Estimation of a Regression Function with a Sharp Change Point Using Boundary Wavelets

Cheol-Woo Park, Woo-Chul Kim

Department of Statistics, Seoul National University, Seoul, 151-747, Korea

Abstract

We propose a sharp change point estimator based on the differences between right and left boundary wavelet smoothers. It is constructed by applying a two step procedure to the observed data and has the minimax convergence rate. Next, we estimate the regression function with boundary wavelets in the left and right regions of the estimated jump point separately. This method helps us capture the feature of a discontinuity in practice. Both mean integrated squared error and mean squared error of the estimated function are derived and we then show that these rates of convergence are the same as the case in which a jump point does not exist. Simulated examples demonstrate the improved performance of the proposed methods.

Keywords: Block thresholding; Boundary wavelets; Rate of convergence; Sharp change point problem; Wavelet function estimation.

1. Introduction

In many applications of nonparametric function estimation, one is interested in problems confined to an interval. Both numerical analysis in an interval and image analysis where the domain of interest is the Cartesian product of two intervals, need boundary correction at the edges of the interval. In the wavelet context, to work out boundary problems, several solutions have been proposed and they all correspond to different choices of how to adapt the multiresolution hierarchy to the bounded interval. Some solutions, for example, Cohen, Daubechies and Vial (1993) who use interior and boundary wavelets, give unconditional bases for the $C^l([0,1])$ spaces, with $l < m$ if wavelet functions are elements of C^m.

Consider the problem of estimating a regression function whose domain is restricted to an interval. The response measurements Y_i at fixed design points $x_i = i/n,\ i = 1, \cdots, n$ satisfy

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

where g is a regression function whose domain is an interval $[0,1]$ and ε_i's are independent errors with $E(\varepsilon_i) = 0$, $Var(\varepsilon_i) = \sigma^2 < \infty$. The function g in (1) may be very smooth, or it can have sharp change points such as cusps or jumps.

A wavelet method is well known for its good local adaptivity. Even when dealing with smooth functions which contain discontinuities, wavelet estimators still exhibit good convergence rates (see, e.g., Hall and Patil (1995)). However, in practice, the Gibbs phenomenon still exists, see Fig. 1 (a) in Section 4, for example. We adapt boundary wavelets to solve such a problem. First we consider a method which allows us to accurately estimate the location of an abrupt change and then propose a function estimator which can be adjusted to the estimated jump point.

In this paper, we deal with the detection of a sharp change point and the estimation of a regression function with a single jump point. Recent works related to change point analysis include Müller (1992) and Loader (1996), who adapt kernel smoothers; Wang (1995), Raimondo (1998), and Kim and Park (2003), who adapt empirical wavelet coefficients. In the context of the estimation of regression functions with discontinuities, Müller (1992) and Kang, Koo and Park (2000) use a kernel method, and Kim and Park (2003) use a wavelet method.

Section 2 gives a sharp change point estimator based on the moving average type of boundary wavelets smoothers. As in Raimondo (1998), we use a two step procedure and obtain the convergence rate of the proposed estimator. In Section 3, a wavelet block thresholding estimator is proposed when a regression function has a single jump point. With the estimated jump point, we split the domain of a regression function into two intervals and estimate each curve separately. We then obtain both Mean Integrated Squared Error (MISE) and Mean Squared Error (MSE). Section 4 includes numerical examples that compare the performance of the proposed estimators with existing ones; the proof of asymptotic result is given in Section 5.

2. Sharp change point detection

2.1. Sharp change point problem

We consider a class of functions on $[0, 1]$ with either a single jump point or a 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', $0 < \alpha' \le 1$.

(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) $\mathcal{G}_\alpha$ $(\alpha \ge 1)$ 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 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 Sharp change point detection

In this subsection, a sharp change point estimator is proposed with a two step procedure as in Raimondo (1998). Using spline boundary wavelets which can be found in Chui (1997), we construct a preliminary estimator, and then enhance its properties.

Let Ψ^+ denote the left boundary scaling function (ϕ^+) when $N=0$ or the left boundary wavelet function (ψ^+) when $N \geq 1$ with its support $[0, M_\Psi]$, and Ψ^- denote the right boundary scaling function (ϕ^-) when $N=0$ or the right boundary wavelet function (ψ^-) when $N \geq 1$ with its support $[-M_\Psi, 0]$. From the property of boundary wavelet functions, $\psi^\pm$ satisfy

$$\int x^l \psi^\pm(x)dx = 0, \quad l = 0, 1, \cdots, N-1.$$

Then, test functions searching the location of a sharp change point are given by

$$\hat{\Delta}^l_{j_l,k} = \frac{1}{n}\sum_{(l)} Y_i \Psi^l_{j_l,k}(x_i), \quad l = 0, 1, \tag{2}$$

where $\sum_{(l)}$, $l = 0, 1$ denotes the sum over all odd and even indices $i = 1, \cdots, n$, respectively, and

$$\begin{aligned} \xi(k) &= (k+1)M_\Psi, \\ \Psi^0_{j_0,k} &= \Psi^+_{j_0,\xi(k+1)} - \Psi^-_{j_0,\xi(k)}, \\ \Psi^1_{j_1,k} &= \Psi^+_{j_1,d+\xi(k)} - \Psi^-_{j_1,d-\xi(k)}, \end{aligned} \tag{3}$$

for some integer d, j_0 and j_1 to be specified. To avoid stochastic dependence between the two steps of the procedure, we split the initial sample into two (odd and even) subsamples.

In the first step, the location of the maximum of the absolute value of the $\hat{\Delta}^0_{j_0,k}$ in (2) at a sufficiently large level j_0 will be a reasonable estimator for the location of a sharp change point. Thus, the preliminary estimator $\hat{\tau}^p$ of τ can be constructed as

$$\hat{\tau}^p = M_\Psi(\hat{k}_n + 1.5)/2^{j_0}, \tag{4}$$

where

$$\hat{k}_n = \arg\max_{k=0,\cdots,(2^{j_0}-3M_\Psi)/M_\Psi} |\hat{\Delta}^0_{j_0,k}|.$$

In (3), the index $\xi(k+1)$ might be replaced by $\xi(k)$ for Ψ^+, and then the sharp change point estimator (4) would be redefined as $\hat{\tau}^p = M_\Psi(\hat{k}_n + 1)/2^{j_0}$. The latter seems more natural and intuitive, but, through a limited simulation study, we found that the former is more stable and yields a better result in terms of MSE.

In the second step, we sharpen the value of $\hat{\tau}^p$, with a narrower grid in the neighborhood of this point. Let j_1 be a greater level than j_0 and $d = \left[2^{j_1}\hat{\tau}^p - M_\Psi 2^{j_1-j_0}\right]$, which is the greatest integer that does not exceed $2^{j_1}\hat{\tau}^p - M_\Psi 2^{j_1-j_0}$. Denote U_k, U_{k+1} as the endpoints of the support of $\Psi^+_{j_1,d+\xi(k)}$, $k = -1, \cdots, N_0 - 2$, where $N_0 = 2 \times 2^{j_1-j_0}$ and k_0 is the index which satisfies $\tau \in [U_{k_0}, U_{k_0+1})$. Then one can construct Bernoulli independent random variables using $\hat{\Delta}^1_{j_1,k}$ in (2) for an appropriate choice of levels j_0, j_1 and of constant C_1,

$$\eta_k = 1\{|\hat{\Delta}^1_{j_1,k}(Y)| > C_1 n^{-1/2}\}, \quad k = -1, 0, \cdots, N_0 - 2, \tag{5}$$

which have an abrupt change of parameter at time $k = k_0$. The last step is to determine $\hat{k}_0$ by minimizing the formula:

$$\sum_{k=-1}^{l} \eta_k + \sum_{k=l+1}^{N_0-2} (1 - \eta_k),$$

for $l = -1, 0, \cdots, N_0 - 3$. Then, the final estimator is then given by $\hat{\tau} = U_{\hat{k}_0}$. This achieves the minimax rate of convergence that was shown by Raimondo (1998) with ordinary wavelets.

Theorem 1 *Let $g \in \mathcal{G}_\alpha$ and η satisfies $\alpha < \eta < \alpha'$ where α' is in the definition of $\mathcal{G}_\alpha$ in Section 2.1. Suppose that j_0 and j_1 in the construction of $\hat{\tau}^p$ and $\hat{\tau}$ satisfy*

$$2^{-j_0} \asymp n^{-1/(1+2\eta)} \quad \text{and} \quad 2^{-j_1} \asymp n^{-1/(1+2\alpha)},$$

where the relation $\asymp$ means that the ratios of the two sides are bounded between constants A and B. Then there exists a threshold C_1 in (5), and we have

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

The proof is similar to that of Theorem 4.1 in Raimondo (1998).

The following corollary gives an approximate threshold value to be used to decide a sharp change point. Let

$$S_{j_0}(\tau) = \{k_1 - 1, k_1, k_1 + 1\},$$

where k_1 is the integer such that $\tau \in$ supp $(\Psi^+_{j_0,\xi(k_1+1)})$. Define the threshold by

$$\delta = \frac{\sigma\sqrt{2\log n}}{\sqrt{n}}. \tag{6}$$

Corollary 2 *For $g \in \mathcal{G}_\alpha$, as n goes to infinity,*

1. *for all* $k \notin S_{j_0}(\tau)$, $\quad P\left(|\hat{\Delta}^0_{j_0,k}| \geq \delta\right) \longrightarrow 0.$

2. $P\left(|\hat{\Delta}^0_{j_0,k_1}| \geq \delta\right) \longrightarrow 1.$

This result can be derived in the process of proving Theorem 1. From Corollary 2, a point that makes $|\hat{\Delta}^0_{j_0,k}|$ exceed δ can be considered as a sharp change point.

Remark 1 In the case that a regression function has several sharp change points, the proposed method can be applied with the threshold in (6) and Corollary 2. Suppose g has m sharp change points at τ_i, $i = 1, \cdots, m$, where m is a known integer. If $S_{j_0} = \{k : k = 0, 1, \cdots, (2^{j_0} - 3M_\Psi)/M_\Psi\}$, then the estimate $(\hat{\tau}_1, \cdots, \hat{\tau}_m)$ of $(\tau_1, \cdots, \tau_m)$ can be constructed as follows:

(1) Find the $\hat{k}_{n,1}$ that maximizes $|\hat{\Delta}^0_{j_0,k}|$ for all $k \in S_{j_0}$ and obtain $\hat{\tau}^p_1$ from the equation (4). Application of the second step to $\hat{\tau}^p_1$ yields $\hat{\tau}_1$.

(2) Find the $\hat{k}_{n,2}$ that maximizes $|\hat{\Delta}^0_{j_0,k}|$ for $k \in S_{j_0} \backslash \{\hat{k}_{n,1}, \hat{k}_{n,1} \pm 1\}$ obtain $\hat{\tau}^p_2$ from the equation (4). Application of the second step to $\hat{\tau}^p_2$ yields $\hat{\tau}_2$.

(3) Continue the procedure until $\hat{k}_{n,m}$ that maximizes $|\hat{\Delta}^0_{j_0,k}|$ for $k \in S_{j_0} \backslash \bigcup_{i=1}^{m-1} \{\hat{k}_{n,i}, \hat{k}_{n,i} \pm 1\}$ is found and obtain $\hat{\tau}^p_m$ from the equation (4). Application of the second step to $\hat{\tau}^p_m$ yields $\hat{\tau}_m$.

If m is unknown, the threshold δ in (6) can be used to find the number of sharp change points. The procedure described by (1)–(3) above continues while $|\hat{\Delta}^0_{j_0,k}|$ is not less than δ. A similar idea to estimate several change points and its theoretic properties can be found in Wang (1995).

Remark 2 To deal with an irregular design, including a random design, one can apply a design transformation method in Hall, Park and Turlach (1998). This method converts an irregular design into equispaced one. Kim and Park (2003) dealt with the detection of a sharp change point problem with wavelets in a random design by modifying this method. Theorem 1 is still valid in a random design if a design transformation function is smooth. For more detail, see Kim and Park (2003).

3. Function Estimation

From now on, we restrict our attention to only a regression function with a single jump point. Let us assume that 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). Suppose that a mother wavelet ψ is r-regular with $r > s$, the support of ϕ and ψ is $[-M+1, M]$, and ψ satisfies the order of M moment condition, where M is a positive integer.

The idea of the proposed function estimator is straightforward. Once the estimate of the location of a jump point, $\hat{\tau}$, is given, a function g is estimated separately in the intervals, $[0, \hat{\tau}]$ and $[\hat{\tau}, 1]$. By scale adjustment, each interval can be related to the interval $[0, 1]$, and an ordinary wavelet regression estimator such as BlockJS in Cai (1999) can be used to estimate each function.

For this purpose, we use the complete orthonormal system on $[0, 1]$ proposed by Cohen, Daubechies, and Vial (1993). The M vanishing moments family is adapted in such a way near the edges of the interval that we can obtain unconditional bases for $C^s([0, 1])$. Thus, that system consists of interior scaling and wavelet functions, and boundary scaling and wavelet functions. In this section, 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.

The block thresholding estimator proposed by Cai (1999), called BlockJS, allows threshold decisions to be made more accurately by neighboring empirical coefficients and improves rates of convergence compared to the term-by-term thresholding estimators (see, e.g., Donoho and Johnstone (1994)).

The true function g in (1) can be decomposed into two functions of elements of Besov ball, $B^s_{p,q}(C)$,

$$g(x) = \begin{cases} g_0(x) & \text{for} \quad 0 < x \le \tau, \\ g_1(x) & \text{for} \quad \tau < x \le 1, \end{cases} \tag{7}$$

where

$$g_0(x) \equiv g_0(x; \tau) = \sum_{k=0}^{2^{j_3}-1} a^{0,\tau}_{j_3,k} \frac{1}{\tau} \phi_{j_3,k}\left(\frac{x}{\tau}\right) + \sum_{j=j_3}^{\infty} \sum_{k=0}^{2^j-1} b^{0,\tau}_{j,k} \frac{1}{\tau} \psi_{j,k}\left(\frac{x}{\tau}\right),$$

and

$$g_1(x) \equiv g_1(x; \tau) = \sum_{k=0}^{2^{j_4}-1} a^{1,\tau}_{j_4,k} \frac{1}{1-\tau} \phi_{j_4,k}\left(\frac{x-\tau}{1-\tau}\right) + \sum_{j=j_4}^{\infty} \sum_{k=0}^{2^j-1} b^{1,\tau}_{j,k} \frac{1}{1-\tau} \psi_{j,k}\left(\frac{x-\tau}{1-\tau}\right).$$

In this case, wavelet coefficients are given by

$$a^{0,\tau}_{j_3,k} = \int_0^\tau g_0(x)\phi_{j_3,k}\left(\frac{x}{\tau}\right) dx, \quad b^{0,\tau}_{j,k} = \int_0^\tau g_0(x)\psi_{j,k}\left(\frac{x}{\tau}\right) dx,$$

and

$$a^{1,\tau}_{j_4,k} = \int_\tau^1 g_1(x)\phi_{j_4,k}\left(\frac{x-\tau}{1-\tau}\right) dx, \quad b^{1,\tau}_{j,k} = \int_\tau^1 g_1(x)\psi_{j,k}\left(\frac{x-\tau}{1-\tau}\right) dx.$$

Then, the corresponding BlockJS type estimator is defined as

$$\begin{aligned} \hat{g}(x) &\equiv \hat{g}(x; \hat{\tau}) \equiv \hat{g}_0(x; \hat{\tau}) + \hat{g}_1(x; \hat{\tau}) \\ &= \sum_{k=0}^{2^{j_3}-1} \hat{a}^{0,\hat{\tau}}_{j_3,k} \frac{1}{\hat{\tau}} \phi_{j_3,k}\left(\frac{x}{\hat{\tau}}\right) + \sum_{j=j_3}^{j_5} \sum_{k=0}^{2^j-1} \tilde{b}^{0,\hat{\tau}}_{j,k} \frac{1}{\hat{\tau}} \psi_{j,k}\left(\frac{x}{\hat{\tau}}\right) \end{aligned}$$

$$+ \sum_{k=0}^{2^{j_4}-1} \hat{a}^{1,\hat{\tau}}_{j_4,k} \frac{1}{1-\hat{\tau}} \phi_{j_4,k}\left(\frac{x-\hat{\tau}}{1-\hat{\tau}}\right) + \sum_{j=j_4}^{j_6} \sum_{k=0}^{2^j-1} \tilde{b}^{1,\hat{\tau}}_{j,k} \frac{1}{1-\hat{\tau}} \psi_{j,k}\left(\frac{x-\hat{\tau}}{1-\hat{\tau}}\right)$$

where

$$\begin{aligned}
a^{0,\hat{\tau}}_{j_3,k} &= \frac{1}{n}\sum_{i=1}^{n} Y_i \phi_{j_3,k}\left(\frac{x_i}{\hat{\tau}}\right), \quad \hat{a}^{1,\hat{\tau}}_{j_4,k} = \frac{1}{n}\sum_{i=1}^{n} Y_i \phi_{j_4,k}\left(\frac{x_i-\hat{\tau}}{1-\hat{\tau}}\right), \\
\tilde{b}^{l,\hat{\tau}}_{j,k} &= \left(1 - \lambda L_l \frac{\sigma^2}{n} \Big/ S^{l2}_{(jb)^l}\right)_+ \hat{b}^{l,\hat{\tau}}_{j,k}, \qquad l = 0, 1, \\
\hat{b}^{0,\hat{\tau}}_{j,k} &= \frac{1}{n}\sum_{i=1}^{n} Y_i \psi_{j,k}\left(\frac{x_i}{\hat{\tau}}\right), \quad \hat{b}^{1,\hat{\tau}}_{j,k} = \frac{1}{n}\sum_{i=1}^{n} Y_i \psi_{j,k}\left(\frac{x_i-\hat{\tau}}{1-\hat{\tau}}\right), \\
\lambda &= 4.50524, \quad L_0 = [\log n\hat{\tau}], \quad L_1 = [\log n(1-\hat{\tau})], \\
(jb)^l &= \{(j,k) : (b-1)L_l \le k \le bL_l - 1\}, \quad S^{l2}_{(jb)^l} = \sum_{k \in (jb)^l} (\hat{b}^{l,\hat{\tau}}_{j,k})^2, \quad l = 0, 1,
\end{aligned}$$

j_3 and j_4 are fixed constants, $j_5 = [\log_2 n\hat{\tau}] - 1$, and $j_6 = [\log_2 n(1-\hat{\tau})] - 1$. Here, σ can be estimated by wavelet coefficients (see Donoho and Johnstone (1994)).

The MISE of the proposed function estimator is given by the following theorem. This theorem shows that the proposed estimator results in the same rate of convergence in Cai (1999) even though a regression function is not continuous.

Theorem 3 *Let $M > \max(s, 1)$. If ϕ satisfies Lipchitz condition of order 1, then, for all $C > 0$ and $1 \le q \le \infty$, we have*

$$\begin{aligned}
&\sup_g \; E \int_0^1 (\hat{g}(x) - g(x))^2 dx \\
&\sim \begin{cases} n^{-\frac{2s}{1+2s}} & \text{for} \quad p \ge 2 \\ n^{-\frac{2s}{1+2s}} (\log n)^{\frac{2-p}{p(1+2s)}} & \text{for} \quad 1 \le p < 2 \quad \text{and} \quad sp \ge 1. \end{cases}
\end{aligned}$$

where $\sup_g$ *means* $\sup_{g_0, g_1 \in B^s_{p,q}(C)}$.

Remark 3 Müller (1992) used a kernel type estimator to estimate each regression function in separate intervals. However, since we consider Besov spaces, it is possible that each regression function, which is separated at a jump point, still has sharp cusp points with smoothness index s. In this case, kernel type estimators cannot achieve the rate of convergence in Theorem 3 (see Donoho and Johnstone (1998) for more details). Also, they tend to oversmooth the function around the 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 better performances when a regression function has jump points.

Now, we investigate the local property of the proposed estimator. Assume g_0 and g_1 in (7) are elements of the local Hölder class, $\Lambda^s(K, t_0, \delta)$, which can be found in Cai (1999). The MSE of the proposed function estimator is described in the following theorem and it shows that Cai's result still holds in this case.

Theorem 4 *Under the assumptions of Theorem 3, we have*

$$\sup_g E(\hat{g}(t_0) - g(t_0))^2 \le C \cdot \left(\frac{\log n}{n}\right)^{\frac{2s}{1+2s}}.$$

where $\sup_g$ *means* $\sup_{g_0, g_1 \in \Lambda^s(K, t_0, \delta)}$.

Remark 4 In the case that one or more points turn out to be sharp change points, we should clarify whether they are jump points or not before applying our method. This procedure can be established by estimating the sizes of the jumps and obtaining their asymptotic distributions. From the asymptotic distributions, we can construct asymptotic confidence intervals of the sizes of the jumps and see if they contain 0. A similar idea to estimate the jump sizes can be found in Müller (1992), and Kim and Park (2003). Once the estimate of the locations of jump points, $\hat{\tau}_1, \hat{\tau}_2, \cdots, \hat{\tau}_{m'}$ where m' is a positive integer, are given, a function is estimated separately in the intervals, $[0, \hat{\tau}_1]$ and $[\hat{\tau}_1, \hat{\tau}_2], \cdots, [\hat{\tau}_{m'}, 1]$.

Remark 5 In the case of an irregular design, it can be converted into equispaced one, using a design transformation referred to in Remark 2. In a transformed design, a regression function can be estimated by the proposed method and which can then be transformed back to the original scale in the last step. Kim and Park (2003) also dealt with the wavelet function estimation problem in a random design and obtained asymptotic results. If a design transformation function is smooth and the smoothness index s is greater than 1/2, then Theorem 3 and 4 are still valid in a random design. See Kim and Park (2003) for more detail.

4. Numerical results

A program for implementing our procedures, as they apply to the data, has been written in C. To investigate the practical performance of the proposed estimators defined in Sections 2 and 3, a limited simulation study is carried out. For this 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 $x_i = i/n$ and ϵ_i's are drawn from a normal distribution with standard deviation σ. The sample size for the simulated examples is chosen to be $n = 1024$.

We compare the numerical performance of the proposed sharp change point estimator with Raimondo's through three examples. Two of them (Polynomial-with-cusp and Heavisine-with-kink) are found in Raimondo (1998) and the last one (Polynomial-with-jump) appears in Nason and Silverman (1994) (see Fig. 1 (a)).

Fig. 1 about here.

Since $N = 0$ in the two polynomial examples and $N = 1$ for the Heavisine example, we use spline boundary wavelets and Daubechies' wavelets (see Daubechies (1992)) corresponding to each value of N for our estimator and Raimondo's, respectively. For resolution levels j_0 and j_1, we use the same values for both estimators. In the first step, the j_0's are taken as 6 for both polynomials and 5 for Heavisine, and the j_1's are taken as 10 for both polynomials and 8 for Heavisine in the second step. For a threshold value of C_1 in (5), we use $\sigma\sqrt{(j_1 - j_0)\log 2}$, which is proposed by Raimondo (1998).

We ran the experiment 500 times and Table 1 shows the estimated MSE's and their standard errors for the three examples.

Table 1 about here.

Simulations show that the proposed estimator yields much smaller MSE's than those of Raimondo through all three examples. In terms of the standard errors, the proposed estimator also yields smaller values than those of Raimondo except the case of Polynomial-with-cusp.

Next, we investigate the numerical performance of the proposed function estimator when a regression function has a single jump point. With Polynomial-with-jump, we compare the proposed estimator with BlockJS, which does not apply any adjustment around a jump point. Daubechies' wavelet db4 and corresponding boundary wavelets are used as orthonormal bases. Resolution levels j_3 and j_4 are taken as 3 while j_5 and j_6 are taken as 8. The estimated regression functions are presented in Fig. 1 (c) and (d). Taking BlockJS as the standard, the proposed estimator performs reasonably well in that it actually reduces the Gibbs phenomenon around a jump point.

In the function estimation, we again ran the experiment 500 times with Polynomial-with-jump. In this experiment, we added the kernel estimator proposed by Müller (1992) to our comparison. With an one-sided kernel and an ordinary kernel, it estimates a jump point and then estimates each regression function in separate intervals. An one-sided kernel funtion $K_{-}(x) = 12x(1-x)(3-5x)$ and a kernel function $K(x) = 3(1-x^2)/4$ were used and a bandwidth $h = 0.1$ was chosen.

Table 2 about here.

Table 2 shows the means of estimated Integral Squared Error (ISE) and their standard errors of the three estimators. Even though the standard error of the kernel estimator is smaller than that of the proposed estimator, the estimated MISE of the proposed estimator is much smaller than that of the kernel estimator. This illustrates that the proposed estimator performs better than the kernel estimator in the case of Polynomial-with-jump. If a regression function has both a jump and a cusp, the MISE of the proposed estimator will be much smaller than that of the kernel estimator because wavelet estimators are more adaptive to sharp cusps. In comparison to BlockJS, although BlockJS has a smaller standard error, it has a much larger bias than that of the proposed estimator. This confirms that the proposed method reduces the Gibbs phenomenon around a jump point.

5. Proof of Theorem 3

Denote C by a generic constant that may vary from place to place through the whole section. To prove Theorem 3, we first write,

$$\begin{aligned}
& E\int_0^1 (\hat{g}(x)-g(x))^2\,dx \\
= \; & E\int_0^{\tau\wedge\hat{\tau}} (\hat{g}(x)-g(x))^2\,dx + E\int_{\tau\wedge\hat{\tau}}^{\tau\vee\hat{\tau}} (\hat{g}(x)-g(x))^2\,dx \\
& +E\int_{\tau\vee\hat{\tau}}^1 (\hat{g}(x)-g(x))^2\,dx \\
= \; & (I)+(II)+(III).
\end{aligned}$$

Term (I).

Term (I) also can be written as

$$\begin{aligned}
& E\int_0^{\tau\wedge\hat{\tau}} (\hat{g}(x)-g(x))^2\,dx \\
\leq \; & E\int_0^{\tau\wedge\hat{\tau}} (\hat{g}_0(x;\hat{\tau})-\hat{g}_0(x;\tau))^2\,dx + E\int_0^{\tau\wedge\hat{\tau}} (\hat{g}_0(x;\tau)-g_0(x))^2\,dx \\
= \; & (1)+(2).
\end{aligned}$$

Term (1).

Again, term (1) can be decomposed as

$$\begin{aligned}
& E\int_0^{\tau\wedge\hat{\tau}} (\hat{g}(x;\hat{\tau})-\hat{g}(x;\tau))^2 dx \\
= \; & E\sum_{k=0}^{2^{j_3}-1}\int_0^{\tau\wedge\hat{\tau}} \left[\hat{a}_{j_3,k}^{0,\hat{\tau}}\frac{1}{\hat{\tau}}\phi_{j_3,k}\left(\frac{x}{\hat{\tau}}\right)-\hat{a}_{j_3,k}^{0,\tau}\frac{1}{\tau}\phi_{j_3,k}\left(\frac{x}{\tau}\right)\right]^2 dx \\
& +E\sum_{j=j_3}^{j_5}\sum_{k=0}^{2^j-1}\int_0^{\tau\wedge\hat{\tau}} \left[\tilde{b}_{j,k}^{0,\hat{\tau}}\frac{1}{\hat{\tau}}\psi_{j,k}\left(\frac{x}{\hat{\tau}}\right)-\tilde{b}_{j,k}^{0,\tau}\frac{1}{\tau}\psi_{j,k}\left(\frac{x}{\tau}\right)\right]^2 dx,
\end{aligned}$$

and negligible terms. Since

$$\hat{a}_{j_3,k}^{0,\hat{\tau}} = \hat{a}_{j_3,k}^{0,\tau} + O_p\left(\frac{1}{n}\right),$$

uniformly in k if ϕ satisfies Lipchitz condition, the first term is

$$
\begin{aligned}
& E\sum_{k=0}^{2^{j_3}-1}\int_0^{\tau\wedge\hat\tau}\left[\hat a_{j_3,k}^{0,\hat\tau}\frac{1}{\hat\tau}\phi_{j_3,k}\left(\frac{x}{\hat\tau}\right)-\hat a_{j_3,k}^{0,\tau}\frac{1}{\tau}\phi_{j_3,k}\left(\frac{x}{\tau}\right)\right]^2dx \\
= \; & E\sum_{k=0}^{2^{j_3}-1}\int_0^{\tau\wedge\hat\tau}\left[\hat a_{j_3,k}^{0,\tau}\left\{\frac{1}{\hat\tau}\phi_{j_3,k}\left(\frac{x}{\hat\tau}\right)-\frac{1}{\tau}\phi_{j_3,k}\left(\frac{x}{\tau}\right)\right\}\right]^2dx+O\left(\frac{1}{n^2}\right) \\
= \; & O\left(\frac{1}{n^2}\right).
\end{aligned}
$$

In the second term, note that

$$
\hat b_{j,k}^{0,\hat\tau} \;=\; O_p(n^{-1/2}),
$$

if $2^j\asymp n^{1/(1+2M)}$ and

$$
\int_0^{\tau\wedge\hat\tau}\left\{\frac{1}{\hat\tau}\psi_{j,k}\left(\frac{x}{\hat\tau}\right)-\frac{1}{\tau}\psi_{j,k}\left(\frac{x}{\tau}\right)\right\}^2dx=O_p\left(2^{3j}|\hat\tau-\tau|^3\right),
$$

for $k=0,\cdots,2^j-1$. Therefore, the second term of Term (1) is

$$
\begin{aligned}
& E\sum_{j=j_3}^{j_5}\sum_{k=0}^{2^j-1}\int_0^{\tau\wedge\hat\tau}\left[\tilde b_{j,k}^{0,\hat\tau}\frac{1}{\hat\tau}\psi_{j,k}\left(\frac{x}{\hat\tau}\right)-\tilde b_{j,k}^{0,\tau}\frac{1}{\tau}\psi_{j,k}\left(\frac{x}{\tau}\right)\right]^2dx \\
\leq \; & 2E\sum_{j=j_3}^{j_5}\sum_{k=0}^{2^j-1}\left[\int_0^{\tau\wedge\hat\tau}\left[\tilde b_{j,k}^{0,\tau}\left\{\frac{1}{\hat\tau}\psi_{j,k}\left(\frac{x}{\hat\tau}\right)-\frac{1}{\tau}\psi_{j,k}\left(\frac{x}{\tau}\right)\right\}\right]^2dx\right. \\
& \left.+\int_0^{\tau\wedge\hat\tau}\left[(\tilde b_{j,k}^{0,\hat\tau}-\tilde b_{j,k}^{0,\tau})\frac{1}{\hat\tau}\psi_{j,k}\left(\frac{x}{\hat\tau}\right)\right]^2dx\right] \\
= \; & o\left(n^{-\frac{2s}{2s+1}}\right)
\end{aligned}
$$

if $M>s$.

Term (2).

$$
\begin{aligned}
& E\int_0^{\tau\wedge\hat\tau}\left(\hat g_0(x;\tau)-g_0(x)\right)^2dx \\
\leq \; & E\int_0^{\tau}\left(\hat g_0(x;\tau)-g_0(x)\right)^2dx \\
= \; & \begin{cases} O\left(n^{-\frac{2s}{1+2s}}\right) & \text{for } p\geq 2 \\ O\left(n^{-\frac{2s}{1+2s}}(\log n)^{\frac{2-p}{p(1+2s)}}\right) & \text{for } 1\leq p<2 \text{ and } sp\geq 1, \end{cases}
\end{aligned}
$$

by Cai (1999).

Term (III).

Similar to Term (I).

Term (II).

Note that

$$
\sup_{x\in(\tau\wedge\hat\tau,\tau\vee\hat\tau)}|\hat g(x)|\leq\sup_x\left|\sum_{k=0}^{2^{j_-}-1}\hat a_{j_-,k}^{*,\hat\tau}\phi_{j_-,k}^*(x;\hat\tau)\right|+\sup_x\left|\sum_{j=j_-}^{j^+}\sum_{k=0}^{2^j-1}\tilde b_{j,k}^{*,\hat\tau}\psi_{j,k}^*(x;\hat\tau)\right|,
$$

where

$$\hat{a}^{*,\hat{\tau}}_{j_-,k} = \begin{cases} \hat{a}^{0,\hat{\tau}}_{j_3,k} & \text{if} \quad \hat{\tau} > \tau, \\ \hat{a}^{1,\hat{\tau}}_{j_4,k} & \text{if} \quad \hat{\tau} \leq \tau, \end{cases}$$

$$\tilde{b}^{*,\hat{\tau}}_{j,k} = \begin{cases} \tilde{b}^{0,\hat{\tau}}_{j,k} & \text{if} \quad \hat{\tau} > \tau, \\ \tilde{b}^{1,\hat{\tau}}_{j,k} & \text{if} \quad \hat{\tau} \leq \tau, \end{cases}$$

$$\phi^*_{j_-,k}(x;\hat{\tau}) = \begin{cases} \frac{1}{\hat{\tau}}\phi_{j_3,k}\left(\frac{x}{\hat{\tau}}\right) & \text{if} \quad \hat{\tau} > \tau, \\ \frac{1}{1-\hat{\tau}}\phi_{j_4,k}\left(\frac{x-\hat{\tau}}{1-\hat{\tau}}\right) & \text{if} \quad \hat{\tau} \leq \tau, \end{cases}$$

$$\psi^*_{j,k}(x;\hat{\tau}) = \begin{cases} \frac{1}{\hat{\tau}}\psi_{j,k}\left(\frac{x}{\hat{\tau}}\right) & \text{if} \quad \hat{\tau} > \tau, \\ \frac{1}{1-\hat{\tau}}\psi_{j,k}\left(\frac{x-\hat{\tau}}{1-\hat{\tau}}\right) & \text{if} \quad \hat{\tau} \leq \tau, \end{cases}$$

$j_- = j_3$ and $j_+ = j_5$ if $\tau < \hat{\tau}$, and $j_- = j_4$ and $j_+ = j_6$ otherwise. Since

$$\sup_{x\in(\tau\wedge\hat{\tau},\tau\vee\hat{\tau})} \left| \sum_{k=0}^{2^{j_-}-1} \hat{a}^{*,\hat{\tau}}_{j_-,k}\phi^*_{j_-,k}(x;\hat{\tau}) \right| = O_p(1),$$

and

$$\sup_{x\in(\tau\wedge\hat{\tau},\tau\vee\hat{\tau})} \left| \sum_{j=j_-}^{j_+} \tilde{b}^{*,\hat{\tau}}_{j,k}\psi^*_{j,k}(x;\hat{\tau}) \right| = O_p(1),$$

if $M > 1$, then we have

$$\begin{aligned} E\int_{\tau\wedge\hat{\tau}}^{\tau\vee\hat{\tau}} (\hat{g}(x) - g(x))^2 dx &= O\left(E|\hat{\tau} - \tau|\right) \\ &= O\left(\frac{1}{n}\right). \end{aligned}$$

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

References

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

Chui, C., 1997. Wavelets: A Mathematical Tool for Signal Processing. Society for Industrial and Applied Mathematics.

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

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

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.

Hall, P., Patil, P., 1995. Formulae for mean integrated squared error of nonlinear wavelet-based density estimators. Ann. Statist. 23, 905–928.

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.

Kim, W. C., Park, C. W., 2003. Wavelet estimation of a regression function with a sharp change point in a random design. submitted.

Loader, C. R., 1996. Change point estimation using nonparametric regression. Ann. Statist. 24, 1667–1678.

Müller, H. G., 1992. Change-points in nonparametric regression analysis. Ann. Statist. 20, 737–761.

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.

	Polynomial-with-cusp	Heavisine	Polynomial-with-jump
Proposed estimator	9.1732 (1.7138)	16.9063 (0.5411)	1.6309 (0.0962)
Raimondo	11.4130 (1.4125)	27.7750 (0.7443)	7.9434 (0.3632)

Table 1: Comparison of two sharp change point estimators. Each cell contains MSE's (their standard errors) for three examples ($\times 10^{-3}$).

	Polynomial-with-jump
Proposed estimator	6.6775 (0.1752)
Kernel etimator	8.7085 (0.1633)
BlockJS	24.4521 (0.0747)

Table 2: Comparison of three function estimators. Each cell contains MISE's (their standard errors) for Polynomial-with-jump ($\times 10^{-4}$).

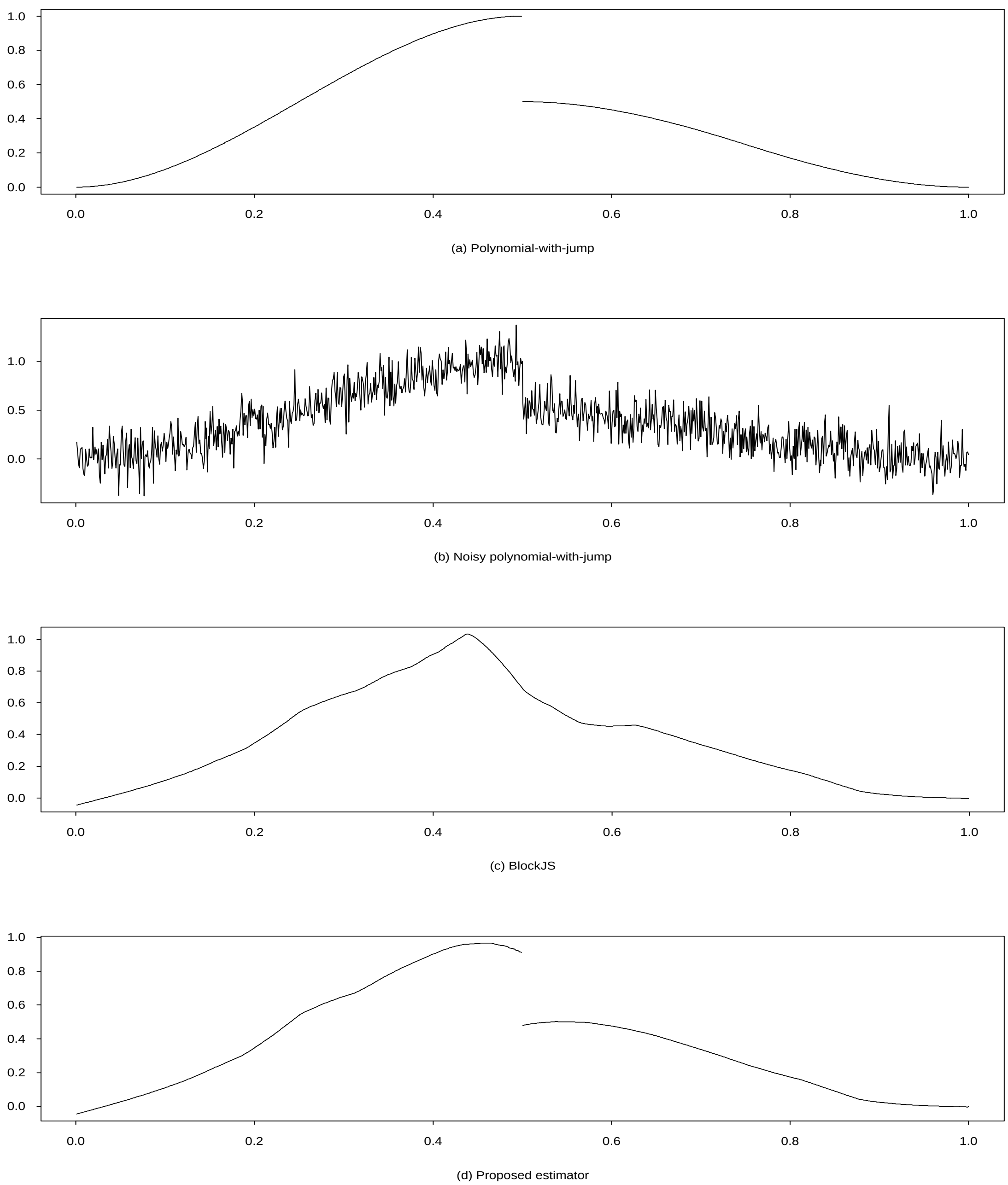

Figure 1: (a) Original signal with a jump point at 0.5. (b) Noisy data with n=1024 and $\sigma = 0.15$. Function estimation by (c) BlockJS and (d) the proposed estimator.