)}$ is small, in order to maintain sufficiently large power, constant c cannot be too large, which affects confidence level of test.

3. Estimation for dependent data

In this section, we concentrate on transformation function estimation with weakly dependent data. We first provide convergence and uniform convergence result and then we will use subsampling algorithm to construct point-wise confidence interval. Theorem 5 and 6 are generalization of theorem 1. Linear processes, including ARMA model, are widely used in modelling dependent data, especially in time series analysis. In this section, we focus on linear process in the following analysis.

Theorem 5. *Assume A2, B1, B2, then for given x, if $f_Z(x) > 0$, then we have $\widehat{g}(x) \to_{a.s.} g(x)$. If in addition we suppose $[c,d]$ being interval such that $\inf_{c\leq x\leq d} f_Z(x) > 0$, then we have $\widehat{g}(x) \to g(x)$ almost surely and uniformly on $[c,d]$.*

Remark 3. *We only need uniform bound on f_ϵ and $f_\epsilon^{'}$ to prove point-wise convergence. For uniform convergence in theorem 5, in addition we need uniform bound on $f_\epsilon^{''}$.*

Theorem 6 proves consistency of subsampling point-wise confidence intervals. Subsampling involves calculating statistics with sequential portions of data and deriving asymptotic valid confidence intervals based on those statistics [12]. Since the portions of data are also realizations of random variables with same joint distribution, as long as asymptotic distribution of the statistics exists, the portions of data catch the dependent structure of underlying random variables. Therefore, subsampling is a useful tool to deal with dependent data.

Theorem 6. *Suppose B1, B2, A2 and A3, and suppose x is a given constant such that $\exists c < x < d$ and $\inf_{c\leq y\leq d} f_Z(y) > 0$, $g(x)^{'} > 0$. Define η being a positive constant. For $b = b(n)$ satisfying: $b/n \to 0$ and $b \to \infty$, we define statistics*

$$S_{n,b}(\eta, x) = \frac{1}{n-b+1} \Sigma_{i=1}^{n-b+1} \mathbf{1}\left\{\sqrt{b}|\widehat{g}_{b,i}(x) - \widehat{g}(x)| \leq \eta\right\} \tag{3.1}$$

Here, $\widehat{g}_{b,i}(x) = \widehat{\xi}_{Y,b,i}(F_Z(x))$ *with* $\widehat{\xi}_{Y,b,i}(p)$ *being sample quantile with sample* $\{Y_i, Y_i+1, ..., Y_i+b-1\}$ *Then, we have:*

1) $S_{n,b}(\eta, x) \to P(\sqrt{n}|\widehat{g}(x) - g(x)| \le \eta)$ *in probability.*

2) Suppose $d(1-\alpha) = \inf\{\eta | S_{n,b}(\eta, x) \ge 1-\alpha\}$*, then*

$$P(\sqrt{n}|\widehat{g}(x) - g(x)| \le d(1-\alpha)) \to 1-\alpha \tag{3.2}$$

In example 3, with the help of theorem 5 and 6, we can make sure that estimator $\widehat{g}(x)$ converges almost surely to the true transformation function and for every given x, theorem 6 can be used to quantify the influence of randomness on estimation.

4. Numerical Experiments and Examples

In this section, we demonstrate finite sample performance on the aforementioned estimator. We divide this section into two parts. In the first part, we apply this estimator to several constructed data and show what happens when conditions are violated. In the second part, we will apply the aforementioned theories to study a real problem. In this problem, we want to know how well the primary settler of an urban waste water treatment plant cleans the organics in waste water (detail explanation and data can be gathered at[7] and the reference therein).

4.1. Finite sample behavior of statistics on constructed data

For the independent cases, we will use false rate, which is defined as the ratio of the number of cases in which true value is outside confidence intervals and the number of all cases, to evaluate accuracy of confidence intervals. For a 95% confidence interval, ideal false rate should be no large than 0.05. For confidence intervals, we fix a point and see how false rate changes with different sample size. For confidence bands, we randomly choose $x \in \mathbf{R}$ satisfying normal distribution and see whether $g(x)$ is outside confidence band or not.

Example 4 (i.i.d data with normal distribution)**.**

In this example, we suppose $Z_i,\ i = 1, 2, ..., n$ *satisfy standard normal distribution. Notice that, for large* x*,*

$$\begin{aligned} \frac{1-F_Z(x)}{f_Z(x)} = \int_x^\infty \exp\left(\frac{1}{2}x^2 - \frac{1}{2}t^2\right) dt = \int_0^\infty \exp\left(-\frac{1}{2}y^2 - xy\right) dy \\ \le \int_0^\infty \exp(-xy) dy = \frac{1}{x} \end{aligned} \tag{4.1}$$

Similarly, for $x \to -\infty$, $\frac{F_Z(x)}{f_Z(x)} = O\left(\frac{1}{|x|}\right)$*. We also have* $|f_Z'(x)/f_Z(x)| = x$*. Constraint* $x \in [-2, 2]$*, and choose* $g(x)$ *as 1)* $(x+4)^2$*, 2)* $\log(x+5)$*, 3)* x^3*.*

Table 3

Finite sample performance of estimated confidence intervals under normal data and different transformation functions, confidence level is 0.95, ci means confidence interval and cb means confidence band

function	x	sample size	false rate for ci	false rate for cb
$(x+4)^2$	4.0	300	0.479	0.121
		600	0.110	0.077
		1200	0.050	0.051
$\log(x+5)$	3.5	300	0.135	0.122
		600	0.073	0.075
		900	0.048	0.054
x^3	3.5	300	0.346	0.116
		600	0.092	0.079
		900	0.046	0.048

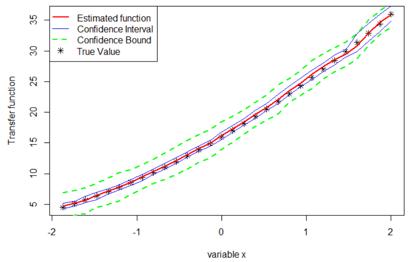
Notice that for $g(x) = x^3$, it has 0 derivative at $x = 0$ and $g^{''}/g^{'}$ is of order $1/|x|$, which tends to infinity as $x \to 0$. This violates assumption B3. Figure 3(c) and 2(d) show that confidence band will be wide when B3 is violated. Other functions all satisfy assumption B3. Main results are demonstrated in figure 4 and table 3. In table 3, confidence level is chosen as 0.95 and number of iteration is 3000.

According to figure 4, when derivative of $g(x)$ is not close to 0, confidence bands will be tight and close to confidence intervals, and when $|g^{'}(x)|$ is small, the performance of confidence bands will be inferior. When assumption B3 is violated, width of confidence bands will be enlarged significantly. The width of confidence intervals is not sensitive for small $|g^{'}(x)|$. However, large $|g^{'}|$ will affect the width of confidence intervals. Table 3 shows that, false rates of confidence intervals and bands are about 0.05 with sample size is about 1000.

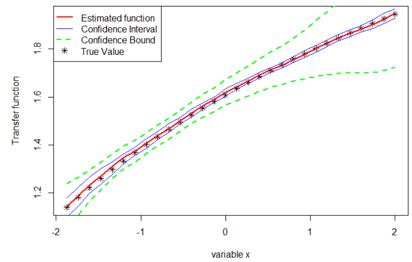
In the test problem, we evaluate performance of tests by ratio of correct test, which is defined as the ratio of the number of tests making correct decisions and the number of all tests. In ideal situation, ratio of correct test should be close to confidence level $1-\alpha$ under null hypothesis and close to 1 under alternatives asymptotically.

Example 5 (Testing for equivalence of transformation function)**.**

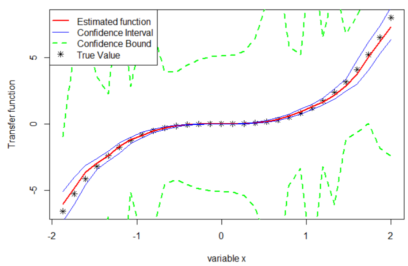
In this example, we examine finite sample performance of test under different $g(x)$ and different perturbations. We suppose sample size is $n_1 = 1000$, $n_2 = 10000$ and $h(x)$ equals 1) $(x+4)^2$, 2) $\log(x+5)$ and 3) e^x. Random variable $Z_i, i = ..., -1, 0, 1, ...$ satisfy standard normal distribution. Also, we suppose the underlying $g(x)$ satisfies: 1) $g(x) = h(x)$, 2) $g(x) = h(x) + \frac{x}{n^{1/8}}$, 3) $g(x) = h(x) + \frac{x}{\sqrt{n}}$. We suppose $H_0 : g(x) = h(x)$, perform tests for 200 times and calculate the ratio of correct tests to evaluate performance of tests. Under assumption 1), correct test should accept H_0 to avoid first kind error and under assumption 2) and 3), correct test should reject H_0 to avoid second kind error. Confidence level is set as 0.85. The result is demonstrated in table 4 and 5. From the experiment, when difference of $h(x)$ and underlying $g(x)$ is of $O\left(\frac{1}{\sqrt{n}}\right)$, whether or not the test can separate h and g depends on the form of perturbations and function g.



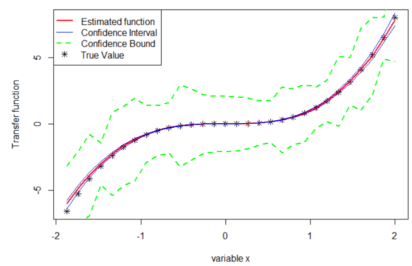
(a) $g(x) = (x + 4)^2$



(b) $g(x) = \log(x + 5)$



(c) $g(x) = x^3$



(d) $g(x) = x^3$, sample size is 20000

FIG 2. *2(a) to 2(c) respectively demonstrates behavior of estimator for different $g(x)$ when sample size is 1500 and confidence level is 0.95, and 2(d) demonstrates performance of estimator with 20000 samples. Red line, blue line and green dashed line respectively represents the estimated transformation function, point-wise confidence interval and confidence band for transformation function. Stars represent true values of transformation function. Notice that $g(x) = x^3$ violates assumption B3. From 2(c) and 2(d) we can see that confidence band will be significantly enlarged when assumption B3 is not satisfied.*

TABLE 4

Ratio of correct test(example 5) under different $g(x)$ and $h(x)$, sample size is 1000, confidence level is 0.85

$h(x)/g(x)$	$h(x)$	$h(x) + x/n^{1/8}$	$h(x) + x/\sqrt{n}$	$h(x) + log(x + 5)/\sqrt{n}$
$(x + 4)^2$	0.84	0.32	0.165	0.14
$\log(x + 5)$	0.87	1.0	0.985	1.0
e^x	0.91	1.0	0.49	0.665

TABLE 5

Ratio of correct test(example 5) under different $g(x)$ and $h(x)$, sample size is 10000, confidence level is 0.85

$h(x)/g(x)$	$h(x)$	$h(x) + x/n^{1/8}$	$h(x) + x/\sqrt{n}$	$h(x) + log(x + 5)/\sqrt{n}$
$(x + 4)^2$	0.89	0.97	0.17	0.615
$\log(x + 5)$	0.885	1.0	0.99	1.0
e^x	0.895	1.0	0.455	0.855

TABLE 6
Finite sample performance of estimated confidence intervals for dependent data

function	x	sample size	lag	false rate for confidence interval
$(x+4)^2$	1.0	3000	5	0.078
	1.0	6000	5	0.061
	1.0	6000	10	0.096
	2.0	6000	5	0.069
$\log(x+10)$	1.0	3000	5	0.071
	1.0	6000	5	0.064
	1.0	6000	10	0.066
	2.0	6000	5	0.052
x^3	1.0	3000	5	0.016
	1.0	6000	5	0.025
	1.0	18000	5	0.051
	1.5	6000	5	0.03
	1.0	15000	10	0.038

For dependent situation, we also apply false rate to evaluate performance of estimator. When assumption A2 is violated, we give an example and it shows that subsampling point-wise confidence intervals fail to be correct under this situation.

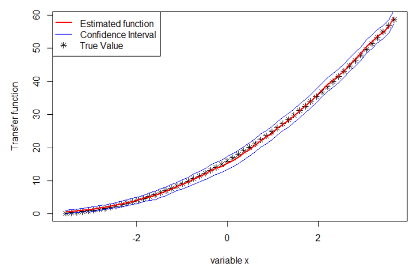
Example 6 (Transformation function estimation with MA data)**.**
In this example, we suppose that $Z_i,\ i = 1, 2, ..., n$ are $MA(m)$ normal data. That is, we suppose i.i.d innovations $\epsilon_i,\ i = ..., -1, 0, 1, ...$ satisfy standard normal distribution $N(0, \sigma^2)$ for some $\sigma > 0$ and let $Z_i = \Sigma_{k=0}^{m} \alpha_k \epsilon_{i-k},\ \alpha_0 = 1$. Notice that, marginal distribution of Z_i is normal distribution $N(0, \sigma^2 \Sigma_{k=0}^{m} \alpha_k^2)$. $MA(m)$ sequence is strong mixing (definition can be seen in [1]) for Z_t and $Z_{t+s},\ s > k$ are independent. Therefore, condition A2 is satisfied for $MA(m)$ sequence with normal innovation.

For a normal example, we choose sample size $n = 3000$ and $m = 10$ with coefficients $\alpha_k = 0.90^k,\ k = 1, 2, ..., 10$ and $\epsilon_i \sim N(0, 1)$. For a counter example, we choose sample size $n = 3000$, $m = 50000$ and coefficients $\alpha_k = 1,\ k = 1, 2, ..., 50000$, $\epsilon_i \sim N(0, 10^{-6})$. Since sample size is only 3000, the second example has strong dependency and large $\sup_{x\in\mathbf{R}} f_\epsilon$ as well.

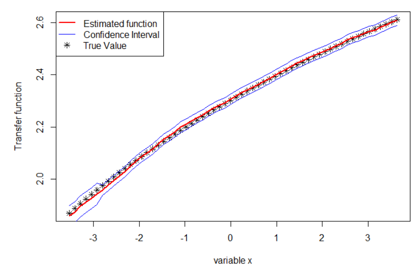
Similar as example 4, $g(x)$ is chosen as 1) $(x+4)^2$, 2)$\log(x+10)$, 3)x^3. We choose $b(n)$ in theorem 6 as $n^{3/5}$. For the first case, we also compute false rate defined in example 4 and the result is demonstrated in table 6 From figure 6 and table 6, we see that dependency affects accuracy of confidence intervals. As dependency becomes stronger, we need more data to construct a precise confidence interval.

4.2. Numerical study on water treatment plant data

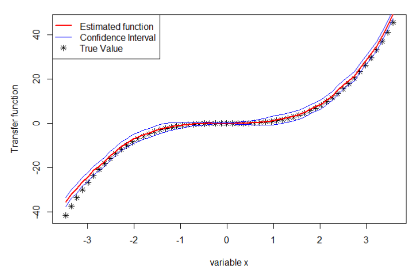
In this section, we apply results mentioned above to study relationship between chemical demand of oxygen in input waste water (DQO-E) and the chemical demand of oxygen in water that has passed the primary settler (DQO-D) in a waste water treatment plant[7]. This index is always used to quantify amount



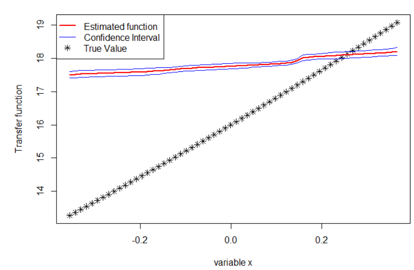
(a) $g(x) = (x + 4)^2$



(b) $g(x) = \log(x + 10)$



(c) $g(x) = x^3$



(d) $g(x) = (x + 4)^2$, (1.7) is violated

FIG 3. *3(a) to 3(c) respectively demonstrates behavior of estimator for different $g(x)$ when sample size is 3000 and confidence level is 0.95. Red line, blue line and stars respectively represents the estimated transformation function, point-wise confidence interval and true values of transformation function. 3(d) is a counter example which is strong dependent and has large* $\sup_{x\in\mathbf{R}} f_\epsilon$.

of organics in water. Instead of regression model, here we will treat DQO-E in wasted water as a random variable and suppose primary settler as a function g that decreases the concentration of organics in the waste water. Thus, the remaining organics (quantified by DQO-D) is equal to $g(DQO-E)$. Intuitively, heavier the input water is polluted, more organics will be remained after the water is cleaned. Thus, it is safe to assume that g is strictly increasing. Q-Q plot of gamma distribution and DQO-E shows that gamma distribution is a suitable approximation for DQO-E. Through maximum likelihood estimate, shape and scale parameter are estimated as 10.97 and 37.10, so we suppose that DQO-E has gamma distribution $\Gamma(10.97, 0.0270)$. Notice that gamma distribution with shape and rate $\alpha > 1$ and β has density $\frac{\beta^\alpha}{\Gamma(\alpha)}x^{\alpha-1}\exp(-\beta x)$. Thus, we have

$$\frac{f_Z^{'}(x)}{f_Z(x)} = \frac{\alpha-1}{x} - \beta \tag{4.2}$$

When $|x|$ is sufficiently small, $f_Z(x) > 0$ is increasing according to (4.2). From mean value theorem, $F_Z(x) = xf_Z(s_x) \leq xf_Z(x)$, here $0 \leq s_x \leq x$. Therefore, as long as

$$\frac{g^{''}(x)x}{g^{'}(x)} = O\left(1\right), \ x \to 0 \tag{4.3}$$

condition B3 is satisfied when $x \to 0$. On the other hand, notice that, as x being large

$$\begin{aligned}\frac{1-F_Z(x)}{f_Z(x)} &= x\int_0^1 (z+1)^{\alpha-1}\exp(-\beta xz)dz + x\int_1^\infty (z+1)^{\alpha-1}\exp(-\beta xz)dz \\ &\leq x\int_0^1 2^{\alpha-1}\exp(-\beta xz)dz + 2^{\alpha-1}\int_0^\infty \frac{z^{\alpha-1}}{x^{\alpha-1}}\exp(-\beta z)dz \\ &= \frac{2^{\alpha-1}}{\beta}(1-\exp(-\beta x)) + \frac{2^{\alpha-1}}{\beta^\alpha x^{\alpha-1}}\mathbf{\Gamma}(\alpha)\end{aligned} \tag{4.4}$$

Here, $\mathbf{\Gamma}(\alpha)$ is gamma function and since $\alpha > 0$, gamma function converges absolutely. Thus, as long as

$$\frac{g^{''}(x)}{g^{'}(x)} = O\left(1\right), \ x \to \infty \tag{4.5}$$

condition B3 is satisfied as $x \to \infty$. We suppose transformation function g in the example satisfies condition (4.3) and (4.5).

We apply the test introduced in theorem 3 to test whether gamma distribution suits DQO-E data or not (that is, we suppose DQO-E is a function h of a $\Gamma(10.97, 0.0270)$ random variable and test $h(x) = x$). In order to avoid bias introduced by estimated shape and scale parameters, we use Monte Carlo method presented in Julian and Peter [3] to calculate p-value. The result is demonstrated in table 7. Figure 5 demonstrates the relation between DQO-E and DQO-D. Slope of g will decrease as input demand of oxygen in waste water increases, so we can make conclusion that primary settler is efficient in cleaning organics when there is high concentration of organic matters in waste water.

TABLE 7
Test for fitting gamma distribution of chemical demand of oxygen in input waste water

Null assumption	statistics	P-value
$h(x) = x$	0.768	0.546

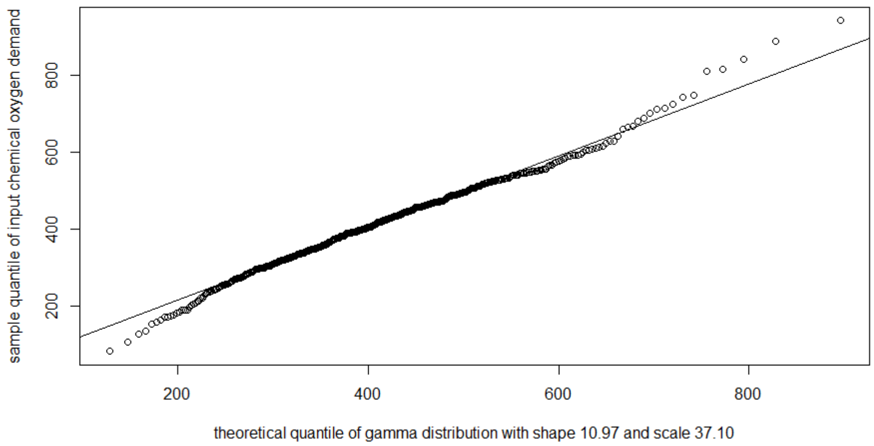


FIG 4. *Q-Q plot for chemical demand of oxygen in input waste water*

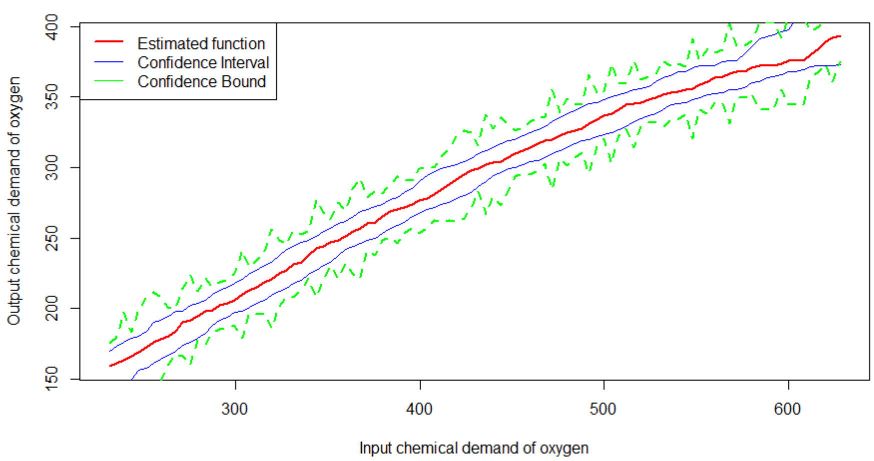


FIG 5. *Relation between DQO-E and DQO-D (definition see 4.2), sample size is 518 and confidence level is 0.99*

5. Conclusion

In this paper, we focus on model $Y_i = g(Z_i),\ i = 1, 2, ...$ with Z_i being random variables with known distribution and $g(x)$ being an unknown strictly monotonic function. We try to estimate $g(x)$ in this model. For i.i.d data, we propose an estimator of $g(x)$ and construct point-wise confidence intervals as well as confidence bands. For short-range dependent data, we prove the consistency of the proposed estimator and use a resampling method to create confidence intervals. Moreover, a goodness of fit test for correctness of $g(x)$ is presented and an alternative of this test is discussed as well.

In numerical part, we study finite sample performance of proposed estimator and test for different $g(x)$ and alternatives. width of confidence bands are sensitive with $g^{'}(x)$. If $g^{'}(x)$ is close to 0, then confidence bands will be much wider than point-wise confidence intervals and if $g^{'}$ is relatively large, then confidence bounds will be close to confidence intervals. On the contrary, small derivative of g will not severely affect point-wise confidence intervals.

In reality, this model can be applied to study relations between input signals with known distribution and responses with unknown distribution, such as correspondence between quality of materials and quality of products, electricity signals with white noises and power of motors, significance of a symptom and concentration of toxic materials in the atmosphere, etc.

6. Appendix

Proofs of the main theorems will be demonstrated here.

Proof of theorem 1. For the 1st part, according to [4],

$$\widehat{g}(x) \rightarrow_{a.s.} g(x) \Leftrightarrow \Sigma_{n=1}^{\infty} \mathbf{1}_{|\widehat{g}(x)-g(x)|>\epsilon} < \infty \text{ for } \forall \epsilon > 0 \tag{6.1}$$

From definition of sample quantile, we have

$$\mathbf{1}_{\widehat{g}(x)-g(x)>\epsilon} \leq \mathbf{1}_{\widehat{F}_Y(\xi_Y(F_Z(x))+\epsilon)<F_Z(x)} \tag{6.2}$$

From strong law of large number (theorem 6.2 in [4]), we have $\widehat{F}_Y(\xi_Y(F_Z(x)) + \epsilon) \rightarrow_{a.s.} F_Y(\xi_Y(F_Z(x)) + \epsilon) > F_Z(x)$, thus

$$\Sigma_{i=1}^{\infty} \mathbf{1}_{\widehat{F}_Y(\xi_Y(F_Z(x))+\epsilon)<F_Z(x)} < \infty \tag{6.3}$$

Also, similarly we can get that $\Sigma_{n=1}^{\infty} \mathbf{1}_{\widehat{g}(x)-g(x)<-\epsilon} < \infty$ and we prove the result.

For the second part, we prove that

$$\lim_{n\rightarrow\infty} \sup P(\widehat{\xi}_Y(c_1) > g(x)) \leq \alpha/2,\ \lim_{n\rightarrow\infty} \sup P(\widehat{\xi}_Y(c_2) < g(x)) \leq \alpha/2 \tag{6.4}$$

For large n, $c_1 + \frac{1}{n} < 1$ and according to definition of $\widehat{\xi}_Y(c_1)$ and c_1, $\widehat{\xi}_Y(c_1) > g(x) \Rightarrow \widehat{F}_Y(g(x)) \leq \widehat{F}_Y(\widehat{\xi}_Y(c_1)) \Rightarrow \widehat{F}_Y(g(x)) < c_1 + \frac{1}{n}$. For $Z_i, i = 1, 2, ..., n$

obeys A1, $Y_i = g(Z_i)$, the observed data Y_i are i.i.d and correspondingly $\mathbf{1}_{Y_i \le x}$ are i.i.d observations. From central limit theorem, and lemma 1, we have

$$\sqrt{n}(\widehat{F}_Y(g(x)) - F_Z(x)) \to_D N(0, F_Z(x)(1 - F_Z(x))) \tag{6.5}$$

Thus,

$$\begin{aligned} &\lim_{n\to\infty} \sup P(\widehat{\xi}_Y(c_1) > g(x)) \\ \le \lim_{n\to\infty} P &\left(\frac{\sqrt{n}(\widehat{F}_Y(g(x)) - F_Z(x))}{\sqrt{F_Z(x)(1-F_Z(x))}} \le \zeta(\alpha/2) + \frac{1}{\sqrt{n(F_Z(x)(1-F_Z(x)))}} \right) \\ &= \alpha/2 \end{aligned} \tag{6.6}$$

Similarly, we have $\lim_{n\to\infty} \sup P(\widehat{\xi}_Y(c_2) < g(x)) \le \alpha/2$ and theorem 1 is proved. □

Proof of theorem 2. Because of B3, then according to [6], since $Y = g(Z), Z \in [a, b]$, and g strictly increasing, then $Y \in [g(a), g(b)]$ and according to lemma 1, we have $f_Y(g(x))g^{'}(x) = f_Z(x)$, $f_Y^{'}(g(x))g^{'}(x)^2 + f_Y(g(x))g^{''}(x) = f_Z^{'}(x)$, thus suppose $z = g(x)$ and

$$\begin{aligned} &\sup_{g(a)<z<g(b)} F_Y(z)(1 - F_Y(z))|\frac{f_Y^{'}(z)}{f_Y^2(z)}| \\ = &\sup_{a<x<b} \frac{F_Z(x)(1-F_Z(x))}{f_Z^2(x)}|f_Z^{'}(x) - f_Z(x)\frac{g^{''}(x)}{g^{'}(x)}| \le \gamma \end{aligned} \tag{6.7}$$

There exists a version of Kiefer process(definition see[6]), such that

$$\begin{aligned} \sup_{\delta_n \le F_Z(x) \le 1-\delta_n} &|n(\widehat{g}(x) - g(x))f_Y(g(x)) - K(F_Z(x), n)| \\ &=_{a.s.} O((n \log\log n)^{1/4}(\log n)^{1/2}) \end{aligned} \tag{6.8}$$

For sufficiently large n, $\delta_n < F_Z(c) < F_Z(d) < 1 - \delta_n$ and the estimation above holds for $\forall x \in [c, d]$. On the other hand, for $\phi^{'}$ is continuous on its support $[d_1, d_2]$ and equal to 0 outside its support, define $\phi_m = \max_{x\in[d_1,d_2]} |\phi^{'}|$, from mean value theorem, we have, $\exists \eta \in \mathbf{R}$ such that

$$|\widehat{f}_Y(\widehat{g}(x)) - \widehat{f}_Y(g(x))| = |\widehat{f}_Y(\eta)^{'}(\widehat{g}(x) - g(x)| \le \frac{\phi_m}{h^2}|\widehat{g}(x) - g(x)| \tag{6.9}$$

We next consider $\widehat{f}_Y(g(x)) - f_Y(g(x))$. From integral transformation, we have

$$\begin{aligned} \widehat{f}_Y(g(x)) - f_Y(g(x)) &= \frac{1}{h}\int_{-\infty}^{\infty} \widehat{F}_Y(g(x) - hy)\phi^{'}(y)dy - f_Y(g(x)) \\ &= \frac{1}{h}\int_{-\infty}^{\infty} (\widehat{F}_Y(g(x) - hy) - F_Y(g(x) - hy))\phi^{'}(y)dy \\ &\quad + \int_{-\infty}^{\infty} \phi(y)(f_Y(g(x) - hy) - f_Y(g(x)))dy \end{aligned} \tag{6.10}$$

From theorem A in [6], since $F_Y(Y_i)$ are uniform random variable, we pick $y = F_Y(g(x) - hz)$ in that theorem, suppose that $c \le x \le d$ and h sufficiently small such that $g(a) < g(c) - hd_2,\ g(b) > g(d) - hd_1$, use lemma 1 and we have

$$\sup_{c\le x\le d,\ d_1\le z\le d_2} |n(\widehat{F}_Y(g(x)-hz)-F_Y(g(x)-hz))-K(F_Z(x),n)| =_{a.s.} O(\log^2 n) \tag{6.11}$$

For n sufficiently large and $h < h_0$, h_0 sufficiently small, we have $g(a) < g(c) - h_0d_2 < g(d)-h_0d_1 < g(b)$ and since $f_Y(x)$ is continuous differentiable according to B3, its derivative at $[g(c) - h_0d_2, g(d) - h_0d_1]$ is bounded, and suppose $f_0 = \max_{x\in[g(c)-h_0d_2,g(d)-h_0d_1]} |f_Y^{'}(x)|$. From the law of iterated logarithm [6], we have

$$\limsup_{n\to\infty} \sup_{0\le y\le 1} |K(y,n)|/(2n\log\log n)^{1/2} =_{a.s.} 1/2 \tag{6.12}$$

From assumption B3, $[c,d]$ is a closed interval and $\min_{x\in[c,d]} |f_Y(g(x))| > 0$. According to (6.8), for sufficiently large n,

$$\begin{aligned}\sup_{x\in[c,d]} |\widehat{g}(x) - g(x)| &\le \sup_{x\in[c,d]} |\frac{K(F_Z(x),n)}{n|f_Y(g(x))|}| + O_{a.s.}(\frac{(\log\log n)^{1/4}(\log n)^{1/2}}{n^{3/4}}) \\ &= O_{a.s.}(\frac{(\log\log n)^{1/2}}{\sqrt{n}})\end{aligned} \tag{6.13}$$

Besides, *supp* $\phi \subseteq [d_1, d_2]$, from (6.11) and (6.10), we have

$$\begin{aligned} &\sup_{c\le x\le d} |\widehat{f}_Y(g(x)) - f_Y(g(x))| \\ &\le \frac{1}{h}\int_{-\infty}^{\infty} |(\widehat{F}_Y(g(x) - hy) - F_Y(g(x) - hy))\phi^{'}(y)|dy \\ &\quad + \int_{-\infty}^{\infty} \phi(y)|(f_Y(g(x) - hy) - f_Y(g(x)))|dy \\ &\le (\sup_{0\le y\le 1} |\frac{K(y,n)}{n}| + O_{a.s.}(\frac{\log^2 n}{n}))\frac{\phi_m(d_2-d_1)}{h} + f_0h\int_{d_1}^{d_2} |y\phi(y)|dy \\ &=_{a.s.} O(h + \frac{(\log\log n)^{1/2}}{h\sqrt{n}})\end{aligned} \tag{6.14}$$

To prove theorem 2, from triangle inequality and (6.8), (6.9) and (6.14),

$$\begin{aligned}
&\sup_{c\le x\le d}|\sqrt{n}(\widehat{g}(x)-g(x))\widehat{f}_Y(\widehat{g}(x))-\frac{K(F_Z(x),n)}{\sqrt{n}}| \\
&\le \sup_{c\le x\le d}|\sqrt{n}(\widehat{g}(x)-g(x))f_Y(g(x))-\frac{K(F_Z(x),n)}{\sqrt{n}}| \\
&\quad+\sup_{c\le x\le d}\sqrt{n}|\widehat{g}(x)-g(x)||\widehat{f}_Y(\widehat{g}(x))-\widehat{f}_Y(g(x))| \\
&\quad+\sup_{c\le x\le d}\sqrt{n}|\widehat{g}(x)-g(x)||\widehat{f}_Y(g(x))-f_Y(g(x))| \\
&\le O_{a.s.}(\frac{(\log\log n)^{1/4}(\log n)^{1/2}}{n^{1/4}})+\sup_{c\le x\le d}\frac{\sqrt{n}\phi_m}{h^2}|\widehat{g}(x)-g(x)|^2 \\
&\quad+O_{a.s.}((\log\log n)^{1/2}h+\frac{(\log\log n)}{h\sqrt{n}}) \\
&=O_{a.s.}(\frac{(\log\log n)^{1/4}(\log n)^{1/2}}{n^{1/4}})+O_{a.s.}(\frac{\log\log n}{\sqrt{n}h^2}) \\
&\quad+O_{a.s.}((\log\log n)^{1/2}h+\frac{(\log\log n)}{h\sqrt{n}})
\end{aligned} \tag{6.15}$$

Thus, let $(\log\log n)^{1/2}h\to 0$ and $\frac{\sqrt{n}h^2}{\log\log n}\to\infty$, we prove the result. □

Proof of corollary 1. Define $A_n=\Big\{\sup_{c\le x\le d}|\sqrt{n}(\widehat{g}(x)-g(x))\widehat{f}_Y(\widehat{g}(x))|>c\Big\}$. From triangle inequality,

$$\begin{aligned}
&\sup_{c\le x\le d}|\sqrt{n}(\widehat{g}(x)-g(x))\widehat{f}_Y(\widehat{g}(x))| \\
&\le \sup_{0\le y\le 1}|\frac{K(y,n)}{\sqrt{n}}|+\sup_{c\le x\le d}|\sqrt{n}(\widehat{g}(x)-g(x))\widehat{f}_Y(\widehat{g}(x))-\frac{K(F_Z(x),n)}{\sqrt{n}}|
\end{aligned} \tag{6.16}$$

For $\forall\,\epsilon>0$ given and sufficiently large n, $P(\sup_{c\le x\le d}|\sqrt{n}(\widehat{g}(x)-g(x))\widehat{f}_Y(\widehat{g}(x))-\frac{K(F_Z(x),n)}{\sqrt{n}}|<\epsilon)=1$. Therefore, $P(A_n)\le P(\sup_{0\le y\le 1}|\frac{K(y,n)}{\sqrt{n}}|>c-\epsilon)$ for large n. According to [5], $K(y,n)/\sqrt{n}$ is a Brownian bridge and according to [5], $P(\sup_{0\le y\le 1}|B(y)|\le c)=1-\Sigma_{k\neq 0}(-1)^{k+1}\exp(-2k^2c^2)$. Thus, from continuity of measure,

$$\lim_{n\to\infty}\sup P(A_n)\le\lim_{\epsilon\to 0}P(\sup_{0\le y\le 1}|\frac{K(y,n)}{\sqrt{n}}|>c-\epsilon)=P(\sup_{0\le y\le 1}|\frac{K(y,n)}{\sqrt{n}}|\ge c) \tag{6.17}$$

The final thing is to prove that $P(\sup_{0\le y\le 1}|\frac{K(y,n)}{\sqrt{n}}|=c)=0$. From continuity of measure, for $c>0$,

$$\begin{aligned}
P(\sup_{0\le y\le 1}|\frac{K(y,n)}{\sqrt{n}}|=c)&=\lim_{n\to\infty}P(c-1/n<\sup_{0\le y\le 1}|\frac{K(y,n)}{\sqrt{n}}|\le c+1/n) \\
&\le\lim_{n\to\infty}\Sigma_{k\neq 0}\exp(-2k^2(c-1/n)^2)-\exp(-2k^2(c+1/n)^2)
\end{aligned} \tag{6.18}$$

From mean value theorem, there exists $\eta_k \in [c-1/k, c+1/k]$, such that $\exp(-2k^2(c-1/n)^2) - \exp(-2k^2(c+1/n)^2) = -8\exp(-2k^2\eta_k^2)k^2\eta_k/n$, so

$$\Sigma_{k\neq 0}\exp(-2k^2(c-1/n)^2) - \exp(-2k^2(c+1/n)^2) = O(1/n) \tag{6.19}$$

and the result is proved. □

Suppose $s(c) = \Sigma_{k\neq 0}(-1)^{k+1}\exp(-2k^2c^2),\ c>0$, then we have, for ϵ sufficiently close to 0, $c-|\epsilon| > c/2$ and

$$\begin{aligned} |s(c+\epsilon) - s(c)| &\le 4\Sigma_{k\neq 0}|\exp(-2k^2\eta_k^2)k^2\eta_k\epsilon| \\ &\le 8\Sigma_{k\neq 0}|\exp(-2k^2(c/2)^2)k^2c\epsilon| = O(\epsilon) \end{aligned} \tag{6.20}$$

Here η_k belongs to c and $c+\epsilon$. This shows that s is continuous.

Proof of theorem 3. Define events

$$M_n = \sup_{\delta_n \le F_Z(x) \le 1-\delta_n} \sqrt{n}\frac{f_Z(x)}{h'(x)}|\widehat{g}(x) - h(x)| > c \tag{6.21}$$

According to triangular inequality,

$$\begin{aligned} \sup_{\delta_n \le F_Z(x) \le 1-\delta_n} \sqrt{n}\frac{f_Z(x)}{h'(x)}|\widehat{g}(x) - h(x)| &\le \sup_{0\le y\le 1}|\frac{K(y,n)}{\sqrt{n}}| \\ + \sup_{\delta_n \le F_Z(x) \le 1-\delta_n} |\sqrt{n}\frac{f_Z(x)}{h'(x)}(\widehat{g}(x) - h(x)) &- \frac{K(F_Z(x),n)}{\sqrt{n}}| \end{aligned} \tag{6.22}$$

For $\forall\ \epsilon > 0$, and sufficiently large n, according to (6.8) and since $f_Y(h(x)) = f_Z(x)/h'(x)$, we have

$$P(\sup_{\delta_n \le F_Z(x) \le 1-\delta_n} |\sqrt{n}\frac{f_Z(x)}{h'(x)}(\widehat{g}(x) - h(x)) - \frac{K(F_Z(x),n)}{\sqrt{n}}| < \epsilon) = 1 \tag{6.23}$$

and

$$\lim_{n\to\infty}\sup P(M_n) \le \lim_{n\to\infty}\sup P(\sup_{0\le y\le 1}|\frac{K(y,n)}{\sqrt{n}}| > c-\epsilon) = P(\sup_{0\le y\le 1}|B(y)| > c-\epsilon) \tag{6.24}$$

For $\forall\epsilon > 0$. From continuity of measure, $\lim_{n\to\infty}\sup P(M_n) \le P(\sup_{0\le y\le 1}|B(y)| \ge c)$. On the other hand,

$$\begin{aligned} \sup_{\delta_n \le F_Z(x) \le 1-\delta_n} \sqrt{n}\frac{f_Z(x)}{h'(x)}|\widehat{g}(x) - h(x)| &\ge \\ \sup_{\delta_n \le F_Z(x) \le 1-\delta_n} |\frac{K(F_Z(x),n)}{\sqrt{n}}| & \\ - \sup_{\delta_n \le F_Z(x) \le 1-\delta_n} |\frac{K(F_Z(x),n)}{\sqrt{n}} - \sqrt{n}\frac{f_Z(x)}{h'(x)}(\widehat{g}(x) - h(x))| & \end{aligned} \tag{6.25}$$

Also using (6.8), and continuity of measure,

$$\begin{aligned}\lim_{n\to\infty}\inf P(M_n > c) \geq \lim_{n\to\infty}\inf P(&\sup_{\delta_n\leq y\leq 1-\delta_n}|\frac{K(y,n)}{\sqrt{n}}| > c+\epsilon)\\ &= P(\sup_{0<y<1}|\ B(y)| > c+\epsilon)\end{aligned} \tag{6.26}$$

Since B is a Brownian bridge, when $y = 0, 1$, $B(y) = 0$ $a.s.$. For $c > 0$, $P(\sup_{0<y<1}|\ B(y)| > c+\epsilon) = P(\sup_{0\leq y\leq 1}|\ B(y)| > c+\epsilon)$. Also, since ϵ is arbitrary, continuity of measure shows that

$$\lim_{n\to\infty}\inf P(M_n) \geq P(\sup_{0\leq y\leq 1}|\ B(y)| > c) \tag{6.27}$$

From (6.19), we know that, when $c > 0$,

$$P(\sup_{0\leq y\leq 1}|B(y)| \geq c) = P(\sup_{0\leq y\leq 1}|\ B(y)| > c) \tag{6.28}$$

and the theorem is proved. □

Proof of theorem 4. Since F_Z is strictly increasing and $x_0 \in (a, b)$, for sufficiently large n, $\delta_n < F_Z(x_0) < 1-\delta_n$. Thus we have

$$\begin{aligned}&\sup_{\delta_n\leq F_Z(x)\leq 1-\delta_n}\sqrt{n}\frac{f_Z(x)}{h'(x)}|\widehat{g}(x)-h(x)|\\ \geq &\sup_{\delta_n\leq F_Z(x)\leq 1-\delta_n}\sqrt{n}\frac{f_Z(x)}{h'(x)}(|g(x)-h(x)|-|\widehat{g}(x)-g(x)|)\\ \geq &\sqrt{n}\frac{f_Z(x_0)}{h'(x_0)}|g(x_0)-h(x_0)|-\sqrt{n}\frac{f_Z(x_0)}{h'(x_0)}|\widehat{g}(x_0)-g(x_0)|\end{aligned} \tag{6.29}$$

Use M_n to represent the same thing as in proof of theorem 3,

$$P(M_n) \geq P(\sqrt{n}\frac{f_Z(x_0)}{h'(x_0)}|g(x_0)-h(x_0)|-c > \sqrt{n}\frac{g'(x_0)}{h'(x_0)}\frac{f_Z(x_0)}{g'(x_0)}|\widehat{g}(x_0)-g(x_0)|) \tag{6.30}$$

According to (6.8), $\sqrt{n}\frac{f_Z(x_0)}{g'(x_0)}(\widehat{g}(x_0)-g(x_0)) \to_{a.s.} K(F_Z(x_0), n)/\sqrt{n}$ so it is $O_p(1)$. On the other hand, $\sqrt{n}|g(x_0)-h(x_0)| \to \infty$, we know that $P(M_n) \to 1$ and the first part is proved.

For the second part, notice that

$$\begin{aligned}&\sup_{\delta_n\leq F_Z(x)\leq 1-\delta_n}\sqrt{n}\frac{f_Z(x)}{g'(x)(1+\frac{s'(x)}{\sqrt{n}g'(x)})}|\widehat{g}(x)-g(x)-\frac{s(x)}{\sqrt{n}}|\\ \leq &\sup_{\delta_n\leq F_Z(x)\leq 1-\delta_n}\sqrt{n}\frac{f_Z(x)}{g'(x)}|\widehat{g}(x)-g(x)|+\sup_{a<x<b}\frac{f_Z(x)|s(x)|}{g'(x)}\end{aligned} \tag{6.31}$$

And we get the result. □

Proof of theorem 5. According to theorem 1 in [18], we choose p in that theorem as $F_Z(x)$, since $f_Z(x) > 0$, $\xi_Z(F_Z(x)) = x$. This is because, on one hand, from definition of ξ_Z, $F_Z(x) \geq F_Z(x) \Rightarrow x \geq \xi_Z(F_Z(x))$. On the other hand, for $y < x$ close to x, from definition of derivative, $F_Z(y) < F_Z(x) - \frac{1}{2}f_Z(x)(x-y) < F_Z(x)$. Since F_Z is increasing and right continuous at point x, from definition of ξ_Z, $F_Z(\xi_Z(F_Z(x))) \geq F_Z(x) \Rightarrow \xi_Z(F_Z(x)) \geq x$, thus the equality holds. We have

$$\xi_n^Z(F_Z(x)) - x = \frac{F_Z(x) - \widehat{F}_Z(x)}{f_Z(x)} + O_{a.s.}(n^{-3/4}(\log\log n)\log^{3/4}(n)) \tag{6.32}$$

According to proposition 1 and lemma 3 in [18], we have that $\exists 0 < \sigma_1 < \infty$ being a constant such that

$$\limsup_{n\to\infty} |\frac{\sqrt{n}(\widehat{F}_Z(x) - F_Z(x))}{\sqrt{2\log\log n}}| =_{a.s.} \sigma_1 \Rightarrow \ \widehat{F}_Z(x) - F_Z(x) = O_{a.s.}\left(\frac{\sqrt{\log\log n}}{\sqrt{n}}\right) \tag{6.33}$$

Thus, in particular, $\widehat{\xi}_Z(F_Z(x)) \to_{a.s.} x$. Since $\widehat{\xi}_Y(F_Z(x)) = g(\widehat{\xi}_Z(F_Z(x)))$ and g is continuous, we have

$$\widehat{g}(x) = \widehat{\xi}_Y(F_Z(x)) \to_{a.s.} g(x) \tag{6.34}$$

For the second part, according to theorem 2 and remark 6 in [18], under condition A2, notice that $f_Z(x) > 0,\ x \in [c,d] \Rightarrow F_Z(x)$ being strictly increasing and thus,

$$\sup_{c\leq x\leq d} |\widehat{\xi}_Z(F_Z(x)) - x| = o_{a.s.}\left(\frac{c_q(n)}{\sqrt{n}}\right) \tag{6.35}$$

Here, $c_q(n) = (\log n)^{1/q}(\log\log n)^{2/q}$ if $q > 2$ and $(\log n)^{3/2}(\log\log n)$ if $q = 2$. Since $[c,d]$ is a closed interval and g is continuous, thus is uniform continuous on $[c,d]$. Therefore, uniformly convergence is proved. □

Proof of theorem 6. According to [12], the only thing to prove is that $\sqrt{n}(\widehat{g}(x) - g(x))$ converges to a non-degenerated distribution. According to lemma 11 in [18], for $q \geq 2$, for sufficiently large n, $|a_n| < 1$ (otherwise the summation will not converge), then

$$\Sigma_{i=n}^{\infty}|a_i|^{\min(1,\alpha/2)} \leq \Sigma_{i=n}^{\infty}|a_i|^{\min(1,\alpha/q)} < \infty \tag{6.36}$$

Thus, we have

$$\sqrt{n}(\widehat{F}_Z(x) - F_Z(x)) \to_d N(0,\sigma^2) \tag{6.37}$$

weakly. N is a normal distribution with unknown variance. Therefore, according to [18], similar with theorem 5, we have

$$\widehat{\xi}_Z(F_Z(x)) - x = \frac{F_Z(x) - \widehat{F}_Z(x)}{f_Z(x)} + O_{a.s.}(n^{-3/4}(c_q(n)\log n)^{1/2}) \tag{6.38}$$

Thus, $\sqrt{n}(\widehat{\xi}_Z(F_Z(x)) - x) \to_d N(0, \sigma^2/f_Z^2(x))$. Since g is differentiable at x, according to lemma 2 and delta method, we have

$$\sqrt{n}(\widehat{g}(x) - g(x)) = \sqrt{n}((g(\widehat{\xi}_Z(F_Z(x)))) - g(x)) \to_d N(0, g^{'}(x)^2\sigma^2/f_Z^2(x)) \tag{6.39}$$

Since $g^{'}, f_Z(x) \neq 0$, the result is proved. □

Acknowledgements

We appreciate UCI Machine Learning Repository for providing water treatment plant data set [7] on our numerical experiment. We are grateful to editors and reviewers for their insightful comments, which help us improve this paper a lot.

References

[1] Kolmogorov, A. N. and Rozanov, Y. A. (1959). On Strong Mixing Conditions for Stationary Gaussian Processes. *Theory of Pobability and its Application* **5** 204–208. MR0133175

[2] Bahadur, R. R. (1966). A Note on Quantiles in Large Samples. *The Annals of Mathematical Statistics* **37** 577–580. MR0189095

[3] BESAG, J. and CLIFFORD, P. (1989). Generalized Monte Carlo significance tests. *Biometrika* **76** 633-642. MR1041408

[4] Cinlar, E. (2011). *Probability and Stochastics*. Springer-Verlag New York. MR2767184

[5] Csorgo, M. (1983). *Quantile Processes with Statistical Applications*. Society for Industrial and Applied Mathematics.

[6] Csorgo, M. and Revesz, P. (1978). Strong Approximations of the Quantile Process. *The Annals of Statistics* **6** 882–894. MR0501290

[7] Dheeru, D. and Karra Taniskidou, E. (2017). UCI Machine Learning Repository.
http://archive.ics.uci.edu/ml/datasets/Water+Treatment+Plant.

[8] Dietz, E. J. and Killeen, T. J. (1981). A Nonparametric Multivariate Test for Monotone Trend with Pharmaceutical Applications. *Journal of the American Statistical Association* **76** 169-174. MR0608189

[9] Helfenstein, U. (2005). *Encyclopedia of Biostatistics(eds P. Armitage and T. Colton)* Intervention Analysis in Time Series. American Cancer Society.

[10] Lehmann, E. L. and Romano, J. P. (2005). *Testing Statistical Hypothesis*. Springer-Verlag New York. MR2135927

[11] Mukerjee, H. (1988). Monotone Nonparametric Regression. *Ann. Statist.* **16** 741–750.

[12] Politis, D. N., Romano, J. P. and Wolf, M. (1999). *Subsampling*. Springer-Verlag New York. MR1707286

[13] Quarteroni, A., Sacco, R. and Saleri, F. (2007). *Numerical Mathematics*. Springer-Verlag Berlin Heidelberg.

[14] Rocke, D. M. and Lorenzato, S. (1995). A Two-Component Model for Measurement Error in Analytical Chemistry. *Technometrics* **37** 176-184.

[15] Rosenblatt, M. (1952). Remarks on a Multivariate Transformation. *The Annals of Mathematical Statistics* **23** 470–472. MR0049525

[16] Sun, S. and Lahiri, S. N. (2006). Bootstrapping the Sample Quantile of a Weakly Dependent Sequence. *Sankhya: The Indian Journal of Statistics (2003-2007)* **68** 130–166. MR2301568

[17] Trustrum, K. and Jayatilaka, A. D. S. (1979). On estimating the Weibull modulus for a brittle material. *Journal of Materials Science* **14** 1080–1084.

[18] Wu, W. B. (2005). On the Bahadur representation of sample quantiles for dependent sequences. *The Annals of Statistics* **33** 1934–1963. MR2166566

[19] Zhao, O. and Woodroofe, M. (2012). Estimating a monotone trend. *Statistica Sinica* **22** 359-378. MR2933180