Electronic Journal of Statistics
Vol. 7 (2013) 2477–2515
ISSN: 1935-7524
DOI: 10.1214/13-EJS845

Asymptotics for p-value based threshold estimation in regression settings

Atul Mallik*, **Moulinath Banerjee***

Department of Statistics
University of Michigan
Ann Arbor, Michigan 48109
e-mail: atulm@umich.edu; moulib@umich.edu

and

Bodhisattva Sen†

Department of Statistics
Columbia University
New York, New York 10027
e-mail: bodhi@stat.columbia.edu

Abstract: We investigate the large sample behavior of a p-value based procedure for estimating the threshold level at which a regression function takes off from its baseline value – a problem that frequently arises in environmental statistics, engineering and other related fields. The estimate is constructed via fitting a "stump" function to approximate p-values obtained from tests for deviation of the regression function from its baseline level. The smoothness of the regression function in the vicinity of the threshold determines the rate of convergence: a "cusp" of order k at the threshold yields an optimal convergence rate of $n^{-1/(2k+1)}$, n being the number of sampled covariates. We show that the asymptotic distribution of the normalized estimate of the threshold, for both i.i.d. and short range dependent errors, is the minimizer of an integrated and transformed Gaussian process. We study the finite sample behavior of confidence intervals obtained through the asymptotic approximation using simulations, consider extensions to short-range dependent data, and apply our inference procedure to two real data sets.

AMS 2000 subject classifications: Primary 62G20, 62G86; secondary 62G30.
Keywords and phrases: Baseline value, change-point, integral of a transformed Gaussian process, least squares, nonparametric estimation, stump function.

Received March 2013.

1. Introduction

Consider a data generating model of the form $Y = \mu(X) + \epsilon$, where μ is a *continuous* function on $[0,1]$ such that $\mu(x) = \tau$ for $x \leq d_0$, and $\mu(x) > \tau$ for $x > d_0$. The covariate X may arise from a random or a fixed design setting and we as-

*Supported by NSF Grant DMS-1007751, NSA grant H98230-11-1-0166 and a Sokol Faculty Award, University of Michigan.
†Supported by NSF Grants DMS-0906597 and AST-1107373.

sume that ϵ has mean zero with finite positive variance. The function μ need not be monotone and the baseline value τ is not necessarily known. We are interested in estimating and constructing confidence intervals (CI's) for the threshold d_0, $d_0 \in (0,1)$, the point from where the function starts to deviate from its baseline value.

Recently, Mallik et al. (2011) introduced a novel method for estimating the threshold d_0 using a p-value based criterion function but did not provide a recipe for constructing confidence intervals (CI's) for the point d_0. In this paper, we address the *inference problem*: we study the asymptotic properties of their procedure in the regression setting and use these results to construct asymptotically valid CI's for the threshold, both in simulation settings and for two key motivating examples from Mallik et al. (2011). The problem, which falls within the sphere of non-regular M-estimation is rather hard, and involves non-trivial applications of techniques from modern empirical processes, as well as results from martingale theory and the theory of Gaussian processes. Along the way, we also deduce results on the large sample behavior of a kernel estimator at local points (see Lemma 2 and Proposition 4) that are of independent interest. In most of the literature, kernel estimates are considered at various fixed points and are asymptotically independent (Csörgő and Mielniczuk, 1995a,b; Robinson, 1997). Hence, they do not admit a functional limit. However, these estimates, when considered at local points, deliver an invariance principle; see Lemma 2 and the proof of Proposition 4.

It is instructive to note that our problem of interest has natural connections with change-point analysis, though it is *not* a change-point problem for the regression function itself. Under mild assumptions, estimating d_0 can be treated as a problem of detecting a change-point in the derivative of a certain order of μ. The literature on change-point detection is enormous; earlier work includes Hinkley (1970), Korostelëv (1987), Dümbgen (1991), Müller (1992), Korostelëv and Tsybakov (1993), Loader (1996), Müller and Song (1997). We also refer the reader to Brodsky and Darkhovsky (1993), Bhattacharya (1994) and Csörgő and Horváth (1997) for an overview of results on change-point estimation in various settings. Change-points have been extensively studied in time-series and sequential problems as well; see Horváth, Horváth and Hušková (2008), Hušková et al. (2008), Gut and Steinebach (2009), Steland (2010) and the references therein. There has also been some work in settings with gradually changing regression functions (local alternative type formulations) but under certain parametric assumptions (Hušková, 1998; Jarušková, 1998).

In particular, the problem of estimating change-points in derivatives within the context of regression has been addressed by a number of authors, e.g. Müller (1992), Cheng and Raimondo (2008), Wishart (2009), Wishart and Kulik (2010). We compare and contrast our approach to the work of these authors in the next section. We further note that our problem can be viewed as a special case of the more general problem of identifying the region where a function, defined on some multi-dimensional space, assumes its baseline (minimum or maximum) value. This problem is relevant to applications in fMRI and image detection (Willett and Nowak, 2007).

In what follows we show that the smoothness of the function in the vicinity of d_0 determines the rate of convergence of our estimator: for a "cusp" of order k at d_0, the best possible rate of convergence turns out to be $n^{-1/(2k+1)}$. The limiting distribution of an appropriately normalized version of the estimator is that of the minimizer of the integral of a transformed Gaussian process. The limiting process is new, and while the uniqueness of the minimizer remains unclear (and appears to be an interesting nontrivial exercise in probability), we can bypass the lack of uniqueness and still provide a thorough mathematical framework to construct honest CI's. Under the assumption of uniqueness, which appears to be a reasonable conjecture based on extensive simulations, we establish auxiliary results to construct asymptotically exact CI's.

The paper is organized thus: we briefly discuss the estimation procedure and the basic assumptions in Section 2. The rate of convergence and the asymptotic distribution of the estimated threshold for a particular version of our procedure, along with some auxiliary results for constructing CI's, are deduced in Sections 3.1 and 3.2, assuming *a known* τ. Asymptotic results for variants of the procedure are discussed in Section 3.3 and extensions of these results to the situation with *an unknown* τ are presented in Section 4. We study the coverage performance of the resulting CI's through simulations in Section 5. The applicability of our approach to short-range dependent data is the content of Section 6. We implement our procedure to two data examples in Section 7 and end with a discussion in Section 8. The proofs of several technical results are available in the Appendix.

2. The method

For simplicity, we first consider the uniform fixed design regression model of the form:

$$Y_i = \mu\left(\frac{i}{n}\right) + \epsilon_i, \qquad 1 \le i \le n, \tag{2.1}$$

with ϵ_i's i.i.d. having variance σ_0^2. Although we suppress the dependence on n, Y_i and ϵ_i must be viewed as triangular arrays. Let K be a symmetric probability density (kernel) and $h_n = h_0 n^{-\lambda}$ denote the smoothing bandwidth, for some $\lambda \in (0,1), h_0 > 0$. Then an estimate of the regression function, at stage n, is given by

$$\hat{\mu}(x) = \frac{1}{nh_n}\sum_{i=1}^{n} Y_i K\left(\frac{x - i/n}{h_n}\right). \tag{2.2}$$

For $x < d_0$, the statistic $\sqrt{nh_n}(\hat{\mu}(x) - \tau)$, whose variance is

$$\Sigma_n^2(x) = \Sigma_n^2(x, \sigma_0) = \mathrm{Var}(\sqrt{nh_n}\hat{\mu}(x)) = \frac{\sigma_0^2}{nh_n}\sum_{i=1}^{n} K^2\left(\frac{x - i/n}{h_n}\right), \tag{2.3}$$

converges to a normal distribution with zero mean and variance $\Sigma^2(x) = \sigma_0^2 \bar{K^2}$ with $\bar{K^2} = \int K^2(u)du$. Let $\hat{\sigma}$ be an estimator of σ_0. Mallik et al. (2011) estimate

d_0 by constructing p-values for testing the null hypothesis $H_{0,x} : \mu(x) = \tau$ against the alternative $H_{1,x} : \mu(x) > \tau$. The approximate p-values are

$$p_n(x,\tau) = 1 - \Phi\left(\frac{\sqrt{nh_n}(\hat{\mu}(i/n) - \tau)}{\Sigma_n(i/n, \hat{\sigma})}\right).$$

To the left of d_0, the null hypothesis holds and these approximate p-values converge weakly to the Uniform(0,1) distribution which has mean 1/2. Moreover, to the right of d_0, where the alternative is true, the p-values converge in probability to 0. This dichotomous behavior of the p-values motivates minimizing

$$\left[\sum_{i:i/n\le d}\left\{p_n(i/n,\tau) - \frac{1}{2}\right\}^2 + \sum_{i:i/n>d}\{p_n(i/n,\tau)\}^2\right] \tag{2.4}$$

over values of d in $(0,1)$ to yield an estimate of d_0. Simple calculations show that this is equivalent to minimizing

$$\begin{aligned}\widetilde{\mathbb{M}}_n(d) &\equiv \widetilde{\mathbb{M}}_n(d,\hat{\sigma}) \\ &= \frac{1}{n}\sum_{i=1}^{n}\left\{\Phi\left(\frac{\sqrt{nh_n}(\hat{\mu}(i/n) - \tau)}{\Sigma_n(i/n,\hat{\sigma})}\right) - \gamma\right\} 1\left(\frac{i}{n} \le d\right).\end{aligned} \tag{2.5}$$

Here, $\gamma = 3/4$. We refer to the above approach as **Method 1**.

We next describe an approach which avoids estimating σ_0 altogether. Relying upon the simple fact that $E[\Phi(Z)] = 1/2$ for a normally distributed Z with zero mean and *arbitrary* variance, it can be seen that $E[\Phi(\sqrt{nh_n}(\hat{\mu}(x) - \tau))]$ converges to 1/2 for $x < d_0$, while for $x > d_0$, it converges to 1. So, the desired dichotomous behavior is preserved even without normalization by the estimate of the variance. To this end, let

$$\mathbb{M}_n(d) = \frac{1}{n}\sum_{i=1}^{n}\left[\Phi\left(\sqrt{nh_n}\left(\hat{\mu}\left(\frac{i}{n}\right) - \tau\right)\right) - \gamma\right] 1\left(\frac{i}{n} \le d\right). \tag{2.6}$$

Then, an estimate of d_0 is given by

$$\hat{d}_n = \underset{d\in[0,1]}{\text{sargmin}}\, \mathbb{M}_n(d),$$

where sargmin denotes the smallest argmin of the criterion function, which does not have a unique minimum. We refer to this approach as **Method 2** and study its limiting behavior in the paper. Analyzing this method is useful in illustrating the core ideas while avoiding some of the tedious details encountered in analyzing **Method 1**.

Remark 1. The above methods are based on a known τ. When τ is unknown, a plug-in estimate can be substituted in its place (more about this in Section 4). Also, for any choice of $\gamma \in (1/2, 1)$ in (2.5) and (2.6), the estimator of d_0 is consistent. The proof follows along the lines of arguments in Mallik et al. (2011, pp. 898–900).

2.1. Variants

If the covariate X is random, one could still use the Nadaraya-Watson estimator to construct p-values. More precisely, let $\mathbb{P}_n$ denote the empirical measure of $(Y_i, X_i), i = 1, \ldots, n$, which are independent realizations from the model $Y = \mu(X) + \epsilon$ with $\sigma_0^2(x) = \text{Var}(\epsilon \mid X = x) > 0$ and X having a continuous positive density f on $[0, 1]$. Then the Nadaraya-Watson estimator is:

$$\tilde{\mu}(x) = \frac{\mathbb{P}_n[YK((x - X)/h_n)]}{\mathbb{P}_n[K((x - X)/h_n)]}.$$

and a consistent estimator of d_0 is:

$$\tilde{d}_n = \operatorname*{sargmin}_{d \in (0,1)} \mathbb{P}_n \left[\left\{ \Phi\left(\sqrt{nh_n} (\tilde{\mu}(X) - \tau) \right) - \gamma \right\} 1(X \leq d) \right] \tag{2.7}$$

(see Mallik et al. (2011, pp. 892)). An estimate based on p-values normalized by the estimate of variance can also be constructed but is computationally more complicated as an estimate of the variance function is needed.

2.2. Basic assumptions

We adhere to the setup of Section 2, i.e., we assume the errors to be independent and homoscedastic and consider a fixed design for the regression setting. The smoothness of the function in the vicinity of d_0 plays a crucial role in determining the rate of convergence. Throughout this paper, we make the following assumptions.

1. Assumptions on μ:
   (a) μ is continuous on $[0, 1]$. We additionally assume that μ is Lipschitz continuous of order α_1 with $\alpha_1 \in (1/2, 1]$.
   (b) μ has a cusp of order k, k being a *known* positive integer, at d_0, i.e., $\mu^{(l)}(d_0) = 0, 1 \leq l \leq k - 1$, and $\mu^{(k)}(d_0+) > 0$, where $\mu^{(l)}(\cdot)$ denotes the l-th derivative of μ. Also, the k-th derivative, $\mu^{(k)}(x)$ is assumed to be continuous and bounded for $x \in (d_0, d_0 + \zeta_0]$ for some $\zeta_0 > 0$.
2. The error ϵ possesses a continuous positive density on an interval.
3. Assumptions on the kernel K:
   (a) K is a symmetric probability density.
   (b) $K(u)$ is non-increasing in $|u|$.
   (c) K is compactly supported, i.e., $K(x) = 0$ when $|x| \geq L_0$, for some $L_0 > 0$.
   (d) K is Lipschitz continuous of order $\alpha_2 \in (1/2, 1]$.

As a consequence of these assumptions, μ and K are bounded, $\bar{K^2} = \int K^2(u)du < \infty$ and $E|W|^k < \infty$, where W has density K. Also, both μ and K are Lipschitz continuous of order $\alpha = \min(\alpha_1, \alpha_2)$. These facts are frequently used in the paper. Common kernels such as the Epanechnikov kernel and the triangular kernel conveniently satisfy the assumptions mentioned above. The results in the next section are developed for a $\gamma \in (1/2, 1)$ (cf. Remark 1) and a known τ. It will be seen in Section 4 that τ can be estimated at a sufficiently fast rate; consequently, even if τ is unknown, appropriate estimates can be substituted in its place to construct the p-values that are instrumental to the methods of this paper, without changing the limit distributions. Without loss of generality, we take $\tau \equiv 0$ in the next section, as one can work with $(Y_i - \tau)$s in place of Y_is.

Comparison with existing approaches. Under the above assumptions, d_0 is a 'change-point' in the k-th derivative of μ. Our procedure for estimating this change-point relies on the discrepancy of p-values, the construction of which requires a kernel-smoothed estimate (or if one desires, local polynomial estimate) of μ. As noted in the Introduction, the estimation of a change-point in the derivative of a regression function has been studied by a number of authors using kernel-based strategies. However, the approaches in these papers are quite different from ours and more importantly, our problem cannot be solved by these methods without making stronger model assumptions than those above. In Müller (1992), the change-point is obtained by direct estimation of the k-th derivative (k corresponds to ν in that paper) on either side of the change-point via one-sided kernels and measuring the difference between these estimates. In contrast, our approach does not rely on derivative estimation. We use an ordinary kernel function to construct a smooth estimate of μ which is required for the point wise testing procedures that lead to the p-values. In fact, a consistent estimate that attains the same rate of convergence as our current estimate could have been constructed using a simple regressogram estimator with an appropriate bin-width, in contrast to the approach in Müller (1992) which uses a k–times differentiable kernel. Müller (1992) also assumes that the k-th derivative of the regression function is at least twice continuously differentiable at all points except d_0 – see, pages 738–739 of that paper – which is stronger than our continuity assumption on $\mu^{(k)}$ (1(b) above). Cheng and Raimondo (2008) develop kernel methods for optimal estimation of the first derivative building on an idea by Goldenshluger, Tsybakov and Zeevi (2006), which is followed up in the context of dependent errors by Wishart and Kulik (2010), and Wishart (2009), but these papers do not consider the case $k > 1$. We also note that our method is fairly simple to implement.

3. Main results

We state and prove results on the limiting behavior of the estimator obtained from Method 2. Results on the variant of the procedure discussed in Section 2.1 follow analogously and are stated in Section 3.3. We consider the model stated in (2.1) with homoscedastic errors and uniform fixed design, and study the limiting behavior of $\hat{d}_n$ which minimizes (2.6). Recall that τ is taken to be zero without loss of generality.

3.1. Rate of convergence

We first consider the population equivalent of $\mathbb{M}_n$, given here by $M_n(d) = E\{\mathbb{M}_n(d)\}$, and study the behavior of its smallest argmin. Let

$$Z_{in} = \frac{1}{\sqrt{nh_n}} \sum_{l=1}^{n} \epsilon_l K\left(\frac{i/n - l/n}{h_n}\right),$$

for $i = 1, \ldots, n$, and Z_0 be a standard normal random variable independent of Z_{in}'s. Also, let

$$\bar{\mu}(x) = \frac{1}{nh_n} \sum_{l=1}^{n} \mu\left(\frac{l}{n}\right) K\left(\frac{x - l/n}{h_n}\right). \tag{3.1}$$

Note that $\sqrt{nh_n}\hat{\mu}(i/n) = \sqrt{nh_n}\bar{\mu}(i/n) + Z_{in}$ and $\mathrm{Var}(Z_{in}) = \Sigma^2(i/n)$ with $\Sigma(\cdot)$ as in (2.3). We have

$$\begin{aligned} E\left[\Phi\left(\sqrt{nh_n}\hat{\mu}(i/n)\right)\right] &= E\left[\Phi\left(\sqrt{nh_n}\bar{\mu}(i/n) + Z_{in}\right)\right] \\ &= E\left[1\left(Z_0 \le \sqrt{nh_n}\bar{\mu}(i/n) + Z_{in}\right)\right] \\ &= \Phi_{i,n}\left(\frac{\sqrt{nh_n}\bar{\mu}(i/n)}{\sqrt{1 + \Sigma_n^2(i/n)}}\right), \end{aligned} \tag{3.2}$$

where $\Phi_{i,n}$ denotes the distribution function of $(Z_0 - Z_{in})/\sqrt{1 + \Sigma_n^2(i/n)}$. Hence,

$$M_n(d) = \frac{1}{n} \sum_{i=1}^{n} \left\{\Phi_{i,n}\left(\frac{\sqrt{nh_n}\bar{\mu}(i/n)}{\sqrt{1 + \Sigma_n^2(i/n)}}\right) - \gamma\right\} 1\left(\frac{i}{n} \le d\right).$$

For $L_0 h_n \le i/n \le 1 - L_0 h_n$, $\Phi_{i,n}$'s and $\Sigma_n(i/n)$'s do not vary with i. We denote them by $\tilde{\Phi}_n$ and $\tilde{\Sigma}_n$ for convenience. Using Corollary 1 and (A.1) from the Appendix, $\tilde{\Sigma}_n$ converges to $\sigma_0\sqrt{\bar{K^2}}$. Also, for such i's, any $\eta > 0$ and sufficiently large n,

$$\frac{1}{nh_n\tilde{\Sigma}_n^2} \sum_{l:|l-i| \le L_0 nh_n}^{n} E\left[\epsilon_l^2 K^2\left(\frac{(i-l)/n}{h_n}\right) 1\left(\frac{|\epsilon_l| K\left((i-l)/(nh_n)\right)}{\sqrt{nh_n}\tilde{\Sigma}_n(l/n)} > \eta\right)\right]$$

is bounded from above by

$$\frac{2\lceil 2L_0 nh_n \rceil \|K\|_\infty^2}{nh_n(\sigma_0^2 \bar{K^2})} E\left[\epsilon_1^2 1\left(\frac{2\|K\|_\infty}{nh_n(\sigma_0\sqrt{\bar{K^2}})}|\epsilon_1| > \eta\right)\right],$$

which converges to zero. Hence, by Lindeberg–Feller CLT, $Z_{in}/\tilde{\Sigma}_n$ and consequently, $\tilde{\Phi}_n$ converge weakly to Φ. In fact, for any i, we can also show that $\Phi_{i,n}$ converges weakly to Φ.

Let $d_n = \mathrm{sargmin}_d M_n(d)$. As mentioned earlier, sargmin denotes the smallest argmin of the objective function M_n which does not have a unique minimizer. The following lemma provides the rate at which d_n converges to d_0.

Lemma 1. *Let* $\nu_n = \min(h_n^{-1}, (nh_n)^{1/2k})$. *Then* $\nu_n(d_n - d_0) = O(1)$.

Proof. It can be shown by arguments analogous to Mallik et al. (2011, pp. 898–900) that $(d_n - d_0)$ is $o(1)$. As d_0 is an interior point of [0,1], $d_n \in (L_0 h_n, 1 - L_0 h_n)$ and corresponds to a local minima of M_n for sufficiently large n, i.e., d_n satisfies

$$\tilde{\Phi}_n\left(\frac{\sqrt{nh_n}\bar{\mu}(d_n)}{\sqrt{1+\tilde{\Sigma}_n^2}}\right) \leq \gamma \text{ and } \tilde{\Phi}_n\left(\frac{\sqrt{nh_n}\bar{\mu}(d_n + 1/n)}{\sqrt{1+\tilde{\Sigma}_n^2}}\right) > \gamma. \tag{3.3}$$

By Pólya's theorem, $\tilde{\Phi}_n$ converges uniformly to Φ. Consequently,

$$0 \leq \sqrt{nh_n}\bar{\mu}(d_n) \leq \Phi^{-1}(\gamma)\sqrt{1 + \sigma_0^2 \bar{K^2}} + o(1). \tag{3.4}$$

Note that $\bar{\mu}(x) = 0$ for $x < d_0 - L_0 h_n$ and $\tilde{\Phi}_n(0)$ converges to $\Phi(0) = 0.5 < \gamma$. So, if $d_n < d_0$, then for (3.3) to hold, $d_n + 1/n + L_0 h_n > d_0$ for large n and thus $h_n^{-1}(d_n - d_0) = O(1)$ which gives the result. Also, when $d_0 < d_n \leq d_0 + L_0 h_n$, the result automatically holds. So, it suffices to consider the case $d_n > d_0 + L_0 h_n$.

Let $u_n(x, v) = (1/h_n)\mu(v)K((x - v)/h_n)$ for $x \in [0, 1]$ and $v \in \mathbb{R}$. By Lemma 4 from the Appendix,

$$\left|\bar{\mu}(d_n) - \int_0^1 u_n(d_n, v)dv\right| = O\left(\frac{1}{(nh_n)^\alpha}\right).$$

By a change of variable, $\int_0^1 u_n(d_n, v)dv = \int_{-L_0}^{L_0} \mu(d_n + uh_n)K(u)du$ for large n. As $d_n > d_0 + L_0 h_n$, the first part of the integrand, $\mu(d_n + uh_n)$, is positive for $u \in [-L_0, L_0]$. Let $[-L_1, L_1]$ be an interval where K is positive. Such an interval exists due to assumptions 4(a) and 4(b). Hence, $\int_{-L_1}^{L_1} \mu(d_n + uh_n)K(u)du = 2L_1\mu(d_n + \xi_n h_n)K(\xi_n) \leq \int u_n(d_n, v)dv$, where ξ_n is some point in $[-L_1, L_1]$. Using Taylor expansion around d_0, $\mu(d_n + \xi_n h_n) = \{\mu^{(k)}(\zeta_n)/k!\}(d_n + \xi_n h_n - d_0)^k$, for some ζ_n lying between d_0 and $d_n + \xi_n h_n$. By (3.4), we get

$$2L_1 \frac{\mu^{(k)}}{k!}(\zeta_n)(d_n + \xi_n h_n - d_0)^k K(\xi_n) \quad = \quad O((nh_n)^{-1/2}).$$

As $d_n \to d_0$, $\mu^{(k)}(\zeta_n)$ converges to $\mu^{(k)}(d_0+)$, which is positive. Also, as $\xi_n \in [-L_1, L_1]$, $K(\xi_n)$ is bounded away from zero, and thus $(d_n + \xi_n h_n - d_0) = O((nh_n)^{-1/2k})$, which yields the result. □

As $\hat{d}_n$ is, in fact, estimating d_n, its rate of convergence for d_0 can at most be ν_n^{-1}. Fortunately, ν_n^{-1} turns out to be the exact rate of convergence of $\hat{d}_n$.

Theorem 1. *Let* ν_n *be as defined in Lemma 1. Then* $\nu_n(\hat{d}_n - d_0) = O_p(1)$.

The proof is given in Section A.1 of the Appendix. It involves coming up with an appropriate distance ρ_n based on the behavior of M_n near d_0 (Lemma 5) and then establishing a modulus of continuity bound for $\mathbb{M}_n - M_n$ with respect to ρ_n. As the summands that constitute $\mathbb{M}_n$ are dependent, the latter cannot be handled directly through VC or bracketing results (Theorems 2.14.1 or 2.14.2 of van der Vaart and Wellner (1996)); rather, we require a blocking argument followed by an application of Doob's inequality to the blocks.

The optimal rate is attained when $h_n^{-1} \sim (nh_n)^{1/(2k)}$ and corresponds to $h_n = h_0 n^{-1/(2k+1)}$ and $\nu_n = n^{1/(2k+1)}$. We now deduce the asymptotic distribution for this particular choice of bandwidth.

3.2. Asymptotic distribution

With $h_n = h_0 n^{-1/(2k+1)}$, we study the limiting behavior of the process

$$Z_n(t) = h_n^{-1} \left[\mathbb{M}_n(d_0 + th_n) - \mathbb{M}_n(d_0)\right], \quad t \in \mathbb{R}, \tag{3.5}$$

where $\mathbb{M}_n$ is defined in (2.6). The process $Z_n(t)$ is minimized at $h_n^{-1}(\hat{d}_n - d_0)$. At the core of the process $Z_n(t)$ lies the estimator $\hat{\mu}$, computed at local points $d_0 + th_n$. Let

$$W_n(t) = \sqrt{nh_n}\hat{\mu}(d_0 + th_n) \tag{3.6}$$

and $B_{loc}(\mathbb{R})$ denote the space of locally bounded functions on $\mathbb{R}$, equipped with the topology of uniform convergence on compacta. We have the following lemma on the limiting behavior of W_n.

Lemma 2. *There exists a Gaussian process $W(t), t \in \mathbb{R}$, with almost sure continuous paths and drift*

$$m(t) = E(W(t)) = \frac{h_0^{k+1/2}\mu^{(k)}(d_0+)}{k!} \int_{-\infty}^{t} (t-v)^k K(v)\, dv$$

and covariance function $Cov(W(t_1), W(t_2)) = \sigma_0^2 \int K(t_1 + u)K(t_2 + u)du$ such that the process $W_n(\cdot)$ converges weakly to $W(\cdot)$ in $B_{loc}(\mathbb{R})$.

The proof is given in Section A.2 of the Appendix. For brevity, $-\int_y^x$ is written as $\int_x^y$ whenever $x > y$.

Theorem 2. *For $h_n = h_0 n^{-1/(2k+1)}$ and $t \in \mathbb{R}$, the process $Z_n(t)$ converges weakly to the process*

$$Z(t) = \int_0^t [\Phi(W(y)) - \gamma]dy$$

in $B_{loc}(\mathbb{R})$.

Proof. Split $Z_n(t)$ as $I_n(t) + II_n(t)$, where

$$\begin{aligned} I_n(t) \quad &= \quad \frac{1}{nh_n}\sum_{i=1}^{n}\Bigg[\left\{\Phi\left(\sqrt{nh_n}\hat{\mu}(i/n)\right) - \gamma\right\} \\ &\qquad \times \left\{1\left(\frac{i}{n} \le d_0 + th_n\right) - 1\left(\frac{i}{n} \le d_0\right)\right\}\Bigg] \end{aligned}$$

$$- \frac{1}{h_n} \int_{d_0}^{d_0 + t h_n} \left(\Phi \left(\sqrt{n h_n} \hat{\mu}(x) \right) - \gamma \right) dx$$

and $II_n = h_n^{-1} \int_{d_0}^{d_0+th_n} (\Phi(\sqrt{nh_n}\hat{\mu}(x)) - \gamma)dx$. Fix $T > 0$ and let $t \in [-T, T]$. Using arguments almost identical to those for proving Lemma 4 in the Appendix, we have

$$\begin{aligned} |I_n(t)| \quad &\leq \sum_{\substack{|d_0 - i/n| \\ \leq T h_n}} \int_{i/n}^{(i+1)/n} \frac{1}{h_n} \left| \Phi \left(\sqrt{n h_n} \hat{\mu}(i/n) \right) - \Phi \left(\sqrt{n h_n} \hat{\mu}(x) \right) \right| dx \\ &\quad + O \left(\frac{1}{n h_n} \right) + \frac{\gamma}{n h_n} (\lfloor n(d_0 + t h_n) \rfloor - \lfloor n(d_0) \rfloor) - \gamma t, \end{aligned}$$

where the $O(1/(nh_n))$ factor accounts for the boundary terms. Using the fact that $x - 1 \leq \lfloor x \rfloor \leq x + 1$, the term $\frac{\gamma}{nh_n}(\lfloor n(d_0 + th_n) \rfloor - \lfloor n(d_0) \rfloor) - \gamma t$ is bounded by $2\gamma(1/(nh_n) + T/n)$ which goes to zero. The sum of integrals in the above display is further bounded by

$$\begin{aligned} &\frac{\lceil 2 T n h_n \rceil}{n h_n} \sup_{\substack{|x-y| < 1/n \\ x, y \in [d_0 - T h_n, d_0 + T h_n]}} \left| \Phi \left(\sqrt{n h_n} \hat{\mu}(x) \right) - \Phi \left(\sqrt{n h_n} \hat{\mu}(y) \right) \right| \\ &\qquad \leq \frac{\lceil 2 T n h_n \rceil}{2 \pi n h_n} \sup_{\substack{|u-v| < 1/(n h_n) \\ u, v \in [-T, T]}} |W_n(u) - W_n(v)| . \end{aligned}$$

The above display goes in probability to zero due to the asymptotic equicontinuity of the process W_n and hence the term I_n converges in probability to zero uniformly in t over compact sets. Further, we have

$$\begin{aligned} II_n(t) &= h_n^{-1} \int_{d_0}^{d_0 + t h_n} \left(\Phi \left(\sqrt{n h_n} \hat{\mu}(x) \right) - \gamma \right) dx \\ &= \int_0^t \left[\Phi \left(\sqrt{n h_n} \hat{\mu}(d_0 + y h_n) \right) - \gamma \right] dy \\ &= \int_0^t \left[\Phi \left(W_n(y) \right) - \gamma \right] dy. \end{aligned}$$

As the mapping $W(\cdot) \mapsto \int_0^{\cdot} \Phi(W(y))dy$ from $B_{loc}(\mathbb{R})$ to $B_{loc}(\mathbb{R})$ is continuous, using Lemma 2, the term II_n converges weakly to the process $\int_0^t [\Phi(W(y)) - \gamma]dy$, $t \in \mathbb{R}$. This completes the proof. □

A conservative asymptotic CI for d_0 can be obtained using the following result.

Theorem 3. *The process $Z(t)$ goes to infinity almost surely (a.s.) as $|t| \to \infty$. Moreover, let ξ_0^s and ξ_0^l denote the smallest and the largest minimizers of the process Z. Also, let $c_{\alpha/2}^s$ and $c_{1-\alpha/2}^l$ be the $(\alpha/2)$th and $(1 - \alpha/2)$th quantiles of ξ_0^s and ξ_0^l respectively. For $h_n = h_0 n^{-1/(2k+1)}$, we have*

$$\liminf_{n \to \infty} P[c_{\alpha/2}^s < h_n^{-1}(\hat{d}_n - d_0) < c_{1-\alpha/2}^l] \geq 1 - \alpha.$$

Note that ξ_0^s and ξ_0^l are indeed well defined by continuity of the sample paths of Z and the fact that $Z(t)$ goes to infinity as $|t| \to \infty$. Also, they are Borel measurable as, say for ξ_0^s, the events $[\xi_0^s \leq a]$ and the measurable event $[\inf_{t\leq a} Z(t) \leq \inf_{t>a} Z(t)]$ are equivalent for any $a \in \mathbb{R}$. Hence $c^s_{\alpha/2}$ and $c^l_{1-\alpha/2}$ are well defined. The proof of the result is given in Section A.3 of the Appendix.

A minimum of the underlying limiting process lies in the set $\{y : \Phi(W(y)) = \gamma\}$. As any fixed number has probability zero of being in this set, the distributions of ξ_0^s and ξ_0^l are continuous. The process $\{W(y) : y \in \mathbb{R}\}$ has zero drift for $y < -L_0$ and is therefore stationary to the left of $-L_0$. Hence, it must cross γ infinitely often implying that Z has multiple local extrema. On the other hand, simulations *strongly* suggest that Z has a unique argmin though a theoretical justification appears intractable at this point. The issue of the uniqueness of the argmin of a stochastic process has mostly been addressed in context of Gaussian processes (Lifshits, 1982; Kim and Pollard, 1990; Ferger, 1999), certain transforms of compound Poisson processes (Ermakov, 1976; Pflug, 1983) and set-indexed Brownian motion (Müller and Song, 1996). These techniques do not apply to our setting; in fact, an analytical justification of the uniqueness of the minimizer of Z appears non-trivial. As the simulations provide strong evidence in support of a unique argmin, we use the following result for constructing CI's in practice.

Theorem 4. *Assuming that the process Z has a unique argmin, we have*

$$h_n^{-1}(\hat{d}_n - d_0) \xrightarrow{d} \underset{t\in\mathbb{R}}{argmin}\{Z(t)\},$$

for $h_n = h_0 n^{-1/(2k+1)}$.

Note that when the argmin is unique, Theorem 3 and Theorem 4 yield the same CI. The proof of Theorem 4 is a direct application of the argmin(argmax)-continuous mapping theorem; see Kim and Pollard (1990, Theorem 2.7) or van der Vaart and Wellner (1996, Theorem 3.2.2).

3.3. Limit distributions for variants of the procedure

The rates of convergence and asymptotic distributions can be obtained similarly for most of the variants of the procedure that were discussed in Section 2.1. In what follows, we state the limiting distributions for some of these variants.

Results analogous to Theorem 3 can be shown to hold in the setting with heteroscedastic errors, i.e., $\text{Var}(\epsilon_i) = \sigma_0^2(i/n)$, where $\sigma_0^2(\cdot)$ is a positive continuous function. The process Z has the same form as in Theorem 2 apart from the fact that the σ_0^2 involved in the covariance kernel of the process W that appears in the definition of Z is replaced by $\sigma_0^2(d_0)$. When normalized p-values are used to estimate d_0, we have the following form for the limiting distribution; an outline of its proof is given in Section A.4 of the Appendix.

Proposition 1. *Consider the setting with homoscedastic errors and covariates sampled from the fixed uniform design, as discussed in Section 2. Let $\hat{d}_n^1$ denote* *the estimate obtained from* **Method 1** *by minimizing* $\widetilde{\mathbb{M}}_n$ *defined in* (2.5). *Let* $h_n = h_0 n^{-1/(2k+1)}$ *and* $W^1(t), t \in \mathbb{R}$, *be a Gaussian process with drift*

$$E(W^1(t)) = \frac{h_0^{k+1/2}\mu^{(k)}(d_0+)}{k!\sigma_0\sqrt{\bar{K^2}}}\int_{-\infty}^{t}(t-v)^k K(v)\,dv$$

and covariance function $Cov(W^1(t_1), W^1(t_2)) = (\bar{K^2})^{-1}\int K(t_1+u)K(t_2+u)du$. *Let* $Z^1(t) = \int_0^t\{\Phi(W^1(y)) - \gamma\}dy$, *for* $t \in \mathbb{R}$. *If* $\hat{\sigma}$ *is a* $\sqrt{n}$*-consistent estimate of* σ_0, *then* $h_n^{-1}(\hat{d}_n^1 - d_0)$ *is* $O_p(1)$. *For* Z^1 *possessing a unique argmin a.s., we have*

$$h_n^{-1}(\hat{d}_n^1 - d_0) \xrightarrow{d} \underset{t\in\mathbb{R}}{argmin}\, Z^1(t).$$

When the covariate is sampled from a random design with heteroscedastic errors, the result extends as follows for the estimate based on non-normalized p-values. A sketch of the proof is given in Section A.5 of the Appendix.

Proposition 2. *Consider the setting with covariates sampled from a random design with design density* f *and heteroscedastic errors, as discussed in Section 2.1. The variance function* $\sigma_0^2(x) = Var(\epsilon \mid X = x)$ *is assumed to be continuous and positive. Let* $h_n = h_0 n^{-1/(2k+1)}$ *and* $\tilde{W}(t), t \in \mathbb{R}$, *be a Gaussian process with drift*

$$E(\tilde{W}(t)) = \frac{h_0^{k+1/2}\mu^{(k)}(d_0+)}{k!}\int_{-\infty}^{t}(t-v)^k K(v)\,dv$$

and covariance function $Cov(\tilde{W}(t_1), \tilde{W}(t_2)) = \frac{\sigma_0^2(d_0)}{f(d_0)}\int K(t_1+u)K(t_2+u)du$. *Let* $\tilde{Z}(t) = \int_0^t\{\Phi(\tilde{W}(y)) - \gamma\}dy$, *for* $t \in \mathbb{R}$. *For* $\tilde{d}_n$ *defined as in* (2.7), *assume that* $h_n^{-1}(\tilde{d}_n - d_0)$ *is* $O_p(1)$. *For* $\tilde{Z}$ *possessing a unique argmin a.s., we have*

$$h_n^{-1}(\tilde{d}_n - d_0) \xrightarrow{d} \underset{t\in\mathbb{R}}{argmin}\, \tilde{Z}(t).$$

4. The case of an unknown τ

Although most of the results have been deduced under the assumption of a known τ, in real applications τ is generally not known. In this situation, one would need to impute an estimate of τ in the objective function to carry out the procedure. It can be shown that the rate of convergence and the limit distribution does not change as long as we have a $\sqrt{n}$-consistent estimator of τ. The following result makes this formal; its proof is given in Section A.6 of the Appendix.

Proposition 3. *Let* $\hat{d}_n$ *now denote the minimizer of*

$$\mathbb{M}_n(d, \hat{\tau}) = \frac{1}{n}\sum_{i=1}^{n}\left[\Phi\left(\sqrt{nh_n}\left(\hat{\mu}\left(\frac{i}{n}\right) - \hat{\tau}\right)\right) - \gamma\right]1\left(\frac{i}{n} \leq d\right),$$

where $\sqrt{n}(\hat{\tau}-\tau) = O_p(1)$ *and* $h_n = h_0 n^{-1/(2k+1)}$. *Then* $h_n^{-1}(\hat{d}_n - d_0)$ *is* $O_p(1)$. *Assuming that the process* Z *defined in Theorem 2 has a unique argmin, we have*

$$h_n^{-1}(\hat{d}_n - d_0) \xrightarrow{d} \underset{t\in\mathbb{R}}{argmin}\{Z(t)\}.$$

Quite a few choices are possible for estimating τ. If d_0 can be safely assumed to be larger than some η, then a simple averaging of the observations below η would yield a $\sqrt{n}$-consistent estimator of τ. *If a proper choice of* η *is not available*, one can obtain an initial (consistent) estimate of τ using the method proposed in Section 2.4 of Mallik et al. (2011) (see (5.1)), compute $\hat{d}_n$ and then average the responses from, say, $[0, c\hat{d}_n]$, $c \in (0,1)$, to obtain a $\sqrt{n}$-consistent estimator of τ. This leads to an iterative procedure which we discuss in more detail in Section 5. In what follows, we justify that such an estimate of τ is indeed $\sqrt{n}$-consistent.

Lemma 3. *Let* $0 < c < 1$. *For any consistent estimator* d_n' *of* d_0, *define*

$$\hat{\tau} := \frac{1}{\lfloor ncd_n' \rfloor}\sum_{i=1}^{n} Y_i 1\left(\frac{i}{n} \le cd_n'\right).$$

We have $\sqrt{n}(\hat{\tau}-\tau) = O_p(1)$.

Proof. Note that for $T > 0$ and $0 < \kappa < \min(c,(1-c))d_0$,

$$\begin{aligned} P\left[\sqrt{n}|\hat{\tau}-\tau| > T\right] \quad \le \quad & P\left[\sqrt{n}|\hat{\tau}-\tau| > T, \kappa < cd_n' < d_0 - \kappa\right] \\ & + P\left[d_n' - d_0 < (\kappa - cd_0)/c\right] \\ & + P\left[d_n' - d_0 > ((1-c)d_0 - \kappa)/c\right]. \end{aligned}$$

The second and the third term on the right side of the above display both converge to zero. Also,

$$\begin{aligned} & E\left[n(\hat{\tau}-\tau)^2 1\left(\kappa < cd_n' < d_0-\kappa\right)\right] \\ = \quad & nE\left[\left(\frac{1}{\lfloor ncd_n'\rfloor}\sum_{i=1}^{\lfloor ncd_n'\rfloor}\epsilon_i\right)^2 1\left(\kappa < cd_n' < d_0-\kappa\right)\right] \\ \le \quad & \frac{n}{(n\kappa-1)^2}E\left[\sup_{a\le d_0-\kappa}\left(\sum_{i=1}^{\lfloor na\rfloor}\epsilon_i\right)^2\right] \\ \le \quad & \frac{4n}{(n\kappa-1)^2}E\left[\left(\sum_{i=1}^{\lfloor n(d_0-\kappa)\rfloor}\epsilon_i\right)^2\right] \le \frac{4n(n(d_0-\kappa)+1)\sigma_0^2}{(n\kappa-1)^2}. \end{aligned}$$

Here, the penultimate step followed from Doob's inequality. Hence, $E[n(\hat{\tau}-\tau)^2 1(\kappa < cd_n' < d_0-\kappa)] = O(1)$. Thus, by Chebyshev's inequality,

$$P\left[\sqrt{n}|\hat{\tau}-\tau| > T, \kappa < cd_n' < d_0-\kappa\right] \le O(1)/T^2$$

which can be made arbitrarily small by choosing T large. This completes the proof. □

5. Simulations

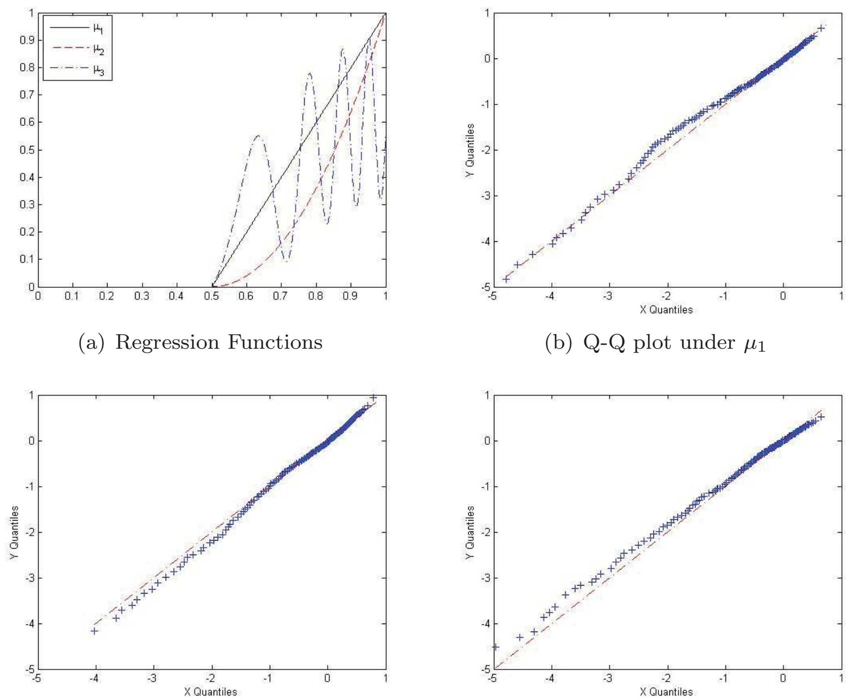
We consider three choices for the underlying regression function $\mu_k(x) = [2(x - 0.5)]^k 1(x > 0.5), x \in [0,1],\ k = 1,2$ and $\mu_3(x) = [(x-0.5) + (1/5)\sin(5(x-0.5)) + 0.3\sin(100(x-0.5)^2)]1(x > 0.5)$. All these functions are at their baseline value 0 up to $d_0 = 0.5$. The functions μ_1 (linear) and μ_2 (quadratic) both rise to 1 while μ_3 exhibits non-isotonic sinusoidal behavior after rising at d_0. The right derivative at d_0, a factor that appears in the limiting process Z, is the same for μ_1 and μ_3. The functions are plotted in the upper left panel of Figure 1. The functions μ_1 and μ_2 are paired up with normally distributed errors having mean 0 and standard deviation $\sigma_0 = 0.1$, while the noise added with μ_3 is from a t-distribution with 5 degrees of freedom, scaled to have the standard deviation σ_0. The three models, μ_1 with normal errors, μ_2 with normal errors and μ_3 with t-distributed errors, are referred to by the the name of their regression functions only. We work with $\gamma = 3/4$ as extreme values of γ (close to 0.5 or 1) tend to cause instabilities.

We construct the estimate of d_0 using the normalized p-values as they exhibit better finite sample performance and study the coverage performance of the approximate CI's obtained from the limiting distributions with estimated nuisance parameters. The error variance σ_0^2 is estimated in a straightforward manner using $\hat{\sigma}^2 = (1/n)\sum_i \{Y_i - \hat{\mu}(i/n)\}^2$. More sophisticated estimates of the error variance are also available (Gasser, Sroka and Jennen-Steinmetz, 1986; Hall, Kay and Titterington, 1990) but we avoid them for the sake of simplicity. We use the Epanechnikov kernel for constructing the estimate of μ. For moderate samples, the bad behavior of kernel estimates near the boundary affects the coverage performance. In order to correct for this, we only consider the terms between h_n to $1 - h_n$ in our objective function, i.e., for $d \in (h_n, 1-h_n)$,

$$\mathbb{M}_n(d,\tau) = \frac{1}{n} \sum_{h_n \le \frac{i}{n} \le 1-h_n} \left\{ \Phi\left(\frac{\sqrt{nh_n}(\hat{\mu}(i/n) - \tau)}{\Sigma_n(i/n, \hat{\sigma})} \right) - \gamma \right\} 1\left(\frac{i}{n} \le d \right).$$

The asymptotic distribution of the minimizer of this restricted criterion function still has the same form as in *Proposition 1*. A good choice for h_0 in the optimal bandwidth $h_n = h_0 n^{-1/(2k+1)}$ can be obtained through minimizing the MSE of $\hat{\mu}(d_0)$. Standard calculations show that

$$\begin{aligned} \text{Bias}(\hat{\mu}(d_0)) &= \frac{\mu^{(k)}(d_0+)}{k!} h_n^k E[W^k 1(W > 0)] + o(h_n^k) + O\left(\frac{1}{(nh_n)^\alpha}\right), \text{ and} \\ \text{Var}(\hat{\mu}(d_0)) &= \frac{\sigma_0^2}{nh_n}\bar{K}^2 + o\left(\frac{1}{nh_n}\right), \end{aligned}$$



(a) Regression Functions (b) Q-Q plot under μ_1

(c) Q-Q plot under μ_2 (d) Q-Q plot under μ_3

FIG 1. *Regression functions and Q-Q plots.*

where W has density K. The MSE is minimized at $h_n = h_0^{opt} n^{-1/(2k+1)}$ where

$$h_0^{opt} = \left[\frac{\sigma_0^2 \bar{K^2}(k!)^2}{2k\{\mu^{(k)}(d_0+)E[W^k 1(W>0)]\}^2}\right]^{-1/(2k+1)}.$$

This bandwidth goes to 0 at the right rate needed for estimating d_0. Moreover, efficient estimation of μ in the vicinity of d_0 is likely to aid in estimating d_0. Hence, we advocate the use of this choice of h_0 for our procedure.

With the above mentioned choice of h_0, we compare the distribution of $h_n^{-1}(\hat{d}_n - d_0)$ for $n = 1000$ data points over 2000 replications with the deduced asymptotic distribution. As τ is assumed unknown, we implement an iterative scheme. We obtain an initial estimate of τ using the method prescribed in Mallik et al. (2011), i.e.,

$$\hat{\tau}_{init} = \underset{\tilde{\tau}\in\mathbb{R}}{\operatorname{argmin}} \sum \left\{\Phi\left(\frac{\sqrt{nh_n}(\hat{\mu}(i/n) - \tilde{\tau})}{\Sigma_n(i/n, \hat{\sigma})}\right) - \frac{1}{2}\right\}^2. \qquad (5.1)$$

This estimate of τ, based on h_0^{opt}, is used to compute $\hat{d}_n$. We re-estimate τ by averaging the responses for which $i/n \in [0, 0.9\hat{d}_n]$, use this new estimate of τ

to update the estimate of d_0, and proceed thus. The Q-Q plots are shown in Figure 1 which show considerable agreement between the two distributions.

Next, we explore the coverage performance of the CI's constructed by imputing estimates of the nuisance parameters in the limiting distribution. Computing h_0 requires the knowledge of the k-th derivative of μ at d_0 which we also need to generate from the limit process. To estimate $\mu^{(k)}(d_0+)$, first observe that $\mu(x) = \mu^{(k)}(d_0+)(x-d_0)^k/k! + o((x-d_0)^k)$ for $x > d_0$. Hence, an estimate of $\mu^{(k)}(d_0+)$ can be obtained by fitting a k-th power of the covariate to the right of $\hat{d}_n$. More precisely, an estimate of $\xi_0 \equiv \mu^{(k)}(d_0+)/k!$ is given by

$$\begin{aligned} \hat{\xi} &= \underset{\xi}{\operatorname{argmin}} \sum_{i=1}^{n} \{Y_i - \xi(i/n - \hat{d}_n)^k\}^2 1(i/n \in (\hat{d}_n, \hat{d}_n + b_n]) \\ &= \frac{\sum Y_i (i/n - \hat{d}_n)^k 1(i/n \in (\hat{d}_n, \hat{d}_n + b_n])}{\sum (i/n - \hat{d}_n)^{2k} 1(i/n \in (\hat{d}_n, \hat{d}_n + b_n])}, \end{aligned}$$

where $b_n \downarrow 0$ and $nb_n^{2k+1} \to \infty$. For the optimal h_n, this provides a good estimate of ξ_0.

We include this in our iterative method where we start with an arbitrary choice of h_0 and compute $\hat{\tau}_{init}$. We use $\hat{\tau}_{init}$ to compute $\hat{d}_n$ and $\hat{\mu}^{(k)}(d_0+)$. The parameter $\hat{\mu}^{(k)}(d_0+)$ is estimated using a reasonably wide smoothing bandwidth b_n, $b_n = 5(n/\log n)^{-1/(2k+1)}$. These initial estimates are used to compute the next level estimate of h_0 using the expression for h_0^{opt}. We re-estimate τ by averaging the responses for which $i/n \in [0, 0.9\hat{d}_n]$ and proceed thus. On average, the estimates stabilize within 7 iterations. The coverage performance over 5000 replications is given below in Table 1. The approximate CI's mostly exhibit over-coverage for moderate sample sizes for μ_1 and μ_3 but converge to the desired confidence levels for large n. Also, the limiting distribution is same under models μ_1 and μ_3 which is evident from the coverages and the length of CI's for large n.

6. Dependent data

We briefly discuss the extension of Method 2 to dependent data in this section. Our problem is relevant to applications from time series models (see Section 7) where it is not reasonable to assume that the errors ϵ_i's are independent. A data generating model of the form (2.1) can be assumed here with the exception that the errors now arise from a stationary sequence $\{\ldots, \epsilon_{-1}, \epsilon_0, \epsilon_1, \ldots\}$ and exhibit short-range dependence in the sense of Robinson (1997). As with (2.1), the dependence of Y_i's and ϵ_i's on n is suppressed but they must be viewed as triangular arrays. The extension to this setting would work along the following lines. The estimate of μ with dependent errors still has the same form as (2.2). With additional assumptions (Assumptions 1–5 of Robinson (1997)), it is guaranteed that $\sqrt{nh_n}(\hat{\mu}(x_i) - \mu(x_i))$, $x_i \in (0,1)$ and $x_1 \neq x_2$, converge jointly in distribution to independent normals with zero mean – a fact that justifies the consistency of our p-value based estimates in this setting (Mallik et al., 2011). Hence, $\hat{d}_n$, defined using (2.6), can still be used to estimate the threshold. The limiting distribution would be of the same form as in Lemma 2 but with a dif-

TABLE 1
Coverage probabilities and length of the CI (in parentheses) using the true parameters (T) and the estimated parameters (E) for different sample sizes under μ_1, μ_2 and μ_3

n	90% CI		95% CI	
	T	E	T	E
30	0.949 (0.462)	0.961 (0.614)	0.989 (0.588)	0.987 (0.659)
50	0.943 (0.420)	0.951 (0.539)	0.971 (0.547)	0.978 (0.625)
100	0.921 (0.357)	0.939 (0.448)	0.965 (0.483)	0.972 (0.559)
500	0.914 (0.218)	0.922 (0.258)	0.961 (0.299)	0.965 (0.346)
1000	0.907 (0.173)	0.911 (0.197)	0.955 (0.237)	0.959 (0.265)
2000	0.900 (0.137)	0.903 (0.153)	0.951 (0.188)	0.954 (0.205)
μ_1				
n	90% CI		95% CI	
	T	E	T	E
30	0.957 (0.544)	0.849 (0.651)	0.992 (0.624)	0.899 (0.665)
50	0.948 (0.539)	0.876 (0.615)	0.973 (0.620)	0.908 (0.627)
100	0.933 (0.519)	0.883 (0.602)	0.964 (0.617)	0.917 (0.616)
500	0.917 (0.415)	0.889 (0.477)	0.962 (0.548)	0.934 (0.555)
1000	0.907 (0.385)	0.894 (0.424)	0.957 (0.511)	0.944 (0.525)
2000	0.904 (0.350)	0.899 (0.384)	0.951 (0.471)	0.948 (0.490)
μ_2				
n	90% CI		95% CI	
	T	E	T	E
30	0.960 (0.461)	0.968 (0.620)	0.992 (0.590)	0.994 (0.672)
50	0.949 (0.424)	0.959 (0.541)	0.977 (0.548)	0.982 (0.630)
100	0.925 (0.358)	0.941 (0.472)	0.970 (0.482)	0.976 (0.539)
500	0.915 (0.218)	0.925 (0.304)	0.961 (0.299)	0.966 (0.348)
1000	0.906 (0.173)	0.914 (0.199)	0.954 (0.237)	0.958 (0.264)
2000	0.901 (0.138)	0.904(0.154)	0.950 (0.188)	0.954 (0.204)
μ_3				

ferent scaling factor that appears in the covariance function of the process W. We outline the form of the limiting distribution below. The technical details are more involved in the sense of tedium but the approach in deriving the limiting distribution remains the same at the conceptual level.

To precisely state the limiting distribution, let $\rho(i,j) = \rho(i-j)$ denote the covariance between ϵ_i and ϵ_j and let ψ denote the underlying spectral density defined through the relation $\sigma_0^2\rho(l) = \int_{-\pi}^{\pi}\psi(u)\exp(\imath l u)du, l \in \mathbb{Z}$. Let $\bar{W}$ be a Gaussian process with drift $m(\cdot)$ (defined in Lemma 2) and covariance function

$$\mathrm{Cov}(\bar{W}(t_1), \bar{W}(t_2)) = 2\pi\psi(0)\int K(t_1+u)K(t_2+u)du.$$

It is not uncommon for the spectral density at zero, $\psi(0) = (2\pi)^{-1}\sigma_0^2\sum_{j\in\mathbb{Z}}\rho(j)$, to appear in settings with short range dependence (Robinson, 1997; Anevski and Hössjer, 2006).

Proposition 4. *Consider the setup of Method 2 with the errors now exhibiting short-range dependence as discussed above. Assume that for $h_n = h_0 n^{-1/(2k+1)}$, the resulting estimate $\hat{d}_n$ obtained using (2.6) satisfies $h_n^{-1}(\hat{d}_n - d_0) = O_p(1)$ and that the process $\bar{Z}(t) = \int_0^t[\Phi(\bar{W}(y)) - \gamma]dy$, $t \in \mathbb{R}$, has a unique minimum a.s. Then*

$$h_n^{-1}(\hat{d}_n - d_0) = \underset{t\in\mathbb{R}}{argmin}\,\bar{Z}(t).$$

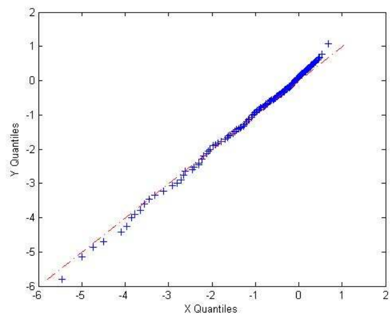


Fig 2. *Regression setting with dependent errors: Q-Q plot under* μ_1.

The proof is outlined in Section A.7 of the Appendix.

An illustration of the above phenomenon is shown through a Q-Q plot (Figure 2), where we generate ϵ_i's from an AR(1) model $\epsilon_i = 0.25\epsilon_{i-1} + z_i$. Here, z_i's are mean 0 normal random variables with variance 0.0094 so that ϵ_i's have variance $(0.1)^2$. The Q-Q plot shows considerable agreement between the empirical quantiles, obtained from samples of size $n = 1000$, with the theoretical quantiles.

7. Data analysis

We now apply our procedure to two interesting examples from Mallik et al. (2011).

The first data set involves measuring concentration of mercury in the atmosphere through a LIDAR experiment. There are 221 observations with the predictor variable 'range' varying from 390 to 720 and there is visible evidence of heteroscedasticity. The observed covariates can be considered to have arisen from a random design and the threshold d_0 corresponds to the distance at which there is a sudden rise in the concentration of mercury. See pages 2 and 10 of Mallik et al. (2011) for more details.

We employ a variant of Method 2 based on the Nadaraya–Watson estimator without normalizing by the estimate of the variance. It is reasonable to assume here that the function is at its baseline till range value 480. The estimate of τ is obtained by taking the average of observations until range reaches 480, which gives $\hat{\tau} = -0.0523$. The estimate $\hat{d}_n$ is obtained through the iterative approach described in Section 5. The expression for the approximate bias of the Nadaraya-Watson estimator turns out to be the same as that for the fixed design kernel estimator at d_0 while the approximate variance turns out to be $(\sigma_0^2 \bar{K^2})/(nh_n f(d_0))$ and the optimal value of h_0 is adjusted accordingly. The limiting distribution, as well as the optimal h_0, involves the parameter $\sigma_0(d_0)$, which we estimate using $\hat{\sigma}(\hat{d}_n)$ where $\hat{\sigma}^2(x) = [\mathbb{P}_n(Y - \hat{\mu}(X))^2 K((x-X)/h_n)]/[\mathbb{P}_n K((x-X)/h_n)]$.

The estimate $\hat{d}_n$ has an inherent bias which is a recurring feature in boundary estimation problems. A simple but effective way to reduce this bias is to subtract the median of the limiting distribution with imputed parameters, say $\hat{q}_{0.5}$, from our crude estimate, after proper normalization (so that the limiting median is zero). More precisely, $\hat{d}_n - n^{-1/(2k+1)}\hat{q}_{0.5}$ is our final estimate. Assuming k to be 1, the resulting estimate of d_0 is 551.05 which appears reasonable (see Figure 1 of Mallik et al. (2011)). Moreover, the CI's are $[550.53, 555.17]$ and $[549.75, 557.82]$ for confidence levels of 90% and 95%, respectively, which also seem reasonable.

Our second data set, which comes from the last example in the Introduction of Mallik et al. (2011) (see pages 3 and 10 of that paper for more details), involves the measurement of annual global temperature anomalies, in degree Celsius, over the years 1850 to 2009. The depiction of the data (see Figure 1 of Mallik et al. (2011)) suggests a trend function which stays at its baseline value for a while followed by a nondecreasing trend. We follow the approach of Wu, Woodroofe and Mentz (2001) and Zhao and Woodroofe (2012), and model the data as having a non-parametric trend function and short-range dependent errors. The flat stretch at the beginning is also noted in Zhao and Woodroofe (2012), where isotonic estimation procedures are considered in settings with dependent data. They also provide evidence for the errors to be arising from a lower order auto-regressive process. A comprehensive approach would incorporate a cyclical component as well (Schlesinger and Ramankutty, 1994), which we do not pursue in our paper.

The estimate of the baseline value, after averaging the anomalies up to the year 1875, is $\hat{\tau} = -0.3540$. Using this estimate of τ, we employ Method 2 with non-normalized p-values (see (2.6)) in this example with the optimal h_0 chosen through an iterative approach. Constructing the CI involves estimating an extra parameter $\psi(0)$ for which we use the estimates computed in Wu, Woodroofe and Mentz (2001, pp. 800) (the parameter σ^2 estimated in that paper is precisely $2\pi\psi(0)$). Assuming k to be 1, the estimate of the threshold d_0 after bias correction, which signifies the advent of global warming, turns out to be 1912. The CI's are $[1908, 1917]$ and $[1906, 1919]$ for confidence levels 90% and 95% respectively. This is compatible with the observation on page 2 of Zhao and Woodroofe (2012) that global warming does not appear to have begun until 1915.

8. Conclusion

We conclude with a discussion of some open problems that can provide avenues for further investigation into this problem.

Adaptivity. In this paper we have provided a comprehensive treatment of the asymptotics of a p-value based procedure to estimate the threshold d_0 at which an unknown regression function μ takes off from its baseline value, with the aim of constructing CI's for d_0. We have assumed knowledge of the order of the 'cusp' of μ at d_0, which we need to achieve the optimal rate of convergence (and construct the corresponding CI's), though not for consistency. When k is unknown, ideas from multiscale testing procedures for white noise models

(Dümbgen and Spokoiny, 2001; Dümbgen and Walther, 2008) can conceivably be used to develop adaptive procedures in our model. This is a hard open problem and will be a topic of future research.

Resampling. A natural alternative to using the limit distribution (with estimated nuisance parameters) to construct CI's for d_0 would be to use bootstrap/resampling methods. Drawing from results obtained in similar change-point and non-standard problems (see e.g., Sen, Banerjee and Woodroofe (2010); Seijo and Sen (2011)) it is very likely that the usual bootstrap method will be inconsistent in our setup. However, model based bootstrap procedures have recently been studied in the change-point context and have been shown to work (Seijo and Sen, 2011). Similar ideas may work for our problem as well, but a thorough understanding of such bootstrap procedures is beyond the scope of the present paper. Subsampling can be proven to be consistent in our setting, but its finite sample properties were seen to be rather dismal.

Behavior near the boundary. For our simulations, we concentrated on the case $d_0 = 0.5$. The estimate also performs well (in terms of MSE's) in settings where d_0 is close to the boundary as long as there are sufficiently many observations on either side of d_0 (see Section 2.3 of Supplementary material of Mallik et al. (2011)). We do not address the cases where d_0 is exactly at the boundary, e.g, $d_0 = 0$. This leads to a testing problem (flat stretch vs. no flat stretch) which goes beyond the scope of our discussion. However, we would like to point out that $\hat{d}_n$ would be consistent even when $d_0 = 0$ or 1 regardless of the bad behavior of the kernel estimates near the boundary.

Simultaneous estimation of d_0 and τ. When τ is unknown, we have provided a procedure which estimates the threshold d_0 and the baseline value τ simultaneously through an iterative scheme (see Section 5); however, our method requires the use of two objective functions, one for updating the estimate of d_0 and the other that of τ. While estimating τ and d_0 from a *single* objective function, say by minimizing (2.4) over putative values of both d_0 and τ would be ideal, this optimization problem is hard to solve. In fact, it is unclear whether a tractable solution that provides consistent estimates can be obtained. For example, note that minimizing the least squares criterion in (2.4) naively over choices of τ (say $\tilde{\tau}$) does not necessarily yield meaningful estimates as the criterion goes to 0 when $\tilde{\tau} \to -\infty$ and $d \to 0$, so constraining the optimization would appear necessary.

Of course, modifications to the above least squares criterion are possible. As the average of the responses up to d_0 yields an estimate of τ_0, one can alternatively minimize

$$\left[\sum_{i:i/n \le d} \left\{p_n(i/n, \tau_d) - \frac{1}{2}\right\}^2 + \sum_{i:i/n > d} \{p_n(i/n, \tau_d)\}^2\right],$$

where $\tau_d = (1/\lfloor nd \rfloor) \sum_{i \le n} Y_i 1[i/n \le d]$. This is expected to yield consistent estimates for d_0 and τ ($\hat{\tau} = \tau_{\hat{d}}$). However, the estimate of τ may be biased and it is far from clear whether the estimate of d_0 will exhibit the same rate of convergence. Note that the above criterion runs over the data twice (once while computing τ_d and the second time through the sum up to d) and is harder to handle analytically. One can also consider a slightly different criterion by replacing τ_d by $c\tau_d$ ($c \in (0,1)$) in the above display, which does not estimate τ as efficiently as its predecessor but avoids the bias issue. Finally, it is unclear whether other useful criteria can be formulated to simultaneously estimate d_0 and τ in this non-parametric setup, mainly owing to the fact that d_0 is a feature of the covariate domain while τ is a feature of the response domain.

Minimaxity. The estimators studied in our paper attain the convergence rate of $n^{-1/(2k+1)}$. This leads to a natural question as to whether this is the best possible rate of convergence. When μ is monotone increasing, d_0 is precisely $\mu^{-1}(\tau)$, where μ^{-1} is the right continuous inverse of μ. Wright (1981) (Theorem 1) shows that the rate of convergence of the isotonic least squares estimate μ at a point, x_0, where the first $k-1$ derivatives vanish but the kth does not, is precisely $n^{-k/(2k+1)}$. A slightly more general result establishing a process convergence is stated in Fact 1 of Banerjee (2009). Using this in conjunction with the techniques for the proof of Theorem 1 in Banerjee and Wellner (2005), it can be deduced that the rate of convergence of the isotonic estimate of μ^{-1} at $\mu(x_0)$ is $n^{-1/(2k+1)}$, which matches the rate attained by our approach. Hence, we expect this rate to be minimax in our setting. We note that this rate is not the same as the faster rate $\min(n^{-2/(2k+3)}, n^{-1/(2k+1)})$ obtained in Neumann (1997) for a change-point estimation problem in a density deconvolution model and also observed in the convolution white noise models of Goldenshluger, Tsybakov and Zeevi (2006) and Goldenshluger et al. (2008). These models are related to our setting; e.g., Problem 1 in Goldenshluger et al. (2008) is a Gaussian white noise model where the underlying regression function also has a cusp of a known order at an unknown point of interest. The convolution white noise model considered in Goldenshluger, Tsybakov and Zeevi (2006) (Problem 2 in Goldenshluger et al. (2008)) is equivalent to this problem for a particular choice of the convolution operator; see Goldenshluger, Tsybakov and Zeevi (2006, pp. 352–353) and Goldenshluger et al. (2008, pp. 790–791) for more details. Besides these being white noise models, they differ from our setting through an additional smoothness condition (Goldenshluger, Tsybakov and Zeevi, 2006, pp. 354–355), which translates, in our setting, to assuming that $\mu^{(k)}$ is Lipschitz outside any neighborhood of d_0, an assumption *not made* in this paper. Hence, Neumann's rate need not be minimax for our setting. The faster rate of Neumann (1997) was also observed for $k = 1$ in Cheng and Raimondo (2008) but once again under the assumption that the derivative of the regression function is at least twice differentiable away from the change-point, again an assumption *not made* in this paper.

Acknowledgments

We thank Michael Woodroofe for numerous stimulating discussions that helped develop several ideas and simplify some of the proofs in the paper.

Appendix A: Appendix

We use the notations '$\lesssim$' and '$\gtrsim$' to imply that the corresponding inequalities hold up to some positive constant multiple, and E^* to denote the outer expectation with respect to the concerned probability measure.

We start with proving a few auxiliary results that are repeatedly used in the paper. Recall that K and μ are Lipschitz continuous of order $\alpha \in (1/2, 1]$. Let $u_n(x, v) = (1/h_n)\mu(v)K((x-v)/h_n)$ for $x \in [0,1]$ and $v \in \mathbb{R}$.

Lemma 4. *For $\bar{\mu}(\cdot)$ as in (3.1), we have*

$$\sup_{x\in[0,1]} \left|\bar{\mu}(x) - \int_0^1 u_n(x,v)dv\right| = O\left(\frac{1}{(nh_n)^\alpha}\right).$$

Proof. Note that $\bar{\mu}(x) = (1/n)\sum_i u_n(x, i/n)$ and $u_n(x,v) = 0$ whenever $|x - v| \geq L_0 h_n$. Moreover, the difference between

$$\bar{\mu}(x) - \int_0^1 u_n(x,v)dv$$

and

$$\sum_{\substack{1\leq i\leq n\\ |x-i/n|\leq L_0 h_n}} \int_{i/n}^{(i+1)/n} \{u_n(x,i/n) - u_n(x,v)\}dv$$

is at most

$$\int_0^{1/n} |u_n(x,v)|dv + \int_{x-L_0h_n}^{x-L_0h_n+1/n} |u_n(x,v)|dv$$

which is bounded by $(1/n)\sup_{x,v} u_n(x,v) \leq \|\mu\|_\infty \|K\|_\infty/(nh_n)$. Hence,

$$\begin{aligned}&\left|\bar{\mu}(x) - \int_0^1 u_n(x,v)dv\right| \\ \leq\; & O\left(\frac{1}{nh_n}\right) + \sum_{\substack{1\leq i\leq n\\ |x-i/n|\leq L_0 h_n}} \int_{i/n}^{(i+1)/n} |u_n(x,i/n) - u_n(x,v)|dv.\end{aligned}$$

For $v_1, v_2 \in \mathbb{R}$, $h_n|u_n(x,v_1) - u_n(x,v_2)| \leq |\mu(v_1) - \mu(v_2)|K((x-v_1)/h_n) + |\mu(v_2)||K((x-v_1)/h_n) - K((x-v_2)/h_n)|$. As K and μ are Lipschitz continuous of order α, $|u_n(x,v_1) - u_n(x,v_2)| \lesssim 1/h_n^{1+\alpha}|v_1 - v_2|^\alpha$. Also, the cardinality of the set $\{i : 1 \leq i \leq n,\ |x - i/n| \leq L_0 h_n\}$ is at most $2L_0 nh_n + 2$ and therefore, the above display is further bounded (up to a positive constant multiple) by

$$O\left(\frac{1}{nh_n}\right) + \sum_{\substack{1\leq i\leq n\\ |x-i/n|\leq L_0 h_n}} \int_{i/n}^{(i+1)/n} \frac{|i/n - v|^\alpha}{h_n^{1+\alpha}} dv \leq O\left(\frac{1}{nh_n}\right) + \frac{2L_0 nh_n + 2}{(\alpha+1)(nh_n)^{1+\alpha}},$$

which is $O(1/(nh_n)^\alpha)$. Here, the final bound does not depend on x and thus, we get the desired result. □

Note that the above result holds for generic functions μ and K, satisfying assumptions 1(a), 4(c) and 4(d). Letting $\mu(x) \equiv \sigma_0^2$ and substituting K^2 for K, we get:

Corollary 1. *Let $z_n(x,v) = (\sigma_0^2/h_n)K^2\left((x-v)/h_n\right)$. Then,*

$$\sup_{x\in[0,1]}\left|\Sigma_n^2\left(x\right) - \int_0^1 z_n(x,v)dv\right| = O\left(\frac{1}{(nh_n)^\alpha}\right).$$

As a consequence, when $i/n \in [L_0h_n, 1-L_0h_n]$,

$$\begin{aligned}\Sigma_n^2\left(i/n\right) &= \int_0^1 z_n(i/n,v)dv + o(1)\\ &= \sigma_0^2 \min_j \int_{(i-n)/(nh_n)}^{i/(nh_n)} K^2(u)du + o(1)\\ &= \sigma_0^2\bar{K^2} + o(1). \qquad (A.1)\end{aligned}$$

A.1. Proof of Theorem 1

To prove Theorem 1, we use Theorem 3.2.5 of van der Vaart and Wellner (1996) (see also Theorem 3.4.1) which requires coming up with a non-negative map $d \mapsto \rho_n(d,d_n)$ such that

$$M_n(d) - M_n(d_n) \geq \rho_n^2(d,d_n).$$

Then a bound on the modulus of continuity with respect to ρ_n is needed, i.e.,

$$\begin{aligned}&E\left[\sqrt{n}\sup_{\rho_n(d,d_n)<\delta}\left|(\mathbb{M}_n(d)-\mathbb{M}_n(d_n)) - (M_n(d)-M_n(d_n))\right|\right] \qquad (A.2)\\ &= E\left[\sqrt{n}\sup_{\rho_n(d,d_n)<\delta}\left|(\mathbb{M}_n-M_n)(d) - (\mathbb{M}_n-M_n)(d_n)\right|\right] \lesssim \phi_n(\delta),\end{aligned}$$

where the map $\delta \mapsto \phi_n(\delta)/\delta^\alpha$ is decreasing for some $\alpha < 2$. The rate of convergence is then governed by the behavior of ϕ_n. We start with the following choice for ρ_n.

Lemma 5. *Fix $\eta > 2L_0 > 0$. Let $d \mapsto \rho_n(d,d_n)$ be a map from $(0,1)$ to $[0,\infty)$ such that*

$$\begin{aligned}\rho_n^2(d,d_n) = {}& (K_1/n)\left\{\left|\lfloor nd\rfloor - \lfloor n(d_0 - L_0h_n)\rfloor\right| 1(d \leq d_0 - L_0h_n)\right.\\ &\left.+ \left|\lfloor nd\rfloor - \lfloor n(d_n + \eta/\nu_n)\rfloor\right| 1(d > d_n + \eta/\nu_n)\right\},\end{aligned}$$

for some $K_1 > 0$. Then K_1 and $\kappa > 0$ can be chosen such that for sufficiently large n and $\rho_n(d,d_n) < \kappa$, we have

$$M_n(d) - M_n(d_n) \geq \rho_n^2(d,d_n).$$

We first provide the proof of Theorem 1 using Lemma 1. By the above Lemma, there exists $A < \infty$ such that for sufficiently large n and any $\delta > 0$, $\{\rho_n(d, d_n) < \delta\} \subset \{|d - d_n| < \delta^2/K_1 + A/\nu_n + 2/n\}$. Let $d > d_n$ and

$$\begin{aligned} U(i, d) &= \left\{\Phi\left(\sqrt{nh_n}(\hat{\mu}(i/n))\right) - \Phi_{i,n}\left(\frac{\sqrt{nh_n}(\bar{\mu}(i/n))}{\sqrt{1 + \Sigma_n^2(i/n)}}\right)\right\} \\ &\quad \times 1\left(d_n < \frac{i}{n} \le d\right) \end{aligned}$$

where $\hat{\mu}$ is defined in (3.1). By (3.2), $E\{U(i, d)\} = 0$. Also, for $1 \le i, j \le n$, $U(i, d)$ and $U(j, d)$ are independent whenever $|i - j| \ge 2L_0 nh_n$. Let $j_1^i = i$ and $j_l^i = j_{l-1}^i + \lceil 2L_0 nh_n \rceil$. Then,

$$S(i, d) := \frac{1}{n} \sum_{l : j_l^i \le n} U(j_l^i, d),$$

a sum of at most $\lceil (d - d_n)/(2L_0 h_n) \rceil$ non-zero independent terms, is a martingale in d, $d \ge d_n$, with right continuous paths. As $|U(\cdot, d)| \le 1$, $E\{U^2(\cdot, d)\}$ is at most 1. Using Doob's inequality, we get

$$\begin{aligned} & E\left[\sup_{\substack{|d - d_n| < \delta^2/K_1 + A/\nu_n + 2/n \\ d \ge d_n}} |S(i, d)|\right] \\ \le\ & \left\{ES^2(i, d_n + \delta^2/K_1 + A/\nu_n + 2/n)\right\}^{1/2} \\ =\ & \frac{1}{n}\left[\sum_{l : j_l^i \le n} E\{U^2(j_l^i, d_n + \delta^2/K_1 + A/\nu_n + 2/n)\}\right]^{1/2} \\ \le\ & \frac{1}{L_0 n h_n^{1/2}} (\delta^2/K_1 + A/\nu_n + 2/n)^{1/2}. \end{aligned}$$

As $(\mathbb{M}_n - M_n)(d) - (\mathbb{M}_n - M_n)(d_n) = \sum_{i=1}^{\lceil 2L_0 nh_n \rceil - 1} S(i, d)$, for sufficiently large n,

$$\begin{aligned} & E\left[\sqrt{n} \sup_{\rho_n(d, d_n) < \delta, d > d_n} |(\mathbb{M}_n - M_n)(d) - (\mathbb{M}_n - M_n)(d_n)|\right] \\ \le\ & E\left[\sqrt{n} \sup_{\substack{|d - d_n| < \delta^2/K_1 + A/\nu_n + 2/n \\ d \ge d_n}} |(\mathbb{M}_n - M_n)(d) - (\mathbb{M}_n - M_n)(d_n)|\right] \\ \le\ & \sqrt{n}(2L_0 nh_n) \frac{1}{L_0 n h_n^{1/2}} (\delta^2/K_1 + A/\nu_n + 2/n)^{1/2} \lesssim \phi_n(\delta), \end{aligned} \tag{A.3}$$

where $\phi_n(\delta) = \sqrt{nh_n}(\delta^2 + \nu_n^{-1} + n^{-1})^{1/2}$. This bound can also be shown to hold when $d \le d_n$. Also, $\phi_n(\cdot)$ and $\rho_n(\cdot, d_n)$ satisfy the conditions of Theorem 3.2.5

of van der Vaart and Wellner (1996). Hence, the rate of convergence, say r_n, satisfies

$$r_n^2 \phi_n \left(\frac{1}{r_n}\right) \lesssim \sqrt{n} \;\Rightarrow nh_n(r_n^2 + r_n^4/\nu_n + r_n^4/n) \lesssim n.$$

Note that $r_n^2 = \nu_n$ satisfies the above relation and therefore $\nu_n \rho_n^2(\hat{d}_n, d_n)$ is $O_p(1)$. Consequently, we get $\nu_n(\hat{d}_n - d_0) = O_p(1)$. □

Proof of Lemma 5. Since $\bar{\mu}(x) = 0$ for $x < d_0 - L_0 h_n$, note that $d_n > d_0 - L_0 h_n$ for sufficiently large n. As $\Phi_{i,n}(0)$ converges to $1/2$ uniformly in i, it can be seen that for large n and $d \leq d_0 - L_0 h_n$,

$$\begin{aligned} M_n(d) - M_n(d_n) &\geq M_n(d) - M_n(d_0 - L_0 h_n) \\ &= \sum_{i=1}^{n} \{\gamma - \Phi_{i,n}(0)\} 1 \left(d < \frac{i}{n} \leq d_0 - L_0 h_n\right) \\ &\geq \frac{1}{2}\left(\gamma - \frac{1}{2}\right) \{\lfloor nd \rfloor - \lfloor n(d_0 - L_0 h_n) \rfloor\}/n. \end{aligned} \tag{A.4}$$

Next, we show that

$$\tilde{\Phi}_n \left(\frac{\sqrt{nh_n}\bar{\mu}(d_n + \eta/\nu_n)}{\sqrt{1 + \tilde{\Sigma}_n^2}}\right) - \gamma > K_0, \tag{A.5}$$

for sufficiently large n and some $K_0 > 0$. Using (3.3), note that $\tilde{\Phi}_n(\sqrt{nh_n}\bar{\mu}(d_n)/\sqrt{1 + \tilde{\Sigma}_n^2(d_n)})$ converges to γ and consequently, $\sqrt{nh_n}\bar{\mu}(d_n)/\sqrt{1 + \tilde{\Sigma}_n^2(d_n)}$ is $O(1)$. As $\Sigma_n^2(d_n)$ is also $O(1)$, it suffices to show that $\sqrt{nh_n}(\bar{\mu}(d_n + \eta/\nu_n) - \bar{\mu}(d_n))$ is bounded away from zero. To show this, note that by Lemma 4,

$$\begin{aligned} &\sqrt{nh_n}(\bar{\mu}(d_n + \eta/\nu_n) - \bar{\mu}(d_n)) \\ &\quad = \int_{-L_0}^{L_0} \sqrt{nh_n}\, \{\mu(d_n + \eta/\nu_n + uh_n) - \mu(d_n + uh_n)\}\, K(u) du + o(1). \end{aligned}$$

Choose $\kappa > 0$ such that μ is non-decreasing in $(d_0, d_0 + 3\kappa)$. For sufficiently large n, $d_n + \eta/\nu_n + L_0 h_n < d_0 + 3\kappa$, and hence, the integrand in the above display is non-negative. With L_1 such that $K_{min} = \inf\{K(x) : x \in [-L_1, L_1]\} > 0$, the above display is bounded from below by

$$2L_1 K_{min}\sqrt{nh_n}(\mu(d_n + \eta/\nu_n - L_1 h_n) - \mu(d_n + L_0 h_n)).$$

As $\eta > 2L_0$, note that $d_n + \eta/\nu_n - L_1 h_n > d_n + L_0 h_n > d_0$. With $\zeta_n^{(1)}$ and $\zeta_n^{(2)}$ being some points in $(d_0, d_n + \eta/\nu_n - L_1 h_n)$ and $(d_0, d_n + L_0 h_n)$ respectively, we have

$$\begin{aligned} &\sqrt{nh_n}\{\mu(d_n + \eta/\nu_n - L_0 h_n) - \mu(d_n + L_0 h_n)\} \\ &\quad = \frac{\sqrt{nh_n}}{k!}\{\mu^{(k)}(\zeta_n^{(1)})(d_n + \eta/\nu_n - L_1 h_n - d_0)^k - \mu^{(k)}(\zeta_n^{(2)})(d_n + L_0 h_n - d_0)^k\} \end{aligned}$$

$$
\begin{aligned}
&> \frac{\sqrt{nh_n}\mu^{(k)}(\zeta_n^{(1)})}{k!}[(d_n+\eta/\nu_n - L_1h_n - d_0)^k - (d_n + L_0h_n - d_0)^k] \\
&\quad + \frac{\sqrt{nh_n}}{k!}[\mu^{(k)}(\zeta_n^{(1)}) - \mu^{(k)}(\zeta_n^{(2)})](d_n + L_0h_n - d_0)^k \\
&> \sqrt{nh_n}\left[\frac{\left\{\mu^{(k)}(d_0+) + o(1)\right\}(\eta/\nu_n - 2L_0h_n)^k}{k!} + o(1)(d_n - d_0 + L_0h_n)^k\right].
\end{aligned}
$$

Using Lemma 1, $(d_n - d_0)$ is $O(1/\nu_n)$ and hence, the above display is further bounded from below by

$$
\frac{\sqrt{nh_n}}{\nu_n^k}\left[\frac{\mu^{(k)}(d_0+)}{k!}(\eta - 2L_0)^k + o(1)\right].
$$

As $\sqrt{nh_n}/\nu_n^k \geq 1$, (A.5) holds.

Further, as the kernel $K(u)$ is non-increasing in $|u|$, $\bar{\mu}$ is non-decreasing in $(d_0, d_0 + 2\kappa)$. For $d \in (d_n + \eta/\nu_n, d_0 + 2\kappa)$,

$$
\begin{aligned}
M_n(d) - M_n(d_n) &\geq M_n(d) - M_n(d_0 + \eta/\nu_n) \\
&\geq \sum_{d_0+\eta/\nu_n \leq i/n \leq d} \left\{\Phi_{i,n}\left(\sqrt{nh_n}\bar{\mu}(i/n)/\sqrt{1+\Sigma_n^2(i/n)}\right) - \gamma\right\} \\
&\geq K_0(\lfloor nd \rfloor - \lfloor n(d_n + \eta/\nu_n)\rfloor)/n. \qquad \text{(A.6)}
\end{aligned}
$$

Using Lemma 1, there exists $A_0 < \infty$ such that for sufficiently large n, $\nu_n|d_0 - d_n| \leq A_0$, and hence $\{\rho_n(d, d_n) < \kappa\} \subset \{|d - d_0| < \kappa^2/K_1 + A/\nu_n + 2/n\} \subset \{|d - d_0| < 2\kappa\}$, where $A = 2\max(\eta, L_0, A_0)$. Letting $K_1 = (1/2)\min(\gamma - 1/2, K_0)$ and using (A.4) and (A.6), we get the desired result. □

A.2. Proof of Lemma 2

In order to prove *Lemma 2*, we first justify a few auxiliary results required to prove the tightness of W_n. Recall that

$$
W_n(t) = \sqrt{nh_n}\hat{\mu}(d_0 + th_n).
$$

Let $\bar{\epsilon}_n(\cdot)$ be such that $\bar{\epsilon}_n(t) = W_n(t) - \sqrt{nh_n}\bar{\mu}(d_0 + th_n)$, i.e.,

$$
\bar{\epsilon}_n(t) = \frac{1}{\sqrt{nh_n}}\sum_{i=1}^{n}\epsilon_i K\left(\frac{d_0 - i/n}{h_n} + t\right). \qquad \text{(A.7)}
$$

Lemma 6. *The processes $\sqrt{nh_n}\bar{\epsilon}_n(d_0 + th_n)$, $t \in \mathbb{R}$, are asymptotically tight in $C(\mathbb{R})$.*

Proof. As the kernel K is Lipschitz of order $\alpha > 1/2$, there exists a constant $C_0 > 0$, such that $|K(t) - K(s)| \leq C_0|t - s|^\alpha$. Fix $T > 0$. For $s, t \in [-T, T]$, we have

$$
\begin{aligned}
&E\left[\bar{\epsilon}_n(t)-\bar{\epsilon}_n(s)\right]^2 \\
&= \frac{1}{nh_n}\sum_{i=1}^{n}\sigma_0^2\left|K\left(\frac{d_0-i/n}{h_n}+t\right)-K\left(\frac{d_0-i/n}{h_n}+s\right)\right|^2 \\
&= \frac{1}{nh_n}\sum_{i=1}^{n}\sigma_0^2\left|K\left(\frac{d_0-i/n}{h_n}+t\right)-K\left(\frac{d_0-i/n}{h_n}+s\right)\right|^2 \\
&= \frac{1}{nh_n}\sum_{|d_0-i/n|<(L_0+T)h_n}\sigma_0^2\left|K\left(\frac{d_0-i/n}{h_n}+t\right)-K\left(\frac{d_0-i/n}{h_n}+s\right)\right|^2 \\
&\leq 4(L_0+T+1)\sigma_0^2C_0^2|t-s|^{2\alpha}.
\end{aligned}
$$

Since $\alpha > 1/2$, the result is a consequence of Theorem 12.3 of Billingsley (1968, pp. 95). □

We use a version of the Arzela-Ascoli theorem to prove the next result and thus we state it below for convenience.

Theorem 5 (Arzela-Ascoli)**.** *Let f_n be a sequence of continuous functions defined on a compact set $[a,b]$ such that f_n converge pointwise to f and for any $\delta_n \downarrow 0$ $\sup_{|x-y|<\delta_n}|f_n(x)-f_n(y)|$ converges to 0. Then $\sup_{x\in[a,b]}|f_n(x)-f(x)|$ converges to zero.*

Lemma 7. *The sequence of functions $\sqrt{nh_n}\bar{\mu}(d_0+th_n)$ converges to $m(t)$, uniformly over compact sets in $\mathbb{R}$.*

Proof. The pointwise convergence is evident from Lemma 2. To justify the uniform convergence, let $\bar{z}_n(x,t) = (1/h_n)\mu(x)K((d_0-x)/h_n+t)$. By arguments similar to those for Lemma 4, $|\bar{z}_n(x,t)-\bar{z}_n(y,t)| \lesssim 1/h_n^{1+\alpha}|x-y|^\alpha$ and consequently, for $t \in [-T,T]$,

$$
\begin{aligned}
&\left|\bar{\mu}(d_0+th_n)-\int \bar{z}_n(x,t)dx\right| \\
&\quad\leq O\left(\frac{1}{nh_n}\right)+\sum_{\substack{1\leq i\leq n\\ |d_n-i/n|\leq(L_0+T)h_n}}\frac{|i/n-x|^\alpha}{h_n^{1+\alpha}}dx = O\left(\frac{1}{(nh_n)^\alpha}\right).
\end{aligned}
$$

As the above bound does not depend on t and $\alpha > 1/2$, for $s,t \in [-T,T]$, and $\delta > 0$,

$$
\begin{aligned}
&\sup_{|t-s|<\delta}\left|\sqrt{nh_n}\bar{\mu}(d_0+th_n)-\sqrt{nh_n}\bar{\mu}(d_0+sh_n)\right| \\
&\quad= \sup_{|t-s|<\delta}\left|\sqrt{nh_n}\int_{-\infty}^{\infty}\{\bar{z}_n(x,t)-\bar{z}_n(x,s)\}dx\right|+o(1) \\
&\quad\leq \sqrt{nh_n}\int_{-\infty}^{\infty}\mu(d_0+uh_n)\left|K(t-u)-K(s-u)\right|du+o(1)
\end{aligned}
$$

$$\leq \quad \sqrt{nh_n} \int_0^{L_0+T} \frac{\mu^{(k)}(\zeta_u)}{k!} (uh_n)^k \left| K(t-u) - K(s-u) \right| du + o(1),$$

where ζ_u is some intermediate point between d_0 and d_0+uh_n. The k-th derivative of μ is bounded on $(d_0, d_0 + (L_0 + T)h_n)$ for sufficiently large n and $h_n^k\sqrt{nh_n}$ equals $h_0^{k+1/2}$. As K is uniformly continuous, the above display goes to zero as $\delta \to 0$ by DCT. Hence, by the Arzela–Ascoli theorem we get the desired result. □

We now continue with the proof of Lemma 2. For $(a_i, t_i) \in \mathbb{R}^2$, $i = 1, \dots l$, we have

$$\sum_{i,j} a_i a_j \mathrm{Cov}(W(t_i), W(t_j)) = \int \left\{ \sum_i a_i K(t_i + u) \right\}^2 du \geq 0.$$

Hence, the defined covariance function is non-negative definite and by Kolmogorov consistency, the Gaussian process W exists.

Let $r(h) = \{\int K(h+u)K(u)du\}/\bar{K}^2$ denote the correlation function of W. For W to have a continuous modification, by Hunt's theorem (e.g., see Cramér and Leadbetter (1967, pp. 169–171)), it suffices to show that $r(h)$ is $1-O((log(h))^{-\delta})$ for some $\delta > 3$ as $h \to 0$. Note that the kernel K is Lipschitz continuous of order α and hence, we have

$$\begin{aligned} |(1-r(h))(log(h))^\delta| &= \left| \frac{h^\alpha (log(h))^\delta}{\int K^2(u)du} \left(\int_{-L_0}^{L_0} \frac{(K(u) - K(h+u))}{h^\alpha} K(u)du \right) \right| \\ &\lesssim \frac{1}{\int K^2(u)du} \left| h^\alpha (log(h))^\delta \right| \to 0. \end{aligned}$$

Thus, W has a continuous modification. Next, we justify weak convergence of the process W_n to W.

As a consequence of Lemmas 6 and 7, the process W_n is asymptotically tight. To justify finite dimensional convergence, it suffices to show that:

$$\begin{pmatrix} W_n(t_1) \\ W_n(t_2) \end{pmatrix} \stackrel{d}{\to} \begin{pmatrix} W(t_1) \\ W(t_2) \end{pmatrix}, \tag{A.8}$$

where $t_1, t_2 \in \mathbb{R}$. Let $x_j = d_0 + t_j h_n$, $j = 1$ and 2. Then,

$$\begin{aligned} \bar{\mu}(x_j) &= \sqrt{nh_n} \left\{ \int_0^1 \frac{1}{h_n} \mu(x) K\left(\frac{x_j - x}{h_n} \right) dx + O\left(\frac{1}{(nh_n)^\alpha} \right) \right\} \\ &= \sqrt{nh_n} \int_{(d_0-1)/h_n+t_j}^{d_0/h_n+t_j} \mu(d_0 + (t_j - v)h_n) K(v)\, dv + o(1) \\ &= \sqrt{nh_n} \int_{(d_0-1)/h_n+t_j}^{t_j} \frac{\mu^{(k)}(d_0^+)}{k!} \left((t_j - v)^k h_n^k + o(h_n^k) \right) K(v)\, dv + o(1) \\ &= h_0^{k+1/2} \frac{\mu^{(k)}(d_0+)}{k!} \left(\int_{-\infty}^{t_j} (t_j - v)^k K(v)\, dv + o(1) \right) + o(1) \\ &= m(t_j) + o(1). \end{aligned}$$

The last step follows from DCT as the k-th derivative of μ is bounded in a right neighborhood of d_0 and $\int |v|^k K(v)dv < \infty$. Moreover,

$$\begin{aligned} E\left[\bar{\epsilon}_n(x_j)\right] &= 0, \\ Var\left[\bar{\epsilon}_n(x_j)\right] &= \Sigma_n^2(x_j) \to \sigma_0^2 \bar{K^2}, \end{aligned}$$

and, by a change of variable,

$$\begin{aligned} &\operatorname{Cov}\left[\bar{\epsilon}_n(x_1), \bar{\epsilon}_n(x_2)\right] \\ &= \operatorname{Cov}\left[\frac{1}{\sqrt{nh_n}} \frac{\sum \epsilon_i K\left((x_1 - i/n)/h_n\right)}{\Sigma_n(x_1)}, \frac{1}{\sqrt{nh_n}} \frac{\sum \epsilon_i K\left((x_2 - i/n)/h_n\right)}{\Sigma_n(x_2)}\right] \\ &\quad \to \sigma_0^2 \int K(t_1 + u) K(t_2 + u) du. \end{aligned}$$

Also,

$$\frac{\max_i K^2\left((x_j - i/n)/h_n\right)}{\sum K^2\left((x_j - i/n)/h_n\right)} \leq \frac{\|K\|_\infty^2}{nh_n(\bar{K^2} + o(1))} \to 0.$$

Hence, the Lindeberg–Feller condition is satisfied for $\bar{\epsilon}_n(x_j)$s and by the Cramér-Wold device, (A.8) holds. This justifies the finite dimensional convergence and hence, we have the result. □

A.3. Proof of Theorem 3

In order to prove Theorem 3, an ergodic theorem and Borell's inequality are found useful, which are stated below for convenience. For the proofs of the two results, see, for example, Cramér and Leadbetter (1967, pp. 147), and (Adler and Taylor, 2007, pp. 49–53), respectively. Also, we use Theorem 3 from Ferger (2004) which we state below as well.

Theorem 6. *Consider a real continuous second order stationary process $\xi(t)$ with mean 0 and correlation function $R(t)$. If*

$$\frac{1}{T}\int_0^T R(t)dt = O\left(\frac{1}{T^a}\right)$$

for any $a > 0$, then ξ satisfies the law of large numbers, i.e., $T^{-1}\int_0^T \xi(t)dt$ converges a.s. to zero as $T \to \infty$.

Theorem 7 (Borell's inequality)**.** *Let ξ be a centered Gaussian process, a.s. bounded on a set I. Then $E\{\sup_{u\in I} \xi(u)\} < \infty$ and for all $x > 0$,*

$$P\left\{\sup_{u\in I} \xi(u) - E\left(\sup_{u\in I} \xi(u)\right) > x\right\} \leq \exp\left(\frac{-x^2}{2\sigma_I^2}\right),$$

where $\sigma_I^2 = \sup_{u\in I} Var\{\xi(u)\}$.

Theorem 8 (Ferger (2004))**.** *Let $\mathbb{V}_n$, $n \geq 0$, be stochastic processes in $D(\mathbb{R})$, defined on a common probability space $(\Omega, \mathcal{A}, P)$. Let ξ_n be a Borel-measurable minimizer of $\mathbb{V}_n$. Suppose that:*

(i) $\mathbb{V}_n$ converges weakly to $\mathbb{V}_0$ in $D[-C, C]$ for each $C > 0$.
(ii) The trajectories of $\mathbb{V}_0$ almost surely possess a smallest and a largest minimizer ξ_0^s and ξ_0^l respectively, which are Borel measurable.
(iii) The sequence ξ_n is uniformly tight.

Then for every $x \in \mathbb{X}$,

$$P[\xi_0^l < x] \leq \liminf_{n\to\infty} P_*[\xi_n < x] \leq \limsup_{n\to\infty} P^*[\xi_n \leq x] \leq P[\xi_0^s \leq x].$$

Here, $\mathbb{X} = \{x \in \mathbb{R} : P[\mathbb{V}_0$ is continuous at $x] = 1\}$.

We now continue with the proof of Theorem 3. Let

$$W_0(t) = W(t) - m(t).$$

This is a mean zero stationary process and thus, so is the process $D(t) = \Phi(W_0(t)) - 1/2$ with correlation function $R(t)$, say. As, K is supported on $[-L_0, L_0]$, $W(t_1)$ and $W(t_2)$ are independent whenever $|t_1 - t_2| \geq 2L_0$ and hence $R(t) = 0$ for $t > 2L_0$. So, $(1/t)\int_0^t R(y)dy = O(1/t)$ as $|t| \to \infty$ and therefore, by *Theorem 6*, $Z_1(t) = (1/t)\int_0^t D(y)dy \to 0$ a.s. as $|t| \to \infty$. For $t < 0$, we write $Z(t)$ as

$$Z(t) \quad = \quad t\left[Z_1(t) + (1/2 - \gamma) + (1/t)\int_0^t \{\Phi(W(t)) - \Phi(W_0(t))\}\, dy\right].$$

When $t < -L_0$, $m(t) = 0$, which gives $W(t) = W_0(t)$ and hence the third term in the above display goes to zero and $Z(t) \to \infty$ a.s. as $t \to -\infty$. For $t > 0$, fix $M > 0$ and j be a positive integer. Then

$$\begin{aligned} P\left[\inf_{t\in[j,j+1]} W(t) < M\right] &\leq P\left[\inf_{t\in[j,j+1]} W_0(t) + \inf_{t\in[j,j+1]} m(t) < M\right] \\ &= P\left[\sup_{t\in[j,j+1]} (-W_0(t)) > m(j) - M\right], \end{aligned}$$

as $\inf_{t\in[j,j+1]} m(t) = m(j)$. By Borell's inequality, the above probability is bounded by $\exp[\{-m(n) - L_0 - E\sup_{t\in[j,j+1]}(-W_0(t))\}^2]$, where by stationarity, $E\sup_{t\in[j,j+1]}(-W_0(t)) = E\sup_{t\in[0,1]}(-W_0(t))$ which is finite, again due to Borell's inequality. Also, it can be seen that $m(j) \gtrsim (j - L_0)^k$ and hence $\sum_{j\geq 1} P[\sup_{t\in[j,j+1]}(-W_0(t)) > m(j) - M] < \infty$. Using Borel–Cantelli lemma, we get $P[\liminf_{t\to\infty} W(t) > M] = 1$. As M can be made arbitrarily large, we get that $W(t)$ diverges to ∞ a.s. as $t \to \infty$ and consequently so does $Z(t)$.

Note that Z_n (defined in (3.5)) converges weakly to Z in $B_{loc}(\mathbb{R})$ and consequently, in $D(\mathbb{R})$ as well. Moreover, Z has continuous sample paths with probability 1. As $Z(t) \to \infty$ when $|t| \to \infty$, ξ_0^s and ξ_0^l are well defined and Borel measurable. Further, recall that $h_n^{-1}(\hat{d}_n - d_0)$, the smallest argmin of the process $Z_n(\cdot)$, is determined by the ordering of finitely many random variables and hence, is measurable. Also, by Theorem 1, it is $O_p(1)$. Hence, conditions (i), (ii) and (iii) of Theorem 8 are satisfied with $\mathbb{V}_n = Z_n$ and $\mathbb{V}_0 = Z$, and thus,

$$\begin{aligned}\liminf_{n\to\infty} P[c^s_{\alpha/2} < h_n^{-1}(\hat{d}_n - d_0) < c^l_{1-\alpha/2}] &\geq \liminf_{n\to\infty} P[h_n^{-1}(\hat{d}_n - d_0) < c^l_{1-\alpha/2}] \\ &\quad - \limsup_{n\to\infty} P[h_n^{-1}(\hat{d}_n - d_0) \leq c^s_{\alpha/2}] \\ &\geq 1-\alpha.\end{aligned}$$

Hence, we get the desired result. □

A.4. Outline of the proof of Proposition 1

We assume the rate of convergence for the proof as it is a consequence of arguments similar to that for the proof of Proposition 3 (see Section A.6).

To see that **Method 1** ends up yielding the given limiting distribution, recall that for $\tau = 0$,

$$\hat{d}_n^1 = \operatorname*{sargmin}_{d\in[0,1]} \frac{1}{n}\sum_{i=1}^n \left\{\Phi\left(\frac{\sqrt{nh_n}\hat{\mu}(i/n)}{\Sigma_n(i/n,\hat{\sigma})}\right) - \gamma\right\} 1\left(\frac{i}{n} \leq d\right).$$

Thus, the form of the limit distribution is dictated by the asymptotic behavior of the local process

$$Z_n^1(t) = \frac{1}{nh_n}\sum_{i=1}^n \left\{\Phi\left(\frac{\sqrt{nh_n}\hat{\mu}(i/n)}{\Sigma_n(i/n,\hat{\sigma})}\right) - \gamma\right\}\left(1\left(\frac{i}{n} \leq d_0 + th_n\right) - 1\left(\frac{i}{n} \leq d_0\right)\right).$$

Proceeding as we did in the proof of Theorem 2, Z_n^1 can be split into $I_n^1(t) + II_n^1(t)$, where

$$II_n^1(t) = h_n^{-1}\int_{d_0}^{d_0+th_n} \left(\Phi\left(\frac{\sqrt{nh_n}\hat{\mu}(x)}{\Sigma_n(x,\hat{\sigma})}\right) - \gamma\right) dx \tag{A.9}$$

and the contribution of $I_n^1(t) = Z_n^1(t) - II_n^1(t)$ can be shown to converge to zero. By a change of variable, II_n^1 can be written as

$$II_n^1(t) = \int_0^t \left[\Phi\left(\frac{W_n(y)}{\Sigma_n(d_0+yh_n,\hat{\sigma})}\right) - \gamma\right] f(d_0+yh_n)dy,$$

where, W_n is as defined in (3.6). This term differs from its analogue for **Method 2** (see (2.6)) through the normalizing factor $\Sigma_n(d_0+yh_n,\hat{\sigma})$ which converges in probability to $\sigma_0\sqrt{\bar{K^2}}$. The tightness of the ratio process $W_n(y)/\Sigma_n(d_0+yh_n,\hat{\sigma})$ can be established through calculations similar to those in the proof of Lemma 2. Hence, by a Slutsky-type argument, we get that

$$h_n^{-1}(\hat{d}_n^1 - d_0) \xrightarrow{d} \operatorname*{argmin}_{t\in\mathbb{R}} \int_0^t \left\{\Phi\left(\frac{W(y)}{\sigma_0\sqrt{\bar{K^2}}}\right) - \gamma\right\} dy,$$

for $h_n = h_0 n^{-1/(2k+1)}$. Note that the process on the right side of the above display is precisely Z_1. This completes the proof. □

A.5. Outline of the proof of Proposition 2

Here, we provide a brief outline of the proof to convince the reader about the form of the limiting distribution. Note that this is dictated by the asymptotic behavior of the local process

$$\tilde{Z}_n(t) \quad = \quad \mathbb{P}_n\left[\left\{\Phi\left(\sqrt{nh_n}\tilde{\mu}(X)\right) - \gamma\right\}(1(X \le d_0 + th_n) - 1(X \le d_0))\right]$$

that arises out of the criterion in (2.7) (with $\tau = 0$). As in the proof of Theorem 2, $\tilde{Z}_n$ can be split into $\tilde{I}_n(t) + \tilde{II}_n(t)$, where

$$\tilde{II}_n(t) = h_n^{-1}\int_{d_0}^{d_0+th_n}\left(\Phi\left(\sqrt{nh_n}\tilde{\mu}(x)\right) - \gamma\right)f(x)dx \tag{A.10}$$

and the contribution of $\tilde{I}_n(t) = \tilde{Z}_n(t) - \tilde{II}_n(t)$ can be shown to go to zero. By a change of variable, $\tilde{II}_n$ can be written as

$$\tilde{II}_n(t) = \int_0^t\left[\Phi\left(\tilde{W}_n(y)\right) - \gamma\right]f(d_0 + yh_n)dy,$$

where $\tilde{W}_n(y) = \sqrt{nh_n}\tilde{\mu}(d_0 + yh_n)$. The process $\tilde{W}_n$ can be shown to converge weakly to the process $\tilde{W}$ by an imitation of the arguments in the proof of Lemma 2. Also, $f(d_0 + yh_n)$ converges to $f(d_0) > 0$. Consequently

$$h_n^{-1}(\tilde{d}_n - d_0) \stackrel{d}{\to} \operatorname*{argmin}_{t\in\mathbb{R}}\left\{f(d_0)\tilde{Z}(t)\right\} = \operatorname*{argmin}_{t\in\mathbb{R}}\left\{\tilde{Z}(t)\right\}.$$

□

A.6. Proof of Proposition 3

Recall that

$$\mathbb{M}_n(d,\tilde{\tau}) = \frac{1}{n}\sum_{i=1}^n\left[\Phi\left(\sqrt{nh_n}\left(\hat{\mu}\left(\frac{i}{n}\right) - \tilde{\tau}\right)\right) - \gamma\right]1\left(\frac{i}{n} \le d\right)$$

and $M_n(d,\tilde{\tau}) = E[\mathbb{M}_n(d,\tilde{\tau})]$. We make the dependence on the parameter τ explicit for the analysis. Here $M_n(d,\hat{\tau})$ is interpreted as $M_n(d,\tilde{\tau})$ computed at $\tilde{\tau} = \hat{\tau}$. Now, we extend the proof of Theorem 1 to show that the rate of convergence remains the same.

Rate of convergence. As $\sqrt{n}(\hat{\tau} - \tau) = O_p(1)$, for any $\epsilon > 0$, there exists $V_{\epsilon/2} > 0$ such that $P[\sqrt{n}|\hat{\tau} - \tau| < V_{\epsilon/2}] > 1 - \epsilon$. To show that the rate of convergence does not change, we need to derive a bound on

$$E\left[\sqrt{n}\sup_{\rho_n(d,d_n)<\delta}|(\mathbb{M}_n(d,\hat{\tau}) - \mathbb{M}_n(d,\hat{\tau})) - (M_n(d_n,\tau) - M_n(d_n,\tau))|\right]$$

having the same order as $\phi_n(\delta)$ (see (A.2) and (A.3)). A slight relaxation is possible. For each $\epsilon > 0$, it suffices to find a bound of the form

$$E\left[\sqrt{n}\sup_{\rho_n(d,d_n)<\delta}|(\mathbb{M}_n - M_n)(d,\hat{\tau}) - (\mathbb{M}_n - M_n)(d_n,\tau)|1(\Omega_\epsilon)\right] \leq C_\epsilon\phi_n(\delta), \tag{A.11}$$

where $P[\Omega_\epsilon] > 1-\epsilon$ and $C_\epsilon > 0$; see Banerjee and McKeague (2007, Theorem 5.2). For $\Omega_\epsilon = [\hat{\tau} \in [\tau - V_{\epsilon/2}/\sqrt{n}, \tau + V_{\epsilon/2}/\sqrt{n}]]$, the left side of the above display can be bounded by

$$\begin{aligned}
&E\left[\sqrt{n}\sup_{\rho_n(d,d_n)<\delta, |\tilde{\tau}-\tau|<V_{\epsilon/2}/\sqrt{n}}|(\mathbb{M}_n - M_n)(d,\tilde{\tau}) - (\mathbb{M}_n - M_n)(d_n,\tau)|\right] \\
&\quad \leq E\left[\sqrt{n}\sup_{\rho_n(d,d_n)<\delta}|(\mathbb{M}_n - M_n)(d,\tau) - (\mathbb{M}_n - M_n)(d_n,\tau)|\right] \\
&\qquad + E\left[\sqrt{n}\sup_{\substack{\rho_n(d,d_n)<\delta,\\ |\tilde{\tau}-\tau|<V_{\epsilon/2}/\sqrt{n}}}|(\mathbb{M}_n(d,\tilde{\tau}) - \mathbb{M}_n(d,\tau)) - (\mathbb{M}_n(d_n,\tilde{\tau}) - \mathbb{M}_n(d_n,\tau))|\right].
\end{aligned}$$

The first term on the right side is precisely the term dealt in the case of a known τ (see (A.2)). As for the second term, note that by the Lipschitz continuity of Φ,

$$\begin{aligned}
&|(\mathbb{M}_n(d,\tilde{\tau}) - \mathbb{M}_n(d,\tau)) - (\mathbb{M}_n(d_n,\tilde{\tau}) - \mathbb{M}_n(d_n,\tau))| \\
&\quad \leq \frac{1}{n}\sum_{i=1}^{n}\left\{\left|\Phi\left(\sqrt{nh_n}\left(\hat{\mu}\left(\frac{i}{n}\right) - \tilde{\tau}\right)\right) - \Phi\left(\sqrt{nh_n}\left(\hat{\mu}\left(\frac{i}{n}\right) - \tau\right)\right)\right|\right. \\
&\qquad \left.\times\left|1\left(\frac{i}{n} \leq d\right) - 1\left(\frac{i}{n} \leq d_n\right)\right|\right\} \\
&\quad \lesssim \sqrt{nh_n}|\tilde{\tau} - \tau|\frac{|\lceil nd\rceil - \lceil nd_n\rceil|}{n}.
\end{aligned}$$

Thus, we have

$$\begin{aligned}
&E\left[\sqrt{n}\sup_{\substack{\rho_n(d,d_n)<\delta,\\ |\tilde{\tau}-\tau|<V_{\epsilon/2}/\sqrt{n}}}|(\mathbb{M}_n(d,\tilde{\tau}) - \mathbb{M}_n(d,\tau)) - (\mathbb{M}_n(d_n,\tilde{\tau}) - \mathbb{M}_n(d_n,\tau))|\right] \\
&\qquad\qquad \lesssim \sqrt{n}\frac{\sqrt{nh_n}V_{\epsilon/2}}{\sqrt{n}}(\delta^2 + 2/n) \lesssim V_{\epsilon/2}\phi_n(\delta),
\end{aligned}$$

for $\delta < 1$ and large n. Hence, the expression in (A.11) has the same bound $\phi_n(\cdot)$ (up to a different constant) and thus, we get the same rate of convergence.

Limit distribution. Recall from (3.5) that

$$Z_n(t) = Z_n(t,\tau) = h_n^{-1}\left[\mathbb{M}_n(d_0 + th_n,\tau) - \mathbb{M}_n(d_0,\tau)\right].$$

To show that the limiting distribution of $\hat{d}_n$ remains the same, it suffices to show that

$$\sup_{t\in[-T,T]} |Z_n(t,\hat{\tau}) - Z_n(t,\tau)| \tag{A.12}$$

converges in probability to zero, for any $T > 0$. Again by the Lipschitz continuity of Φ,

$$\begin{aligned} & |Z_n(t,\hat{\tau}) - Z_n(t,\tau)| \\ &= \frac{1}{nh_n}\left|\sum_{i=1}^n \left\{\Phi\left(\sqrt{nh_n}(\hat{\mu}\left(\frac{i}{n}\right) - \hat{\tau})\right) - \Phi\left(\sqrt{nh_n}(\hat{\mu}\left(\frac{i}{n}\right) - \tau)\right)\right.\right. \\ &\qquad \left.\left. \times \left(1\left(\frac{i}{n} \le d_0 + th_n\right) - 1\left(\frac{i}{n} \le d_0\right)\right)\right\}\right| \\ &\lesssim \frac{1}{nh_n}\left|\sum_{i=1}^n \sqrt{nh_n}|\hat{\tau}-\tau|\left(1\left(\frac{i}{n} \le d_0 + Th_n\right) - 1\left(\frac{i}{n} \le d_0\right)\right)\right| \\ &\le h_n\frac{(Tnh_n+2)}{nh_n}\sqrt{n}|\hat{\tau}-\tau|. \end{aligned}$$

As the above bound is uniform in $t \in [-T,T]$ and $\sqrt{n}(\hat{\tau} - \tau)$ is $O_p(1)$, the expression in (A.12) converges in probability to zero and hence, we get the desired result. □

A.7. Proof of Proposition 4

Given what has been done earlier for proving results from Section 3.2, it suffices to show that the process $\bar{\epsilon}_n(t)$, defined in (A.7), converges weakly to a mean zero Gaussian process having the covariance function of $\bar{W}$ in the setup of Section 6. As $W_n(t) = \sqrt{nh_n}\bar{\mu}(d_0 + th_n) + \bar{\epsilon}_n(t)$, Lemma 7 then justifies the weak convergence of W_n to $\bar{W}$. The statement and the proof of Lemma 2 relies on the i.i.d. assumption only through the convergence of W_n's and the form of their limit. Hence, it would follow that the process Z_n (defined in (3.5)) converges to $\bar{Z}$. The result then follows from applying the argmin continuous mapping theorem as in proving Theorem 4.

We start by showing the covariance function of the process $\bar{\epsilon}_n$ converges to that of $\bar{W}$. For $t_1, t_2 \in \mathbb{R}$, let $x_j = d_0 + t_j h_n$, $j = 1, 2$. We have

$$\mathrm{Cov}(\bar{\epsilon}_n(t_1), \bar{\epsilon}_n(t_2)) = \frac{\sigma_0^2}{nh_n}\sum_{l,j}\rho(l-j)K\left(\frac{x_1 - l/n}{h_n}\right)K\left(\frac{x_2 - j/n}{h_n}\right).$$

As $\sigma_0^2\rho(l-j) = \int_{-\pi}^{\pi}\psi(u)\exp(\imath(l-j)u)du$, the above expression reduces to

$$\frac{1}{nh_n}\int_{-\pi}^{\pi}\psi(u)\hat{K}_{x_1}(u)\hat{K}_{x_2}(-u)du,$$

where for $x, u \in \mathbb{R}$, $\hat{K}_x(u) = \sum_j K(h^{-1}\{x - j/n\})e^{\imath ju}$. Under short range dependence, Assumption 1 of Robinson (1997) requires ψ to be an even nonnegative function which is continuous and positive at 0. Using this assumption, it can be shown that the difference between the above display and

$$\frac{\psi(0)}{nh_n}\int_{-\pi}^{\pi}\hat{K}_{x_1}(u)\hat{K}_{x_2}(-u)du$$

goes to zero by calculations almost identical to those in Robinson (1997, pp. 2061–2062). As $\int_{-\pi}^{\pi}\exp(\imath(l-j)u)du = 2\pi\delta_{lj}$, with δ_{lj} being the Kronecker delta, the above expression equals

$$\frac{2\pi\psi(0)}{nh_n}\sum_{l,j}\delta_{lj}K\left(\frac{x_1 - l/n}{h_n}\right)K\left(\frac{x_2 - j/n}{h_n}\right).$$

Following the arguments identical to that in the proof of Lemma 2, this expression can be shown to converge to the covariance function of $\bar{W}$. What remains now is the justification of the asymptotic normality of finite dimensional marginals of $\bar{\epsilon}_n$ and proving tightness.

Justifying asymptotic normality of the finite dimensional marginals of $\bar{\epsilon}_n$ requires showing the asymptotic normality of any finite linear combination of marginals of $\bar{\epsilon}_n$ and then applying the Cramér-Wold device. Given the convergence of the covariances, it suffices to prove that for $(c_r, t_r) \in \mathbb{R}, 1 \le r \le R \in \mathbb{N}$,

$$\frac{1}{\sqrt{v_n}}\sum_{r\le R}c_r\bar{\epsilon}_n(t_r) \xrightarrow{d} N(0,1), \tag{A.13}$$

where $v_n^2 = \ \mathrm{Var}(\sum_{r\le r}c_r\bar{\epsilon}_n(t_r))$. The left hand side equals $\sum_i w_{in}\epsilon_i$ where

$$w_{in} = \frac{1}{\sqrt{nh_n}v_n}\sum_{r\le R}c_rK\left(\frac{d_0 - i/n}{h_n} + t_r\right).$$

As in (Robinson, 1997, Assumption 2), we assume ϵ_i's to be a linear process with martingale innovations and square summable coefficients, i.e, there is a sequence of martingale differences $u_j, j \in \mathbb{Z}$ adapted to $\mathcal{F}_j = \sigma\{u_k : k \le j\}$ with mean 0 and variance 1, such that

$$\epsilon_i = \sum_{j=-\infty}^{\infty}\alpha_j u_{i-j}, \quad \sum_{j=-\infty}^{\infty}\alpha_j^2 < \infty. \tag{A.14}$$

To show asymptotic normality, we justify conditions (2.3) and (2.6) from Robinson (1997). The condition (2.3) is just a normalization requirement which holds in our case as the variance of the left hand side of (A.13) is 1. The condition (2.6) of Robinson (1997) is about justifying the existence of a positive-valued sequence a_n such that as $n \to \infty$,

$$\left(\sum_i w_{in}^2\sum_{|j|>a_n}\alpha_j^2\right)^{1/2} + \max_{1\le i\le n}|w_{in}|\sum_{|j|\le a_n}|\alpha_j| \to 0. \tag{A.15}$$

For a_n such that $a_n \to \infty$ and $nh_n/a_n \to \infty$, $\sum_{|j|>a_n} \alpha_j^2 = o(1)$, due to (A.14). Also, by Cauchy-Schwartz, $\sum_{|j|\leq a} |\alpha_j| = O(\sqrt{a_n})$. By the compactness of the kernel and the fact that $v_n = O(1)$, $\sum_i w_{in}^2 = O(1)$. As the kernel K is bounded, $\max_{1\leq i\leq n} |w_{in}| = O(1/\sqrt{nh_n})$. Hence, the left hand side of (A.15) is $o(1) + O(\sqrt{a_n/nh_n})$ which is $o(1)$. This shows convergence of the finite dimensional marginals.

For tightness, recall that for $t \in [-T, T]$

$$\bar{\epsilon}_n(t) = \frac{1}{\sqrt{nh_n}} \sum_{i:|d_0-i/n|\leq (L_0+T)h_n} \epsilon_i K\left(\frac{d_0 - i/n}{h_n} + t\right).$$

We have

$$E\left[\bar{\epsilon}_n(t_1) - \bar{\epsilon}_n(t_2)\right]^2 = \frac{1}{nh_n}\int_{-\pi}^{\pi} \psi(u)(\hat{K}_{x_1}(u) - \hat{K}_{x_2}(u))(\hat{K}_{x_1}(-u) - \hat{K}_{x_2}(-u))du.$$

As ψ is a bounded function, the above expression is bounded up to a constant, due to Cauchy-Schwartz, by

$$\frac{1}{nh_n}\int_{-\pi}^{\pi} |\hat{K}_{x_1}(u) - \hat{K}_{x_2}(u)|^2 du.$$

As $\hat{K}_{x_1}(u) = \sum_j K((x_1 - j/n)/h_n)e^{\iota ju}$,

$$|\hat{K}_{x_1}(u) - \hat{K}_{x_2}(u)|^2 \lesssim nh_n|t_1 - t_2|^{2\alpha}$$

due to Lipschitz continuity of K. Hence,

$$E\left[\bar{\epsilon}_n(t_1) - \bar{\epsilon}_n(t_2)\right]^2 \quad \lesssim \quad |t_1 - t_2|^{2\alpha}$$

The tightness follows from Theorem 12.3 of Billingsley (1968, pp. 95). □

References

Adler, R. J. and Taylor, J. E. (2007). *Random fields and geometry. Springer Monographs in Mathematics.* Springer, New York. MR2319516

Anevski, D. and Hössjer, O. (2006). A general asymptotic scheme for inference under order restrictions. *Ann. Statist.* **34** 1874–1930. MR2283721

Banerjee, M. (2009). Inference in exponential family regression models under certain shape constraints using inversion based techniques. In *Advances in multivariate statistical methods. Stat. Sci. Interdiscip. Res.* **4** 249–271. World Sci. Publ., Hackensack, NJ. MR2603933

Banerjee, M. and McKeague, I. W. (2007). Confidence sets for split points in decision trees. *Ann. Statist.* **35** 543–574. MR2336859

Banerjee, M. and Wellner, J. A. (2005). Confidence intervals for current status data. *Scand. J. Statist.* **32** 405–424. MR2204627

Bhattacharya, P. K. (1994). Some aspects of change-point analysis. In *Change-point problems (South Hadley, MA, 1992). IMS Lecture Notes Monogr. Ser.* **23** 28–56. Inst. Math. Statist., Hayward, CA. MR1477912

Billingsley, P. (1968). *Convergence of probability measures*. John Wiley & Sons Inc., New York. MR0233396

Brodsky, B. E. and Darkhovsky, B. S. (1993). *Nonparametric methods in change-point problems*. *Mathematics and its Applications* **243**. Kluwer Academic Publishers Group, Dordrecht. MR1228205

Cheng, M.-Y. and Raimondo, M. (2008). Kernel methods for optimal change-points estimation in derivatives. *J. Comput. Graph. Statist.* **17** 56–75. MR2424795

Cramér, H. and Leadbetter, M. R. (1967). *Stationary and related stochastic processes. Sample function properties and their applications*. John Wiley & Sons Inc., New York.

Csörgő, M. and Horváth, L. (1997). *Limit theorems in change-point analysis*. *Wiley Series in Probability and Statistics*. John Wiley & Sons Ltd., Chichester. With a foreword by David Kendall. MR2743035

Csörgő, S. and Mielniczuk, J. (1995a). Close short-range dependent sums and regression estimation. *Acta Sci. Math. (Szeged)* **60** 177–196. MR1348687

Csörgő, S. and Mielniczuk, J. (1995b). Nonparametric regression under long-range dependent normal errors. *Ann. Statist.* **23** 1000–1014. MR1345211

Dümbgen, L. (1991). The asymptotic behavior of some nonparametric change-point estimators. *Ann. Statist.* **19** 1471–1495. MR1126333

Dümbgen, L. and Spokoiny, V. G. (2001). Multiscale testing of qualitative hypotheses. *Ann. Statist.* **29** 124–152. MR1833961

Dümbgen, L. and Walther, G. (2008). Multiscale inference about a density. *Ann. Statist.* **36** 1758–1785. MR2435455

Ermakov, M. S. (1976). On the asymptotic behavior of statistical estimates for samples having a density with singularities. *Theory Probab. Appl.* **21** 649–651.

Ferger, D. (1999). On the uniqueness of maximizers of Markov-Gaussian processes. *Statist. Probab. Lett.* **45** 71–77. MR1718353

Ferger, D. (2004). A continuous mapping theorem for the argmax-functional in the non-unique case. *Statist. Neerlandica* **58** 83–96. MR2042258

Gasser, T., Sroka, L. and Jennen-Steinmetz, C. (1986). Residual variance and residual pattern in nonlinear regression. *Biometrika* **73** 625–633. MR0897854

Goldenshluger, A., Tsybakov, A. and Zeevi, A. (2006). Optimal change-point estimation from indirect observations. *Ann. Statist.* **34** 350–372. MR2275245

Goldenshluger, A., Juditsky, A., Tsybakov, A. B. and Zeevi, A. (2008). Change-point estimation from indirect observations. I. Minimax complexity. *Ann. Inst. Henri Poincaré Probab. Stat.* **44** 787–818. MR2453845

Gut, A. and Steinebach, J. (2009). Truncated sequential change-point detection based on renewal counting processes. II. *J. Statist. Plann. Inference* **139** 1921–1936. MR2497549

Hall, P., Kay, J. W. and Titterington, D. M. (1990). Asymptotically optimal difference-based estimation of variance in nonparametric regression. *Biometrika* **77** 521–528. MR1087842

Hinkley, D. V. (1970). Inference about the change-point in a sequence of random variables. *Biometrika* **57** 1–17. MR0273727

Horváth, L., Horváth, Z. and Hušková, M. (2008). Ratio tests for change point detection. In *Beyond parametrics in interdisciplinary research: Festschrift in honor of Professor Pranab K. Sen. Inst. Math. Stat. Collect.* **1** 293–304. Inst. Math. Statist., Beachwood, OH. MR2462213

Hušková, M. (1998). Estimators in the location model with gradual changes. *Comment. Math. Univ. Carolin.* **39** 147–157. MR1623002

Hušková, M., Kirch, C., Prášková, Z. and Steinebach, J. (2008). On the detection of changes in autoregressive time series. II. Resampling procedures. *J. Statist. Plann. Inference* **138** 1697–1721. MR2400475

Jarušková, D. (1998). Change-point estimator in gradually changing sequences. *Comment. Math. Univ. Carolin.* **39** 551–561. MR1666790

Kim, J. and Pollard, D. (1990). Cube root asymptotics. *Ann. Statist.* **18** 191–219. MR1041391

Korostelëv, A. P. (1987). Minimax estimation of a discontinuous signal. *Teor. Veroyatnost. i Primenen.* **32** 796–799. MR0927265

Korostelëv, A. P. and Tsybakov, A. B. (1993). *Minimax theory of image reconstruction. Lecture Notes in Statistics* **82**. Springer-Verlag, New York. MR1226450

Lifshits, M. A. (1982). On the absolute continuity of the distributions of functionals of stochastic processes. *Theory Probab. Appl.* **27** 600–607. MR0673927

Loader, C. R. (1996). Change point estimation using nonparametric regression. *Ann. Statist.* **24** 1667–1678. MR1416655

Mallik, A., Sen, B., Banerjee, M. and Michailidis, G. (2011). Threshold estimation based on a p-value framework in dose-response and regression settings. *Biometrika* **98** 887–900. MR2860331

Müller, H.-G. (1992). Change-points in nonparametric regression analysis. *Ann. Statist.* **20** 737–761. MR1165590

Müller, H.-G. and Song, K.-S. (1996). A set-indexed process in a two-region image. *Stochastic Process. Appl.* **62** 87–101. MR1388764

Müller, H.-G. and Song, K.-S. (1997). Two-stage change-point estimators in smooth regression models. *Statist. Probab. Lett.* **34** 323–335. MR1467437

Neumann, M. H. (1997). Optimal change-point estimation in inverse problems. *Scand. J. Statist.* **24** 503–521. MR1615339

Pflug, G. C. (1983). The limiting log-likelihood process for discontinuous density families. *Z. Wahrsch. Verw. Gebiete* **64** 15–35. MR0710646

Robinson, P. M. (1997). Large-sample inference for nonparametric regression with dependent errors. *Ann. Statist.* **25** 2054–2083. MR1474083

Schlesinger, M. E. and Ramankutty, N. (1994). An oscillation in the global climate system of period 65-70 years. *Nature* **367** 723-726.

Seijo, E. and Sen, B. (2011). Change-point in stochastic design regression and the bootstrap. *Ann. Statist.* **39** 1580–1607. MR2850213

Sen, B., Banerjee, M. and Woodroofe, M. (2010). Inconsistency of bootstrap: the Grenander estimator. *Ann. Statist.* **38** 1953–1977. MR2676880

Steland, A. (2010). A surveillance procedure for random walks based on local linear estimation. *J. Nonparametr. Stat.* **22** 345–361. MR2662597

van der Vaart, A. W. and Wellner, J. A. (1996). *Weak convergence and empirical processes: with applications to statistics*. *Springer Series in Statistics*. Springer-Verlag, New York. MR1385671

Willett, R. M. and Nowak, R. D. (2007). Minimax optimal level set estimation. *IEEE Trans. Image Process.* **16** 2965–2979. MR2472804

Wishart, J. (2009). Kink estimation with correlated noise. *J. Korean Statist. Soc.* **38** 131–143. MR2649688

Wishart, J. and Kulik, R. (2010). Kink estimation in stochastic regression with dependent errors and predictors. *Electron. J. Stat.* **4** 875–913. MR2721037

Wright, F. T. (1981). The asymptotic behavior of monotone regression estimates. *Ann. Statist.* **9** 443–448. MR0606630

Wu, W. B., Woodroofe, M. and Mentz, G. (2001). Isotonic regression: another look at the changepoint problem. *Biometrika* **88** 793–804. MR1859410

Zhao, O. and Woodroofe, M. (2012). Estimating a monontone trend. *Statist. Sinica* **22** 359–378. MR2933180