Wavelet estimation of a regression function with a sharp change point in a random design

Cheolwoo Park[a] **Woo-Chul Kim**[b,*]

[a] *Department of Statistics, University of Florida, Gainesville, FL 32611, USA*

[b] *Department of Statistics, Seoul National University, Seoul, 151-747, South Korea*

and Department of Statistics, University of California at Davis, Davis, CA 95616, USA

Abstract

In a random design nonparametric regression model, this paper deals with the detection of a sharp change point and the estimation of a regression function with a single jump point. A method based on design transformation and binning is used in order to convert a random design into an equispaced design whose number of points is a power of 2. Using the continuous wavelet transform of the data, we construct a sharp change point estimator and obtain its rate of convergence. Wavelet methods are well known for their good adaptivity around sudden local changes; however, in practice, the Gibbs phenomenon still exists. This difficulty is overcome by suitably adjusting the data with preliminary estimators for the location and the size of discontinuity. Global and local asymptotic results of the proposed method are obtained. The method is also tested on simulated examples and the results show that the proposed method alleviates the Gibbs phenomenon.

Keywords: Block thresholding; Continuous wavelet transform; Design transformation and binning; Random design; Rate of convergence; Sharp change point problem; Wavelet function estimation.

---

1. Introduction

We observe independent data pairs (X_i, Y_i), for $i = 1, \ldots, n$, under a random design regression model:

$$Y_i = g(X_i) + \varepsilon_i, \quad 1 \leq i \leq n, \tag{1}$$

where g is a regression function and ε_i's are independent normal errors with $E(\varepsilon_i) = 0$, Var$(\varepsilon_i) = \sigma^2 < \infty$. The design points X_i's are assumed to be supported in the interval $[0, 1]$.

This paper provides an analysis of a sharp change point which describes a sudden localized change, and the estimation of a regression function with a single jump point. These problems have been consid-

*Corresponding author.

E-mail address: cpark@stat.ufl.edu (C. Park), wckim@stats.snu.ac.kr (W. Kim)

ered in many other contexts under an equispaced design, that is $X_i = i/n$ in (1). In this work,we assume a general setting, a random design as in Korostelev and Tsybakov (1993).

Fig. 1 about here.

Fig. 1 (a) shows one example of target functions of interest, which was contrived by Nason and Silverman (1994), and Fig. 1 (b) shows the corresponding noisy data. The $n = 1000$ design points are generated from a uniform distribution and the errors are generated from a normal distribution with $\sigma = 0.15$. An estimate of this data with the BlockJS estimator, which was proposed by Cai (1999), is in Fig. 1 (c). Wavelet estimators are known to have fast convergence rates and good practical performance even if a function contains discontinuities. In practice, however, the Gibbs phenomenon still exists, as illustrated in Fig. 1 (c), and a proper method that can adjust properly to such a problem is needed. Fig. 1 (d) shows a function estimate of the proposed method with the same data and it reveals the elimination of wiggles around a jump point.

Section 2 deals with a sharp change point problem under a random design. A sharp change point estimator based on the continuous wavelet transform is provided, which was developed in Wang (1995). A similar simulation to that of Raimondo (1998) is introduced to explain why we adapt the continuous wavelet transform to a sharp change point detection. Also, the convergence rate of the estimator is obtained. In Section 3, we assume that there is a single jump point in a regression function and provide a wavelet function estimator reducing the Gibbs phenomenon. This is done by adjusting the data using estimators for a location of a jump point and a corresponding jump size. Mean Integrated Squared Error (MISE) and Mean Squared Error (MSE) are obtained. Numerical examples are presented in Section 4 and some concluding remarks are provided in Section 5. The proofs of asymptotic results are given in Section 6.

2. Sharp change point detection

Section 2.1 defines the sharp change point problem being considered in this paper. In Section 2.2, a motivation of a sharp change point estimator based on the continuous wavelet transform is explained by a simulated example. A design transformation dealing with a random design and an estimator detecting a sharp change point are described in Section 2.3.

2.1. Sharp change point problem

We consider a class of functions on $[0, 1]$ with either a single jump point or a single cusp point as follows :

(a) $\mathcal{G}_0$ is a class of functions g on $[0, 1]$ such that

   (i) $\liminf_{h\to 0} \mid g(\tau + h) - g(\tau - h) \mid > 0$ for a unique $\tau \in (0, 1)$.

(ii) $\sup_{0<x<y<\tau} |g(x) - g(y)|/|x-y|^{\alpha'} < \infty$ and
$\sup_{0<\tau<x<y} |g(x) - g(y)|/|x-y|^{\alpha'} < \infty$ for some α', $0 < \alpha' \leq 1$.

(b) $\mathcal{G}_\alpha$ $(0 < \alpha < 1)$ is a class of functions g on $[0,1]$ such that

(i) $\liminf_{h\to 0} \mid g(\tau+h) - g(\tau-h) \mid /|h|^\alpha > 0$ for a unique $\tau \in (0,1)$.

(ii) g is differentiable on $(0,1)$ except at τ.

(c) $\mathcal{G}_\alpha$ $(\alpha \geq 1)$ is a class of functions g on $[0,1]$ such that

(i) g is N times differentiable on $(0,1)$ where N is the integer part of α.

(ii) $g^{(N)} \in \mathcal{G}_{\alpha-N}$.

For $g \in \mathcal{G}_\alpha$ $(\alpha \geq 0)$, the estimation of a single jump point or a single cusp point $\tau = \tau(g)$ satisfying

$$\liminf_{h\to 0} |g^{(N)}(\tau+h) - g^{(N)}(\tau-h)|/|h|^{\alpha-N} > 0$$

was called the sharp change point problem by Raimondo (1998).

2.2. Discussion of Wang's estimator

Wang (1995) proposed a method to detect jumps and sharp cusps in a function, which is observed with noise, by checking if the continuous wavelet transform of the data has significantly large absolute values across fine scale levels. In a white noise regression model,

$$Y(dx) = g(x)dx + \zeta W(dx), \quad x \in [0,1],$$

where W is a standard Wiener process and ζ is a formal noise level parameter. The continuous wavelet transform of Y, which is a function of the scale parameter a and the location parameter b, is defined by

$$TY(a,b) = \int \psi_a(b-u)Y(du), \tag{2}$$

where $\psi_a(x) = a^{-1/2}\psi(x/a)$. If the function g has a sharp change at b, then, for some fine scales a_ζ, $|TY(a_\zeta, b)|$ is expected to be larger than the others.

However, the existence of an irregularity at point τ for the otherwise smooth function g does not imply that the wavelet coefficients of g near τ will be large for arbitrary fine scales. Mallat and Hwang (1992) showed that when a function has fast oscillations, its worst singular behavior is observed outside the neighborhood of τ. Raimondo (1998) also gave a numerical example of lack of translation invariance of the discrete wavelet transform, with $a = 2^{-j}$ and $b = 2^{-j}k$ in (2).

Related to this issue, we argue that empirical wavelet coefficients by the continuous spatial parameter b with the scale parameter $a = 2^{-j}$ in (2) would yield better results in practice than those by the discrete wavelet transform. To demonstrate this argument, we conduct a similar experiment to that of Raimondo (1998) with integrated "Brownian-path-with-jumps".

Fig. 2 about here.

Fig. 2 displays the function (solid line), which is Hölder continuous with exponent $\alpha < 3/2$, except at points $x = 1/4, 1/2$ and $3/4$ where there is a "kink" of size 0.1. Empirical wavelet coefficients for $j = 7, 8$ and 9 by the discrete wavelet transform (DWT) and the continuous wavelet transform (CWT) with db2 are depicted in Fig. 3. Here, the CWT means the wavelet transform with $a = 2^{-j}$ and the continuous spatial parameter b in (2), and db2 means the Daubechies' wavelet with order 2 (see, e.g., Daubechies (1992)).

Fig. 3 about here.

Fig. 3 shows that, while the DWT version of Wang's estimator does not detect three "kink" points simultaneously at all three levels, the CWT performs quite well for all $j = 7, 8$ and 9. This leads us to confirm that empirical wavelet coefficients by the CWT can be a useful tool in detecting sharp change points.

2.3. Design transformation and sharp change point detection

Hall, Park and Turlach (1998) pointed out that a design transformation allows wavelet estimators to be used with irregularly spaced design without degrading transients by oversmoothing, and to produce curve estimates with lower variability in comparison to the alternative approaches.

Dealing with a random design, we first transform it to an equispaced design. Let $X_{(1)} < \cdots < X_{(n)}$ be the order statistics of the sample $\{X_1, \ldots, X_n\}$ and put $t_i = i/n$ for $1 \le i \le n$. Denote the distribution function of the design points X_i by F and $\hat{F}$ which maps $[0, X_{(1)}]$ linearly to $[0, t_1]$, $[X_{(i-1)}, X_{(i)}]$ to $[t_{i-1}, t_i]$ for $2 \le i \le n-1$, and $[X_{(n-1)}, 1]$ to $[t_{n-1}, 1]$.

Now, assume that the g in (1) has a sharp change point at τ, $0 < \tau < 1$, and is smooth otherwise as in Section 2.1. Suppose that a mother wavelet ψ has the support $[-N, N+1]$ and satisfies the order of $(N+1)$ moment condition, that is,

$$\int u^l \psi(u) du = 0, \quad l = 0, 1, \cdots, N.$$

Define a test statistic as

$$\hat{\Delta}_{j_0}(t) = \frac{2^{j_0/2}}{n} \sum_{i=1}^{n} Y_{[i]} \psi(2^{j_0}(t_i - t)), \tag{3}$$

where $Y_{[i]}$'s are concomitants for $1 \le i \le n$ and j_0 is a suitably chosen integer to be specified in Theorem 1 in Section 3. The location of the supremum of absolute magnitude of these empirical wavelet coefficients will be a reasonable estimator for the location of a sharp change point in the transformed design. Let $Q \subset (0,1)$ be a closed interval such that $\tau \in Q$. Define $\hat{\theta}$, an estimator of the transformed sharp change point $\theta \equiv F(\tau)$, by

$$\hat{\theta} \equiv \inf\{z \in Q : |\hat{\Delta}_{j_0}(z)| = \sup_{t \in Q} |\hat{\Delta}_{j_0}(t)|\}. \tag{4}$$

In the original design, we can construct $\hat{\tau}$, an estimator of τ, by

$$\hat{\tau} \equiv \hat{F}^{-1}(\hat{\theta}). \tag{5}$$

Suppose that $g \in \mathcal{G}_\alpha$. While the result of Wang (1995) holds for small noninteger values of α, $0 \leq \alpha < 1$, we have the following theorem for $\alpha \geq 0$.

Theorem 1 *Suppose that j_0 in the construction (4) of $\hat{\theta}$ satisfies*

$$2^{-j_0} \asymp (n^{-1}(\log n)^\eta)^{1/(1+2\alpha)}$$

where η is any real number greater than one, and the relation $\asymp$ means that the ratios of the two sides are bounded between constants A and B. If F^{-1} satisfies the Lipschitz condition, then we have

$$|\hat{\tau} - \tau| = O_p\left((n^{-1}(\log n)^\eta)^{1/(1+2\alpha)}\right). \tag{6}$$

Raimondo (1998) obtained the minimax rate of convergence, $n^{-1/(1+2\alpha)}$, by considering a two step procedure. The rate of convergence in (6) has penalty in terms of the logarithmic term, however it is fast enough not to affect the rate of convergence of a function estimator (see Theorem 3 in Section 3.2).

Remark 1 The j_0 can be selected practically by comparing $\hat{\Delta}_{j_0}(t)$ in (3) at all levels with the threshold, $C_n = \sigma(2\log n/n)^{\frac{1}{2}}$. See Wang (1995) for the details.

3. Estimation of a regression function with a single jump point

So far, we considered the estimation of a sharp change point when a regression function has either a cusp point or a jump point. From now on,we restrict our attention to only a regression function with a single jump in order to focus on how to avoid the Gibbs phenomenon that the usual wavelet estimators exhibit. Wavelet function estimators perform well around cusp points, but show the Gibbs phenomenon around jump points as in Figure 1 (c). Therefore, we propose an estimator to avoid the Gibbs phenomenon near a jump point while keeping adaptive wavelet cusp estimation.

Assume that the g in (1) has a jump ($\alpha = 0$ in Section 2.1) at τ, $0 < \tau < 1$, and is smooth otherwise in the sense that the function is an element of Besov ball, $B^s_{p,q}(C)$ (see, e.g., DeVore and Popov (1988) for more details). If α' satisfies $s - 1/p \leq \alpha' \leq 1$ in condition (a)–(ii) in Section 2.1, then there exists a function $f \in B^s_{p,q}(C)$ such that the function g can be written as

$$g(x) = f(x) + \Delta 1_{[\tau,1]}(x), \quad 0 \leq x \leq 1, \tag{7}$$

where Δ is a jump size.

The basic idea of the proposed method is as follows. First, we transform a random design to an equispaced one by the transformation described in Section 2.3. In the transformed model, define $Z_i = Y_{[i]} - \Delta 1_{[\theta,1]}(t_i)$ for $i = 1, \ldots, n$. If Z_i's are available, then one can use an ordinary wavelet regression estimator such as the BlockJS in Cai (1999) to estimate f in (7). Suppose that we have estimators of θ and Δ with good performance. If we subtract our jump size estimate from the observations after the jump point estimate, the adjusted data may be considered as observations from a regression function with a smoothness index s. Using the adjusted data, we can estimate the function f by an ordinary wavelet regression estimator. After that, a final estimate can be obtained by adding the jump size estimate to the estimate of f and transforming back to the original scale.

3.1. Preliminary estimators

Let us assume that we have an equispaced design after transforming a random design. The transformed jump point can be estimated by $\hat{\theta}$ defined in (4).

A jump size Δ in (7) can be estimated by (3) in an equispaced design:

$$\hat{\Delta} \equiv \hat{\Delta}(\hat{\theta}) = 2^{j_1/2}\hat{\Delta}_{j_1}(\hat{\theta}) \Big/ \int_0^{N+1} \psi(u)du. \tag{8}$$

Proposition 2 *Let s be a smoothness index of Besov ball $B^s_{p,q}(C)$ and assume that a mother wavelet ψ is r-regular with $r \geq s$. If $r > 1/2$ and $j_1 \equiv j_1(n)$ satisfies $2^{j_1(1+2s)}/n \to \infty$, then we have*

$$(n/2^{j_1})^{\frac{1}{2}} \left(\hat{\Delta} - \Delta\right) \longrightarrow N\left(0, \sigma^2 \Big/ \left\{\int_0^{N+1} \psi(u)du\right\}^2\right),$$

and

$$|\hat{\Delta} - \Delta| = o_p\left(n^{-\frac{s}{1+2s}}\right). \tag{9}$$

This result can be derived in the process of proving Theorem 1. This proposition provides asymptotic confidence intervals for the size of the jump. These intervals enable us to distinguish a jump point from cusps by checking if intervals include 0 (e.g., see Müller (1992)). When the asymptotic confidence intervals are constructed, σ can be estimated by wavelet coefficients, for example, see Donoho and Johnstone (1994).

3.2. Estimation of a function

Since we have the preliminary estimators of a jump point τ (θ in the transformed model) and a jump size Δ defined in (4), (5) and (8), we obtain the adjusted data by

$$Y^*_{[i]} = Y_{[i]} - \hat{\Delta} 1_{[\hat{\theta},1]}(t_i), \quad 1 \leq i \leq n.$$

Given $Y^*_{[i]}$'s, we apply a binning method to make the number of transformed design points a power of 2. Put $\log_2 m = \lfloor \log_2 n \rfloor$, which is the greatest integer that does not exceed $\log_2 n$. For $1 \leq l \leq m$, let

$v_l = (l - \frac{1}{2})/m$ and

$$W_l = \frac{\sum_{i=1}^n Y^*_{[i]} 1_{((l-1)/m, l/m]}(t_i)}{\sum_{i=1}^n 1_{((l-1)/m, l/m]}(t_i)}.$$

Given (v_l, W_l) for $1 \le l \le m$, we construct a block thresholding estimator $\hat{\gamma}$ of the mean function $\gamma \equiv f(F^{-1})$ as follows:

$$\hat{\gamma}(x) = \sum_{k=0}^{2^{j_2}-1} \hat{a}_{j_2,k}\phi_{j_2,k}(x) + \sum_{j=j_2}^{J-1}\sum_{k=0}^{2^j-1} \tilde{b}_{j,k}\psi_{j,k}(x), \tag{10}$$

where

$$\begin{aligned}
\hat{a}_{j_2,k} &= \frac{1}{m}\sum_{l=1}^m W_l \phi_{j_2,k}(v_l), \quad \hat{b}_{j,k} = \frac{1}{m}\sum_{l=1}^m W_l\psi_{j,k}(v_l), \\
\tilde{b}_{j,k} &= \left(1 - \lambda L \frac{\sigma^2}{m} \Big/ S^2_{(jb)}\right)_+ \hat{b}_{j,k}, \quad \lambda = 4.50524, \quad L = \lfloor \log m \rfloor, \\
(jb) &= \{(j,k) : (b-1)L \le k \le bL - 1\}, \quad S^2_{(jb)} = \sum_{k \in (jb)} \hat{b}^2_{j,k},
\end{aligned}$$

j_2 is a fixed constant, and $J = \lfloor \log_2 m \rfloor$. Here, σ can be estimated by wavelet coefficients (see Donoho and Johnstone (1994)). Then, the proposed regression function estimator of g is

$$\begin{aligned}
\hat{g}(x) &= \hat{\gamma}(\hat{F}(x)) + \hat{\Delta}1_{[\hat{\theta},1]}(\hat{F}(x)) \\
&= \hat{f}(x) + \hat{\Delta}1_{[\hat{\tau},1]}(x),
\end{aligned}$$

where $\hat{F}$ is a linear map in Section 2.3 and $\hat{f} \equiv \hat{\gamma}(\hat{F})$.

Kang, Koo, and Park (2000) applied this translation idea to an ordinary kernel estimator, but since we consider a regression function in Besov spaces, it cannot achieve the rate of convergence in Theorem 3 (see Donoho and Johnstone (1998) for more details). Furthermore, it tends to oversmooth a function around cusp points in practice, which has already been verified in several contexts (e.g., see Chapter 10 of Härdle et al. (1998)). Therefore, the proposed estimator not only maintains the adaptive properties of wavelet estimators around cusp points, but also gives a better performance when a regression function has jump points.

Remark 2 To avoid the boundary problems, we use the complete orthonormal system on $[0, 1]$ proposed by Cohen, Daubechies, and Vial (1993). It consists of interior scaling and wavelet functions, and boundary scaling and wavelet functions. In this subsection and Section 6.2, for any integer j and k, $\phi_{j,k}(x) = 2^{j/2}\phi(2^j x - k)$ denote both interior and boundary scaling functions, and $\psi_{j,k}(x) = 2^{j/2}\psi(2^j x - k)$ denote both interior and boundary wavelet functions without any confusion.

Remark 3 The block thresholding estimator used in this paper was proposed by Cai (1999), called the BlockJS. It is known to improve the rate of convergence compared to the term-by-term thresholding estimators (see, e.g., Donoho and Johnstone (1994)).

MISE of the proposed estimator is given by the following theorem. This theorem shows that the proposed estimator results in the same rate of convergence as in Cai (1999) even if a regression function has a jump point.

Theorem 3 *Let us assume that a mother wavelet ψ is r-regular with $r \geq s$. If F^{-1} satisfies the Lipschitz condition, then, for all $C > 0$, $\frac{1}{2} < s < r$, and $1 \leq q \leq \infty$, we have*

$$\sup_g \; E \; \left(\int_0^1 (\hat{g}(x) - g(x))^2 dx \right) \sim \begin{cases} n^{-\frac{2s}{1+2s}} & \text{for} \quad p \geq 2 \\ n^{-\frac{2s}{1+2s}} (\log n)^{\frac{2-p}{p(1+2s)}} & \text{for} \quad 1 \leq p < 2 \quad \text{and} \quad sp \geq 1, \end{cases}$$

where $\sup_g$ *means* $\sup_{f \in B^s_{p,q}(C)}$ *with f being the smooth part of g as in (7).*

Now, we investigate a local property of the proposed estimator. Assume f in (7) is an element of the local Hölder class, $\Lambda^s(K, t_0, \delta)$, which can be found in Cai (1999). MSE of the proposed estimator is derived in the following theorem and it shows that Cai's result still holds in this case.

Theorem 4 *Let $x_0 \in (0, 1)$ be fixed. Then, under the assumptions of Theorem 3,we have*

$$\sup_g E(\hat{g}(x_0) - g(x_0))^2 \leq C \cdot \left(\frac{\log n}{n} \right)^{\frac{2s}{1+2s}},$$

where $\sup_g$ *means* $\sup_{f \in \Lambda^s(K, t_0, \delta)}$ *with f being the smooth part of g as in (7).*

4. Numerical results

To investigate the practical performance of the proposed estimator defined in Section 2 and 3, a limited simulation study is carried out. For the simulation study, n response-predictor pairs (X_i, Y_i) are generated according to the prescription

$$Y_i = g(X_i) + \epsilon_i, \quad i = 1, \cdots, n,$$

where X_i's are drawn from a uniform distribution and ϵ_i's are drawn from a normal distribution with mean 0 and standard deviation σ. The sample size for the simulated examples is chosen to be $n = 1000$.

The first regression function (see Fig 1 (a)) is given by

$$g_1(x) = \begin{cases} 4x^2(3 - 4x) & \text{for } 0 \leq x \leq \frac{1}{2}, \\ \frac{4}{3}x(4x^2 - 10x + 7) - \frac{3}{2} & \text{for } \frac{1}{2} < x \leq \frac{3}{4}, \\ \frac{16}{3}x(x-1)^2 & \text{for } \frac{3}{4} < x \leq 1. \end{cases}$$

In this case, g_1 has a jump of size 0.5 at $x = 0.5$. Gaussian white noise with $\sigma = 0.15$ is added to produce the simulated data (see Fig 1 (b)).

The second regression function has both a jump and a cusp, and is given by

$$g_2(x) = 2 - 2|x - 0.26|^{1/5}1(0 \le x \le 0.26) - 2|x - 0.26|^{3/5}1(0.26 < x \le 1) + 1(0.78 \le x \le 1),$$

which appeared in Wang (1995). In this case, g_2 has an unbalanced cusp at $x = 0.26$ and a jump of size 1 at $x = 0.78$. The standard deviation of the Gaussian noise is $\sigma = 0.2$.

Using the two simulated examples above, we compare the numerical performance of the proposed estimator with those of the BlockJS estimator and the Translation-Invariant (TI) estimator proposed by Coifman and Donoho (1995), which is known to reduce the Gibbs phenomenon. A program for implementing the proposed procedure and the BlockJS has been written in C. The Wavelab toolbox has been used for the TI estimator (see Coifman and Donoho (1995) for the details).

For both examples, we use the same wavelet functions and the same resolution levels. For the proposed sharp change point estimator, the Haar wavelet function is chosen and a resolution level j_0 is taken as 6. We take $j_1 = 6$ and choose the Haar wavelet for the estimation of the jump size. In the function estimation, Daubechies' wavelet db4 and corresponding boundary wavelets are used as orthonormal bases, and j_2 and J are taken as 4 and 9 respectively for the proposed estimator and the BlockJS estimator. For the TI estimator, we use the Haar wavelet functions and apply a hard thresholding rule, which provided the best results in Coifman and Donoho (1995).

We ran the experiment 500 times and Table 1 shows the means of estimated Integral Squared Error (ISE) and their standard errors of the three estimators for the two examples.

Table 1 about here.

In both examples, the proposed estimator outperforms the BlockJS in terms of the bias and the standard error. This illustrates that the proposed method actually reduces the Gibbs phenomenon around a jump point (compare Fig. 1 (c) and (d)). Compared to the TI estimator, it has a smaller bias, but a slightly larger standard error in both examples. However, much smaller biases in both cases confirm that the proposed method reduces the Gibbs phenomenon more efficiently around a jump point.

Since the proposed function estimator depends on the behavior of the jump point estimator, one needs to investigate the performance of the jump point estimator. In the second example, which is a more challenging problem, only 1.8% of 500 experiments selected the cusp point $x = 0.26$ as a jump point. This might elevate the standard error of the proposed function estimator, but it still shows quite good performance.

5. Concluding Remarks

We have established theoretical properties of the sharp change point estimator and the wavelet block thresholding estimator of a regression function with a jump point. We also considered practical aspects of the proposed estimator via numerical study. As in Theorem 1, the proposed sharp change point estimator achieved near-minimax (in terms of the logarithmic term) rate of convergence. In the function estimation, according to a simulation study, the proposed method can actually alleviate the Gibbs phenomenon, while it still has good asymptotic results, as described in Theorem 3 and 4.

We have described a wavelet regression estimator when a regression function has a single jump point. Wang (1995) also proposed a detection procedure in cases where a regression function has multiple discontinuities. The method can be applied when the number of jump points is unknown. The proposed estimator can also be extended to cases with multiple jump points. We will leave this work for a future study.

6. Proofs

Denote C by a generic constant that may vary from place to place through the whole section.

6.1. Proof of Theorem 1

$$\begin{aligned} |\hat{\tau} - \tau| &= |\hat{F}^{-1}(\hat{\theta}) - F^{-1}(\theta)| \\ &\leq |\hat{F}^{-1}(\hat{\theta}) - F^{-1}(\hat{\theta})| + C|\hat{\theta} - \theta|. \end{aligned}$$

Since $\max_i |X_{(i+1)} - X_{(i)}| = O_p(n^{-1})$ and the distributions of $F(X_{(i)})$ are $Beta(i, n-i+1)$ for $1 \leq i \leq n$, the first term is

$$|\hat{F}^{-1}(\hat{\theta}) - F^{-1}(\hat{\theta})| = O_p(n^{-1}). \tag{11}$$

For the second term, decompose $\hat{\Delta}_{j_0}(t)$ in (3) as

$$\begin{aligned} \hat{\Delta}_{j_0}(t) &= \frac{2^{j_0/2}}{n} \sum_{i=1}^{n} \xi(t_i)\psi(2^{j_0/2}(t_i - t)) + \frac{2^{j_0/2}}{n} \sum_{i=1}^{n} e_i \psi(2^{j_0/2}(t_i - t)) \\ &\quad + \frac{2^{j_0/2}}{n} \sum_{i=1}^{n} \left(g(X_{(i)}) - \xi(t_i)\right) \psi(2^{j_0/2}(t_i - t)) \\ &= \frac{2^{j_0/2}}{n} \sum_{i=1}^{n} U_i \psi(2^{j_0/2}(t_i - t)) + \frac{2^{j_0/2}}{n} \sum_{i=1}^{n} \left(g(X_{(i)}) - \xi(t_i)\right) \psi(2^{j_0/2}(t_i - t)), \end{aligned} \tag{12}$$

where $\xi \equiv g(F^{-1})$, e_i's are rearrangement of ϵ_i's, and $U_i \equiv \xi(t_i) + e_i$, for $1 \leq i \leq n$. Since

$$E(X_{(i)}) = F^{-1}(i/n) + O(n^{-1}), \tag{13}$$

and

$$E\left(|X_{(i)} - E(X_{(i)})|^r\right) = O\left(n^{-r/2}\right), \tag{14}$$

for some $r > 0$ (see David and Johnson (1954)), the second part of (12) can be ignored.

A similar procedure of proving Theorem 2 in Wang (1995) can be applied to the first part of (12), and we can obtain

$$|\hat{\theta} - \theta| = O_p\left((n^{-1}(\log n)^{\eta})^{1/(1+2\alpha)}\right),$$

if $2^{-j_0} \asymp (n^{-1}(\log n)^{\eta})^{1/(1+2\alpha)}$ for any real number η greater than one. Thus (6) is proved.

6.2. Proof of Theorem 3

Let $\sum_i 1_{((l-1)/m,l/m]}(t_i) \equiv \nu_l$. Then, $\hat{a}_{j_2,k}$ in (10) can be decomposed as

$$\begin{aligned}
\hat{a}_{j_2,k} &= \frac{1}{m}\sum_{l=1}^{m}\gamma(v_l)\phi_{j_2,k}(v_l) + \frac{1}{m}\sum_{l=1}^{m} e_l\phi_{j_2,k}(v_l) + \frac{1}{m}\sum_{l=1}^{m}\left[\frac{1}{\nu_l}\sum_i f(X_{(i)}) - \gamma(v_l)\right]\phi_{j_2,k}(v_l) \\
&\quad + \frac{1}{m}\sum_{l=1}^{m}\frac{1}{\nu_l}\sum_i \left(\Delta 1_{[\tau,1]}(X_{(i)}) - \hat{\Delta} 1_{[\hat{\theta},1]}(t_i)\right)\phi_{j_2,k}(v_l) \\
&= \frac{1}{m}\sum_{l=1}^{m} Z_l\phi_{j_2,k}(v_l) + \frac{1}{m}\sum_{l=1}^{m} R_l\phi_{j_2,k}(v_l) + \frac{1}{m}\sum_{l=1}^{m} V_l\phi_{j_2,k}(v_l) \\
&\equiv \hat{a}^Z_{j_2,k} + \hat{a}^R_{j_2,k} + \hat{a}^V_{j_2,k},
\end{aligned}$$

where $e_l \equiv 1/\nu_l \sum_i \epsilon_i$, $\gamma \equiv f(F^{-1})$, $Z_l \equiv \gamma(v_l) + e_l$, $R_l \equiv 1/\nu_l \sum_i \left(\Delta 1_{[\tau,1]}(X_{(i)}) - \hat{\Delta} 1_{[\hat{\theta},1]}(t_i)\right)$, and $\sum_i$ represents the summation over $(l-1)n/m < i \le ln/m$, for $1 \le l \le m$. Similarly, $\hat{b}_{j,k}$ can be written as

$$\hat{b}_{j,k} \equiv \hat{b}^Z_{j,k} + \hat{b}^R_{j,k} + \hat{b}^V_{j,k}.$$

To prove Theorem 3, we write

$$E(||\hat{g} - g||_2^2) \le E(||\hat{f} - f||_2^2) + E(||(\hat{\Delta} - \Delta)1_{[\tau,1]}||_2^2) + E(||\hat{\Delta}(1_{[\hat{\tau},1]} - 1_{[\tau,1]})||_2^2). \qquad (15)$$

By (6) and (9),

$$\begin{aligned}
E(||\hat{\Delta}(1_{[\hat{\tau},1]} - 1_{[\tau,1]})||_2^2) &= O\left(n^{-1}(\log n)^{\eta}\right), \\
E(||(\hat{\Delta} - \Delta)1_{[\tau,1]}||_2^2) &= o\left(n^{-\frac{2s}{1+2s}}\right),
\end{aligned}$$

which are negligible.

For the first term of (15), we write $\hat{f}$ as

$$\begin{aligned}
\hat{f}(x) &= \sum_k \hat{a}^Z_{j_2,k}\phi_{j_2,k}(\hat{F}(x)) + \sum_j\sum_k \tilde{b}^Z_{j,k}\psi_{j,k}(\hat{F}(x)) \\
&\quad + \sum_j\sum_k (\tilde{b}^{ZT}_{j,k} - \tilde{b}^Z_{j,k})\psi_{j,k}(\hat{F}(x)) + \sum_k \hat{a}^R_{j_2,k}\phi_{j_2,k}(\hat{F}(x)) + \sum_j\sum_k \tilde{b}^R_{j,k}\psi_{j,k}(\hat{F}(x)) \\
&\quad + \sum_k \hat{a}^V_{j_2,k}\phi_{j_2,k}(\hat{F}(x)) + \sum_j\sum_k \tilde{b}^V_{j,k}\psi_{j,k}(\hat{F}(x)),
\end{aligned}$$

where

$$\begin{aligned}
\tilde{b}^Z_{j,k} &= (1 - \lambda L\sigma^2/mS^{Z^2}_{(jb)})_+\hat{b}^Z_{j,k}, \quad \tilde{b}^{ZT}_{j,k} = (1 - \lambda L\sigma^2/mS^2_{(jb)})_+\hat{b}^Z_{j,k}, \\
\tilde{b}^R_{j,k} &= (1 - \lambda L\sigma^2/mS^2_{(jb)})_+\hat{b}^R_{j,k}, \quad \tilde{b}^V_{j,k} = (1 - \lambda L\sigma^2/mS^2_{(jb)})_+\hat{b}^V_{j,k},
\end{aligned}$$

and $S^{Z^2}_{(jb)} = \sum_{(jb)} \hat{b}^{Z^2}_{j,k}$.

Let

$$\hat{\gamma}^Z(\hat{F}(x)) \equiv \sum_k \hat{a}^Z_{j_2,k}\phi_{j_2,k}(\hat{F}(x)) + \sum_j \sum_k \tilde{b}^Z_{j,k}\psi_{j,k}(\hat{F}(x)),$$

then the first term of (15) can be written as

$$\begin{aligned} E(||\hat{f} - f||_2^2) &= E(||\hat{\gamma}^Z(\hat{F}) - \gamma(\hat{F})||_2^2) + E(||\gamma^Z(\hat{F}) - \gamma(F)||_2^2) \\ &+E(\sum_j \sum_k (\tilde{b}^{ZT}_{j,k} - \tilde{b}^Z_{j,k})^2) + E(\sum_k \hat{a}^{R^2}_{j_2,k}) + E(\sum_j \sum_k \tilde{b}^{R^2}_{j,k}) \\ &+E(\sum_k \hat{a}^{V^2}_{j_2,k}) + E(\sum_j \sum_k \tilde{b}^{V^2}_{j,k}) + (\text{negligible terms}). \end{aligned} \quad (16)$$

The first term of (16) is

$$E(||\hat{\gamma}^Z(\hat{F}) - \gamma(\hat{F})||_2^2) \leq \begin{cases} Cn^{-\frac{2s}{1+2s}} & \text{for} \quad p \geq 2, \\ Cn^{-\frac{2s}{1+2s}}(\log n)^{\frac{2-p}{p(1+2s)}} & \text{for} \quad 1 \leq p < 2 \quad \text{and} \quad sp \geq 1, \end{cases}$$

by Cai (1999) and the second term of (16) is negligible by (11). The fourth and fifth terms of (16) are also negligible if $s > \frac{1}{2}$ by (13) and (14). The other terms of (16) are easy to prove to be negligible by (6) and (9).

We omit a proof of Theorem 4 since it is similar to that of Theorem 3.

Acknowledgement

We would like to thank the referees for suggesting the second regression function and the TI estimator in Section 4.

References

Cai, T. T., 1999. Adaptive wavelet estimation: a block thresholding and oracle inequality approach. Ann. Statist. 27, 898–924.

Cohen, A., Daubechies, I., Vial, P., 1993. Wavelets on the interval and fast wavelet transforms. Appl. Computat. Harmon. Anal. 1, 54–81.

Coifman, R. R., Donoho, D. L., 1995. Translation-invariant de-noising. Antoniadis, A., Oppenheim, G. (eds), Wavelets and Statistics, Lecture Notes in Statistics 103, Springer-Verlag, New-York, 125–150.

Daubechies, I., 1992. Ten Lectures on Wavelets. Society for Industrial and Applied Mathematics.

David, F. N., Johnson, N. L., 1954. Statistical treatment of censored data. Part I. Fundamental formulae. Biometrika. 41, 228–240.

DeVore, R., Popov, V., 1988. Interpolation of Besov spaces. Trans. Amer. Math. Soc. 305, 397–414.

Donoho, D., Johnstone, I. M., 1994. Ideal spatial adaptation by wavelet shrinkage. Biometrika. 81, 425–455.

Donoho, D., Johnstone, I. M., 1998. Minimax estimation via wavelet shrinkage. Ann. Statist. 26, 879–921.

Hall, P., Park, B. U., Turlach, B. A., 1998. A note on design transformation and binning in nonparametric curve estimation. Biometrika. 85, 469–476.

Härdle, W., Kerkyacharian, G., Picard, D., Tsybakov, A., 1998. Wavelets, approximation, and statistical applications. Springer-Verlag, New York.

Kang, K. H., Koo, J. Y., Park, C. W., 2000. Kernel estimation of discontinuous regression functions. Statist. Prob. Lett. 47, 277–285.

Korostelev, A. P., Tsybakov, A. B., 1993. Minimax Theory of Image Reconstruction. Springer-Verlag, New York.

Mallat, S. G., Hwang, W. L., 1992. Singularity detection and processing with wavelets. IEEE Trans. on Information Theory. 38, 617–643.

Nason, G. P., Silverman, B. W., 1994. The discrete wavelet transform in S. J. Comput. Graph. Statist. 3, 163–191.

Raimondo, M., 1998. Minimax estimation of sharp change points. Ann. Statist. 26, 1379–1397.

Wang, Y., 1995. Jump and sharp cusp detection by wavelets. Biometrika. 82, 385–397.

Table 1 Mean (standard error) of ISE's for the two examples ($\times 10^{-4}$)

	Proposed estimator	BlockJS	TI
Polynomial-with-jump	5.5489 (0.1271)	29.5056 (0.7595)	12.9898 (0.0691)
Polynomial-with-jump and cusp	30.5333 (0.4737)	63.2650 (0.8199)	75.0565 (0.2017)

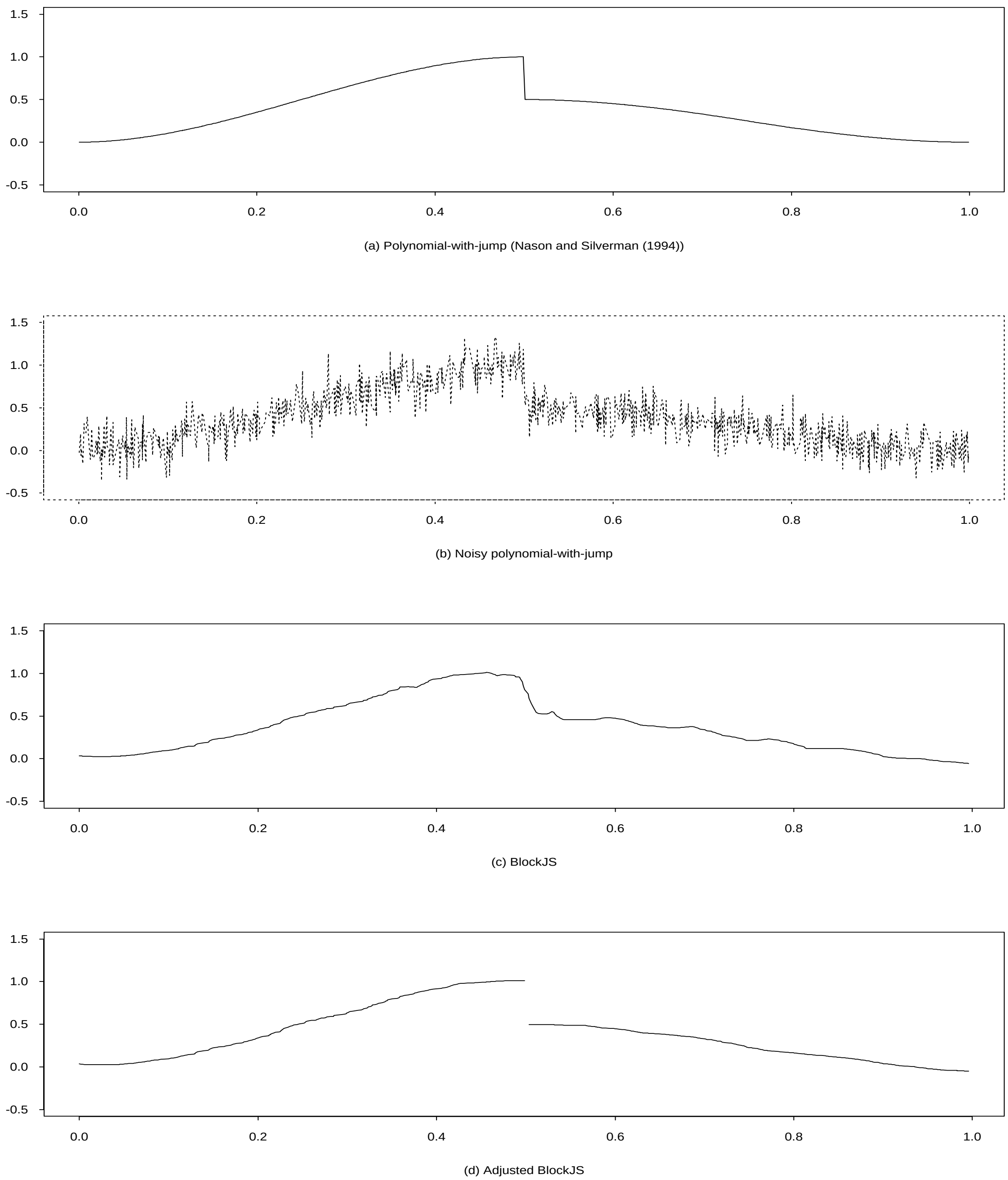

Figure 1: (a) Original signal: Polynomial-with-jump (Nason and Silverman (1994)) and (b) noisy data with $n = 1000$ and $\sigma = 0.15$. The results of the wavelet function estimation by (c) the BlockJS estimator and (d) the proposed estimator. The proposed method can alleviate the Gibbs phenomenon, while the BlockJS estimator still exhibits it.

Figure 2: Integrated "Brownian-path-with-jumps". Original signal (solid line) and its derivative (dotted line) with jumps of size 0.1 at $x = 1/4, 1/2$ and $3/4$ ($n = 1024$).

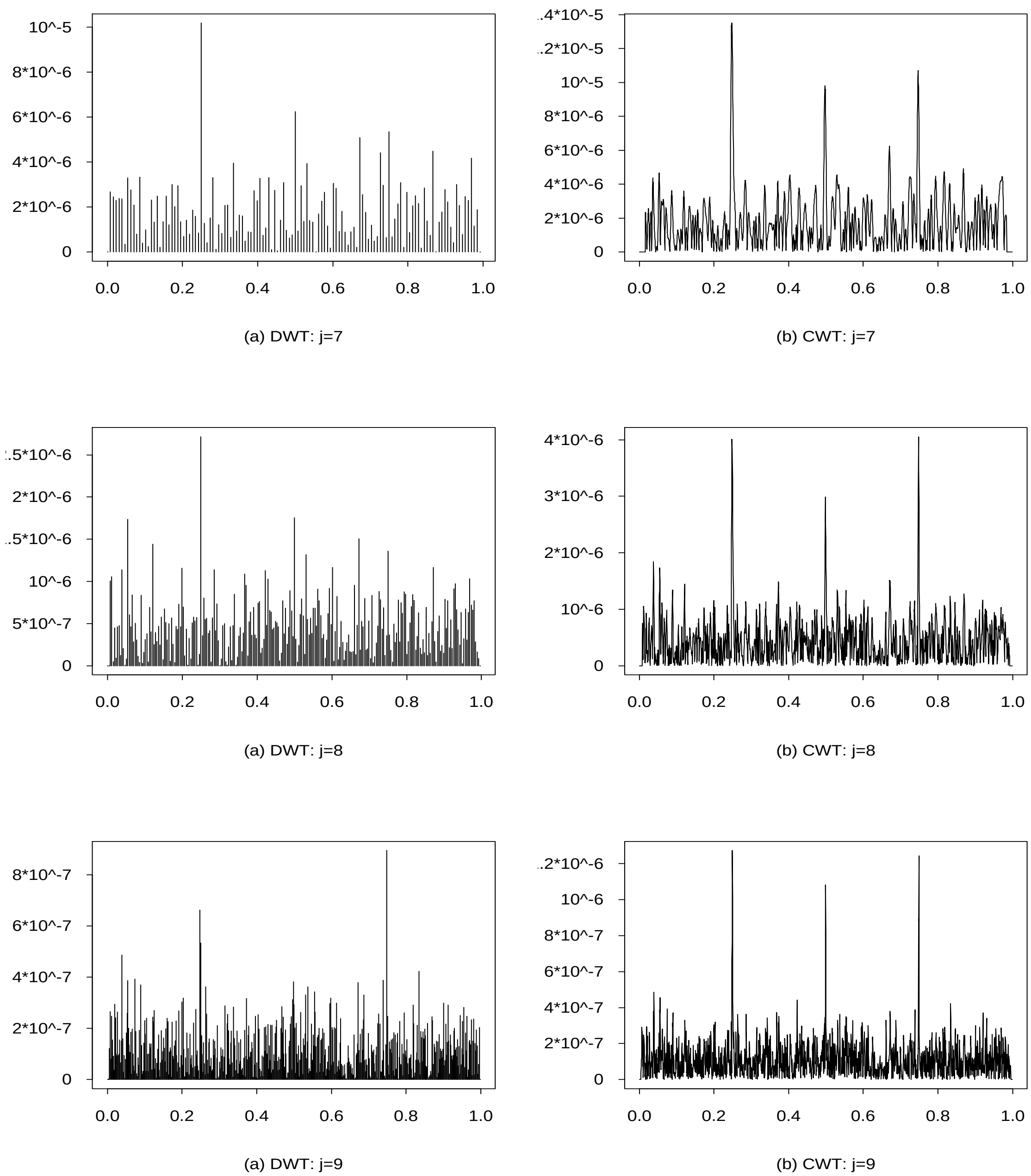

Figure 3: Comparison of two versions (DWT and CWT) with integrated "Brownian-path-with-jumps" for $j = 7, 8$ and 9.