Biometrika (2012), **99**, 1, *pp*. 1–18

Printed in Great Britain
Advance Access publication on 31 July 2012

Supplementary file for "Jump Information Criterion for Statistical Inference in Estimating Discontinuous Curves"

BY ZHIMING XIA
School of Mathematics, Northwest University, Xi'an, Shaanxi, 710127, P. R. China
statxzm@nwu.edu.cn

AND PEIHUA QIU
Department of Biostatistics, University of Florida, Gainesville, Florida, 32608, U.S.A.
pqiu@phhp.ufl.edu

1. SEVERAL LEMMAS

LEMMA 1. *If the kernel function K_j, for $j = l, c, r$, is uniformly Lipschitz-1 continuous, then*

$$\left\|\tilde{h}_j(x,s) - h_j(\tau)\right\|_{L^\infty[-1/2,1/2]} = \mathcal{O}\left(\frac{1}{nh_n}\right) \quad (j = l, c, r),$$

where $\|\cdot\|$ is with respect to $x = s + \tau h_n$ when $\tau \in [-1/2, 1/2]$, s is a given point in $(0, 1)$, $\tilde{h}_j(x,s) = \sum_{i=1}^{n} K_j^*\left(\frac{x_i - x}{h_n}\right) I_{\{x_i > s\}} - I_{\{x>s\}}$ and

$$h_l(\tau) = -I_{\{\tau\in[0,1/2]\}} \int_{-1/2}^{-\tau} K_l^*(u)du,$$

$$h_c(\tau) = I_{\{\tau\in[-1/2,0)\}} \int_{-\tau}^{1/2} K_c^*(u)du - I_{\{\tau\in[0,1/2]\}} \int_{-1/2}^{-\tau} K_c^*(u)du,$$

$$h_r(\tau) = I_{\{\tau\in[-1/2,0)\}} \int_{-\tau}^{1/2} K_r^*(u)du.$$

Proof. The proof is quite straightforward by using the uniform Lipschitz continuity property of the kernel function K_j. It is therefore omitted here. □

LEMMA 2. *Let $g(\tau) = h_r(\tau) - h_l(\tau)$, where $\tau \in [-1/2, 1/2]$, and h_r and h_l are the same as those in Lemma* 1. *Then, under Assumption 2,*

(i) for a given small number $\epsilon > 0$, $g'(\tau)$ is strictly increasing in the interval $[-\epsilon, 0]$, strictly decreasing in the interval $[0, \epsilon]$, and always satisfies $0 < K_r^*(\epsilon) \le |g'(\tau)| \le K_r^*(0)$;

(ii) the function $g(\tau)$ is strictly increasing in the intervals $\left[-\frac{v_{r,2}}{v_{r,1}}, 0\right]$ and $\left[-\frac{v_{l,2}}{v_{l,1}}, 1/2\right]$, and strictly decreasing in the intervals $\left[0, -\frac{v_{l,2}}{v_{l,1}}\right]$ and $\left[-1/2, -\frac{v_{r,2}}{v_{r,1}}\right]$, where $v_{j,k} = \int v^k K_j(v)dv$, for $j = l, r, c, k = 0, 1, 2, 3$;

(iii) the function $g(\tau)$, for $\tau \in [-1/2, 1/2]$, has a unique maximum point at $\tau = 0$ with maximum value 1.

Proof. (i). It can be checked that

$$g'(\tau) = \begin{cases} \dfrac{K_r(-\tau)\,(v_{r,2} + v_{r,1}\tau)}{v_{r,0}v_{r,2} - v_{r,1}^2}, & \tau \in [-1/2, 0), \\ \dfrac{-K_l(-\tau)\,(v_{l,2} + v_{l,1}\tau)}{v_{l,0}v_{l,2} - v_{l,1}^2}, & \tau \in [0, 1/2]. \end{cases} \tag{1}$$

By the triangle inequality, we know that $v_{r,0}v_{r,2} - v_{r,1}^2 > 0$. Therefore, the conclusions are valid by using the properties of $K_r(-\tau)$ and $K_l(-\tau)$.

(ii). This is a direct conclusion of formula (1).

(iii). This is a direct conclusion of the result in (ii). □

LEMMA 3. *Under the assumptions of Theorem 1,*

(i)

$$\|\widehat{a}_j - E(\widehat{a}_j)\|_{L^\infty(\overline{D}_{h_n/2,1\to m_0})} = \mathcal{O}\left\{\left(\frac{\ln n}{nh_n}\right)^{1/2}\right\} \quad (j = c, l, r),$$

(ii)

$$\|E(\widehat{a}_j) - f\|_{L^\infty(\overline{D}_{h_n/2,1\to m_0})} = \mathcal{O}(h_n^2) \quad (j = c, l, r),$$

(iii)

$$\|\widehat{a}_j - f\|_{L^\infty(\overline{D}_{h_n/2,1\to m_0})} = \mathcal{O}\left\{h_n^2 + \left(\frac{\ln n}{nh_n}\right)^{1/2}\right\} \quad (j = c, l, r),$$

$$\|\widehat{a}_j(s_k + \tau h_n) - f(s_k + \tau h_n) - d_k h_j(\tau)\|_{L^\infty[-1/2,1/2]} = \mathcal{O}\left\{h_n^2 + \left(\frac{\ln n}{nh_n}\right)^{1/2}\right\}$$

$$(j = c, l, r),$$

where s_k is the kth jump position.

Proof. (i). The proof of this result is the same of Theorem 3.1 in Gijbels et al. (2007). It is therefore omitted here.

(ii). Without loss of generality, we prove result (iii) in the case when $j = r$. Then

$$\begin{aligned} E\left\{\widehat{a}_r(x)\right\} &= \sum_{i=1}^{n} K_r^*\left(\frac{x_i - x}{h_n}\right) f(x_i) \\ &= \sum_{i=1}^{n} K_r^*\left(\frac{x_i - x}{h_n}\right)\left\{f(x+) + \sum_{j=1}^{\infty}\frac{f^{(j)}(x+)}{j!}(x_i - x)^j\right\} \\ &= f(x+) + \sum_{j=1}^{\infty}\frac{f^{(j)}(x+)}{j!}\sum_{i=1}^{n} K_r^*\left(\frac{x_i - x}{h_n}\right)(x_i - x)^j \\ &= f(x+) + \frac{h_n^2 f^{(2)}(x+)}{2}\frac{v_{r,2}^2 - v_{r,1}v_{r,3}}{v_{r,0}v_{r,2} - v_{r,1}^2} + o(h_n^2). \end{aligned}$$

Because $f^{(2)}(x+)$ is assumed to be uniformly bounded in $\overline{D}_{h_n/2,1\to m_0}$, the result (ii) can be obtained from the above formula.

(iii). First, we prove the first formula in (iii). By the triangle inequality and the results (i) and (ii), we have

$$\begin{aligned}\|\widehat{a}_j - f\|_{L^\infty(\overline{D}_{h_n/2,1\to m_0})} &\le \|\widehat{a}_j - E\widehat{a}_j\|_{L^\infty(\overline{D}_{h_n/2,1\to m_0})} + \|E\widehat{a}_j - f\|_{L^\infty(\overline{D}_{h_n/2,1\to m_0})} \quad (2)\\ &= \mathcal{O}\left\{h_n^2 + \left(\frac{\ln n}{nh_n}\right)^{1/2}\right\} \quad (j = c, l, r).\end{aligned}$$

Now, we prove the second formula in (iii). Let $Y_i' = Y_i - d_k I_{\{x_i > s_k\}} = f(x_i) - d_k I_{\{x_i>s_k\}} + \varepsilon_i$, for $x_i \in D_{h_n/2,k}$. The one-sided kernel estimator based on the new data $\{Y_i'\}$ is defined as follows

$$\begin{aligned}\widehat{a}_j'(x) &= \sum_{i=1}^n Y_i' K_j^*\left(\frac{x_i - x}{h_n}\right)\\ &= \sum_{i=1}^n \left\{f(x_i) - d_k I_{\{x_i>s_k\}} + \varepsilon_i\right\} K_j^*\left(\frac{x_i - x}{h_n}\right)\\ &= \widehat{a}_j(x) - d_k \sum_{i=1}^n K_j^*\left(\frac{x_i - x}{h_n}\right) I_{\{x_i>s_k\}}, \quad j = l, r, c, \quad x \in D_{h_n/2,k}.\end{aligned}$$

Because $f(x) - d_k I_{\{x>s_k\}}$ does not have any jumps in the region $D_{h_n/2,k}$, by the formula (2), we have

$$\begin{aligned}\left\|\widehat{a}_j'(x) - \left(f(x) - d_k I_{\{x>s_k\}}\right)\right\|_{L^\infty(D_{h_n/2,k})} &= \left\|\widehat{a}_j(x) - f(x) - d_k \tilde{h}_j(x, s_k)\right\|_{L^\infty(D_{h_n/2,k})}\\ &= \mathcal{O}\left(h_n^2 + \left(\frac{\ln n}{nh_n}\right)^{1/2}\right),\\ &\quad j = r, l; \quad k = 1, \ldots, m_0\end{aligned}$$

almost surely, where $\tilde{h}_j(x, s_k) = \sum_{i=1}^n K_j^*\left(\frac{x_i - x}{h_n}\right) I_{\{x_i>s_k\}} - I_{\{x>s_k\}}$.

By combining this result with the ones in Lemma 1, we have

$$\begin{aligned}&\|\widehat{a}_j(s_k + \tau h_n) - f(s_k + \tau h_n) - d_k h_j(\tau)\|_{L^\infty[-1/2,1/2]}\\ \le &\left\|\widehat{a}_j(s_k + \tau h_n) - f(s_k + \tau h_n) - d_k \tilde{h}_j(x, s_k)\right\|_{L^\infty[-1/2,1/2]}\\ &+ d_k \left\|\tilde{h}_j(x, s_k) - h_j(\tau)\right\|_{L^\infty[-1/2,1/2]}\\ \le &\,\mathcal{O}\left\{h_n^2 + \left(\frac{\ln n}{nh_n}\right)^{1/2} + \frac{1}{nh_n}\right\} = \mathcal{O}\left\{h_n^2 + \left(\frac{\ln n}{nh_n}\right)^{1/2}\right\} \quad (j = c, l, r). \qquad \square\end{aligned}$$

LEMMA 4. *Under the same assumptions of Theorem 1,*

$$\lim_{n\to\infty} \mathrm{pr}\left(D_{h_n/2,i} \subset \widehat{D}_{\epsilon h_n,i}\right) = 1 \quad (i = 1, \ldots, m_0),$$

where $\epsilon > 1/2$ is a constant.

Proof. For an arbitrary positive number $\epsilon > 1/2$, by the first formula in Theorem 2, we have

$$|\widehat{s}_i(m) - s_i| < (\epsilon - 1/2)\delta_n < (\epsilon - 1/2)h_n, \quad (i = 1, \ldots, m_0), \tag{3}$$

almost surely when n is large enough, where $\delta_n = \{(h_n \ln n)/n\}^{1/2-\delta}$, for $\delta \in (0, 1/2)$, is defined in Theorem 2. Expression (3) implies that

$$\widehat{s}_i(m) - \epsilon h_n < s_i - h_n/2, \;\; \widehat{s}_i(m) + \epsilon h_n > s_i + h_n/2,$$

almost surely when n is large enough. Thus,

$$[\widehat{s}_i(m) - \epsilon h_n, \widehat{s}_i(m) + \epsilon h_n] \supset [s_i - h_n/2, s_i + h_n/2]. \qquad \square$$

LEMMA 5. *Let $\widehat{\Delta}_n(m) = JIC(m) - JIC(m_0)$ and*

$$\Delta_n(m) = \begin{cases} n h_n C \sum_{j=m+1}^{m_0} d_j^2 + P(n) \sum_{j=m+1}^{m_0} \dfrac{1}{|d_i|^{\gamma}}, & m < m_0, \\ C_0 \left(n h_n \ln n\right)^{1/2} + (m - m_0) C_m P(n) \left(\dfrac{n h_n}{\ln n}\right)^{\gamma/2}, & m \geq m_0, \end{cases} \tag{4}$$

where $C_0, C_m > 0$ are positive constants, C_m depends only on m, $\gamma \geq 0$, and

$$C = \int_{-1/2}^{0} \left\{ \int_{-\tau}^{1/2} K_c^*(u) \mathrm{d}u \right\}^2 \mathrm{d}\tau + \int_{0}^{1/2} \left\{ \int_{-1/2}^{-\tau} K_c^*(u) \mathrm{d}u \right\}^2 \mathrm{d}\tau. \tag{5}$$

Then, under the assumptions of Theorem 3, we have almost surely

$$\sup_{m \geq 0} \left| \widehat{\Delta}_n(m) - \Delta_n(m) \right| = \mathcal{O}\left(R_n\right), \tag{6}$$

where $R_n = n h_n^2 + P(n)\{\ln n/(n h_n)\}^{1/2}$.

Proof. To study the properties of JIC(m), we first notice that

$$\begin{aligned} \mathrm{SSR}(m) &= \sum_{i=1}^{n} \left\{ Y_i - \widehat{f}_m(x_i) \right\}^2 \\ &= \sum_{i=1}^{n} \left\{ \widehat{f}_m(x_i) - f(x_i) \right\}^2 + \sum_{i=1}^{n} \varepsilon_i^2 - 2 \sum_{i=1}^{n} \varepsilon_i \left\{ \widehat{f}_m(x_i) - f(x_i) \right\} \\ &= B_1(m) + B_2 + B_3(m). \end{aligned}$$

When $m < m_0$, by the expression (28) and similar arguments to those when discussing A_{31} in the proof of Theorem 2, we have

$$\begin{aligned} B_1(m) = &\sum_{j=1}^{m} \sum_{x_i \in D_{h_n/2,j}} \left\{ \widehat{f}_m(x_i) - f(x_i) \right\}^2 + \sum_{j=m+1}^{m_0} \sum_{x_i \in D_{h_n/2,j}} \left\{ \widehat{f}_m(x_i) - f(x_i) \right\}^2 \\ &+ \sum_{x_i \in \overline{D}_{h_n/2,1 \to m_0}} \left\{ \widehat{f}_m(x_i) - f(x_i) \right\}^2 = B_{11}(m) + B_{12}(m) + B_{13}(m), \end{aligned}$$

where

$$\begin{aligned}
B_{11}(m) &= \sum_{j=1}^{m} \sum_{x_i \in D_{h_n/2,j}} \left\{ \mathcal{O}\left(\frac{\ln n}{nh_n}\right) + d_j^2 \left(I_{\{x > \widehat{s}_j(m)\}} - I_{\{x > s_j\}} \right)^2 \right\} \\
&= \mathcal{O}\left(\ln n\right) + \mathcal{O}\left\{ (nh_n \ln n)^{1/2} \right\} = \mathcal{O}\left\{ (nh_n \ln n)^{1/2} \right\}, \\
B_{12}(m) &= \sum_{j=m+1}^{m_0} \sum_{x_i \in D_{h_n/2,j}} \left\{ d_j^2 h_c^2 \left(\frac{x - s_j}{h_n}\right) + \mathcal{O}\left(\frac{\ln n}{nh_n}\right) \right\} \\
&= nh_n C \sum_{i=m+1}^{m_0} d_j^2 + \mathcal{O}\left(\ln n\right), \\
B_{13}(m) &= \sum_{x_i \in \overline{D}_{h_n/2,1 \to m_0}} \mathcal{O}\left(\frac{\ln n}{nh_n}\right) = \mathcal{O}\left(\frac{\ln n}{h_n}\right).
\end{aligned}$$

The constant C above is defined in (5).

We can discuss the case $m \geq m_0$ similarly. So, we have the following result:

$$B_1(m) = \begin{cases} nh_n C \displaystyle\sum_{i=m+1}^{m_0} d_j^2 + \mathcal{O}\left\{ (nh_n \ln n)^{1/2} \right\}, & m < m_0 \\ \mathcal{O}\left\{ (nh_n \ln n)^{1/2} \right\}, & m \geq m_0. \end{cases} \tag{7}$$

Now,

$$\text{SSR}(m) - \text{SSR}(m_0) = \{B_1(m) - B_1(m_0)\} - 2\sum_{i=1}^{n} \varepsilon_i w_i, \tag{8}$$

where $w_i = \widehat{f}_m(x_i) - \widehat{f}_{m_0}(x_i)$. The first term $B_1(m) - B_1(m_0)$ on the right-hand-side of (8) can be handled by (7). Next, we focus on the second term $\sum_{i=1}^{n} \varepsilon_i w_i$. In the case when $m < m_0$, by (24) we have

$$\begin{aligned}
w_i &= \left\{ \widehat{f}_m(x_i) - f(x_i) \right\} - \left\{ \widehat{f}_{m_0}(x_i) - f(x_i) \right\} \\
&= \sum_{j=m+1}^{m_0} \widehat{d}_j \left\{ \sum_{k=1}^{n} I_{\{x_k > \widehat{s}_j(m)\}} K_c^* \left(\frac{x_k - x_i}{h_n}\right) - I_{\{x_i > \widehat{s}_j(m)\}} \right\} \\
&= d_j h_c(\tau_i) + \mathcal{O}\left\{ \left(\frac{\ln n}{nh_n}\right)^{1/2} \right\} - d_j \left\{ I_{\{\tau_i > (\widehat{s}_j(m) - s_j)/h_n\}} - I_{\{\tau_i > 0\}} \right\},
\end{aligned}$$

where $x_i = s_j + \tau_i h_n/2 \in D_{h_n/2,j}, \tau_i \in [-1/2, 1/2]$, and $j = m+1, \ldots, m_0$. In the above expression, we have used the result that $w_i = 0$ when $x_i \in \overline{D}_{h_n/2,m+1 \to m_0}$, which can be checked easily by some calculations of $\widehat{f}_m(x_i), \widehat{f}_{m_0}(x_i)$. Correspondingly,

$$\sum_{i=1}^{n} \varepsilon_i w_i = \sum_{j=m+1}^{m_0} \left(F_{j1} + F_{j2} + F_{j3}\right),$$

where

$$
\begin{aligned}
F_{j1} &= d_j \sum_{\tau_i \in [-/2,1/2]} h_c(\tau_i)\varepsilon_i, \\
F_{j2} &= \sum_{\tau_i \in [-/2,1/2]} \mathcal{O}\left\{ \left(\frac{\ln n}{nh_n} \right)^{1/2} \right\} \varepsilon_i, \\
F_{j3} &= -d_j \sum_{\tau_i \in [-/2,1/2]} \left\{ I_{\{\tau_i > (\widehat{s}_j(m) - s_j)/h_n\}} - I_{\{\tau_i > 0\}} \right\} \varepsilon_i.
\end{aligned}
$$

By the triangle inequality,

$$
\begin{aligned}
|F_{j1}| &= \mathcal{O}\left\{ (nh_n)^{1/2} \right\}, \\
|F_{j2}| &\le \left\{ \sum_{\tau_i \in [-/2,1/2]} \mathcal{O}\left(\frac{\ln n}{nh_n} \right) \right\}^{1/2} \left(\sum_{\tau_i \in [-/2,1/2]} \varepsilon_i^2 \right)^{1/2} = \mathcal{O}\left\{ (nh_n \ln n)^{1/2} \right\}, \\
|F_{j3}| &\le |d_j| \left[\sum_{\tau_i \in [-/2,1/2]} \left\{ I_{\{\tau_i > (\widehat{s}_j(m) - s_j)/h_n\}} - I_{\{\tau_i > 0\}} \right\}^2 \right]^{1/2} \left(\sum_{\tau_i \in [-/2,1/2]} \varepsilon_i^2 \right)^{1/2} \\
&= \mathcal{O}\left\{ (nh_n \ln n)^{1/2} \right\}
\end{aligned}
$$

uniformly and almost surely. Thus, almost surely

$$
\sum_{i=1}^{n} \varepsilon_i w_i = \mathcal{O}\left\{ (nh_n \ln n)^{1/2} \right\}. \tag{9}
$$

In the case when $m \ge m_0$, we have

$$
\begin{aligned}
w_i &= \sum_{j=m_0+1}^{m} \widehat{d}_j \left\{ \sum_{k=1}^{n} I_{\{x_k > \widehat{s}_j(m)\}} K_c^* \left(\frac{x_k - x_i}{h_n} \right) - I_{\{x_i > \widehat{s}_j(m)\}} \right\} \\
&= \mathcal{O}\left\{ \left(\frac{\ln n}{nh_n} \right)^{1/2} \right\} h_c \left\{ \tau_i | \widehat{s}_j(m) = s_j^* \right\} + \mathcal{O}\left\{ \left(\frac{\ln n}{nh_n} \right)^{1/2} \right\},
\end{aligned}
$$

where $\widehat{s}_j(m)$, for $j = m_0 + 1, \ldots, m$, are $m - m_0$ spurious jumps and s_j, for $j = m_0 + 1, \ldots, m$, are $m - m_0$ points in $\overline{D}_{h_n/2, 1 \to m_0}$. It can also be checked in this case that $w_i = 0$ when $x_i \in \overline{D}_{h_n/2, m_0+1 \to m}$. Therefore, we have

$$
\sum_{i=1}^{n} \varepsilon_i w_i = \sum_{j=m_0+1}^{m} (G_{j1} + G_{j2}),
$$

where

$$
\begin{aligned}
G_{j1} &= \sum_{\tau_i \in [-/2,1/2]} \mathcal{O}\left\{ \left(\frac{\ln n}{nh_n} \right)^{1/2} \right\} h_c \left\{ \tau_i | \widehat{s}_j(m) = s_j^* \right\} \varepsilon_i \\
G_{j2} &= \sum_{\tau_i \in [-/2,1/2]} \mathcal{O}\left\{ \left(\frac{\ln n}{nh_n} \right)^{1/2} \right\} \varepsilon_i.
\end{aligned}
$$

By some similar arguments to those in the case when $m < m_0$, we have

$$\begin{aligned}|G_{j1}| &\leq \left[\sum_{\tau_i \in [-/2,1/2]} \mathcal{O}\left(\frac{\ln n}{nh_n}\right) h_c^2 \left\{\tau_i | \widehat{s}_j(m) = s_j^*\right\}\right]^{1/2} \left(\sum_{\tau_i \in [-/2,1/2]} \varepsilon_i^2\right)^{1/2} \\ &= \mathcal{O}\left\{(nh_n \ln n)^{1/2}\right\}, \\ |G_{j2}| &\leq \left\{\sum_{\tau_i \in [-/2,1/2]} \mathcal{O}\left(\frac{\ln n}{n^2 h_n^2}\right)\right\}^{1/2} \left(\sum_{\tau_i \in [-/2,1/2]} \varepsilon_i^2\right)^{1/2} = \mathcal{O}\left\{(\ln n)^{1/2}\right\}.\end{aligned}$$

Thus,

$$\sum_{i=1}^{n} \varepsilon_i w_i = \mathcal{O}\left\{(nh_n \ln n)^{1/2}\right\} \tag{10}$$

almost surely.

After combining formulas (7)-(10), we almost surely have

$$\text{SSR}(m) - \text{SSR}(m_0) = \begin{cases} nh_n C \sum_{i=m+1}^{m_0} d_j^2 + (nh_n \ln n)^{1/2} \mathcal{O}(1), & m < m_0, \\ (nh_n \ln n)^{1/2} \mathcal{O}(1), & m \geq m_0, \end{cases} \tag{11}$$

and

$$\frac{\text{SSR}(m)}{n} = \begin{cases} \sigma^2 + \mathcal{O}(h_n), & m < m_0, \\ \sigma^2 + \mathcal{O}\left\{\left(\frac{h_n \ln n}{n}\right)^{1/2}\right\}, & m \geq m_0. \end{cases} \tag{12}$$

Furthermore, by the formula (12) and Taylor expansion,

$$\begin{aligned}\widehat{\Delta}_n(m) &= n \log\left\{\frac{\text{SSR}(m)}{\text{SSR}(m_0)}\right\} + P(n) \sum_{j=m\wedge m_0+1}^{m\vee m_0} \frac{1}{\left|\widehat{d}_j(m)\right|^{\gamma}} \\ &= n\left[\frac{\text{SSR}(m) - \text{SSR}(m_0)}{\text{SSR}(m_0)} + \left\{\frac{\text{SSR}(m) - \text{SSR}(m_0)}{\text{SSR}(m_0)}\right\}^2 \mathcal{O}(1)\right] \\ &\quad + P(n) \sum_{j=m\wedge m_0+1}^{m\vee m_0} \frac{1}{\left|\widehat{d}_j(m)\right|^{\gamma}} \\ &= \frac{\text{SSR}(m) - \text{SSR}(m_0)}{\sigma^2} + P(n) \sum_{j=m\wedge m_0+1}^{m\vee m_0} \frac{1}{\left|\widehat{d}_j(m)\right|^{\gamma}} + \mathcal{O}(R_m),\end{aligned}$$

where R_m is a remainder that may depend on m.

When $m < m_0$, $R_m = nh_n^2$. In such cases, by the formula (11), we have

$$\begin{aligned}\widehat{\Delta}_n(m) &= \left\{ nh_n \frac{C}{\sigma^2} \sum_{j=m+1}^{m_0} d_j^2 + (nh_n \ln n)^{1/2} \mathcal{O}(1) \right\} \\ &\quad + P(n) \sum_{j=m+1}^{m_0} \left\{ \frac{1}{|d_j|^\gamma} + \left(\frac{\ln n}{nh_n} \right)^{1/2} \mathcal{O}(1) \right\} + \mathcal{O}(nh_n^2) \\ &= \Delta_n(m) + \mathcal{O}(R_n^{(l)}),\end{aligned}$$

where $R_n^{(l)} = nh_n^2 + P(n) \left(\frac{\ln n}{nh_n} \right)^{1/2}$. Similarly, when $m \geq m_0$, $R_m = h_n \ln n$. In such cases, by the formula (11), we have

$$\begin{aligned}\widehat{\Delta}_n(m) &= (nh_n \ln n)^{1/2} \mathcal{O}(1) + (m - m_0) P(n) \left(\frac{nh_n}{\ln n} \right)^{\gamma/2} \sum_{i=m_0+1}^{m} \frac{1}{C_i} + h_n \ln n \mathcal{O}(1) \\ &= \Delta_n(m) + \mathcal{O}(R_n^{(r)}),\end{aligned}$$

where $C_i = \mathcal{O}(1) > 0$, for $i = m_0 + 1, \ldots, m$, and $R_n^{(r)} = h_n \ln n$. Let

$$R_n = \max\left(R_n^{(l)}, R_n^{(r)} \right) = R_n^{(l)} = nh_n^2 + P(n) \left(\frac{\ln n}{nh_n} \right)^{1/2}.$$

Then,

$$\sup_{m \geq 0} \left| \widehat{\Delta}_n(m) - \Delta_n(m) \right| = \mathcal{O}(R_n). \qquad \square$$

2. PROOF OF THEOREM 1

By Lemma 3, we almost surely have

$$\begin{aligned}\left(\frac{nh_n}{\ln n} \right)^{1/2} \|M_n(x)\|_{L^\infty(\overline{D}_{h_n/2,1\to m_0})} &= \left(\frac{nh_n}{\ln n} \right)^{1/2} \|\widehat{a}_r(x) - \widehat{a}_l(x)\|_{L^\infty(\overline{D}_{h_n/2,1\to m_0})} \\ &\leq \left(\frac{nh_n}{\ln n} \right)^{1/2} \|\widehat{a}_r(x) - f(x)\|_{L^\infty(\overline{D}_{h_n/2,1\to m_0})} \\ &\quad + \left(\frac{nh_n}{\ln n} \right)^{1/2} \|\widehat{a}_l(x) - f(x)\|_{L^\infty(\overline{D}_{h_n/2,1\to m_0})} \\ &= \mathcal{O}(1).\end{aligned}$$

Similarly, we have

$$
\begin{aligned}
&\left(\frac{nh_n}{\ln n}\right)^{1/2} \|M_n(s_j+\tau h_n) - d_j g(\tau)\|_{L^\infty[-1/2,1/2]} \\
&\le \left(\frac{nh_n}{\ln n}\right)^{1/2} \|\widehat{a}_r(s_j+\tau h_n) - f(s_j+\tau h_n) - d_j h_r(\tau)\|_{L^\infty[-1/2,1/2]} \\
&\quad + \left(\frac{nh_n}{\ln n}\right)^{1/2} \|\widehat{a}_l(s_j+\tau h_n) - f(s_j+\tau h_n) - d_j h_l(\tau)\|_{L^\infty[-1/2,1/2]} \\
&= \mathcal{O}(1) \quad (j=1,2,\ldots,m_0),
\end{aligned}
$$

where the equation holds almost surely and $\{h_j : j = r, l\}$ are defined in Lemma 1. This finishes the proof.

3. Proof of Proposition 1

First, we have

$$
\begin{aligned}
H &= H_C\,(I - H_J) + H_J \\
&= H_C + H_J - H_C H_J.
\end{aligned} \tag{13}
$$

Therefore, $\mathrm{tr}\,(H) = \mathrm{tr}\,(H_C) + \mathrm{tr}\,(H_J) - \mathrm{tr}\,(H_C H_J)$. The three terms on the right-hand-side of this expression will be calculated separately below. First, from the expression of the local linear estimator (cf., the expression given immediately after Assumption 2 in the paper), we have

$$
\widehat{a}_j(x) = l_j^T(x) Y, \quad j = l, r, c; \;\; x \in [h_n/2, 1 - h_n/2], \tag{14}
$$

where $l_j(x) = (l_{1,j}(x), \cdots, l_{n,j}(x))^T$, and $l_{i,j}(x) = K_j^*\left(\frac{x_i - x}{h_n}\right)$. Then, we can decompose H_J as $H_J = \sum_{j=1}^{m} H_{J,j}$, where

$$
H_{J,j} = [0, \ldots, 0, l_r\{\widehat{s}_j(m)\} - l_l\{\widehat{s}_j(m)\}, \ldots, l_r\{\widehat{s}_j(m)\} - l_l\{\widehat{s}_j(m)\}]^T\,,
$$

the 0 elements of $H_{J,j}$ correspond to $x_i \le \widehat{s}_j(m)$, and the remaing elements correspond to $x_i > \widehat{s}_j(m)$. By the properties of the local linear kernel estimators, we have $\mathrm{tr}\,(H_{J,j}) = \sum_{i=1}^{n} l_{i,r}\{\widehat{s}_j(m)\} = 1$, and the summation of all rows of $H_{J,j}$ equals 0. Furthermore, we can obtain

$$
\mathrm{tr}\,(H_J) = \sum_{j=1}^{m} \mathrm{tr}\,(H_{J,j}) = m, \tag{15}
$$

$$
\mathrm{tr}\,(H_C H_J) = 0. \tag{16}
$$

By formula (14), we get

$$
\mathrm{tr}\,(H_C) = \sum_{i=1}^{n} l_{i,c}(x_i).
$$

By some similar arguments as those in Lemma 1, we have

$$
l_{i,c}(x_i) = \frac{1}{nh_n} K_c^*(0) + o\left(\frac{1}{nh_n}\right).
$$

So,

$$\operatorname{tr}(H_C) = \frac{K_c^*(0)}{h_n} + o\left(\frac{1}{h_n}\right). \tag{17}$$

Finally by formulas (13), (15), (16) and (17), we have the result in the proposition.

4. PROOF OF THEOREM 2

Proof of part (1). First, we focus on proving the first expression in the theorem in the case when $j = 1$. By Theorem 1, we have

$$\left(\frac{nh_n}{\ln n}\right)^{1/2} \|M_n(x) - m_n(x)\|_{L^\infty(D)} = \mathcal{O}(1) \tag{18}$$

almost surely, where

$$m_n(x) = \sum_{j=1}^{m_0} g_j\left(\frac{x - s_j}{h_n}\right) I_{\{x \in D_{h_n/2,j}\}}, x \in D,$$

and $g_j(\tau) = d_j g(\tau)$. Equivalently, we have

$$\left(\frac{nh_n}{\ln n}\right)^{1/2} \|\widehat{g}^*(\tau) - g^*(\tau)\|_{L^\infty(\Pi)} = \mathcal{O}(1), a.s., \tag{19}$$

where $\Pi = [-1/2, 1/2]$, $\widehat{g}^*(\tau) = |M_n(s_1 + \tau h_n)|$, and $g^*(\tau) = m_n(s_1 + \tau h_n)$. Actually, there exists a 1-1 mapping from $x \in [0,1]$ to $\tau \in \Pi$ by the linear transformation $\tau = (x - s_1)/h_n$, which shows the equivalence between (18) and (19). In the case when $\tau \in [-1/2, 1/2]$, $x \in [s_1 - h_n/2, s_1 + h_n/2]$ and $g^*(\tau) = g_1(\tau) = d_1 g(\tau)$.

By Theorem 1, we can find a set A such that $P(A) = 1$ and for all $\omega \in A$, the equality in the formula (18) or (19) is true. Let $\epsilon > 0$ be the same quantity as the one in Lemma 2 and let

$$W_{n,\epsilon} = \left\{ s_1 - C_1\left(\frac{\ln n}{nh_n}\right)^{1/2}, s_1 + C_1\left(\frac{\ln n}{nh_n}\right)^{1/2} \right\},$$

and $C_1 = C/K_r^*(\epsilon)$. Then, when n is large enough, $W_{n,\epsilon} \subseteq [s_1 - \epsilon, s_1 + \epsilon]$. Also, there exists a constant $\mathrm{D}_n > 0$ such that

$$\begin{aligned} \mathrm{D}_n &= g^*(0) - \sup_{\tau \in [-1/2,1/2] \cap W_{n,\epsilon}^c} g^*(\tau) \\ &= g_1(0) - g_1\left\{-C_1\left(\frac{\ln n}{nh_n}\right)^{1/2}\right\} \\ &= g_1'(\tau^*)\left\{C_1\left(\frac{\ln n}{nh_n}\right)^{1/2}\right\} \geq C\left\{\left(\frac{\ln n}{nh_n}\right)^{1/2}\right\}, \end{aligned} \tag{20}$$

where $\tau^* = -\theta C_1\{\ln n/(nh_n)\}^{1/2}$ and $\theta \in (0,1)$. On the other hand, by Theorem 1 and the definition of $\widehat{\tau} = \{\hat{s}_1(m) - h_n/2\}/h_n$, for $\omega \in A$, there always exist a positive integer $N =$

$N(\omega)$ and a constant $C > 0$ such that when $n > N$,

$$\begin{aligned} g^*(0) - g^*(\widehat{\tau}) &= \{g^*(0) - \widehat{g}^*(0)\} + \{\widehat{g}^*(0) - \widehat{g}^*(\widehat{\tau})\} + \{\widehat{g}^*(\widehat{\tau}) - g^*(\widehat{\tau})\} \\ &\leq |g^*(0) - \widehat{g}^*(0)| + \{\widehat{g}^*(0) - \widehat{g}^*(\widehat{\tau})\} + |\widehat{g}^*(\widehat{\tau}) - g^*(\widehat{\tau})| \\ &< \frac{C}{2}\left(\frac{\ln n}{nh_n}\right)^{1/2} + 0 + \frac{C}{2}\left(\frac{\ln n}{nh_n}\right)^{1/2} = C\left(\frac{\ln n}{nh_n}\right)^{1/2}. \end{aligned} \tag{21}$$

By combining (20) and (21), we have

$$\begin{aligned} g^*(\widehat{\tau}) &> \sup_{\tau \in [-1/2,1/2] \cap W^c_{n,\epsilon}} g^*(\tau) \\ &= g_1\left\{C\left(\frac{\ln n}{nh_n}\right)^{1/2}\right\}, \quad n > N, \omega \in A. \end{aligned}$$

This result indicates that $\widehat{\tau} \in W_{n,\epsilon}$ when $n > N$ and $\omega \in A$. Therefore, for each $\omega \in A$, we can find $C > 0$ such that

$$\begin{aligned} \left(\frac{nh_n}{\ln n}\right)^{1/2} |\widehat{\tau}| &< C, \\ \left(\frac{nh_n}{\ln n}\right)^{1/2} \left|\frac{\widehat{s}_1(m) - s_1}{h_n}\right| &< C, \\ \left(\frac{n}{h_n \ln n}\right)^{1/2} |\widehat{s}_1(m) - s_1| &< C. \end{aligned} \tag{22}$$

So, $\{n/(h_n \ln n)\}^{1/2} |\widehat{s}_j(m) - s_j| = \mathcal{O}(1)$ almost surely when $j = 1$. According to our algorithm, we'll delete a "tie" $[\widehat{s}_1(m) - (1/2+\epsilon)h_n, \widehat{s}_1(m) + (1/2+\epsilon)h_n], \epsilon > 0$. By Lemma 4, we know d_2 will be the largest jump size in the rest region almost surely. Similarly the corresponding results when $j = 2, \ldots, m$ can be proved in the same way.

Next, we prove the second equation in the theorem. First, we have

$$\begin{aligned} &\left(\frac{nh_n}{\ln n}\right)^{1/2} \left|\widehat{d}_j(m) - d_j\right| \\ &\leq \left(\frac{nh_n}{\ln n}\right)^{1/2} \left|\widehat{d}_j(m) - g_j\left\{\frac{\widehat{s}_j(m) - s_j}{h_n}\right\}\right| + \left(\frac{nh_n}{\ln n}\right)^{1/2} \left|g_j\left\{\frac{\widehat{s}_j(m) - s_j}{h_n}\right\} - d_j\right| \\ &= B_1 + B_2. \end{aligned}$$

By Theorem 1, we have $B_1 = \mathcal{O}(1)$ almost surely, and by Lemma 2, we have

$$\begin{aligned} \left|g_j\left\{\frac{\widehat{s}_j(m) - s_j}{h_n}\right\} - d_j\right| &= \left|g_j\left\{\frac{\widehat{s}_j(m) - s_j}{h_n}\right\} - g_j(0)\right| \\ &\leq d_j K^*_r(0) \left|\frac{\widehat{s}_j(m) - s_j}{h_n}\right|, \end{aligned} \tag{23}$$

where $g_j(\tau) = d_j g(\tau)$, and the last inequality is due to Lemma 2. After combining formulas (22) and (23), we have

$$B_2 \leq d_j K^*_r(0) \left(\frac{nh_n}{\ln n}\right)^{1/2} \left|\frac{\widehat{s}_j(m) - s_j}{h_n}\right| \leq C.$$

So, the second equation is true.

Next, we prove the last two equations. First,

$$\begin{aligned}\widehat{f}_m(x) - f(x) &= \sum_{i=1}^{n} Y_i' K_c^*\left(\frac{x_i - x}{h_n}\right) + \widehat{f}_{J,m}(x) - f(x)\\ &= \sum_{i=1}^{n}\left\{f_C(x_i) + f_J(x_i) + \varepsilon_i - \widehat{f}_{J,m}(x_i)\right\} K_c^*\left(\frac{x_i - x}{h_n}\right)\\ &\quad + \widehat{f}_{J,m}(x) - [f_C(x) + f_J(x)]\\ &= A_1(x) + A_2(x) + A_3(x),\end{aligned} \tag{24}$$

where

$$\begin{aligned}A_1(x) &= \sum_{i=1}^{n}\{f_C(x_i) + \varepsilon_i\} K_c^*\left(\frac{x_i - x}{h_n}\right) - f_C(x),\\ A_2(x) &= \sum_{j=1}^{m_0} d_j\left\{\sum_{i=1}^{n} I_{\{x_i > s_j\}} K_c^*\left(\frac{x_i - x}{h_n}\right) - I_{\{x > s_j\}}\right\},\\ A_3(x) &= -\sum_{j=1}^{m} \widehat{d}_j(m)\left\{\sum_{i=1}^{n} I_{\{x_i > \widehat{s}_j(m)\}} K_c^*\left(\frac{x_i - x}{h_n}\right) - I_{\{x > \widehat{s}_j(m)\}}\right\}.\end{aligned}$$

Furthermore, by Lemmas 1 and 3, and by Assumption 1,

$$A_1(x) = \left\{h_n^2 + \left(\frac{\ln n}{nh_n}\right)^{1/2}\right\}\mathcal{O}(1), \quad x \in D, \tag{25}$$

$$A_2(x) = \begin{cases} d_j h_c\left(\frac{x - s_j}{h_n}\right) + \mathcal{O}\left(\frac{1}{nh_n}\right), & x \in D_{h_n/2,j}, 1 \le j \le m_0\\ 0, & x \in \overline{D}_{h_n/2,1\to m_0}.\end{cases} \tag{26}$$

The term $A_3(x)$ is a little more complicated to handle, and we decompose it into

$$A_3(x) = \begin{cases} A_{31}(x) + A_{32}(x) + A_{33}(x) + A_{34}(x), & x \in D_{h_n/2,j}, 1 \le j \le m\\ 0, & x \in \overline{D}_{h_n/2,1\to m},\end{cases}$$

where

$$\begin{aligned}A_{31}(x) &= -\widehat{d}_j(m)\sum_{i=1}^{n}\left\{I_{\{x_i > \widehat{s}_j(m)\}} - I_{\{x_i > s_j\}}\right\} K_c^*\left(\frac{x_i - x}{h_n}\right) = \mathcal{O}\left\{\left(\frac{\ln n}{nh_n}\right)^{1/2}\right\},\\ A_{32}(x) &= \left\{\widehat{d}_j(m) - d_j\right\}\left\{I_{\{x > \widehat{s}_j(m)\}} - I_{\{x > s_j\}}\right\} = \mathcal{O}\left\{\left(\frac{\ln n}{nh_n}\right)^{1/2}\right\},\\ A_{33}(x) &= d_j\left\{I_{\{x > \widehat{s}_j(m)\}} - I_{\{x > s_j\}}\right\},\\ A_{34}(x) &= -\widehat{d}_j(m)\left\{\sum_{i=1}^{n} I_{\{x_i > s_j\}} K_c^*\left(\frac{x_i - x}{h_n}\right) - I_{\{x > s_j\}}\right\}\\ &= -d_j h_c\left(\frac{x - s_j}{h_n}\right) + \mathcal{O}\left\{\left(\frac{\ln n}{nh_n}\right)^{1/2}\right\}.\end{aligned}$$

So we have

$$A_3(x) = \begin{cases} -d_j h_c\left(\frac{x-s_j}{h_n}\right) + \mathcal{O}\left\{\left(\frac{\ln n}{nh_n}\right)^{1/2}\right\} + d_j\left\{I_{\{x>\widehat{s}_j(m)\}} - I_{\{x>s_j\}}\right\}, & \\ & x \in D_{h_n/2,j}, 1 \le j \le m \\ 0, & x \in \overline{D}_{h_n/2,1\to m}. \end{cases} \quad (27)$$

After combining formulas (24)-(27), we get the following expression

$$\widehat{f}_m(x) - f(x) = \begin{cases} \mathcal{O}\left\{\left(\frac{\ln n}{nh_n}\right)^{1/2}\right\} + d_j\left\{I_{\{x>\widehat{s}_j(m)\}} - I_{\{x>s_j\}}\right\}, & \\ & x \in D_{h_n/2,j}, j = 1,\ldots,m, \\ d_j h_c\left(\frac{x-s_j}{h_n}\right) + \mathcal{O}\left\{\left(\frac{\ln n}{nh_n}\right)^{1/2}\right\}, & x \in D_{h_n/2,j}, j = m+1,\ldots,m_0, \\ \mathcal{O}\left\{\left(\frac{\ln n}{nh_n}\right)^{1/2}\right\}, & x \in \overline{D}_{h_n/2,1\to m_0}, \end{cases} \quad (28)$$

where the equation is valid almost surely and uniformly in x. Let $\epsilon > 0$ be an arbitrarily small positive number. Then, by the first formula in Theorem 2, we have $\widehat{s}_j(m) \in [s_j - \epsilon\delta_n, s_j + \epsilon\delta_n]$ almost surely when n is large enough, where δ_n is defined in Theorem 2. So, the following result is true almost surely and uniformly when n is large enough:

$$I_{\{x>\widehat{s}_j(m)\}} - I_{\{x>s_j\}} = 0, \text{ when } x \in D \backslash D_{\delta_n,j}. \quad (29)$$

By (28) and (29), the last two conclusions in the first part of the theorem.

Proof of part (2). When $m \ge m_0$, by some similar arguments for (28), we have

$$\widehat{f}_m(x) - f(x) = \begin{cases} \mathcal{O}\left\{\left(\frac{\ln n}{nh_n}\right)^{1/2}\right\} + d_j\left\{I_{\{x>\widehat{s}_j(m)\}} - I_{\{x>s_j\}}\right\}, & \\ & x \in D_{h_n/2,j}, j = 1,\ldots,m_0 \\ \mathcal{O}\left\{\left(\frac{\ln n}{nh_n}\right)^{1/2}\right\}, & x \in \overline{D}_{h_n/2,1\to m_0}. \end{cases}$$

The remaining results can be shown easily by similar arguments to those in the case when $m < m_0$.

5. Proof of Theorem 3

By Lemma 5, we can find a set A such that $P(A) = 1$, and for all $\omega \in A$ the equality (6) holds. By the definition of $\widehat{\Delta}_n(m)$ in Lemma 5, functions $\widehat{\Delta}_n$ and JIC(m) have the same minimizer $\widehat{m}$. Further, by the formula (4), given any neighbourhood

$$W_{n,\epsilon} = \left(m_0 - \epsilon\frac{R_n}{K_n}, m_0 + \epsilon\frac{R_n}{K_n}\right),$$

where R_n is the same as in Lemma 5 and $K_n = \min\left\{nh_n, P(n)\left(\frac{nh_n}{\ln n}\right)^{\gamma/2}\right\}$, there exists a constant $\mathrm{D}_n > 0$ such that

$$\begin{aligned}
\mathrm{D}_n &= \inf_{W_n^c} \Delta_n(m) - \Delta_n(m_0) \\
&= \left\lceil \epsilon \frac{R_n}{K_n} \right\rceil [K_n + o(K_n)] \\
&= \epsilon R_n + \epsilon R_n o(1) + \left(\left\lceil \epsilon \frac{R_n}{K_n} \right\rceil - \epsilon \frac{R_n}{K_n} \right) \{K_n + o(K_n)\} \\
&\geq \epsilon R_n,
\end{aligned} \tag{30}$$

where the term $o(K_n)$ is positive according to the definition of Δ_n, and $\lceil x \rceil$ denotes the minimum integer larger than or equal to x.

On the other hand, according to Lemma 5 and the definition of $\widehat{m}$, for $\omega \in A$, there always exists a positive integer $N = N(\omega)$ such that when $n > N$,

$$\begin{aligned}
\Delta_n(\widehat{m}) - \Delta_n(m_0) &= \{\Delta_n(\widehat{m}) - \widehat{\Delta}_n(\widehat{m})\} + \{\widehat{\Delta}_n(\widehat{m}) - \widehat{\Delta}_n(m_0)\} + \{\widehat{\Delta}_n(m_0) - \Delta_n(m_0)\} \\
&\leq \left|\Delta_n(\widehat{m}) - \widehat{\Delta}_n(\widehat{m})\right| + \{\widehat{\Delta}_n(\widehat{m}) - \widehat{\Delta}_n(m_0)\} + \left|\widehat{\Delta}_n(m_0) - \Delta_n(m_0)\right| \\
&< \epsilon R_n/2 + 0 + \epsilon R_n/2 = \epsilon R_n.
\end{aligned} \tag{31}$$

By combining (30) and (31), we have

$$\Delta_n(\widehat{m}) < \inf_{W_n^c} \Delta_n(m), \text{when } n > N, \omega \in A,$$

which implies that $\widehat{m} \in W_{n,\epsilon}$ when $n > N$ and $\omega \in A$. The result in the theorem is then proved.

6. PROOF OF THEOREM 4

The final estimators $\widehat{s}_j, \widehat{d}_j$ and $\widehat{f}(\cdot)$ can be written as the following compound forms:

$$\widehat{s}_j = \sum_{m \in \mathcal{M}} \widehat{s}_j(m) I_{\{\widehat{m}=m\}}, \tag{32}$$

$$\widehat{d}_j = \sum_{m \in \mathcal{M}} \widehat{d}_j(m) I_{\{\widehat{m}=m\}}, \tag{33}$$

$$\widehat{f}(\cdot) = \sum_{m \in \mathcal{M}} \widehat{f}_m(\cdot) I_{\{\widehat{m}=m\}}, \tag{34}$$

where $\mathcal{M} = \{0, 1, 2, \cdots\}$. The first two results in Theorem 4 can be proved in the same way as that in Theorem 2. Next, we will prove the third result.

By Theorems 2-3, we can find a set A satisfying $P(A) = 1$ such that for all $\omega \in A$, the following two equations hold:

$$\widehat{m}(\omega) = m_0, \quad n \geq N(\omega),\ \omega \in A, \tag{35}$$

$$\left(\frac{nh_n}{\ln n}\right)^{1/2} \left\|\widehat{f}_{m_0}(x) - f(x)\right\|_{L^\infty(\overline{D}_{\delta_n, 1\to m_0})} = \mathcal{O}(1), \tag{36}$$

where $N(\omega)$ is a positive integer relying on $\omega \in A$. By combining formulas (34) and (35), we have

$$\widehat{f}(x) = \sum_{m\in\mathcal{M}} \widehat{f}_m(x) I_{\{\widehat{m}=m\}} = \widehat{f}_{m_0}(x), \quad n > N(\omega),\ \omega \in A.$$

By combining this result and the one in the formula (36), we have

$$\left(\frac{nh_n}{\ln n}\right)^{1/2} \left\|\widehat{f}(x) - f(x)\right\|_{L^\infty(\overline{D}_{\delta_n,1\to m_0})} = \mathcal{O}(1), \quad \omega \in A.$$

7. DERIVATION OF THE BAYESIAN INFORMATION CRITERION (15)

The Bayesian information criterion was originally suggested for parametric models as an asymptotic approximation to a transformation of the Bayesian posterior probability of a candidate model (cf., Schwarz, 1978; Yao, 1988; Zhang & Siegmund, 2007; Hannart & Naveau, 2012). For the nonparametric jump regression model (1) with an unknown number of jumps, as shown in Proposition 1 and the related discussion afterwards, it seems impossible to define this criterion in the same way as in parametric cases. So, we define the criterion here for the nonparametric model (1) by focusing on the jump estimation part alone and by treating the resulting model as a parametric model, as described below.

As described in Subsection 2.2, our procedure for estimating model (1) consists of three steps in cases when the number of jumps m is assumed known. First, the jump location estimators $\widehat{s}_j(m)$ and the corresponding jump size estimators $\widehat{d}_j(m)$ are defined based on the two one-sided estimators $\widehat{a}_l(\cdot)$ and $\widehat{a}_r(\cdot)$. Second, the estimator of the continuity part $\widehat{f}_{C,m}(x)$ is obtained using the data $Y_{i,m}, i = 1, \ldots, n$. Third, the estimator of $f(x)$ is defined as the summation of the estimated continuity and jump parts obtained in the first two steps. Here, we notice that estimators $\widehat{s}_j(m)$ and $\widehat{f}_{C,m}(x)$ are based on $\widehat{a}_j(x), j = l, c, r$. By Proposition 1, the resulting model complexity is in the order of $\mathcal{O}(1/h_n)$. However, if we focus on the estimation of the jump size d_j alone and treating the estimators $\widehat{s}_j(m)$ and $\widehat{f}_{C,m}(x)$ as given beforehand, then the estimation problem becomes parametric. Next, we will derive the BIC criterion in such cases.

Let $Y^*_{i,m} = Y_i - \widehat{f}_{C,m}(x)$, for each i. Because all estimators $\widehat{s}_j(m)$ and $\widehat{f}_{C,m}(x)$ are consistent estimators (cf., Theorem 2), we have the following expression:

$$Y^*_{i,m} = \sum_{j=1}^{m} d_j I_{\{x_i > \widehat{s}_j(m)\}} + \epsilon_{i,n} \quad (i = 1, \ldots, n), \tag{37}$$

where $\epsilon_{i,n}$ are random errors that may depend on n and have an asymptotic mean of 0. Next, we partition the interval $[0, 1]$ into $M = [1/h_n]$ mutually disjoint intervals $D_{h_n/2,r_k} = [r_k - h_n/2, r_k + h_n/2]$, for $k = 1, \ldots, M$, where there are m intervals containing a true jump point each. Without loss of generality, we assume that the first m intervals contain jumps and $r_j = \widehat{s}_j$, for $j = 1, \ldots, m$. So, for design points in the first m intervals, model (37) becomes

$$Y^*_{i,m} = g(r_j) + d_j I_{\{x_i > r_j\}} + \epsilon_{i,n}, \quad x_i \in D_{h_n/2,r_j}, j = 1, \ldots, m, \tag{38}$$

where $g(x) = \sum_{j=1}^{m} d_j I_{\{x > \widehat{s}_j(m)\}}$. In cases when the design points belong to the last $M - m$ intervals, model (37) becomes

$$Y^*_{i,m} = g(r_j) + \epsilon_{i,n}, \quad x_i \in D_{h_n/2,r_j}, j = m + 1, \ldots, M. \tag{39}$$

It should be pointed out that the d_j's are estimated sequentially. Namely, at the time when we estimate d_j, $\{d_1, \ldots, d_{j-1}\}$ have all be estimated. So, in (38), the term $g(r_j)$ can be assumed known.

Let $\mathcal{F}$ be the σ-field expanded from $\{Y^*_{i,m}, \widehat{s}_j(m), 1 \le i \le n, 1 \le j \le m\}$, B denote the event that the model (37) is valid, and B_j denote the event that the jth local model defined by (38) and (39) is true. By the independence among the local models, the posterior probability of B can be expressed as

$$\text{pr}\,(B \mid \mathcal{F}) = \text{pr}\,(\cap B_j \mid \mathcal{F}) = \prod_{j=1}^{M} \text{pr}\,(B_j \mid \mathcal{F})\,.$$

By using the results in Lemma 6 below about the quantity $\text{Pr}\,(B_j \mid \mathcal{F})$, we have

$$\begin{aligned}
-2\log \text{Pr}\,(B \mid \mathcal{F}) &= \sum_{j=1}^{M} -2\log \text{Pr}\,(B_j \mid \mathcal{F}) \\
&\approx \sum_{j=1}^{m} \left\{ nh_n \log(2\pi\sigma^2) + \frac{\text{SSR}_{j,1}}{\sigma^2} + \log nh_n \right\} \\
&\quad + \sum_{j=m+1}^{M} \left\{ nh_n \log(2\pi\sigma^2) + \frac{\text{SSR}_{j,0}}{\sigma^2} \right\} \\
&\approx n\log(2\pi\sigma^2) + \frac{\text{SSR}(m)}{\sigma^2} + m\log(nh_n),
\end{aligned}$$

where

$$\begin{aligned}
\text{SSR}_{j,0} &= \left\{ y^{(j)} - g(r_j)1_{N_h} \right\}^T \left\{ y^{(j)} - g(r_j)1_{N_h} \right\}, \\
\text{SSR}_{j,1} &= \left\{ y^{(j)} - g(r_j)1_{N_h} - X_j\widehat{d}_j \right\}^T \left\{ y^{(j)} - g(r_j)1_{N_h} - X_j\widehat{d}_j \right\},
\end{aligned}$$

$y^{(j)}$ is the vector of all y observations in $D_{h_n/2,r_j}$, 1_{N_h} is the N_h-dimensional vector with all elements equal to 1, N_h is the number of observations in $D_{h_n/2,r_j}$, and $X_j = (I_{\{x_1 > r_j\}}, \ldots, I_{\{x_{N_h} > r_j\}})^T$. After a constant is ignored, a Bayesian information criterion can be defined as

$$\text{BIC}^*(m) = \text{SSR}(m)/\sigma^2 + m\log(nh_n). \tag{40}$$

In practice, because σ^2 is often unknown, the above definition is usually replaced by the following equivalent definition:

$$\text{BIC}(m) = n\log\{\text{SSR}(m)/n\} + m\log(nh_n) \tag{41}$$

The equivalence between $\text{BIC}^*(m)$ and $\text{BIC}(m)$ in (40) and (41) can be briefly explained below. Let $\Delta\text{SSR}(j-1) = \text{SSR}(j) - \text{SSR}(j-1)$, and $\text{LR}(j, j-1) = \Delta\text{SSR}(j-1)/\text{SSR}(j-1)$. Then,

$$\begin{aligned}
\text{BIC}(j) - \text{BIC}(j-1) &= n\log\{1 + \text{LR}(j, j-1)\} + \log(nh_n) \\
&= n\left[\text{LR}(j, j-1) + o\{\text{LR}(j, j-1)\}\right] + \log(nh_n) \\
&\approx \frac{\Delta\text{SSR}(j-1)}{\sigma^2} + \log(nh_n) \\
&= \text{BIC}^*(j) - \text{BIC}^*(j-1).
\end{aligned} \tag{42}$$

In the second equality of the above expressions, results in (12) have been used. Expressions in (42) guarantee that BIC$^*(m)$ and BIC(m) would have the same asymptotical minimum value $\widehat{m}$ and the same convergence rate of $\widehat{m}$. But, BIC(m) does not depend on σ^2 directly, and thus is more convenient to use in practice.

LEMMA 6. *Let $y^{(j)}$ be the vector of observations in $D_{h_n/2,r_j}$, 1_{N_h} be the N_h-dimensional vector with all elements equal to 1, I_{N_h} be the $N_h \times N_h$ identity matrix, N_h is the number of observations in $D_{h_n/2,r_j}$, and $X_j = (I_{\{x_1 > r_j\}}, \ldots, I_{\{x_{N_h} > r_j\}})^T$. If we assume that $y^{(j)} \sim N_{N_h}(g(r_j)1_{N_h} + X_j\theta_j, \sigma^2 I_{N_h})$, where $\theta_j = d_j$, for $1 \le j \le m$, and $\theta_j = 0$, otherwise. Then, we have:*

(1). when $m < j \le M$,

$$-2\log \mathrm{pr}\,(B_j \mid \mathcal{F}) \approx nh_n(2\pi\sigma^2) + \left\{y^{(j)} - g(r_j)1_{N_h}\right\}^T \left\{y^{(j)} - g(r_j)1_{N_h}\right\}/\sigma^2,$$

(2). when $1 \le j \le m$,

$$\begin{aligned} &-2\log \mathrm{pr}\,(B_j \mid \mathcal{F}) \\ &\approx nh_n(2\pi\sigma^2) + \left\{y^{(j)} - g(r_j)1_{N_h} - X_j\widehat{\theta}_j\right\}^T \left\{y^{(j)} - g(r_j)1_{N_h} - X_j\widehat{\theta}_j\right\}/\sigma^2 + \log(nh_n). \end{aligned}$$

Outline of the Proof:

Let $\xi(y^{(j)})$ be the marginal distribution of $y^{(j)}$. Then $\mathrm{pr}\,(B_j \mid \mathcal{F}) = \xi^{-1}(y^{(j)})\mathrm{pr}(y^{(j)} \mid B_j)$. In cases when $m < j \le M$, it is easy to obtain (i) from the normality assumption, after $\xi^{-1}(y^{(j)})$ is ingored. In cases when $1 \le j \le m$, we have

$$\mathrm{pr}(y^{(j)} \mid B_j) = \int L(d_j \mid y)\pi(d_j)\mathrm{d}d_j, \tag{43}$$

where $\pi(d_j)$ is a prior distribution on d_j. First, it can be checked that the maximum likelihood estimate of d_j is $\widehat{d}_j = \sum\limits_{x_i > r_j} y_i/(N_h/2) - g(r_j)$, by maximizing the likelihood function

$$L(d_j \mid y^{(j)}) = \left(2\pi\sigma^2\right)^{-N_h/2} \exp\left[-\frac{\{y^{(j)} - g(r_j)1_{N_h} - X_j d_j\}^T \{y^{(j)} - g(r_j)1_{N_h} - X_j d_j\}}{2\sigma^2}\right]. \tag{44}$$

By Laplace approximation, we have

$$\begin{aligned} \log L(d_j \mid y^{(j)}) &\approx \log L\left(\widehat{d}_j \mid y^{(j)}\right) + \left\{\frac{\partial \log L(d_j \mid y^{(j)})}{\partial d_j}\right\}_{d_j = \widehat{d}_j} \left(d_j - \widehat{d}_j\right) \\ &\quad + \frac{1}{2}\left\{\frac{\partial^2 \log L(d_j \mid y^{(j)})}{\partial d_j^2}\right\}_{d_j = \widehat{d}_j} \left(d_j - \widehat{d}_j\right)^2 \\ &= \log L\left(\widehat{d}_j \mid y^{(j)}\right) - \frac{1}{2}\left\{nh_n \overline{I}(\widehat{d}_j, y^{(j)})\right\}\left(d_j - \widehat{d}_j\right)^2, \end{aligned}$$

where $\overline{I}(\widehat{d}_j, y^{(j)}) = 1/\sigma^2$. Furthermore, we have the approximation

$$\begin{aligned} \textstyle\int L(d_j \mid y^{(j)})\pi(d_j)\mathrm{d}d_j &\approx \int L(\widehat{d}_j \mid y^{(j)}) \exp\left[-\frac{1}{2}\left\{nh_n\overline{I}(\widehat{d}_j, y^{(j)})\right\}\left(d_j - \widehat{d}_j\right)^2\right]\pi(d_j)\mathrm{d}d_j \\ &= L(d_j \mid y^{(j)})\,(2\pi)^{1/2}\left|nh_n\overline{I}(\widehat{d}_j, y^{(j)})\right|^{-1/2}. \end{aligned} \tag{45}$$

After combining formulas (43) and (45), we have

$$-2\log \text{pr}(y^{(j)} \mid B_j) \approx -2\log\left\{L(\widehat{d}_j \mid y^{(j)})\cdot\left(\frac{2\pi}{nh_n}\right)^{1/2}\left|\overline{I}(\widehat{d}_j, y^{(j)})\right|^{-1/2}\right\} \approx -2\log L(\widehat{d}_j \mid y^{(j)}) + \log(nh_n) \approx -2\log \text{pr}(B_j \mid \mathcal{F}). \quad (46)$$

In the last approximation of the above expressions, a term only depending on $y^{(j)}$ has been ignored. Then, the result (ii) can be obtained from the combination of (46) and (44).

REFERENCES

GIJBELS, I., LAMBERT, A. & QIU, P. H. (2007). Jump-preserving regression and smoothing using local linear fitting: a compromise. *Annals of the Institute of Statistical Mathematics* **59**, 235–272.

HANNART, A. & NAVEAU, P. (2012). An improved Bayesian information criterion for multiple change-point models. *Technomerics* **54**, 256–268.

SCHWARZ, G. E. (1978). Estimating the dimension of a model. *Annals of Statistics* **6**, 461–464.

YAO, Y. -C. (1988). Estimating the number of change-points via Schwarz criterion. *Statistics and Probability Letters* **6**, 181–189.

ZHANG, N. R. & SIEGMUND, D. O. (2007). A modified Bayes information criterion with applications to the analysis of comparative genomic hybridization data. *Biometrics* **63**, 22–32.