CHANGE POINT ESTIMATION USING NONPARAMETRIC REGRESSION

Short Title: Change Point Estimation

To appear in: The Annals of Statistics, 1996.

Clive R. Loader
AT&T Bell Laboratories

Abstract

We consider a regression model in which the mean function may have a discontinuity at an unknown point. We propose an estimate of the location of the discontinuity based on one-sided nonparametric regression estimates of the mean function. The change point estimate is shown to converge in probability at rate $O(n^{-1})$, and to have the same asymptotic distribution as maximum likelihood estimates considered by other authors under parametric regression models. Confidence regions for the location and size of the change are also discussed.

AMS 1991 *subject classifications.* Primary 62G07.
Key words and phrases. Boundary crossing, change point, nonparametric regression.

1 Introduction

Let $x_i = i/n, 1, \ldots, n$ and $Y_i = f(x_i) + \epsilon_i$ where the residuals are independent $N(0,1)$ and f is right continuous, and left continuous except at an unknown change point $\tau \in (0,1)$. A second quantity of interest is the size of the change, which we measure by $\Delta = f(\tau_+) - f(\tau_-)$.

If f is assumed to be constant except at the change point, this model reduces to the mean shift model for a sequence of independent normal random variables. The maximum likelihood estimate $\hat{\tau}$ of τ is shown by Hinkley (1970) to converge in probability at rate $O(n^{-1})$. Hinkley also shows the limit distribution of $n(\hat{\tau} - \tau)$ is related to the location of the maximum of a two sided random walk. These results are extended to parametric regression models by Kim and Siegmund (1989).

We make the weaker assumption that f varies smoothly away from the change point. Specifically, we suppose there exists a constant β such that

$$|f(x) - f(y)| \leq \beta |x - y| \text{ whenever } (x - \tau)(y - \tau) > 0. \tag{1}$$

Let $K(u)$ be a weight function defined on $[0, \infty)$ satisfying the following conditions:

1. $K(0) > 0$; $K(u) \geq 0$ for $0 < u < 1$ and $K(u) = 0$ for $u \geq 1$.

2. There exists z such that $|K(u) - K(v)| < z|u - v|$ for all $u, v \geq 0$.

3. $\int_0^1 K(u) du = 1$.

We choose a bandwidth h, with the requirements $h \to 0$ as $n \to \infty$, but $nh/\log n \to \infty$.

For some $i; 1 \leq i \leq n$ we have $x_{i-1} < \tau \leq x_i$. However, the data cannot be used to distinguish possible changes in this interval. For definiteness, we suppose that τ is an event time; $\tau = x_i$. Of course, this requires τ to depend on n although we suppress this dependence.

Left and right local regressions are used to estimate the left and right limits of f at event times. For t such that $m = nt$ is an integer, assign weights $K(j/(nh))$ to observations $Y_{m+j}, j = 0, \ldots, nh$. Then, fit a local polynomial model of degree p by weighted least squares:

$$E(Y_{m+j}) = a_0 + a_1 j + \ldots + a_p j^p; j = 0, \ldots, nh.$$

Define $\hat{f}_+(t) = \hat{a}_0$. An estimate $\hat{f}_-(t)$ is defined similarly, using $Y_{m-j-1}, j = 0, \ldots, nh$. For local constant fitting ($p = 0$) we have explicitly

$$\begin{aligned} \hat{f}_-(t) &= \frac{\sum_{j=0}^{nh} K(\frac{j}{nh}) Y_{m-j-1}}{\sum_{j=0}^{nh} K(\frac{j}{nh})} \\ \hat{f}_+(t) &= \frac{\sum_{j=0}^{nh} K(\frac{j}{nh}) Y_{m+j}}{\sum_{j=0}^{nh} K(\frac{j}{nh})}. \end{aligned} \tag{2}$$

Define $\hat{\Delta}_t = \hat{f}_+(t) - \hat{f}_-(t)$. The estimate $\hat{\tau}$ of τ is that value of t which maximizes $\hat{\Delta}_t^2$ over the range $h \leq t \leq 1-h$. One could also consider the maximizer of $\hat{\Delta}_t$ if it is known $\Delta > 0$.

The choice of order of local polynomial turns out to have little impact on the asymptotic results for $\hat{\tau}$ derived below. In practice, for local constant fitting $\hat{\Delta}_t$ may be quite biased, and local linear fitting, although more variable, is usually preferable. This is related to the 'boundary problem' in nonparametric regression, discussed for example in Fan and Gijbels (1992).

The estimate here is similar in principle to that studied by Müller (1992); however, by imposing different conditions on K our estimate has dramatically different properties. These differences and comparisons with other estimates are explored further in Section 2.

Theorem 1 *Let* $\ldots, \epsilon_{-1}, \epsilon_0, \epsilon_1, \ldots$ *be independent* $N(0,1)$ *random variables. Then*

$$\lim_{n\to\infty} P(n(\hat{\tau} - \tau) = l) = P(L_\Delta = l)$$

where L_Δ *is the location of the maximum of the process*

$$Z_i = \begin{cases} \Delta(\epsilon_1 + \ldots + \epsilon_i) - \frac{i\Delta^2}{2} & i > 0 \\ 0 & i = 0 \\ \Delta(\epsilon_i + \ldots + \epsilon_{-1}) - \frac{|i|\Delta^2}{2} & i < 0 \end{cases}. \tag{3}$$

In the simplest parametric change point model, one assumes $f(t) = \mu + \Delta I(t \geq \tau)$. The maximum likelihood estimate of τ is then

$$\hat{\tau}_{\text{par}} = \text{argmax}_{0<t<1} \frac{m(n-m)}{m} \left(\frac{S_n - S_m}{n-m} - \frac{S_m}{m} \right)^2$$

where $m = nt$ and $S_m = \sum_{i=1}^m Y_i$. Hinkley (1970) derived exactly the limit distribution in Theorem 1 for $n(\hat{\tau}_{\text{par}} - \tau)$. The local regression estimates will require a larger n for the asymptotics to be applicable.

Confidence sets for parametric change point problems have been discussed by several authors; Siegmund (1988) reviews several methods. Kim and Siegmund (1989) discuss confidence sets for change points in parametric regression models. The following theorems adapt the likelihood ratio method of Siegmund (1988) to find asymptotic confidence regions for τ and (τ, Δ). To state the results we need some notation. Let $A(u) = (1 \quad u \quad \ldots \quad u^p)^T$; $\Lambda_j = \int_0^1 K(u)^j A(u) A(u)^T du$ and

$$\begin{aligned} M_1 &= 2K(0)[\Lambda_1^{-1}]_{1,1} \\ M_2 &= 2[\Lambda_1^{-1}\Lambda_2\Lambda_1^{-1}]_{1,1}. \end{aligned}$$

The notation $[\cdot]_{1,1}$ denotes matrix subscripting.

Theorem 2 *Under the assumptions of Theorem 1,*

$$\frac{nh}{2M_1}\sup_t(\hat{\Delta}_t^2 - \hat{\Delta}_\tau^2) \xrightarrow{\mathcal{L}} Z_{L_\Delta}.$$

Let $c_1(\gamma, \Delta)$ *denote the* $1-\gamma$ *quantile of* Z_{L_Δ} *and*

$$I_1 = \{t : \frac{nh}{2M_1}(\hat{\Delta}_{\hat{\tau}}^2 - \hat{\Delta}_t^2) < c_1(\gamma, \hat{\Delta}_t)\}.$$

Then

$$\lim_{n\to\infty} P(\tau \in I_1) = 1 - \gamma. \tag{4}$$

For I_1 to be an asymptotic $1-\gamma$ confidence set would require the convergence in (4) to be uniform in τ, Δ and f. Clearly this cannot hold; however, uniformity can easily be obtained by assuming these parameters lie in suitable compact spaces.

Theorem 3 *Suppose the bandwidth satisfies* $nh^3 \to 0$ *in addition to the existing conditions. Let* U *be a* χ_1^2 *random variable, independent of* Z_{L_Δ}. *Define*

$$l(\tau, \Delta) = \frac{nh}{2M_2}(\hat{\Delta}_\tau - \Delta)^2 + \frac{nh}{2M_1}(\hat{\Delta}_{\hat{\tau}}^2 - \hat{\Delta}_\tau^2).$$

Then

$$l(\tau, \Delta) \xrightarrow{\mathcal{L}} \frac{1}{2}U + Z_{L_\Delta}.$$

Let $c_2(\gamma, \Delta)$ *be the* $(1-\gamma)$ *quantile of* $\frac{1}{2}U + Z_{L_\Delta}$, *and define*

$$I_2 = \{(t, \Delta_0) : l(t, \Delta_0) < c_2(\gamma, \Delta_0)\}.$$

Then

$$\lim_{n\to\infty} P((\tau, \Delta) \in I_2) = 1 - \gamma.$$

We remark the condition $nh^3 \to 0$ ensures the bias of $\hat{\Delta}_\tau$ is small relative to its variance. When $p \geq 1$ and appropriate smoothness conditions on $f(t), t \neq \tau$ this can be weakened.

Approximations to $c_1(\gamma, \Delta)$ and $c_2(\gamma, \Delta)$ are given by Siegmund (1988), equations 7 and 25. In our notation, these are

$$\begin{aligned} P(Z_{L_\Delta} > c_1) &\approx 1 - (1 - \nu(\Delta)e^{-c_1})^2 \\ P(\frac{1}{2}U + Z_{L_\Delta} > c_2) &\approx P(\chi_1^2 > 2c_2) + 4\nu(\Delta)c_2 e^{-c_2}. \end{aligned}$$

These are asymptotic as c_1 and $c_2 \to \infty$. The quantity $\nu(\Delta)$ is defined by Siegmund (1988) eqn. 4; the approximation $\nu(\Delta) \approx e^{-0.583\Delta}$ suffices for most purposes.

Asymptotics for change point estimates based on two sided random walks have been derived for a number of models; Dümbgen (1991) is a recent reference. An important question studied by Ritov (1990) is efficiency; can one do better by aiming for functionals of the random walk other than the maximizer? The answer depends on the loss function; a maximum likelihood type estimate is appropriate for 0-1 loss. For a quadratic loss, Ritov's results suggest considering estimates of the form

$$\tilde{\tau} = \frac{\int t \exp(nh\hat{\Delta}_t^2/2M_1)\mu_n(dt)}{\int \exp(nh\hat{\Delta}_t^2/2M_1)\mu_n(dt)}$$

where μ_n denote counting measure on $\{j/n : |j - n\hat{\tau}| < i_0\}$ and $i_0 \to \infty$ at a suitably slow rate; see (5) below. For detecting a change in the drift of Brownian motion, the asymptotic efficiency of the maximum likelihood estimate is about 73% under quadratic loss; see Ibragimov and Has'minskii (1981).

2 Comparisons

Our change point estimate is illustrated in Figure 1. The data in the top panel is convoluted with the split kernel (middle panel) to obtain $\hat{\Delta}_t$ in the bottom panel. The crucial condition leading to the $O_p(n^{-1})$ convergence in Theorem 1 is $K(0) > 0$. If t is increased from τ (i.e. the kernel in Figure 1 moved to the right) post change observations will switch abruptly from $\hat{f}_+(t)$ to $\hat{f}_-(t)$, and $\hat{f}_-(t)$ responds rapidly to the change. Likewise, when t is decreased from τ, $\hat{f}_+(t)$ responds rapidly to the change. This results in the sharp peak in $\hat{\Delta}_t$ at $t = \tau$. Smoothness of the kernel at points other than 0 is required to minimize spurious noise in the process.

An early related paper is McDonald and Owen (1986), who estimate a regression curve with possible discontinuities using weighted combinations of left, right and central smooths at various bandwidths are combined using a mean squared error criterion. Change point estimates based on the difference of left and right smooths are introduced by Müller (1992) and Hall and Titterington (1992) as part of a more complicated smoothing procedure. However, in both these papers the weights used do not satisfy our conditions and the resulting estimates have different asymptotic performance. The conditions (K4) and (K2) in Müller (1992) jointly imply $K(0) = 0$; this leads to estimates with an inferior rate of convergence.

Hall and Titterington (1992) derive their estimates from different principles, but in their examples $\hat{f}_l$ and $\hat{f}_r$ are our one-sided local linear estimates with the uniform weight function $K(u) = I_{[0,1]}(u)$. A result similar to Theorem 1 still holds in this case, but observations around the discontinuities at $u = \pm 1$ contribute to the limit distribution with a fraction of $1/4$. The ϵ_i in (3) now have variance $1 + 2(0.25)^2 = 1.125$. This slightly reduces the efficiency of the estimate.

We give a simulated example to more fully appreciate the difference between our estimate and Müller's. Consider the model $f(x) = 4\sin(5x) + 3x + I(x \geq 0.7)$, and $n = 1000$. This represents a challenging problem; the change is nearly impossible to detect by eye and a long sequence is required for any estimator to have much chance of detection with $N(0,1)$ residuals.

The left panel of Figure 2 displays results based on 10001 simulations, showing the median and 90th percentile of the distribution of $n|\hat{\tau} - \tau|$. Local linear regression with $K(u) = \frac{3}{2}(1 - u^2)I(0 \leq u \leq 1)$ was used to construct $\hat{f}_-(t)$ and $\hat{f}_+(t)$. The estimate requires quite large bandwidths: $nh \geq 60$ to reliably detect the change with probability 0.5, and $nh \geq 130$ to detect with probability 0.9. The asymptotic MAD of 2 is achieved for $nh \geq 100$, while the asymptotic 90th percentile is never quite achieved. Confidence sets for τ were computed using Theorem 2 with $nh = 150$ and $1 - \gamma = 0.9$; the actual coverage obtained was 91.1% with a median size of 11 observations. The joint confidence region of Theorem 3 achieved an actual coverage of 92.0%; again with a nominal 90% coverage.

The estimate of Müller (1992) is considered in the right panel of Figure 2, using the boundary kernel $K(u) = 12u(1-u)(3-5u)I(0 \leq u \leq 1)$. Similarly large bandwidths are required to detect the change; however, the minimum MAD achieved is 5. The estimate is more sensitive to the bandwidth, with bias sometimes dominating $\hat{\Delta}_t$ for $nh \geq 220$. The asymptotic distribution in this case is $n(\hat{\tau} - \tau) \sim N(0, 0.296nh)$ by Müller (1992 Corollary 3.1).

3 Proofs

To prove Theorem 1, we consider the behavior of $\hat{\Delta}_t$ in two parts. Lemma 1 considers the case $|t - \tau| \geq i_0/n$, where $i_0 \to \infty$ as $n \to \infty$, with

$$i_0 = o(\min(\sqrt{nh/\log n}, h^{-1})). \tag{5}$$

Lemma 2 considers $|t - \tau| < i_0/n$.

Lemma 1 *$\hat{\tau}$ is a consistent estimate of τ;*

$$P(|\hat{\tau} - \tau| > i_0/n) \to 0.$$

Lemma 2 *Let $m = n\tau$. As $n \to \infty$,*

$$\frac{nh}{2M_1}(\hat{\Delta}^2_{\tau+i/n} - \hat{\Delta}^2_\tau) = -\Delta(\epsilon_m + \ldots + \epsilon_{m+i-1}) - \frac{i\Delta^2}{2} + o(1) \tag{6}$$

$$\frac{nh}{2M_1}(\hat{\Delta}^2_{\tau-i/n} - \hat{\Delta}^2_\tau) = \Delta(\epsilon_{m-i} + \ldots + \epsilon_{m-1}) - \frac{i\Delta^2}{2} + o(1). \tag{7}$$

The $o(1)$ term holds uniformly for $1 \le i \le i_0$.

For simplicity, proofs will be for local constant fitting ($p = 0$) only; for general p the results follow by considering the asymptotically equivalent kernels $K^*(u) = K(u)[\Lambda_1^{-1}A(u)]_1$. We first apply the results to prove Theorems 1, 2 and 3. Let $\hat{\tau}'$ be the value of t that maximizes $\hat{\Delta}_t^2$ over $[\tau - i_0/n, \tau + i_0/n]$. Then by Lemma 2,

$$P(n(\hat{\tau}' - \tau) = l) \to P(L_\Delta = l).$$

Since $\hat{\tau} = \hat{\tau}'$ whenever $n|\hat{\tau} - \tau| < i_0$,

$$P(n(\hat{\tau}' - \tau) = l) - P(n|\hat{\tau} - \tau| > i_0) \le P(n(\hat{\tau} - \tau) = l) \le P(n(\hat{\tau}' - \tau) = l)$$

and an application of Lemma 1 completes the proof of Theorem 1. The proof of Theorem 2 is similar.

To prove Theorem 3 we need to show

$$\sqrt{nh}(\hat{\Delta}_\tau - \Delta) \xrightarrow{\mathcal{L}} N(0, M_2) \tag{8}$$

and is asymptotically independent of $nh(\hat{\Delta}_{\hat{\tau}}^2 - \hat{\Delta}_\tau^2)$. Since $\hat{\Delta}_t$ is normally distributed, to establish (8) it suffices to show convergence of moments. Using the continuity of K and f,

$$\begin{aligned} |E\hat{f}_+(\tau) - f(\tau_+)| &\le \frac{\sum_{j=0}^{nh} |K(\frac{j}{nh})(f(\tau + \frac{j}{n}) - f(\tau_+))|}{S(nh)} \\ &\le h\beta \frac{\sum_{j=0}^{nh} \frac{j}{nh} |K(\frac{j}{nh})|}{S(nh)} \\ &= O(h) \end{aligned} \tag{9}$$

where $S(nh) = \sum_{j=0}^{nh} K(\frac{j}{nh})$; note $(nh)^{-1}S(nh) \to \int_0^1 K(u)du = 1$. Treating $\hat{f}(\tau_-)$ similarly and using the assumption $nh^3 \to 0$,

$$\sqrt{nh}E(\hat{\Delta}_\tau - \Delta) + O((nh^3)^{\frac{1}{2}}) \to 0$$

Evaluating the variance of $\hat{\Delta}_\tau$ from (2) is straightforward.

The asymptotic independence follows because $\hat{\Delta}_\tau$ depends on Y_j, $|n\tau - j| < nh$, while $\hat{\Delta}_{\hat{\tau}}^2 - \hat{\Delta}_\tau^2$ depends only on Y_j in the negligible subinterval $|n\tau - j| < i_0$.

It remains to prove Lemmas 1 and 2. The following Lemma is used repeatedly.

Lemma 3 *Let $\eta_{n,j}; j = 1, \ldots, n; n \ge 1$ be a triangular array of $N(0, \sigma_{n,j}^2)$ random variables; not necessarily independent. Let $M_n = \sup_{1 \le j \le n} |\eta_{n,j}|$. If $\sup_{1 \le j \le n} \sigma_{n,j}^2 \log n \to 0$ as $n \to \infty$, then $M_n \to 0$ with probability 1.*

Proof: Fix $\delta > 0$. If n is sufficiently large, then $\sigma_{n,j}^2 \log n \leq \delta^2/5$ for all j. Using Bonferroni's inequality and the bound $1 - \Phi(c) \leq \phi(c)/c$,

$$\begin{aligned} P(M_n > \delta) &\leq \sum_{j=1}^{n} 2(1 - \Phi(\delta/\sigma_{n,j})) \\ &\leq 2n(1 - \Phi(\sqrt{5 \log n})) \\ &\leq \frac{2}{n^{3/2}\sqrt{10\pi \log n}}. \end{aligned}$$

Hence, $\sum_{n=1}^{\infty} P(M_n > \delta) < \infty$ and by the Borel-Cantelli lemma, $M_n > \delta$ only finitely often. Since δ is arbitrary, this implies $M_n \to 0$.

□

Proof of Lemma 1: We assume $\Delta > 0$; the case $\Delta < 0$ is similar. Essentially following (9) and a similar bound for $E(\hat{f}_{-}(t))$,

$$\begin{aligned} \sup_{t \geq \tau + h} |E\hat{\Delta}_t| &= O(h) \to 0; \\ \mathrm{var}(\hat{\Delta}_t) &= \frac{M_2}{nh} + o((nh)^{-1}). \end{aligned}$$

Applying Lemma 3 gives $\sup_t |\hat{\Delta}_t - E\hat{\Delta}_t| \to 0$ with probability 1. Hence, $\sup_{t > \tau + h} \hat{\Delta}_t \to 0$, and $\hat{\Delta}_\tau \to \Delta$. Therefore,

$$P(\hat{\tau} > \tau + h) \leq P(\hat{\Delta}_\tau < \frac{1}{2}\Delta) + P(\sup_{t > \tau + h} \hat{\Delta}_t > \frac{1}{2}\Delta) \to 0. \tag{10}$$

Now,

$$\begin{aligned} P(\tau + i_0/n \leq \hat{\tau} \leq \tau + h) &\leq \sum_{i=i_0}^{nh} P(\hat{\Delta}_{\tau+i/n}^2 \geq \hat{\Delta}_\tau^2) \\ &\leq \sum_{i=i_0}^{nh} P(\hat{\Delta}_{\tau+i/n} - \hat{\Delta}_\tau > 0) + \sum_{i=i_0}^{nh} P(\hat{\Delta}_{\tau+i/n} + \hat{\Delta}_\tau < 0). \end{aligned}$$

We show the first sum converges to 0; the second is easier. Since the terms are normal tail probabilities, we need to approximate the mean and variance of $\hat{\Delta}_{\tau+i/n} - \hat{\Delta}_\tau$. From the definition of $\hat{\Delta}_t$,

$$\begin{aligned} S(nh)E(\hat{\Delta}_{\tau+i/n} - \hat{\Delta}_\tau) &= \sum_{j=0}^{nh} K(\frac{j}{nh})\left(f(\tau + \frac{j+i}{n}) - f(\tau + \frac{j}{n})\right) \\ &\quad - \sum_{j=0}^{i-1} K(\frac{j}{nh})\left(f(\tau + \frac{i-1-j}{n}) - f(\tau - \frac{j+1}{n})\right) \\ &\quad + \sum_{j=i}^{nh} K(\frac{j}{nh})\left(f(\tau + \frac{i-1-j}{n}) - f(\tau - \frac{j+1}{n})\right). \end{aligned}$$

By (1), the difference of the f's can be bounded by $\pm\beta i/n$ in the first and third sums, and $\Delta \pm \beta i/n$ in the second sum. Also, $|K(u)| \leq z$ for all u. Hence,

$$S(nh)E(\hat{\Delta}_{\tau+i/n} - \hat{\Delta}_\tau) \leq -\Delta \sum_{j=0}^{i-1} K(\frac{j}{nh}) + 5hz\beta i.$$

Since $K(0) > 0$, there exists $c > 0$ such that for n sufficiently large and all $i \leq nh$,

$$S(nh)E(\hat{\Delta}_{\tau+i/n} - \hat{\Delta}_\tau) \leq -ci \tag{11}$$

Similarly,

$$\begin{aligned} S(nh)^2 \mathrm{var}(\hat{\Delta}_{\tau+i/n} - \hat{\Delta}_\tau) &= 2\sum_{j=0}^{nh} \left(K(\frac{j+i}{nh}) - K(\frac{j}{nh})\right)^2 + \sum_{j=0}^{i-1} \left(K(\frac{i-j-1}{nh}) + K(\frac{j}{nh})\right)^2 \\ &\leq 3z^2 i. \end{aligned} \tag{12}$$

Combining (11) and (12),

$$\begin{aligned} \sum_{i=i_0}^{nh} P(\hat{\Delta}_{\tau+i/n} - \hat{\Delta}_\tau > 0) &\leq \sum_{i=i_0}^{nh} (1 - \Phi(\frac{c\sqrt{i}}{\sqrt{8}z})) \\ &\leq \sum_{i=i_0}^{nh} \frac{C}{\sqrt{i}} e^{-\lambda i} \end{aligned}$$

where C and λ are positive constants, using the bound $1 - \Phi(x) \leq \phi(x)/x$. The sum is asymptotic to

$$\frac{Ce^{-\lambda i_0}}{\sqrt{i_0}(1 - e^{-\lambda})} \to 0$$

since $i_0 \to \infty$. Hence, $P(\tau + i_0/n < \hat{\tau} < \tau + h) \to 0$. Combining with (10) gives $P(\hat{\tau} > \tau + i_0/n) \to 0$. Treating the left tail similarly completes the proof of Lemma 1.

□

Proof of Lemma 2: We prove only (6); a proof of (7) is similar. Along the lines of the proof of Lemma 1, $(\hat{\Delta}_t + \hat{\Delta}_\tau)/2 \to \Delta$ uniformly for $|t - \tau| < i_0/n$, and it therefore suffices to show

$$\frac{nh}{M_1}(\hat{\Delta}_{\tau+i/n} - \hat{\Delta}_\tau) = -(\epsilon_m + \ldots + \epsilon_{m+i-1}) - \frac{i\Delta}{2} + o(1)$$

where $m = n\tau$. We achieve this by showing

$$S(nh)(\hat{f}_-(\tau + i/n) - \hat{f}_-(\tau)) = iK(0)\Delta + K(0)(\epsilon_m + \ldots + \epsilon_{m+i-1}) + o(1) \tag{13}$$

$$S(nh)(\hat{f}_+(\tau + i/n) - \hat{f}_+(\tau)) = -K(0)(\epsilon_m + \ldots + \epsilon_{m+i-1}) + o(1) \tag{14}$$

where $o(1)$ holds uniformly in $i < i_0$; note $M_1 = 2K(0)$ for local constant fitting.

Using the definition of $\hat{f}_-(t)$,

$$\begin{aligned}
S(nh)(\hat{f}_-(\tau+i/n)-\hat{f}_-(\tau)) &= \sum_{j=-i}^{nh} K(\frac{j+i}{nh})Y_{m-j-1} - \sum_{j=0}^{nh} K(\frac{j}{nh})Y_{m-j-1} \\
&= \sum_{j=-i}^{-1} K(\frac{j+i}{nh})(f(x_{m-j-1})+\epsilon_{m-j-1}) \\
&\quad + \sum_{j=0}^{nh}\left(K(\frac{j+i}{nh})-K(\frac{j}{nh})\right)(f(x_{m-j-1})+\epsilon_{m-j-1}). \quad (15)
\end{aligned}$$

We treat this in four parts. Using the Lipschitz continuity of f and K,

$$\begin{aligned}
\sum_{j=-i}^{-1} K(\frac{j+i}{nh})f(x_{m-j-1}) - iK(0)f(\tau_+) &\leq \sum_{j=-i}^{-1}\left(\frac{K(0)\beta i_0}{n}+\frac{|f(\tau_+)|zi_0}{nh}+\frac{z\beta i_0^2}{n^2h}\right) \\
&\leq \frac{K(0)\beta i_0^2}{n}+\frac{|f(\tau_+)|zi_0^2}{nh}+\frac{\beta z i_0^3}{n^2h}.
\end{aligned}$$

Applying (5) and a similar lower bound shows

$$\sum_{j=-i}^{-1} K(\frac{j+i}{nh})f(x_{m-j-1}) = iK(0)f(\tau_+)+o(1) \quad (16)$$

where $o(1)$ holds uniformly in i, $1 \leq i \leq i_0$. Similarly,

$$\begin{aligned}
&\sum_{j=0}^{nh}\left(K(\frac{j+i}{nh})-K(\frac{j}{nh})\right)f(x_{m-j-1}) \\
&\leq f(\tau_-)\sum_{j=0}^{nh}\left(K(\frac{j+i}{nh})-K(\frac{j}{nh})\right)+\beta h\sum_{j=0}^{nh}\left|K(\frac{j+i}{nh})-K(\frac{j}{nh})\right| \\
&= -iK(0)f(\tau_-)+o(1) \quad (17)
\end{aligned}$$

using $i_0 = o(h^{-1})$.

Turning to the random components of (15),

$$\begin{aligned}
\operatorname{var}\sum_{j=-i}^{-1}(K(\frac{j+i}{nh})-K(0))\epsilon_{m-j-1} &= \sum_{j=-i}^{-1}(K(\frac{j+i}{nh})-K(0))^2 \\
&\leq \frac{i_0^3z^2}{(nh)^2}
\end{aligned}$$

and applying Lemma 3,

$$\sum_{j=-i}^{-1} K(\frac{j+i}{nh})\epsilon_{m-j-1} = K(0)(\epsilon_m+\ldots+\epsilon_{m+i-1})+o(1) \quad (18)$$

uniformly in $i < i_0$. Similarly,

$$\operatorname{var} \sum_{j=0}^{nh} \left(K(\frac{j+i}{nh}) - K(\frac{j}{nh}) \right) \epsilon_j \le \frac{i_0^2 z^2}{nh}$$

and another application of Lemma 3 gives

$$\sum_{j=0}^{nh} \left(K(\frac{j+i}{nh}) - K(\frac{j}{nh}) \right) \epsilon_j = o(1). \tag{19}$$

Substituting (16), (17), (18) and (19) into (15) establishes (13). A similar derivation of (14) completes the proof.

□

REFERENCES

DÜMBGEN, L. (1991). The asymptotic behavior of some nonparametric change-point estimates. *Ann. Statist.* **19**, 1471-1495.

FAN, J. and GIJBELS, I. (1992). Variable bandwidth and local linear regression smoothers. *Ann. Statist.* **20**, 2008-2036.

HINKLEY, D. V. (1970). Inference about the change-point in a sequence of random variables. *Biometrika* **57**, 1-17.

IBRAGIMOV, I. A. and HAS'MINSKII, R. Z. (1981). *Statistical Estimation: Asymptotic Theory.* Springer, New York.

HALL, P. and TITTERINGTON, D. M. (1992). Edge-preserving and peak-preserving smoothing. *Technometrics* **34**, 429-440.

KIM, H.-J. and SIEGMUND, D. O. (1989). The likelihood ratio test for a change point in simple linear regression. *Biometrika* **76**, 409-423.

MCDONALD, J. A. and OWEN, A. B. (1986). Smoothing with split linear fits. *Technometrics* **28**, 195-208.

MÜLLER, H.-G. (1992). Change-points in nonparametric regression analysis. *Ann. Statist.* **20**, 737-761.

RITOV, Y. (1990). Asymptotic efficient estimation of the change point with unknown distributions. *Ann. Statist.* **18**, 1829-1839.

SIEGMUND, D. O. (1988). Confidence sets in change-point problems. *Internat. Statist. Rev.* **56**, 31-48.

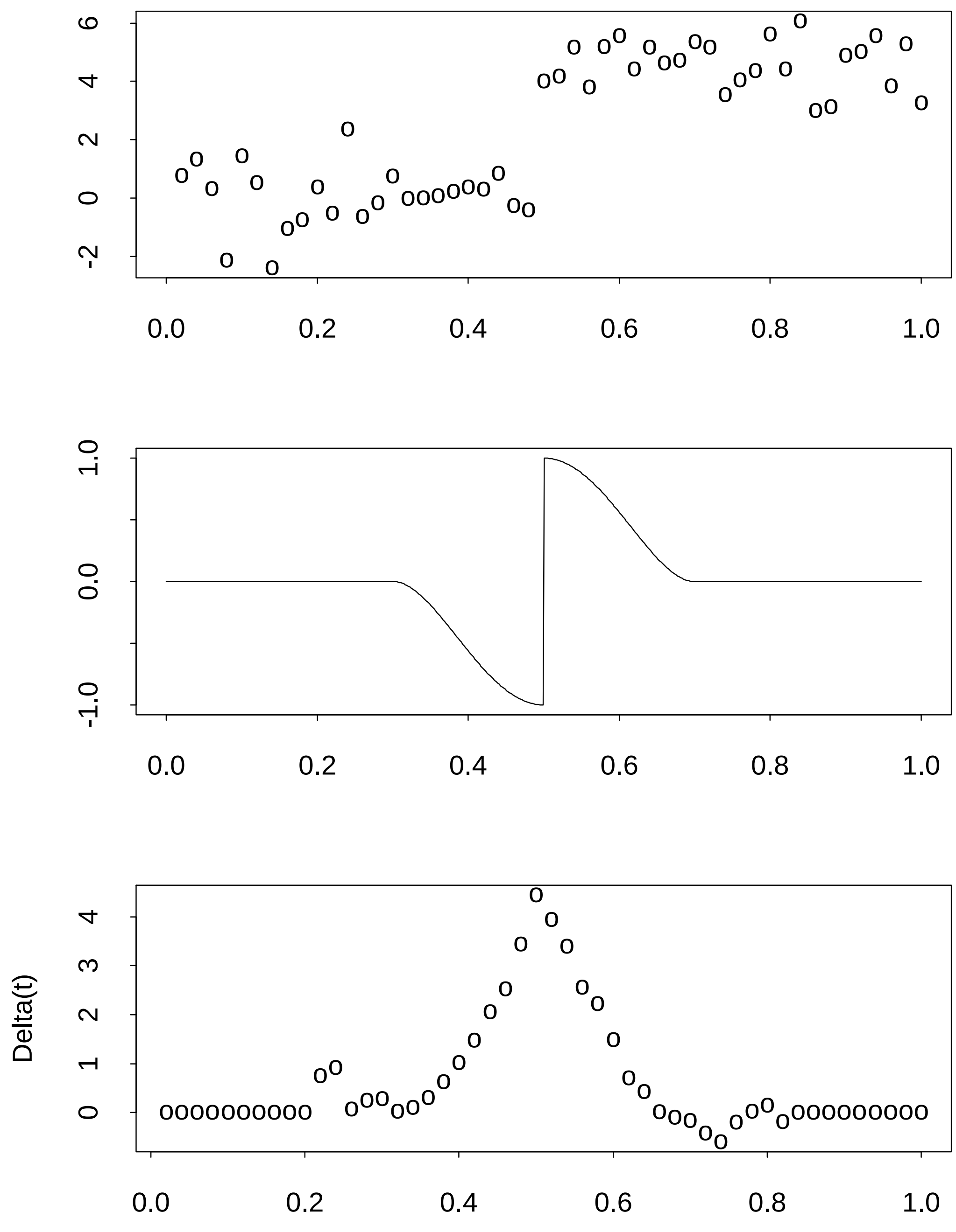

Figure 1: How the change point estimate works. A dataset (top) is convoluted with a split kernel (middle) to produce $\hat{\Delta}_t = \hat{f}_+(t) - \hat{f}_-(t)$ (bottom). The estimate is that value of t which maximizes $\hat{\Delta}_t$.

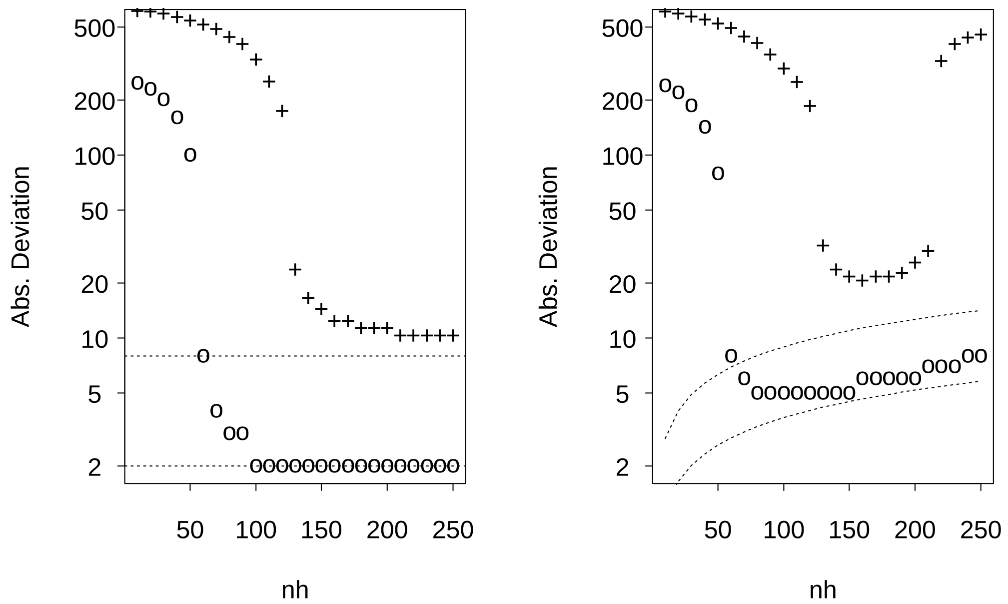

Figure 2: Comparison of our estimate (left) with that of Müller (1992) (right), for a change of size $\Delta = 1$. Plotted are the 50th percentile (o) and 90th percentile (+) of the absolute deviation $n|\hat{\tau} - \tau|$, estimated by Monte Carlo simulation. Dashed lines are the asymptotic approximations to the quantiles.