Asymptotics for p-value based threshold estimation in dose-response settings

Atul Mallik[1], Bodhisattva Sen[2,*], Moulinath Banerjee[1,†] and George Michailidis[1]
atulm@umich.edu, bodhi@stat.columbia.edu, moulib@umich.edu, gmichail@umich.edu
[1]University of Michigan and [2]Columbia University

May 15, 2013

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 arising in dose-response studies, engineering and other related fields. We study the procedure under the so called "dose-response" setting, where several responses can be obtained at each covariate-level. The estimator is constructed via fitting a "stump" function to approximate p-values that test for deviation of the regression function from its baseline level. The smoothness of the regression function in the vicinity of the threshold determines the optimal rate of convergence: a "cusp" of order k at the threshold yields an optimal rate of $N^{-1/(2k+1)}$, where N is the total budget. The asymptotic distribution of the normalized estimator is shown to be the minimizer of a generalized compound Poisson process. A limited simulation study is used to illustrate the method and an application to data from a complex queuing system presented.

Keywords: *Baseline value, change point, least squares, nonparametric estimation, stump function, weighted Poisson process.*

1 Introduction

In diverse applications, the following canonical model generates the available data:

$$Y = \mu(X) + \epsilon, \tag{1}$$

where μ is a function on $[0, 1]$ such that

$$\mu(x) \begin{cases} = \tau & \text{for } x \le d_0 \\ > \tau & \text{for } x > d_0; \end{cases} \tag{2}$$

here $\tau \in \mathbb{R}$, $d_0 \in (0, 1)$, and ϵ has mean zero with finite positive variance. The covariate X may arise from a random or a fixed design and we assume that we have *repeated measurements* on the regression function for each value of the covariate. The function μ need not be monotone and the baseline value τ is not necessarily known.

*Supported by NSF Grants DMS-1150435 and AST-1107373
†Supported by NSF Grant DMS-1007751, NSA grant H98230-11-1-0166 and a Sokol Faculty Award, University of Michigan

This model arises in a number of important contexts, especially dose-response studies, where a number of dose-levels are selected by the experimenter and each dose administered to a group of individuals. The postulated form of μ is a natural one in such experiments: $\mu(x)$ provides information about reaction to dose-level x and is typically at the baseline value up to a certain dose – the d_0 in our model – and changes from there on. In pharmacological contexts, this is typically referred to as the minimum effective dose (MED); see Chen and Chang (2007) and Tamhane and Logan (2002) and the references therein. For applications in toxicology, see e.g., Cox (1987), who uses parametrically specified threshold models, and Mallik et al. (2011), who study a biological application depicting the physiological response of cells from a leukemia rat cell line to different doses of a treatment. It is, therefore, of great interest to identify the threshold $d_0 \in (0, 1)$, the point from where the function starts deviating from its baseline value, and construct confidence intervals (CIs) for the same.

It is important to differentiate the above problem from the classical regression problem where the canonical model would be the same as above but one would only observe a single response for every value of the covariate. We refer to our problem as the 'm, n' setting where m replicates are observed for each of the n levels of the covariate, and m is often allowed to be of comparable order to n. An analytical treatment of regression models – the '$1, n$' setting – typically proceeds by allowing n to go to ∞ whereas in the 'm, n' setting, it is natural to explore the properties of estimators as both m and n grow and understand how the interplay between the two affects their long term behavior. The domains of application of the 'm, n' setting and the classical regression setting are also typically different with the regression setting being more germane to observational data and the former to controlled replicable experiments.

For the posited model (1), Mallik et al. (2011) recently proposed novel and computationally simple procedures for estimating d_0, based on the discrepancy of p-values, in both the classical regression and 'm, n' settings. They established consistency of their estimators under mild conditions, and also studied their finite sample properties. However, the problem of constructing CIs for d_0 was not addressed in that paper. In this work, we address this inference question in the 'm, n' setting by deriving the asymptotic distribution of the estimator, $\hat{d}_{m,n}$ in Mallik et al. (2011) (as well as those of related estimators) as m, n grow to infinity, and demonstrating how to use the quantiles of this distribution to set the limits of the CI. It turns out that the asymptotic behavior of the estimators in the m, n setting is *fundamentally different* from that in the classical regression setting, this latter case having been recently investigated in Mallik et al. (2013). The estimates in the regression setting converge to minimizers of processes with differentiable sample paths that can be written as transforms of Gaussian processes while, as we will see below, those in the m, n setting converge to the minimizers of piecewise-constant processes with jump discontinuities. Thus, many of the tools that play a crucial role in Mallik et al. (2013) are inapplicable in the m, n case (see Remark 6).

It should be noted that the problem of estimating d_0 in different models has received much attention in the statistics literature. If μ is assumed to have a jump discontinuity at d_0, then d_0 corresponds to a usual change-point for μ. Such change-point models are very well understood; see e.g., Hinkley (1970), Korostelëv (1987), Dümbgen (1991), Müller (1992), Korostelëv and Tsybakov (1993), Loader (1996), Müller and Song (1997) and the references therein. Our results, here, are developed for the harder problem that arises when μ is *continuous* at d_0. In particular, the smoother the regression function in a neighborhood of d_0, the greater the challenge in estimating d_0 precisely. We show that if d_0 is a cusp of μ of order k (i.e., the first $k-1$ right derivatives of μ at d_0 equal 0 but the k-th does not, so that d_0 is a change-point in the k-th derivative) and f is locally monotone in a neighborhood of d_0, then $\hat{d}_{m,n} - d_0$ is of order $N^{-1/(2k+1)}$, where $N = m \times n$ is the total budget and m is chosen in some optimal manner (to be specified later) in terms of n.

The limit distribution of $N^{1/(2k+1)}(\hat{d}_{m,n} - d_0)$ is seen to be that of an appropriate minimizer of a jump process drifting off to infinity, that can be viewed as a generalization of a compound Poisson process. The derivation of the asymptotic distribution is complicated owing to the fact that the sample paths of the limit process are piecewise constant, resulting in non-unique minimizers. Hence, the more common continuous mapping arguments that rely on the uniqueness of the extremum of limit processes (see e.g., Theorem 2.7 of Kim and Pollard (1990)) – a phenomenon that shows up often with Gaussian limits and monotone transforms thereof – do not apply, and careful modifications, which rely on the continuity of the argmin functional in spaces of discontinuous functions, are required. In particular, the least squares estimate of d_0 (which is not unique) needs to be carefully picked. Another important challenge lies in deriving the rate of convergence of the estimator, which requires a considerable generalization of the standard rate theorems (see Theorem 4) in the modern empirical processes literature (see e.g., Theorem 3.2.5 of van der Vaart and Wellner (1996)), and the choice of a cleverly constructed dichotomous metric on $\mathbb{R}$ (see Lemma 1) to invoke the generalization. The details are available in the proof of Theorem 1.

The knowledge of k is essential for constructing two-sided CIs based on these limiting results. Although resampling approaches such as subsampling are shown to work (in Section 4.2) for our problem, they do not present a solution for the situation when k is unknown. We do end up providing a partial answer and show that adaptive upper confidence bounds can be constructed in the k-unknown case (Section 4.1).

The remainder of the paper is organized as follows. Section 2 provides a brief discussion of the estimation procedure, its variants and extensions, and the core assumptions. The rates of convergence and the asymptotic distributions are deduced in Section 3, assuming a random design setting. Their implications to constructing CIs in practical applications, along with some auxiliary results on subsampling and adaptivity, are discussed in Section 4. In Section 5, we discuss the large sample behavior of the estimator of d_0 in a fixed design setting. We study the finite sample coverage performance of the CIs through simulations in Section 6 and discuss an application from a complex queuing system. The proofs of several technical results are provided in the Appendix.

2 Problem formulation

For convenience, we formulate the problem and deduce the estimation method in a random design setting. The extension to the fixed design setting is immediate. The expression for the estimator of d_0 is identical to that in the random design with the exception that the covariate X_is would then just be fixed design points. More details on the fixed design scheme are available in Section 5.

We assume the regression model (1) where the covariate X is sampled from a Lebesgue density f, X and ϵ are independent, $E(\epsilon) = 0$ and let $\sigma_0^2 := \mathrm{Var}(\epsilon) > 0$. Consider data $\{(X_i, Y_{ij}) : 1 \le j \le m, 1 \le i \le n\}$, where the X_is are i.i.d. random variables distributed like X, $\{\epsilon_{ij}\}$ are i.i.d. random variables distributed like ϵ, the vectors $\{X_i\}$ and $\{\epsilon_{ij}\}$ are independent, and

$$Y_{ij} = \mu(X_i) + \epsilon_{ij}, \quad 1 \le j \le m, \text{ and } 1 \le i \le n. \tag{3}$$

Here, $N = m \times n$ is the total budget and we assume $m = m_0 n^\beta$ for some $\beta > 0$, to incorporate the scenario that m can be 'large' relative to n, a feature of several dose-response studies.

Let $\bar{Y}_i = \sum_{i=1}^m Y_{ij}/m$ and $\hat{\sigma}^2 = \sum_{i,j}(Y_{ij} - \bar{Y}_i)^2/(nm - n)$. We estimate d_0 by constructing p-values for testing the null hypothesis $H_{0,x} : \mu(x) = \tau$ against the alternative $H_{1,x} : \mu(x) > \tau$ at each dose $X_i = x$. The approximate p-values are

$$p_{m,n}(X_i) = 1 - \Phi(\sqrt{m}(\bar{Y}_i - \tau)/\hat{\sigma}).$$

To the left of d_0, the null hypothesis holds and these approximate p-values converge weakly to a Uniform(0,1) distribution which has mean 1/2. However, 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 suggests proposing

$$\hat{d}_{m,n} = \operatorname*{sargmin}_{d \in [0,1]} \left[\sum_{i: X_i \le d} \left\{ p_{m,n}(X_i) - \frac{1}{2} \right\}^2 + \sum_{i: X_i > d} \{p_{m,n}(X_i)\}^2 \right] \tag{4}$$

as an estimate of d_0. Here, sargmin denotes the smallest argmin of the criterion function, which does not have a unique minimum. In fact, $\hat{d}_{m,n}$ corresponds to an order statistic of X_is and the above criterion is minimized at any point between $\hat{d}_{m,n}$ and the next order statistic. Setting $\gamma = 3/4$ and letting $\mathbb{P}_n$ denote the empirical measure of $(X_i, \bar{Y}_i), i = 1, \ldots, n$, the expression in (4) can be simplified as $\hat{d}_{m,n} = \operatorname{sargmin}_{d \in [0,1]} \mathbb{M}_{m,n}(d)$, where

$$\mathbb{M}_{m,n}(d) \equiv \mathbb{M}_{m,n}(d, \hat{\sigma}) = \mathbb{P}_n \left[\left\{ \Phi\left(\frac{\sqrt{m}(\bar{Y} - \tau)}{\hat{\sigma}} \right) - \gamma \right\} 1(X \le d) \right]. \tag{5}$$

Remark 1. *The above methods are based on a known τ. When τ is unknown, an estimate can be plugged in its place (more about this in Section 3.5). Also, for any choice of $\gamma \in (1/2, 1)$ in (5) 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

Our approach extends readily to the situation with heteroscedastic errors when X and ϵ are no longer independent. In this situation, $E(\epsilon \mid X) = 0$ and $\sigma_0^2(x) = \operatorname{Var}(\epsilon \mid X = x)$ is a non-constant function. The $\{X_i\}$'s are still i.i.d. and given X_i, $\{\epsilon_{ij}\}_{j=1}^m$ are conditionally i.i.d., each being distributed like ϵ conditional on X. If $\hat{\sigma}^2$ is a consistent estimate of $\sigma_0^2(x) = \operatorname{Var}(\epsilon \mid X = x)$, $x \in (0,1)$, an estimator of d_0 is given by

$$\operatorname*{sargmin}_{d \in (0,1)} \mathbb{P}_n \left[\left\{ \Phi\left(\frac{\sqrt{m}(\bar{Y} - \tau)}{\hat{\sigma}(X)} \right) - \gamma \right\} 1(X \le d) \right].$$

A natural candidate for $\hat{\sigma}$ is given by $\hat{\sigma}^2(X_i) := \sum_{j=1}^m (Y_{ij} - \overline{Y}_i)^2/(m-1)$. A variant of the above estimator that completely avoids estimating the variance function can also be constructed. Relying upon the simple fact that $E[\Phi(Z)] = 0.5$ for a normally distributed Z

with zero mean and arbitrary variance, it can be seen that the desired dichotomous behavior is preserved even when we do not normalize by the estimate of the variance and hence, an alternative estimator of d_0 can be given by

$$\tilde{d}_{m,n} = \underset{d\in(0,1)}{\text{sargmin}}\, \mathbb{P}_n \left[\left\{ \Phi\left(\sqrt{m}(\bar{Y} - \tau)\right) - \gamma \right\} 1(X \leq d) \right]. \tag{6}$$

2.2 Basic Assumptions

We study the limiting behavior of $\hat{d}_{m,n}$ assuming that the errors are independent and homoscedastic and consider a random design for the covariate distribution. The smoothness of the function in the vicinity of d_0 plays a crucial role in determining the rate of convergence. For the random design setting we make the following assumptions.

1. The regression function 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 lth 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 errors ϵ possess a continuous positive density on a (finite or infinite) interval.

3. The design density f for the dose-response setting is assumed to be continuous and positive on $[0, 1]$.

Remark 2. *Some words of explanation on why we address the asymptotics for a random design, as opposed to fixed design, are in order. It turns out that there is no limit distribution in this problem when the X_is are the grid-points of a non-random grid, say, the uniform grid of size n, on the domain of the covariate. See Remark 5 for a more technical explanation of this issue. Moreover, note that our data application for the (m, n) setup does come from a random design.*

3 Main Results

We state and prove results on the limiting behavior of the estimator $\hat{d}_{m,n}$ discussed in Section 2. Results on the variants of the procedure discussed in Section 2.1 follow similarly and are stated without proofs in Section 3.3. The results in this section are developed for $\gamma \in (1/2, 1)$ (cf. Remark 1) and a known τ. It will be seen in Section 4.3 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 $\tau \equiv 0$, as one can work with $(Y_{ij} - \tau)$s ($(Y_i - \tau)$s) in place of Y_{ij}s (Y_is).

3.1 Rate of convergence

As $m = m_0 n^\beta$, we consider the asymptotics in the dose-response model as $n \to \infty$. Let P_n denote the measure induced by $(\bar{Y}, X)$ and

$$M_{m,n}(d) = M_{m,n}(d, \sigma_0) = P_n \left[\left\{ \Phi\left(\frac{\sqrt{m}\bar{Y}}{\sigma_0} \right) - \gamma \right\} 1(X \leq d) \right].$$

The process $M_{m,n}$ is the population equivalent of $\mathbb{M}_{m,n}$ defined in (5) and can be simplified as follows. Let

$$Z_{1n} = \frac{1}{\sqrt{m}\sigma_0} \sum_{j=1}^{m} \epsilon_{1j} \tag{7}$$

and Z_0 be a standard normal random variable independent of Z_{1n}s. Then

$$\begin{aligned} E\left[\Phi\left(\frac{\sqrt{m}\bar{Y}_1}{\sigma_0}\right)\middle| X_1 = x\right] &= E\left[\Phi\left(\frac{\sqrt{m}\mu(x)}{\sigma_0} + Z_{1n}\right)\right] \\ &= E\left[E\left[1\left(Z_0 < \frac{\sqrt{m}\mu(x)}{\sigma_0} + Z_{1n}\right)\middle| Z_{1n}\right]\right] \\ &= P\left[\frac{Z_0 - Z_{1n}}{\sqrt{2}} < \frac{\sqrt{m}\mu(x)}{\sqrt{2}\sigma_0}\right] = \Phi_n\left(\frac{\sqrt{m}\mu(x)}{\sqrt{2}\sigma_0}\right), \end{aligned} \tag{8}$$

where Φ_n denotes the distribution function of $(Z_0 - Z_{1n})/\sqrt{2}$. Then, by integrating with respect to the density of X, it can be shown that

$$M_{m,n}(d) = \begin{cases} (\Phi_n(0) - \gamma)\, F(d), & d \le d_0, \\ (\Phi_n(0) - \gamma)\, F(d_0) + \displaystyle\int_{d_0}^{d} \left[\Phi_n\left(\frac{\sqrt{m}\mu(x)}{\sqrt{2}\sigma_0}\right) - \gamma\right] f(x)dx, & d > d_0. \end{cases}$$

Let $d_{m,n} = \text{sargmin}_{d\in(0,1)} M_{m,n}(d)$. We first study the behavior of $d_{m,n}$ which satisfies

$$\Phi_n\left(\tfrac{\sqrt{m}\mu(d_{m,n})}{\sqrt{2}\sigma_0}\right) = \gamma.$$

Let Φ_n^{-1} be the left continuous inverse of Φ_n. By Assumptions 1 and 2, we get

$$\frac{\mu^{(k)}(\zeta_n)}{k!}(d_{m,n} - d_0)^k = \frac{\sqrt{2}\sigma_0\Phi_n^{-1}(\gamma)}{\sqrt{m}}, \tag{9}$$

where ζ_n is some point between d_0 and $d_{m,n}$. As $n \to \infty$, the right-hand side (RHS) of the above display goes to zero. So, $d_{m,n} \to d_0$. Also, Φ_n converges point wise to Φ and the convergence holds for their inverse functions too. Hence,

$$d_{m,n} = d_0 + \left[\frac{k!\sqrt{2}\sigma_0\Phi^{-1}(\gamma)}{\mu^{(k)}(d_0+)}\right]^{1/k} m^{-1/(2k)} + o(m^{-1/(2k)}). \tag{10}$$

This shows that $d_{m,n} - d_0 = O(m^{-1/(2k)}) = O(n^{-\beta/(2k)})$. In a sense, $\hat{d}_{m,n}$, is estimating $d_{m,n}$ instead of d_0, and hence, its rate of convergence to d_0 can be expected to be at most of order $n^{-\beta/(2k)}$. Moreover, $\hat{d}_{m,n}$ is one of the order statistics of X_is and hence, can only be close to d_0 up to an order $1/n$. We next provide a formal statement of the rate of convergence of $\hat{d}_{m,n}$.

Theorem 1. *Let $\alpha = min\,(1, \beta/(2k))$. Then,*

$$n^{\alpha}(\hat{d}_{m,n} - d_0) = m_0^{-\frac{\alpha}{1+\beta}} N^{\frac{\alpha}{1+\beta}}(\hat{d}_{m,n} - d_0) = O_P(1).$$

Remark 3. *The function μ may not satisfy Assumption 1 for any $k \in \mathbb{Z}$ and can still take off at d_0, e.g., $\mu_{(1)}(x) = \exp(-1/(x-d_0))1(x > d_0)$ and $\mu_{(2)}(x) = \exp(-1/(x-d_0)^2)1(x > d_0)$ are two such infinitely differentiable functions with a singularity at d_0. By calculations almost identical to those for deriving (10), it can be shown that $d_{m,n} - d_0 = O\left((\log(n))^{-1/i}\right)$ when $\mu = \mu_{(i)}$, $i = 1, 2$. Hence, we do not expect a universal rate of convergence for $\hat{d}_{m,n}$ when μ is infinitely differentiable at d_0 and adhere to Assumption 1.*

The proof is given in Section A.1 of the Appendix. The optimal rate corresponds to $\alpha = 1$. In terms of the total budget, the best possible rate is achieved when $\beta = 2k$. In that case, $N^{1/(2k+1)}(\hat{d}_{m,n} - d_0) = O_P(1)$. For, $\beta < 2k$, the rate of convergence is $n^{\beta/(2k)}$ or $N^{\beta/\{2k(1+\beta)\}}$.

Remark 4. *The rate $N^{-/1(2k+1)}$ is not surprising as it appears in inverse function estimation: for example, if h is a smooth monotone function, the isotonic regression estimate of $x_0 := h^{-1}(\theta_0)$, where θ_0 is a fixed point in the range of h, converges at rate $S^{-1/(2k+1)}$ (S being sample size) under the assumption that f is (at least) k-times differentiable at x_0, $f^{(k)}(x_0) \neq 0$ and $f^{(l)}(x_0) = 0$ for $1 \leq l < k$, which is the exact analogue of the 'cusp assumption' on d_0 above. The same rate is obtained in the regression version of the problem (where m is identically 1 and the number of sampled covariates equals the budget N) studied in Mallik et al. (2013), and from the discussion on 'Minimaxity' in Section 8 of that paper, is expected to be minimax in the regression setting. As our assumptions on the model are the same as those in the regression setting, we expect this rate to be optimal in the m, n setting as well, even though a formal proof appears difficult and is outside the scope of this paper. For more details on the isotonic estimation of x_0 that appears at the beginning of this remark, see again Section 8 of Mallik et al. (2013).*

3.2 Asymptotic Distribution

We now deduce the asymptotic distribution of $\hat{d}_{m,n}$ for different choices of β, starting with $\beta = 2k$. Note that $n(\hat{d}_{m,n} - d_0) = \text{sargmin}_{t\in\mathbb{R}} \hat{V}_n(t)$ where

$$\hat{V}_n(t) = n\left\{\mathbb{M}_{m,n}(d_0 + t/n, \hat{\sigma}) - \mathbb{M}_{m,n}(d_0, \hat{\sigma})\right\}. \tag{11}$$

We deduce the limit of $\hat{V}_n$ and then apply a special continuous mapping theorem to obtain the asymptotic distribution of $\hat{d}_{m,n}$.

To state the limiting distribution, we introduce the following notation. Let $\{\nu^+(t) : t \geq 0\}$ and $\{\nu^-(t) : t \geq 0\}$ be two independent homogeneous Poisson processes with same intensity $f(d_0)$ but with RCLL (right continuous with left limits) and LCRL (left continuous with right limits) paths, respectively. Let $\{S_i\}_{i\geq 1}$ denote the arrival times for the process ν^+. Further, let $\{Z_i\}_{i\geq 1}$ and $\{U_i\}_{i\geq 1}$ be independent sequences of i.i.d. $N(0,1)$'s and i.i.d. $U(0,1)$'s respectively which are, moreover, independent of the processes ν^+ and ν^-. Define $V(t)$ as:

$$V(t) = \begin{cases} \displaystyle\sum_{j=1}^{\nu^+(t)} \left(\Phi\left(\frac{\sqrt{m_0}\ \mu^{(k)}(d_0+)}{k!\sigma_0} S_j^k + Z_j\right) - \gamma\right), & t \geq 0, \\ \displaystyle\sum_{j=1}^{\nu^-(-t)} (\gamma - U_j), & t < 0, \end{cases} \tag{12}$$

where sum over a null set is taken to be zero. We will show that $\hat{V}_n$ converges weakly to V as processes in $D(\mathbb{R})$, the space of càdlàg functions (right continuous having left limits) on $\mathbb{R}$ equipped with the Skorokhod topology; see Lindvall (1973) for more details on $D(\mathbb{R})$. Moreover, the asymptotic distribution of $\hat{d}_{m,n}$ will be characterized by a minimizer of the process V. The limiting process V does not possess a unique minimizer as it stays at any level it attains for an exponential amount of time. Hence, the usual argmin (argmax) continuous mapping theorem (see for example Theorem 3.2.2 of van der Vaart and Wellner (1996)) does not suffice for deducing the limiting distribution; we also need to show the convergence of the involved jump processes (Lan et al., 2009, pp. 1760–1762).

For convenience, we state a consequence of *Lemmas 3.1*, *3.2* and *3.3* from Lan et al. (2009) which provides a version of the argmin (argmax) continuous mapping theorem required in our setting. Let $\mathcal{S}$ denote the class of piecewise constant functions in $D(\mathbb{R})$ that are continuous at every integer point, assume the value 0 at 0, and possess finitely many jumps on every compact interval $[-C, C]$, where $C > 0$ is an integer. Note that $\mathcal{S}$ is a closed subset of $D(\mathbb{R})$. Also, define the pure jump process, $\tilde{g}$, (of jump size 1) corresponding to the function $g \in D(\mathbb{R})$, as the piecewise constant right continuous function with left limits, such that for any $s > 0$, $\tilde{g}(s)$ counts the number of jumps of the function g in the interval $[0, s]$, while for $s < 0$, $\tilde{g}(s)$ counts the number of jumps in the set $(s, 0)$. We have the following result.

Theorem 2. *Let $\mathbb{V}_n, n \geq 0$, be processes in $D(\mathbb{R})$ such that $\mathbb{V}_n \in \mathcal{S}$, with probability 1. Also, let $\mathbb{J}_n, n \geq 0$, denote the corresponding jump processes and $(\xi_n^s, \xi_n^l), n \geq 0$, be the smallest and largest minimizers for $\mathbb{V}_n$. Suppose that:*

- *(i)* *$(\mathbb{V}_n, \mathbb{J}_n)$ converges weakly to $(\mathbb{V}_0, \mathbb{J}_0)$ as processes in $D[-C, C] \times D[-C, C]$, for each positive integer C.*
- *(ii)* *No two flat stretches of $\mathbb{V}_0(t), t \in [-C, C]$, have the same height a.s., for each positive integer C.*
- *(iii)* *$\{(\xi_n^s, \xi_n^l), n \geq 0\}$ is $O_P(1)$.*

Then $(\xi_n^s, \xi_n^l) \xrightarrow{d} (\xi_0^s, \xi_0^l)$, where $\xrightarrow{d}$ denotes convergence in distribution.

Note that $\hat{V}_n \in \mathcal{S}$ with probability 1. For $t \in \mathbb{R}$, let the function $sgn(t)$ denote the sign of t. Also, let J_n denote the jump process corresponding to $\hat{V}_n(t)$. Then,

$$J_n(t) = sgn(t) \sum_{i=1}^{n} \left[1\left(X_i \leq d_0 + \frac{t}{n} \right) - 1\left(X_i \leq d_0 \right) \right].$$

Further, let J be the jump process associated with $V(t)$, i.e., $J(t) = \nu^+(t)1(t \geq 0) + \nu^-(-t)1(t < 0)$. We have the following result.

Theorem 3. *Let $\beta = 2k$ and $\hat{V}_n$ and V be as defined in (11) and (12) respectively. Then, the conditions (i), (ii) and (iii) of Theorem 2 are satisfied for $\mathbb{V}_n = \hat{V}_n$ and $\mathbb{V}_0 = V$ with J_n and J being the corresponding jump processes. As a consequence,*

$$n(\hat{d}_{m,n} - d_0) \xrightarrow{d} \underset{t \in \mathbb{R}}{sargmin}\, V(t).$$

The proof involves establishing finite dimensional convergence using characteristic functions and justifying a moment condition (see Billingsley (1968, pp. 128)) to prove asymptotic tightness. It is available in Section A.2.

Remark 5. *The counts $\sum_{i\leq n} 1(X_i \in (d_0, d_0 + t/n])$ account for the Poisson process that arises in the limit. If the X_is were drawn from a fixed uniform design, these counts would not converge. Hence, a fixed design setup does not yield a limiting distribution for the underlying processes, and consequently for $\hat{d}_{m,n}$, in the dose-response setting. This fact was also observed in the change point setting of Lan et al. (2009, pp. 1766).*

The limiting random variable $\text{sargmin}_{t\in\mathbb{R}} V(t)$ is continuous by virtue of the fact that the probability of a jump at a particular point for a Poisson process is zero. Its distribution depends upon the parameters m_0, $\mu^{(k)}(d_0+)$, σ_0, $f(d_0)$ and γ. It is clear from the expression for V (see (12)) that a larger m_0, a larger $\mu^{(k)}(d_0+)$ or a smaller σ_0 will skew the limiting distribution more to the left. For the sake of completeness, we state the asymptotics for other choices of β. When $\beta > 2k$, the derivation of the limiting distribution is similar to that of Theorem 3 and is outlined in Section A.3 of the Appendix.

Proposition 1. *Let $\beta > 2k$. Also, let $\left\{\nu_1^+(t) : t \geq 0\right\}$ and $\left\{\nu_1^-(t) : t \geq 0\right\}$ be two independent homogeneous Poisson processes with same intensity $f(d_0)$ but with RCLL and LCRL paths respectively. Let $\left\{\bar{U}_i\right\}_{i\geq 1}$ be a sequence of i.i.d. $U(0,1)$s which is independent of $\{\nu_1^+, \nu_1^-\}$. Define $\bar{V}(t)$ as:*

$$\bar{V}(t) = \begin{cases} (1-\gamma)\nu_1^+(t), & t \geq 0, \\ \displaystyle\sum_{j=1}^{\nu_1^-(-t)} \left(\gamma - \bar{U}_j\right), & t < 0, \end{cases}$$

where sum over a null set is taken to be zero. Then, $n(\hat{d}_{m,n} - d_0) \stackrel{d}{\rightarrow}$ sargmin$_{t\in\mathbb{R}} \bar{V}(t) =$ sargmin$_{t\leq 0} \bar{V}(t)$.

The case $\beta < 2k$ yields a markedly different result from the above two scenarios: we do not get a non-degenerate limiting distribution any longer as the normalized estimator converges to a constant. The proof is given in Section A.4 of the Appendix.

Proposition 2. *Choose $\beta < 2k$. Let*

$$\hat{H}_n(t) = n^{\beta/(2k)} \left\{\mathbb{M}_{m,n}\left(d_0 + \frac{t}{n^{\beta/(2k)}}, \hat{\sigma}\right) - \mathbb{M}_{m,n}(d_0, \hat{\sigma})\right\}$$

and

$$c(t) = \begin{cases} \left(\frac{1}{2} - \gamma\right) f(d_0)t, & t \leq 0 \\ f(d_0) \displaystyle\int_0^t \left\{\Phi\left(\frac{\sqrt{m_0}\ \mu^{(k)}(d_0+)}{\sqrt{2}k!\sigma_0} u^k\right) - \gamma\right\} du, & t > 0. \end{cases}$$

Then, for any $L > 0$,

$$\sup_{t\in[-L,L]} |\hat{H}_n(t) - c(t)| \stackrel{P}{\rightarrow} 0, \tag{13}$$

and

$$n^{\beta/(2k)}(\hat{d}_{m,n} - d_0) \stackrel{P}{\rightarrow} \underset{d\in\mathbb{R}}{\text{argmin}}\{c(t)\} = \left(\frac{\sqrt{2}\ k!\ \sigma_0\ \Phi^{-1}(\gamma)}{\sqrt{m_0}\ \mu^{(k)}(d_0+)}\right)^{1/k}.$$

3.3 Limit distributions for variants of the procedure

The rates of convergence and asymptotic distributions can be obtained similarly for the variants of the procedure that were discussed in Section 2.1. In what follows, we state the limiting distributions, without proofs, for one of the variants.

For heteroscedastic errors, the non-normalized version of the procedure (p-values are not normalized by the estimate of the variance), see (6), yields the following limiting distribution.

Proposition 3. *Consider the dose-response setting as stated in Section 2 but with heteroscedastic errors, i.e.,* $\sigma_0^2(x) = Var(\epsilon \mid X = x)$ *need not be identically* σ_0 *but is assumed to be continuous and positive. Let*

$$\tilde{d}_{m,n} = \underset{d \in (0,1)}{sargmin} \mathbb{P}_n \left[\left\{ \Phi\left(\sqrt{m}(\bar{Y} - \tau)\right) - \gamma \right\} 1(X \leq d) \right],$$

with $m = m_0 n^{2k}$. *Let* $\{\nu^+(t) : t \geq 0\}$ *and* $\{\nu^-(t) : t \geq 0\}$ *be two independent homogeneous Poisson processes with same intensity* $f(d_0)$ *but with RCLL and LCRL paths, respectively. Let* $\{S_i\}_{i \geq 1}$ *denote the arrival times for the process* ν^+. *Further, let* $\left\{Z_i^{(1)}\right\}_{i \geq 1}$ *and* $\left\{Z_i^{(2)}\right\}_{i \geq 1}$ *be independent sequences of i.i.d.* $N(0, \sigma_0^2(d_0))$*'s. Define* $\tilde{V}(t)$ *as:*

$$\tilde{V}(t) = \begin{cases} \displaystyle\sum_{j=1}^{\nu^+(t)} \left\{ \Phi\left(\frac{\sqrt{m_0}\ \mu^{(k)}(d_0+)}{k!} S_j^k + Z_j^{(1)} \right) - \gamma \right\}, & t \geq 0, \\ \displaystyle\sum_{j=1}^{\nu^-(-t)} \left\{ \gamma - \Phi(Z_j^{(2)}) \right\}, & t < 0, \end{cases}$$

Then, $n(\tilde{d}_{m,n} - d_0) \xrightarrow{d} sargmin_{t \in \mathbb{R}} \tilde{V}(t)$.

Remark 6. *As noted in the Introduction, the limit distribution in the '*m, n*' setting is governed by the minimizer of a generalized compound Poisson process, in contrast to the integral of a transformed Gaussian process that appears in the '*$1, n$*' setting considered in (Mallik et al., 2013, Theorem 2). The appearance of the transformed Gaussian process is an outcome of the local spatial averaging of responses needed to construct the p-values in the absence of multiple replications in the regression setting. In terms of total budget, both settings admit the same best-possible rate of* $N^{1/(2k+1)}$*, under an appropriate trade-off between* m *and* n *in the dose-response setting, and an appropriate choice of smoothing bandwidth in the regression setting.*

4 Construction of CIs

As the form of the limit distribution depends upon the allocation of the total budget N between m and n and may involve k, the construction of CIs requires some care. Consider, first, the case that k is assumed known. Writing $m = m_0\, n^{\beta}$, we can set $\beta = 2k$, the optimal choice in terms of the total budget, to solve for m_0, and then construct a CI for d_0 using the result in Theorem 3. This requires estimating nuisance parameters like $f(d_0)$, σ_0 and $\mu^{(k)}(d_0+)$, of which the last is the hardest to estimate. Note that we have already estimated σ_0 in order to construct $\hat{d}_{m,n}$, while the design density at d_0 can be estimated using $\hat{f}(\hat{d}_{m,n})$, where $\hat{f}$ is a standard kernel density estimate of f. As far as $\mu^{(k)}(d_0+)$ is concerned, observe that

$$\mu(x) = \mu^{(k)}(d_0+)(x - d_0)^k/k! + o((x - d_0)^k)$$

for $x > d_0$. An estimate of $\mu^{(k)}(d_0+)$ can, therefore, be obtained by fitting a local polynomial to the right of $\hat{d}_{m,n}$ that involves the k-th power of the covariate. Specifically, an estimate of $\xi_0 \equiv \mu^{(k)}(d_0+)/k!$ is:

$$\begin{aligned} \hat{\xi} &= \operatorname*{argmin}_{\xi} \sum_{i=1}^{n} \{\bar{Y}_i - \xi(X_i - \hat{d}_{m,n})^k\}^2 1(X_i \in (\hat{d}_{m,n}, \hat{d}_{m,n} + b_n]) \\ &= \frac{\sum \bar{Y}_i (X_i - \hat{d}_{m,n})^k 1(X_i \in (\hat{d}_{m,n}, \hat{d}_{m,n} + b_n])}{\sum (X_i - \hat{d}_{m,n})^{2k} 1(X_i \in (\hat{d}_{m,n}, \hat{d}_{m,n} + b_n])}, \end{aligned}$$

where $b_n \downarrow 0$ and $nb_n^{2k+1} \to \infty$. The condition $nb_n^{2k+1} \to \infty$ is typical for estimating the k-th derivative at a known fixed point; see e.g., Gasser and Müller (1984), Härdle and Gasser (1985). The following lemma, whose proof is given in Section A.5 of the Appendix, justifies the consistency of this estimate for the optimal choice of β, thereby providing a way to construct CIs by imputing this estimate in the limiting distribution.

Proposition 4. *Let $\beta = 2k$. Then $\hat{\xi} \xrightarrow{P} \xi_0$.*

Remark 7. *The estimate $\hat{\xi}$ is effectively a kernel estimate with the smoothing kernel being uniform on $(0, 1]$. Alternative consistent estimators of ξ can be obtained using other one sided kernels. To fix ideas, we only use the above mentioned estimate in the paper.*

4.1 Adaptive upper confidence bounds

Note that the above inference strategy is not adaptive to the order of smoothness, k, at d_0. While we have not been able to develop an adaptive method for two-sided CIs, we are able to propose a strategy for one-sided honest CIs for d_0 (which are also of consequence in applications) that avoids knowledge of k. For example, if d_0 represents the minimal effective dose in a pharmacological setting, practitioners would be naturally interested in finding an upper confidence bound for d_0. The following result, whose proof follows along the same lines as that of Proposition 1, is our starting point for building such CIs.

Proposition 5. *Consider the dose-response setting with homoscedastic errors and normalized p–values and define $\underline{d}_{m,n} = \text{sargmin}_{0 \le d \le d_0} \mathbb{M}_{m,n}(d)$, so that $\underline{d}_{m,n} \le \hat{d}_{m,n} = \text{sargmin}_{0 \le d \le 1} \mathbb{M}_{m,n}(d)$. Then, for any $\beta > 0$,*

$$n\left(\underline{d}_{m,n} - d_0\right) \Rightarrow_d \text{sargmin}_{t \le 0} \sum_{j=1}^{\nu^-(t)} (\gamma - U_j),$$

where U_j's and ν^- are as in Theorem 3.

In fact the above result does not require m to grow as a power of n. The condition $\min(m, n) \to \infty$ suffices. Note that the limit distribution above is concentrated on the negative axis (as it must, since $\underline{d}_{m,n} \leq d_0$) and does not depend upon k. Simulating its quantiles requires just an estimate of $f(d_0)$. Let K_α be its α'th quantile. Then,

$$\lim_{n\to\infty} P\left(d_0 \leq \underline{d}_{m,n} - K_\alpha/n\right) = 1-\alpha\,.$$

Now, $\underline{d}_{m,n}$ is obviously unknown, but $\underline{d}_{m,n} \leq \hat{d}_{m,n}$ which is known. It follows easily that:

$$\lim\inf_{n\to\infty} P\left(d_0 \leq \hat{d}_{m,n} - K_\alpha/n\right) \geq 1-\alpha\,.$$

An essentially honest level $1-\alpha$ upper confidence bound for d_0 is therefore given by $[0, \hat{d}_{m,n} - K_\alpha/n]$. For an asymptotic allocation where $\beta > 2k$, by Proposition 1, the limit distributions of $\underline{d}_{m,n}$ and $\hat{d}_{m,n}$ coincide. Hence, these conservative upper confidence bounds are in a sense, minimally conservative, as they are exact for the situation $\beta > 2k$.

4.2 Subsampling

As an alternative to using the limit distribution, subsampling can be used to construct CIs for the case $\beta \geq 2k$. Let q_n be a sequence of integers such that $q_n/n \to 0$ and $q_n \to \infty$. A subsample is constructed by selecting q_n many X_is and $l_n = \lfloor q_n m/n \rfloor$ response values at each selected X_i. The subsamples are denoted by $\mathcal{S}_1, \ldots \mathcal{S}_{N_n}$, where $N_n = \binom{n}{q_n}\left[\binom{m}{l_n}\right]^{q_n}$. Let $\hat{d}_{n,q_n,j}$ denote the estimate of d_0 based on $\mathcal{S}_j$, $j = 1, \ldots, N_n$. Let $G_{n,\beta}$ denote the distribution of $n(\hat{d}_{m,n} - d_0)$. For $\beta \geq 2k$, $G_{n,\beta}$ converges weakly to a continuous limiting distribution, say G_β. The approximation to $G_{n,\beta}$, based on subsampling, is given by

$$L_{n,q}(x) = L_{n,q}(x,\beta) = \frac{1}{N_n}\sum_{j=1}^{N_n} 1\left[q_n(\hat{d}_{n,q_n,j} - \hat{d}_{m,n}) \leq x\right].$$

The following result justifies the use of subsampling in constructing CIs for d_0.

Proposition 6. *Let $\beta \geq 2k$. If $q_n/n \to 0$ and $q_n \to \infty$ then:*

(i) $\sup_x |L_{n,q}(x,\beta) - G_\beta(x)| \xrightarrow{P} 0$.

(ii) $P[c_{n,q,\alpha/2} \leq n(\hat{d}_{m,n} - d_0) \leq c_{n,q,1-\alpha/2}] \to 1-\alpha$, *where* $c_{n,q,\xi} = \inf\{x : L_{n,q}(x) \geq \xi\}$.

The proof follows along the lines of that of Theorem 15.7.1 in Lehmann and Romano (2005). The details are provided in Section A.6 of the Appendix. The usual bootstrap methodology is not expected to be consistent.

4.3 The case of an unknown τ

While our results have been deduced under the assumption of a known τ, in real applications τ is generally not known. In this situation, quite a few extensions are possible. 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 $\sqrt{mn}$-consistent estimator of τ. If a proper choice of η is not available, one can obtain an initial estimate of τ using the method proposed in Section 2.4 of Mallik et al. (2011), compute $\hat{d}_{m,n}$ and then average the responses from, say, $[0, c\hat{d}_{m,n}]$, $c \in (0,1)$, to obtain an estimate of τ, which will also be $\sqrt{mn}$-consistent. Note that this leads to an iterative procedure which we discuss in more detail in Section 6.1. Using a $\sqrt{mn}$-consistent estimate of τ, say $\hat{\tau}$, so that the $\overline{Y}_i$s are centered around $\hat{\tau}$ in the p-values, it can be shown that all the asymptotic results encountered earlier stay unchanged. A brief sketch of the following result is given in Section A.7.

Proposition 7. *Let $\hat{d}_{m,n}$ now denote the smallest minimizer of*

$$\mathbb{M}_{m,n}(d, \hat{\sigma}, \hat{\tau}) = \mathbb{P}_n \left[\left\{ \Phi\left(\frac{\sqrt{m}(\bar{Y} - \hat{\tau})}{\hat{\sigma}} \right) - \gamma \right\} 1(X \leq d) \right],$$

where $\sqrt{mn}\,(\hat{\tau} - \tau) = O_p(1)$. For $m = m_0 n^{\beta}$ and α as defined in Theorem 1, we have $n^{\alpha}(\hat{d}_{m,n} - d_0) = O_P(1)$. Also, when $\beta = 2k$,

$$n(\hat{d}_{m,n} - d_0) \stackrel{d}{\rightarrow} \underset{t \in \mathbb{R}}{\mathit{sargmin}}\, V(t),$$

where the process V is as defined in (12).

A similar extension of Proposition 5 is valid as well.

5 Fixed design setting

As mentioned in Section 2, the estimation procedure does not change when we move over from the random design setting to the fixed design setting. For example, with a data generating model of the form

$$Y_{ij} = \mu\left(\frac{i}{n}\right) + \epsilon_{ij}, \quad 1 \leq i \leq n,\ 1 \leq j \leq m,$$

where ϵ_{ij}s are independent and identically distributed with mean 0 and variance σ_0^2, an estimate for d_0, based on non-normalized p-values, is given by

$$\hat{d}^{FD}_{m,n} = \underset{d \in (0,1)}{\text{sargmin}}\, \mathbb{M}_{m,n}(d),$$

where

$$\mathbb{M}^{FD}_{m,n}(d) = \frac{1}{n} \sum_{i=1}^{n} \left\{ \Phi\left(\sqrt{m}\bar{Y}_i\right) - \gamma \right\} 1\left(\frac{i}{n} \leq d \right).$$

Here, τ is assumed known and taken to be zero without any loss of generality. The following result, whose proof is outlined in Section A.8 of the Appendix, shows that $\hat{d}^{FD}_{m,n}$ attains the same rate of convergence as its counterpart in the random design setting.

Proposition 8. *For $m = m_0 n^{\beta}$ and α as defined in Theorem 1, we have*

$$n^{\alpha}(\hat{d}^{FD}_{m,n} - d_0) = O_P(1).$$

As mentioned in Remark 5, there is no limit distribution available in this setting as the sums of the form $\sum_i [1\,(i/n \leq d_0 + t/n) - 1\,(i/n \leq d_0)]$, $t \in \mathbb{R}$, do not converge. However, the asymptotic distributions obtained in the random design setup can be used for setting approximate CIs for d_0 in such cases. Section 5.1.2 of Lan et al. (2007) investigated this issue through simulations in the related setting of a change-point regression model where the quantiles of the limit distribution (of the least squares estimate of the change-point) in a uniform *random design* setting were used for constructing CIs for the change-point when the data were generated from a uniform *fixed design* setting. The CIs obtained were seen to have comparable lengths to those for data generated from the random setting but were prone to over-coverage, and were therefore honest in the fixed design setting. A similar phenomenon was observed in our problem.

One possible way to obtain limiting distributions in this fixed design setting is to consider a slight variant of the model investigated in this paper, where at stage n, the true population threshold corresponds to the design point just preceding d_0. This gives us a sequence of models changing with n where the population threshold at stage n converges to d_0 at a fast $(1/n)$ rate. For such models, it can be shown that the threshold estimator proposed in this paper attains the same rate and possesses a limit distribution that corresponds to the minimizer of a two-sided random walk with drift. However, due to space constraints, we do not go into the details.

6 Data analysis

6.1 Simulations

We consider the underlying regression function as $\mu(x) = [2(x-0.5)]1(x > 0.5), x \in [0,1]$. This function is at its baseline value 0 up to $d_0 = 0.5$ and then rises to 1. The errors are assumed to be normally distributed with mean 0 and standard deviation $\sigma_0 = 0.1$. We work with $\gamma = 3/4$ as extreme values of γ (close to 0.5 or 1) tend to cause instabilities. We study the coverage performance of the approximate CIs obtained from the limiting distributions with the nuisance parameters estimated.

We generate samples for different choices of m and n, under μ. The covariate X is sampled from $U(0,1)$. For estimation, the factor m_0 is chosen so that the allocation between m and n is optimum. We assume τ to be unknown and get its initial estimate through the p-value based approach proposed in Mallik et al. (2011, equation (5)). An iterative scheme is then implemented where we use this initial estimator of τ to compute $\hat{d}_{m,n}$, re-estimate τ by averaging the responses for which X lies in $[0, 0.9\hat{d}_{m,n}]$ and proceed thus. On average, the estimates stabilize within 5 iterations. Firstly, we compare the distribution of $n(\hat{d}_{m,n} - d_0)$ for $m = n = 500$ data points over 5000 replications with the deduced asymptotic distribution. The Q-Q plot, shown in the left panel of Figure 6.1, reveals considerable agreement between the two distributions. In Table 1 we provide the estimated coverage probabilities of the CIs over 5000 replications for the model μ constructed by imputing estimates of the nuisance parameters (as discussed in Section 4) in the limiting distribution. The limiting process V was generated over a compact set incorporating the fact that $d_0 \in (0,1)$ and consequently $n(\hat{d}_{m,n} - d_0) \in [n(\hat{d}_{m,n} - 1), n\hat{d}_{m,n}]$. The smoothing bandwidth for estimating $\mu^{(k)}(d_0+)$ was chosen to be $5(n/\log n)^{-1/(2k+1)}$. The coverage performance is not very sensitive to the choice of this bandwidth as long as it is reasonably wide. The approximate CIs exhibit over-coverage for small samples but have close to the desired nominal coverage level as the sample size increases. As discussed in Section 4.1, upper confidence bounds can be constructed without the knowledge of k. We provide coverage probabilities and average lengths of the CIs $[0, \hat{d}_{m,n} - K_{\alpha/n}]$, for $\alpha = 0.05$ and 0.10

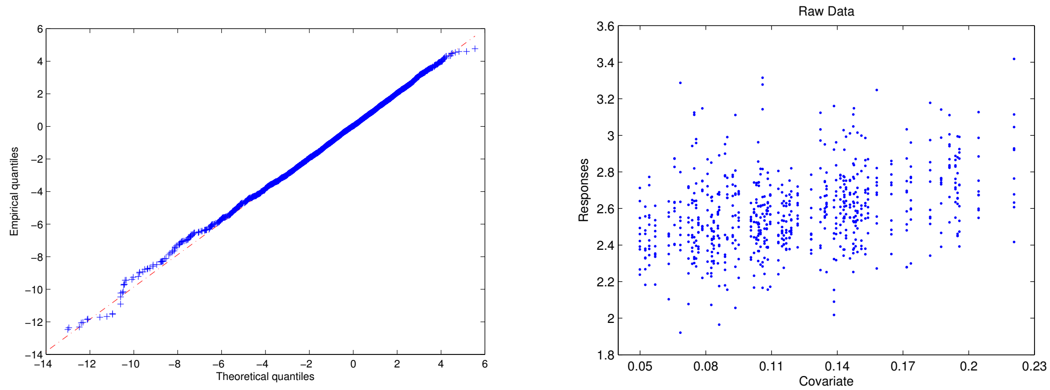

Figure 1: Q-Q plot under μ when $m = n = 500$ over 5000 replications (left plot), and the plot of all the response for the data from the queuing system (right plot).

m	n	90% CI		95% CI	
		T	E	T	E
5	5	0.966 (0.704)	0.860 (0.637)	0.973 (0.764)	0.940 (0.696)
10	10	0.941 (0.454)	0.944 (0.473)	0.970 (0.553)	0.970 (0.568)
15	10	0.924 (0.451)	0.939 (0.472)	0.966 (0.552)	0.966 (0.564)
10	15	0.914 (0.322)	0.935 (0.338)	0.961 (0.408)	0.961 (0.428)
15	15	0.913 (0.320)	0.931 (0.345)	0.959 (0.406)	0.961 (0.435)
20	20	0.910 (0.243)	0.913 (0.254)	0.955 (0.312)	0.960 (0.326)
25	25	0.908 (0.195)	0.910 (0.202)	0.951 (0.252)	0.959 (0.259)
30	30	0.903 (0.163)	0.893 (0.167)	0.951 (0.211)	0.953 (0.215)
50	50	0.901 (0.100)	0.900 (0.100)	0.950 (0.128)	0.951 (0.130)

Table 1: Coverage probabilities and lengths of two-sided CIs (in parentheses) using the true parameters (T) and the estimated parameters (E) for different sample sizes.

in Table 2. The only parameter to estimate for computing the quantile K_α is $f(d_0)$ which, as mentioned earlier, is computed by evaluating a kernel estimate of f at the point $\hat{d}_{m,n}$. As expected, the CIs are conservative but are close to the desired confidence level for large m and n, with their average length converging towards 0.5 (length of the interval $[0, d_0]$).

6.2 Complex queuing system

We consider a complex queuing system comprising multiple classes of customers waiting at infinite capacity queues and a set of processing resources modulated by an external stochastic process. The system employs a resource allocation (scheduling) policy that decides at every time slot which customer class to serve, given the state of the modulating rate process and the backlog of the various queues. In Bambos and Michailidis (2004), a low complexity policy was introduced and its maximum throughput properties established. This canonical system captures the essential features of data/voice transmissions in a wireless network, in multi-product manufacturing systems, and in call centers (for more details see Bambos and Michailidis (2004)). An important quantity of interest to the system's operator is the

m	n	90% CI		95% CI	
		T	E	T	E
5	5	0.951 (0.834)	0.956 (0.865)	0.970 (0.927)	0.971 (0.930)
10	10	0.955 (0.747)	0.978 (0.753)	0.990 (0.851)	0.993 (0.857)
15	10	0.962 (0.747)	0.978 (0.750)	0.990 (0.849)	0.992 (0.855)
10	15	0.933 (0.665)	0.955 (0.672)	0.972 (0.748)	0.991 (0.754)
15	15	0.921 (0.657)	0.959 (0.669)	0.966 (0.741)	0.990 (0.751)
20	20	0.920 (0.618)	0.943 (0.627)	0.962 (0.680)	0.986 (0.690)
25	25	0.921 (0.594)	0.934 (0.598)	0.960 (0.644)	0.972 (0.649)
30	30	0.915 (0.579)	0.935 (0.584)	0.960 (0.620)	0.971 (0.626)
50	50	0.913 (0.548)	0.933 (0.551)	0.958 (0.573)	0.970 (0.576)

Table 2: Coverage probabilities and lengths of one-sided adaptive CIs (in parentheses) using the true parameters (T) and the estimated parameters (E) for different sample sizes.

average delay of jobs (over all classes), which constitutes a key performance metric of the quality of service offered by the system. The average delay of the jobs in a two-class system as a function of its loading under the optimal policy, for a small set of loadings is shown in the right panel of Figure 6.1. These responses were obtained through simulation, since for such complex systems analytic calculations of delays are intractable. More specifically, ten replicates of the response (average delay) were obtained based on 5,000 events per class by simulating the system under consideration and after accounting for a burn-in period of 2,000 per class in order to ensure that it reached its stationary regime. The means per loading, $\bar{Y}_i$s, are shown in the left panel of Figure 6.2. The system operator is interested in identifying the loading beyond which the average delay starts increasing from its initial baseline value. Starting with an initial estimate of τ, using the approach of Mallik et al. (2011),

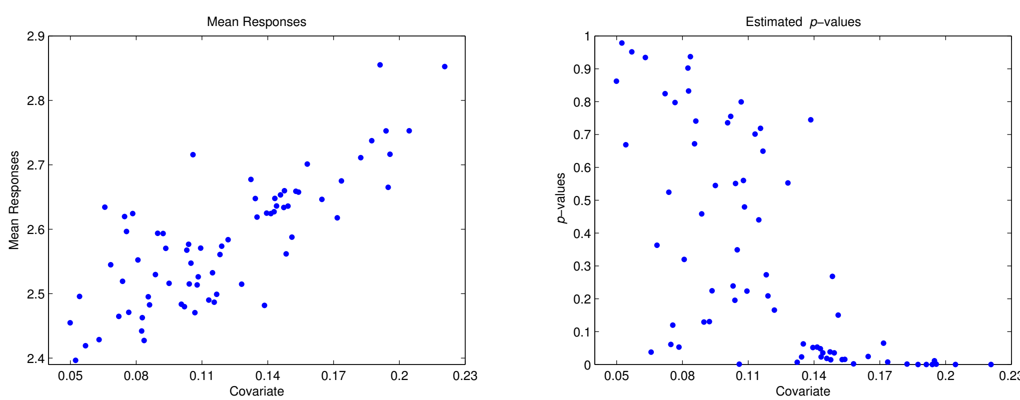

Figure 2: Plot of the average responses $\bar{Y}_i$ (left panel) and the estimated p-values (right panel) for the data from the queuing system.

the iterative approach discussed in the previous sub-section yields the final estimates to be $\hat{d}_{m,n} = 0.1165$ and $\hat{\tau} = 2.5230$, assuming homoscedastic errors. The estimated p-values are plotted in the right panel of Figure 6.2 which illustrates the dichotomy in the behavior of the p-values – they are uniformly distributed to the left of $\hat{d}_{m,n}$, and close to zero beyond

$\hat{d}_{m,n}$. Taking k to be 1, and using the methodology described in the previous sub-section, the 90% and 95% CIs for the threshold turn out to be $[0.1051, 0.1276]$ and $[0.1031, 0.1301]$, respectively. Also, the adaptive upper 90% and 95% confidence bound for the threshold turn out to be 0.1348 and 0.1371, respectively. From the system's operator point of view the average delay of jobs exhibits a markedly increasing trend beyond a loading of 13%.

Acknowledgments

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

A Appendix

We first state a result which is useful in deriving the rate of convergence of our estimators. In the sequel, 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.

Theorem 4. *Let $\{\mathbb{M}_n(d,\sigma),\ n \geq 1\}$ be stochastic processes and $\{M_n(d,\sigma),\ n \geq 1\}$ be deterministic functions, indexed by $d \in \Theta$ and $\sigma \in \Sigma$. Let $d_n \in \Theta$, $\sigma_0 \in \Sigma$ and $\kappa > 0$ be arbitrary, and $d \mapsto \rho_n(d, d_n)$ be an arbitrary map from Θ to $[0,\infty)$. Let $\hat{d}_n$ be a point of minimum of $\mathbb{M}_n(d, \hat{\sigma}_n)$, where $\hat{\sigma}_n$ is random. For each $\epsilon > 0$, suppose that the following hold:*

(a) There exists a sequence of sets $U_{n,\epsilon}$ in Σ which contain σ_0 and $P[\hat{\sigma}_n \notin U_{n,\epsilon}] < \epsilon$.

(b) For all sufficiently large n, $0 < \delta < \kappa$, and d such that $\rho_n(d, d_n) < \kappa$,

$$M_n(d,\sigma_0) - M_n(d_n,\sigma_0) \gtrsim \rho_n^2(d, d_n),$$

$$E^* \sup_{\substack{\rho_n(d,d_n)<\delta \\ \sigma \in U_{n,\epsilon}}} |(\mathbb{M}_n(d,\sigma) - M_n(d,\sigma_0)) - (\mathbb{M}_n(d_n,\sigma) - M_n(d_n,\sigma_0))| \quad \leq \quad C_\epsilon \frac{\phi_n(\delta)}{\sqrt{n}},$$

for a constant $C_\epsilon > 0$ and functions ϕ_n (not depending on ϵ) such that $\delta \mapsto \phi_n(\delta)/\delta^\alpha$ is decreasing for some $\alpha < 2$.

Suppose that r_n satisfies

$$r_n^2\ \phi_n\left(\frac{1}{r_n}\right) \lesssim \sqrt{n},$$

and $\rho_n(\hat{d}_n, d_n)$ converges to zero in probability; then $r_n\ \rho_n(\hat{d}_n, d_n) = O_P(1)$.

This theorem puts together the results in Theorem 3.2.5 in van der Vaart and Wellner (1996) and Theorem 5.2 in Banerjee and McKeague (2007).

A.1 Proof of Theorem 1

The following lemma gives the explicit distance function ρ_n that is used in proving Theorem 1.

Lemma 1. *Fix $\eta > 0$. Let the map $d \mapsto \rho_n^2(d, d_{m,n})$ from $(0,1)$ to $[0,\infty)$ be*

$$K_1 \left[|d - d_0| \, 1\,(d < d_0) + \left| d - d_{m,n} - \frac{\eta}{m^{1/(2k)}} \right| 1 \left(d > d_{m,n} + \frac{\eta}{m^{1/(2k)}} \right) \right], \tag{14}$$

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

$$M_{m,n}(d) - M_{m,n}(d_{m,n}) \geq \rho_n^2(d, d_{m,n}).$$

Using this lemma, we first give a proof of Theorem 1. Note that $\sqrt{mn}(\hat{\sigma} - \sigma_0) = O_P(1)$. So, given $\epsilon > 0$, there exists $L_\epsilon > 0$ such that $P[\sqrt{mn}|\hat{\sigma} - \sigma_0| \leq L_\epsilon] > 1 - \epsilon$. Let $U_{n,\epsilon} = [\sigma_1, \sigma_2] = [\sigma_0 - L_\epsilon/\sqrt{mn}, \sigma_0 + L_\epsilon/\sqrt{mn}]$ and let $\mathbb{G}_n$ denote the empirical process, i.e., $\mathbb{G}_n = \sqrt{n}(\mathbb{P}_n - P_n)$. For κ as in *Lemma 1*, $0 \leq \delta < \kappa$, and ρ_n as defined in (14), consider the expression

$$\begin{aligned}
& E^* \sup_{\substack{\rho_n(d,d_{m,n})<\delta \\ \sigma \in U_{n,\epsilon}}} \sqrt{n} |(\mathbb{M}_{m,n}(d,\sigma) - M_{m,n}(d,\sigma_0)) - (\mathbb{M}_{m,n}(d_{m,n},\sigma) - M_{m,n}(d_{m,n},\sigma_0))| \\
& \leq E^* \sup_{\substack{\rho_n(d,d_{m,n})<\delta \\ \sigma \in U_{n,\epsilon}}} \sqrt{n} |(\mathbb{M}_{m,n}(d,\sigma) - \mathbb{M}_{m,n}(d_{m,n},\sigma)) - (M_{m,n}(d,\sigma) - M_{m,n}(d_{m,n},\sigma))| \\
& + \sup_{\substack{\rho_n(d,d_{m,n})<\delta \\ \sigma \in U_{n,\epsilon}}} \sqrt{n} |(M_{m,n}(d,\sigma) - M_{m,n}(d_{m,n},\sigma)) - (M_{m,n}(d,\sigma_0) - M_{m,n}(d_{m,n},\sigma_0))| \\
& \leq E^* \sup_{\substack{|d - d_{m,n}| < \delta^2/K_1 + Am^{-1/2k} \\ \sigma \in U_{n,\epsilon}}} \left| \mathbb{G}_n \left[\left(\Phi\left(\frac{\sqrt{m}\bar{Y}}{\sigma} \right) - \gamma \right) (1(X \leq d) - 1(X \leq d_{m,n})) \right] \right| \\
& + \sqrt{n} \sup_{\substack{|d - d_{m,n}| < \delta^2/K_1 + Am^{-1/2k} \\ \sigma \in U_{n,\epsilon}}} \left| P_n \left[\left\{ \Phi\left(\frac{\sqrt{m}\bar{Y}}{\sigma} \right) - \Phi\left(\frac{\sqrt{m}\bar{Y}}{\sigma_0} \right) \right\} (1(X \leq d) - 1(X \leq d_{m,n})) \right] \right|.
\end{aligned}$$

The first term in the above display involves an empirical process acting on a class of functions, say $\mathcal{F}$. This class $\mathcal{F}$ is a product of two VC classes, $\{(\Phi(\sqrt{m} \cdot/\sigma) - \gamma) : \sigma \in U_{n,\epsilon}\}$ and $\{1(\cdot \leq d) - 1(\cdot \leq d_{m,n}) : |d - d_{m,n}| < \delta^2/K_1 + Am^{-1/2k}\}$, each with VC-index at most 3. Also, an envelope for this class is given by $G(x) = 1\left[x \in (d_{m,n} - \delta^2/K_1 - Am^{-1/(2k)}, d_{m,n} + \delta^2/K_1 + Am^{-1/(2k)})\right]$ with $(P_n G^2)^{1/2} \lesssim \sqrt{2(\delta^2/K_1 + Am^{-1/(2k)})}$. Hence, the uniform entropy integral for $\mathcal{F}$ is bounded by a constant which only depends upon the VC-indices, i.e., the quantity

$$J(1, \mathcal{F}) = \sup_Q \int_0^1 \sqrt{1 + \log N_C(\epsilon \|G\|_{Q,2}, \mathcal{F}, L_2(Q))} d\epsilon$$

is bounded, where $N_C(\cdot)$ denotes the covering number; see Theorems 9.3 and 9.15 of Kosorok (2008) for more details. Using Theorem 2.14.1 of van der Vaart and Wellner (1996), we have

$$E^* \sup_{\substack{|d - d_{m,n}| < \delta^2/K_1 + Am^{-1/2k} \\ \sigma \in U_{n,\epsilon}}} \left| \mathbb{G}_n \left[\left(\Phi\left(\frac{\sqrt{m}\bar{Y}}{\sigma} \right) - \gamma \right) (1(X \leq d) - 1(X \leq d_{m,n})) \right] \right|$$

$$\leq J(1,\mathcal{F})(P_nG^2)^{1/2} \lesssim \sqrt{2(\delta^2/K_1 + Am^{-1/(2k)})}.$$

Note that for $\sigma \in U_{n,\epsilon} = [\sigma_1, \sigma_2]$, we have

$$\left|\Phi\left(\frac{\sqrt{m}\bar{Y}}{\sigma}\right) - \Phi\left(\frac{\sqrt{m}\bar{Y}}{\sigma_0}\right)\right| \leq \left|\Phi\left(\frac{\sqrt{m}\bar{Y}}{\sigma_1}\right) - \Phi\left(\frac{\sqrt{m}\bar{Y}}{\sigma_2}\right)\right|.$$

Hence, by using the fact that Φ is Lipschitz of order 1, for sufficiently large n, we get

$$\begin{aligned}
\sqrt{n} \sup_{\substack{|d-d_{m,n}|<\delta^2/K_1+Am^{-1/2k} \\ \sigma\in U_{n,\epsilon}}} & \left|P_n\left[\left\{\Phi\left(\frac{\sqrt{m}\bar{Y}}{\sigma}\right) - \Phi\left(\frac{\sqrt{m}\bar{Y}}{\sigma_0}\right)\right\}(1(X\leq d) - 1(X\leq d_{m,n}))\right]\right| \\
&\leq \sqrt{n}P_n\left[\left|\Phi\left(\frac{\sqrt{m}\bar{Y}}{\sigma_1}\right) - \Phi\left(\frac{\sqrt{m}\bar{Y}}{\sigma_2}\right)\right||G(X)|\right] \\
&\leq \sqrt{n}\left[P_n\left|\Phi\left(\frac{\sqrt{m}\bar{Y}}{\sigma_1}\right) - \Phi\left(\frac{\sqrt{m}\bar{Y}}{\sigma_2}\right)\right|^2\right]^{\frac{1}{2}}(P_nG^2)^{\frac{1}{2}} \\
&\lesssim \sqrt{nm}\frac{\sigma_2-\sigma_1}{\sigma_2\sigma_1}(E\bar{Y}^2)^{1/2}\sqrt{2(\delta^2/K_1 + Am^{-1/(2k)})} \\
&\lesssim \frac{4L_\epsilon}{\sigma_0^2}(E\bar{Y}^2)^{1/2}\sqrt{2(\delta^2/K_1 + Am^{-1/(2k)})}.
\end{aligned}$$

As $E(\bar{Y}^2) = (1/m)E\{\mu(X)\}^2 + \sigma_0^2$ is bounded, we have

$$E^* \sup_{\substack{\rho_n(d,d_{m,n})<\delta \\ \sigma\in U_{n,\epsilon}}} \sqrt{n}|(\mathbb{M}_{m,n}(d,\sigma) - M_{m,n}(d,\sigma_0)) - (\mathbb{M}_{m,n}(d_{m,n},\sigma) - M_{m,n}(d_{m,n},\sigma_0))| \leq C_\epsilon\phi_n(\delta), \qquad (15)$$

for some $C_\epsilon > 0$ and $\phi_n(\delta) = \sqrt{\delta^2 + m^{-1/(2k)}}$. Also, $\rho_n^2(d, d_{m,n}) \leq K_1(|d-d_0| + |d_0 - d_{m,n} - \eta m^{-1/(2k)}|) \to 0$, if $|d-d_0| \to 0$. So, $\rho_n(\hat{d}_{m,n}, d_{m,n})$ converges in probability to zero by consistency of $\hat{d}_{m,n}$. Then by *Theorem 4*, the rate of convergence, say r_n, satisfies

$$r_n^2\phi\left(\frac{1}{r_n}\right) \lesssim \sqrt{n} \;\Rightarrow r_n^2 + r_n^4 m^{-1/(2k)} \leq n$$

$$\Rightarrow r_n^2 \lesssim n \wedge \sqrt{n^{1+\beta/(2k)}}. \qquad (16)$$

With $\alpha = \min(1, \beta/(2k)) = \min(1, 1/2 + \beta/(4k), \beta/(2k))$, $r_n^2 = n^\alpha$ satisfies (16). As $m^{-1/(2k)} \lesssim n^{-\alpha}$, we also have $n^\alpha(d_{m,n} + \eta m^{-1/(2k)} - d_0) = O(1)$. So, $n^\alpha\rho_n^2(\hat{d}_{m,n}, d_{m,n}) = O_P(1) \Rightarrow n^\alpha(\hat{d}_{m,n} - d_0) = O_P(1)$. As $m_0 n^{1+\beta} = N$, we get the result. □

Proof of Lemma 1. Let $\epsilon > 0$ be chosen such that μ is increasing on $(d_0, d_0+\epsilon)$. Let $f_0 = \inf_{d:|d-d_0|<\epsilon} f(d) > 0$. For $d \in (d_0 - \epsilon, d_0 + \epsilon)$,

$$\begin{aligned}
M_{m,n}(d) - M_{m,n}(d_{m,n}) \geq\; & 1(d<d_0)f_0\left[(\Phi_n(0)-\gamma)(d-d_0) + M_{m,n}(d_0) - M_{m,n}(d_{m,n})\right] \\
& + 1(d\geq d_0)f_0\int_{d_{m,n}}^{d}\left[\Phi_n\left(\frac{\sqrt{m}\mu(x)}{\sqrt{2}\sigma_0}\right) - \gamma\right]dx \\
\geq\; & 1(d<d_0)f_0\left[|\Phi_n(0)-\gamma|\,|d-d_0|\right] \\
& + 1(d\geq d_0)f_0\int_{d_{m,n}}^{d}\left[\Phi_n\left(\frac{\sqrt{m}\mu(x)}{\sqrt{2}\sigma_0}\right) - \gamma\right]dx.
\end{aligned}$$

Recall that from (10),

$$d_{m,n} = d_0 + \left[\frac{k!\sqrt{2}\sigma_0\Phi^{-1}(\gamma)}{\mu^{(k)}(d_0+)}\right]^{1/k} m^{-1/(2k)} + o(m^{-1/(2k)}). \tag{17}$$

Hence, for sufficiently large n, $d_{m,n} + \eta m^{-1/(2k)} < d_0 + \epsilon$. For such large n's and $d \in (d_{m,n} + \eta m^{-1/(2k)}, d_0 + \epsilon)$, we have:

$$\begin{aligned} M_{m,n}(d) - M_{m,n}(d_{m,n}) &\geq f_0 \int_{d_{m,n}}^{d} \left[\Phi_n\left(\frac{\sqrt{m}\mu(x)}{\sqrt{2}\sigma_0}\right) - \gamma\right] dx \\ &\geq f_0 \int_{d_{m,n}+\eta m^{-1/(2k)}}^{d} \left[\Phi_n\left(\frac{\sqrt{m}\mu(x)}{\sqrt{2}\sigma_0}\right) - \gamma\right] dx \\ &\geq f_0 \left(d - (d_{m,n} + \eta m^{-1/(2k)})\right) \left[\Phi_n\left(\frac{\sqrt{m}\mu(d_{m,n} + \eta m^{-1/(2k)})}{\sqrt{2}\sigma_0}\right) - \gamma\right]. \end{aligned} \tag{18}$$

Next, we show that $[\Phi_n\left(\sqrt{m}\mu(d_{m,n} + \eta m^{-1/(2k)})/(\sqrt{2}\sigma_0)\right) - \gamma]$ is bounded away from zero. By Pólya's theorem, Φ_n converge uniformly to Φ. So, for sufficiently large n,

$$\begin{aligned} &\Phi_n\left(\frac{\sqrt{m}\mu(d_{m,n} + \eta m^{-1/(2k)})}{\sqrt{2}\sigma_0}\right) - \gamma \\ &= \Phi_n\left(\frac{\sqrt{m}\mu(d_{m,n} + \eta m^{-1/(2k)})}{\sqrt{2}\sigma_0}\right) - \Phi_n\left(\frac{\sqrt{m}\mu(d_{m,n})}{\sqrt{2}\sigma_0}\right) \\ &> \frac{1}{2}\left[\Phi\left(\frac{\sqrt{m}\mu(d_{m,n} + \eta m^{-1/(2k)})}{\sqrt{2}\sigma_0}\right) - \Phi\left(\frac{\sqrt{m}\mu(d_{m,n})}{\sqrt{2}\sigma_0}\right)\right]. \end{aligned}$$

As $\Phi\left(\sqrt{m}\mu(d_{m,n})/(\sqrt{2}\sigma_0)\right)$ converges to $\gamma \in (0,1)$, $\sqrt{m}\mu(d_{m,n})$ is $O(1)$. Hence, it suffices to show that the difference $\sqrt{m}\{\mu(d_{m,n}+\eta m^{-1/(2k)}) - \mu(d_{m,n})\}$ is bounded away from zero. With $\tilde{\zeta}_n$ being some point between d_0 and $d_{m,n} + \eta m^{-1/(2k)}$ and ζ_n as defined in (9), we have

$$\begin{aligned} &\sqrt{m}\{\mu(d_{m,n} + \eta m^{-1/(2k)}) - \mu(d_{m,n})\} \\ &= \frac{\sqrt{m}}{k!}\{\mu^{(k)}(\tilde{\zeta}_n)(d_{m,n} + \eta m^{-1/(2k)} - d_0)^k - \mu^{(k)}(\zeta_n)(d_{m,n} - d_0)^k\} \\ &> \frac{\sqrt{m}\mu^{(k)}(\tilde{\zeta}_n)}{k!}[(d_{m,n} + \eta m^{-1/(2k)} - d_0)^k - (d_{m,n} - d_0)^k] \\ &\quad + \frac{\sqrt{m}}{k!}[\mu^{(k)}(\tilde{\zeta}_n) - \mu^{(k)}(\zeta_n)](d_{m,n} - d_0)^k \\ &> \frac{\mu^{(k)}(d_0+)\eta^k}{k!} + o(1). \end{aligned}$$

Hence, we can choose a positive constant K_0 such that

$$\left[\Phi_n\left(\sqrt{m}\mu(d_{m,n} + \eta m^{-1/(2k)})/(\sqrt{2}\sigma_0)\right) - \gamma\right] > K_0$$

for all sufficiently large n and thus, from (18) we get

$$M_{m,n}(d) - M_{m,n}(d_{m,n}) \geq f_0 K_0 \left(d - (d_{m,n} + \eta m^{-1/(2k)})\right) \tag{19}$$

for $d \in (d_{m,n} + \eta m^{-1/(2k)}, d_0 + \epsilon)$. Also, $|\Phi_n(0) - \gamma| > (1/2)\,|1/2 - \gamma|$, for large n. Choose $K_1 = \frac{1}{2} f_0 \ \min[K_0, |\gamma - 1/2|]$ in (14). Then,

$$\begin{aligned}
[\rho_n(d, d_{m,n}) < \kappa] &= [d_0, d_{m,n} + \eta m^{-1/(2k)}] \cup [d < d_0, |d - d_0| < \kappa^2/K_1] \\
&\quad \cup [d > d_{m,n} + \eta m^{-1/(2k)}, |d - d_{m,n} - \eta m^{-1/(2k)}| < \kappa^2/K_1] \\
&\subset [|d - d_{m,n}| < \kappa^2/K_1 + A m^{-1/(2k)}].
\end{aligned}$$

Here, A is a fixed constant chosen such that $A > \max(\eta, m^{-1/(2k)}(d_{m,n} - d_0))$, for all sufficiently large n; this follows from (10). Let κ be chosen such that $\kappa^2/K_1 + 2Am^{-1/(2k)} < \epsilon$ for all sufficiently large n. As $|d_0 - d_{m,n}| < Am^{-1/(2k)}$, this gives $[\rho_n(d, d_{m,n}) < \kappa] \subset (d_0 - \epsilon, d_0 + \epsilon)$. Thus, for large n and d such that $[\rho_n(d, d_{m,n}) < \kappa]$, using the definition of ρ_n and relations (17) and (19), we have the desired result.

A.2 Proof of Theorem 3

In order to deduce the limit of the process $\hat{V}_n$ (see (11)), we first prove a lemma that allows us to work with σ_0 instead of $\hat{\sigma}$.

Lemma 2. *Let $V_n(t) = n\{\mathbb{M}_{m,n}(d_0 + t/n, \sigma_0) - \mathbb{M}_{m,n}(d_0, \sigma_0)\}$. Then, for any $L > 0$,*

$$\sup_{t \in [-L, L]} |\hat{V}_n(t) - V_n(t)| \xrightarrow{P} 0,$$

where $\xrightarrow{P}$ denotes convergence in probability.

Proof. For all $t \in [-L, L]$, we have

$$\begin{aligned}
&|\hat{V}_n(t) - V_n(t)| \\
&= \left| \sum_{i=1}^{n} \left\{ \Phi\left(\frac{\sqrt{m}\bar{Y}_i}{\hat{\sigma}}\right) - \Phi\left(\frac{\sqrt{m}\bar{Y}_i}{\sigma_0}\right) \right\} \left(1\left(X_i \le d_0 + \frac{t}{n}\right) - 1\left(X_i \le d_0\right) \right) \right| \\
&\le \sup_{y \in \mathbb{R}} \left| \Phi\left(\frac{\sqrt{m}y}{\hat{\sigma}}\right) - \Phi\left(\frac{\sqrt{m}y}{\sigma_0}\right) \right| \sum_{i=1}^{n} 1\left(X_i \in \left[d_0 - \frac{L}{n}, d_0 + \frac{L}{n}\right]\right) \\
&\le \sup_{u \in \mathbb{R}} \left| \Phi(u) - \Phi\left(\frac{\hat{\sigma}}{\sigma_0} u\right) \right| \sum_{i=1}^{n} 1\left(X_i \in \left[d_0 - \frac{L}{n}, d_0 + \frac{L}{n}\right]\right).
\end{aligned}$$

Also, $\sigma \mapsto \sup_{u \in \mathbb{R}} |\Phi(u) - \Phi(u\sigma/\sigma_0)|$ can be shown to be continuous; in fact, a closed form expression can be obtained by taking derivatives. It can be seen that for $a \in (0, \infty)$,

$$\sup_{u \in \mathbb{R}} |\Phi(u) - \Phi(au)| = \begin{cases} 0, & a = 1, \\ \left| \Phi\left(\sqrt{\frac{2\ \log a}{a^2 - 1}}\right) - \Phi\left(a\sqrt{\frac{2\ \log a}{a^2 - 1}}\right) \right|, & a \neq 1. \end{cases}$$

This can be shown to be continuous at 1 by elementary calculations. Thus the first term in the bound for $|\hat{V}_n(t) - V_n(t)|$ converges in probability to 0. Moreover, the remaining term is a Binomial random variable $(Bin(n, F(d_0 + L/n) - F(d_0 - L/n)))$ which converges weakly to the Poisson distribution with parameter $2Lf(d_0)$. Thus by Slutsky's theorem, we obtain the desired result. □

We now continue with the proof of Theorem 3. We first prove that $(\hat{V}_n, J_n)$ converges weakly to (V, J) as processes in $D[-C, C] \times D[-C, C]$, for each positive integer C. By *Lemma 2*, it suffices to show that (V_n, J_n) converges weakly to (V, J).

To justify the finite dimensional convergence of (V_n, J_n) to (V, J), first on $[0, \infty)$, let $0 = t_0 \leq t_1 < t_2 < \ldots < t_l$. By Cramér-Wold device, it suffices to show that the characteristic function of

$$(V_n(t_1), J_n(t_1), V_n(t_2) - V_n(t_1), J_n(t_2) - J_n(t_1), \ldots, V_n(t_l) - V_n(t_{l-1}), J_n(t_l) - J_n(t_{l-1}))$$

converges to that of

$$(V(t_1), J(t_1), V(t_2) - V(t_1), J(t_2) - J(t_1), \ldots, V(t_l) - V(t_{l-1}), J(t_l) - J(t_{l-1}))\,.$$

We illustrate this derivation for $l = 2$, the extension to larger ls following in a straightforward manner. For $(c_i, d_i) \in \mathbb{R}^2$, $i = 1, 2$, consider the expression

$$E\left[\exp\left[\imath\left(c_1 V_n(t_1) + d_1 J_n(t_1) + \{c_2(V_n(t_2) - V_n(t_1)) + d_2(J_n(t_2) - J_n(t_1))\}\right)\right]\right]. \tag{20}$$

As $t_0 = 0$, note that

$$\begin{aligned} & c_1 V_n(t_1) + d_1 J_n(t_1) + c_2(V_n(t_2) - V_n(t_1)) + d_2(J_n(t_2) - J_n(t_1)) \\ & \quad = \sum_{j=1}^{n}\sum_{i=1}^{2}\left\{c_i \Phi\left(\frac{\sqrt{m}\bar{Y}_j}{\sigma_0}\right) - c_i\gamma + d_i\right\} 1\left(X_j \in \left(d_0 + \frac{t_{i-1}}{n}, d_0 + \frac{t_i}{n}\right]\right). \end{aligned}$$

The above summands are independent for different js and hence, (20) equals

$$\left[E\left[\exp\left[\imath\sum_{i=1}^{2}\left\{c_i\Phi\left(\frac{\sqrt{m}\bar{Y}_1}{\sigma_0}\right) - c_i\gamma + d_i\right\} 1\left(X_1 \in \left(d_0 + \frac{t_{i-1}}{n}, d_0 + \frac{t_i}{n}\right]\right)\right]\right]\right]^n.$$

Let Z_{1n} be as defined in (7) and $Z \sim N(0, 1)$. Taking iterated expectations (by first conditioning on X_1), the above display equals $(1 + \xi_n/n)^n$, where

$$\begin{aligned} \xi_n &= n\sum_{i=1}^{2}\int_{d_0+t_{i-1}/n}^{d_0+t_i/n}\left[E\left[\exp\left(\imath\left\{c_i\Phi\left(\frac{\sqrt{m}\mu(x)}{\sigma_0} + Z_{1n}\right) - c_i\gamma + d_i\right\}\right)\right] - 1\right] f(x)dx \\ &= \sum_{i=1}^{2}\int_{t_{i-1}}^{t_i}\left[\left[E\left(c_i\exp\left(\imath s\left\{\Phi\left(\frac{\sqrt{m_0}\,\mu^{(k)}(d_0+)}{k!\sigma_0}u^k + o(1) + Z_{1n}\right) - c_i\gamma + d_i\right\}\right)\right) - 1\right]\right. \\ &\qquad \left.\times f\left(d_0 + \frac{u}{n}\right)\right] du. \end{aligned}$$

The $o(1)$ term appearing in the above expression does not depend on u as $\sup_{(d_0, d_0+\zeta_0)} |\mu^{(k)}(x)| < \infty$ by Assumption 1. As $Z_{1n} + o(1)$ converges weakly to Z and $\exp(\imath\cdot)$ is bounded, ξ_n converges to $f(d_0)\xi_0$ where

$$\xi_0 = \sum_{i=1}^{2}\int_{t_{i-1}}^{t_i}\left[E\left(\exp\left(\imath s\left\{\Phi\left(\frac{\sqrt{m_0}\,\mu^{(k)}(d_0+)}{k!\sigma_0}u^k + Z\right) - \gamma\right\}\right)\right) - 1\right] du.$$

So, the expression in (20) converges to $\exp\left(f(d_0)\xi_0\right)$. This is precisely the characteristic function of $(V(t_1), J(t_1), V(t_2) - V(t_1), J(t_2) - J(t_1))$ evaluated at (c_1, d_1, c_2, d_2). To see this, first note that $(V(t_1), J(t_1))$ and $(V(t_2) - V(t_1), J(t_2) - J(t_1))$ are independent by virtue of the fact that the arrival times of events occurring over disjoint sets are independent for a Poisson process. Further, let W_j be i.i.d. $U(0,t)$, for $j \geq 1$, which are independent of $\{Z_j\}_{j\geq 1}$ and ν^+. Using the order statistic characterization of the arrival times of a Poisson process,

$$\begin{aligned} & E\left[\exp\left(\imath\left\{c_1 V(t_1) + d_1 J(t_1)\right\}\right)|\nu^+(t_1)\right] \qquad (21) \\ = \;& E\left[\exp\left(\sum_{j=1}^{\nu^+(t)} \imath\left(c_1\Phi\left(\frac{\sqrt{m_0}\;\mu^{(k)}(d_0+)}{k!\sigma_0}S_j^k + Z_j\right) - c_1\gamma + d_1\right)\right)\middle|\nu^+(t_1)\right] \\ = \;& E\left[\exp\left(\sum_{j=1}^{\nu^+(t)} \imath\left(c_1\Phi\left(\frac{\sqrt{m_0}\;\mu^{(k)}(d_0+)}{k!\sigma_0}W_j^k + Z_j\right) - c_1\gamma + d_1\right)\right)\middle|\nu^+(t_1)\right] \\ = \;& \left[E\;\exp\left(\imath\left(c_1\Phi\left(\frac{\sqrt{m_0}\;\mu^{(k)}(d_0+)}{k!\sigma_0}W_1^k + Z_1\right) - c_1\gamma + d_1\right)\right)\right]^{\nu^+(t_1))} \\ = \;& \left[\frac{g(c_1,d_1,0,t_1)+t_1}{t_1}\right]^{\nu^+(t_1)}, \end{aligned}$$

where for $0 \leq s < t$,

$$g(c_1,d_1,s,t) = \int_s^t \left[E\left(\exp\left(\imath\left\{c_1\Phi\left(\frac{\sqrt{m_0}\;\mu^{(k)}(d_0+)}{k!\sigma_0}u^k + Z\right) - c_1\gamma + d_1\right\}\right)\right) - 1\right]du.$$

Note that the relation in (21) holds even when $\nu^+(t)$ is 0. Thus,

$$E\left[\exp\left(\imath\left\{c_1 V(t_1) + d_1 J(t_1)\right\}\right)\right] = \exp(f(d_0)g(c_1,d_1,0,t_1)).$$

Similarly, it can be deduced that

$$E\left[\exp\left(\imath\left\{c_2(V(t_2)-V(t_1)) + d_2(J(t_2)-J(t_1))\right\}\right)\right] = \exp(f(d_0)g(c_2,d_2,t_1,t_2)).$$

Using the independence between $(V(t_1),J(t_1))$ and $(V(t_2)-V(t_1),J(t_2)-J(t_1))$, we get that the limit of (20) is indeed the characteristic function of $(V(t_1),J(t_1),V(t_2)-V(t_1),J(t_2)-J(t_1))$. Hence, finite dimensional convergence of (V_n,J_n) to (V,J) on $[0,\infty]$ follows from Lévy continuity theorem. The finite dimensional convergence on the entire domain can be deduced analogously.

Next, we complete the proof of weak convergence of (V_n,J_n) to (V,J) by showing asymptotic tightness. For $t_1 < t < t_2$ and sufficiently large n,

$$\begin{aligned} & E[|J_n(t)-J_n(t_1)||J_n(t_2)-J_n(t)|] \\ = \;& E\left[\sum_{i=1}^n\sum_{j=1}^n 1\left(X_i \in \left(d_0+\frac{t_1}{n}, d_0+\frac{t}{n}\right]\right)1\left(X_j \in \left(d_0+\frac{t}{n}, d_0+\frac{t_2}{n}\right]\right)\right] \\ = \;& n(n-1)E\left[\left(X_1 \in \left(d_0+\frac{t_1}{n}, d_0+\frac{t}{n}\right]\right)1\left(X_2 \in \left(d_0+\frac{t}{n}, d_0+\frac{t_2}{n}\right]\right)\right] \\ \leq \;& 2\|f\|_\infty^2(t-t_1)(t_2-t) \leq 2\|f\|_\infty^2(t_2-t_1)^2, \end{aligned}$$

where $\|f\|_\infty < \infty$ by Assumption 2. The above relation shows that the condition stated for tightness in *Theorem 15.6* of Billingsley (1968, pp. 128) is satisfied and hence, the process J_n is asymptotically tight. As $|V_n(t) - V_n(t_1)| \leq |J_n(t) - J_n(t_1)|$, the process V_n is asymptotically tight. As both the marginal processes are tight, (V_n, J_n) is tight and hence, condition (i) of Theorem 2 is satisfied.

Moreover, no two flat stretches of $V(t), t \in [-C, C]$, have the same height (w.p. 1). To see this, let A_C denote this event and define $R_i = \sum_{j=1}^{i} \left(\Phi\left(\sqrt{m_0}\,\mu^{(k)}(d_0+)/(k!\sigma_0)S_j^k + Z_j\right) - \gamma\right)$ when $i > 0$, and $R_i = \sum_{j=1}^{-i} (\gamma - U_j)$ when $i < 0$, and $R_0 = 0$. For non-negative integers n_1 and n_2, $n_1 + n_2 > 0$, we have $P[R_i = R_l | \nu^+(C) = n_1, \nu^-(C) = n_2] = 0$ for $n_1 \geq i > l \geq -n_2$. This is because given $\nu^+(C) = n_1$ and $\nu^-(C) = n_2$, the arrival times for S_js are the order statistics from $U(0, C)$ and thus $R_i - R_l$ is a continuous random variable. Now,

$$
\begin{aligned}
&P[A_C | \nu^+(C) = n_1, \nu^-(C) = n_2] \\
&= 1 - P\left[\bigcup_{n_1 \geq i > l \geq -n_2} [R_i = R_l] \,\middle|\, \nu^+(C) = n_1, \nu^-(C) = n_2 \right] = 1.
\end{aligned}
$$

Also, $P[A_C | \nu^+(C) = 0, \nu^-(C) = 0] = 1$. Hence,

$$
P[A_C] = E[P[A_C | \nu^+(C), \nu^-(C)]] = 1.
$$

Further, let $\hat{h}_l = n(\hat{d}_{m,n} - d_0)$ and $\hat{h}_u$ denote the smallest and largest minimizers of $\hat{V}_n(t)$, respectively. Using Theorem 1, $(\hat{h}_l, \hat{h}_u)$ is $O_P(1)$. Also, let h_l and h_u denote the smallest and largest minimizers for $V(t)$. As $V(0) = 0$ and $V(t) \to \infty$ as $|t| \to \infty$ w.p. 1, we get $(h_l, h_u) = O_P(1)$. To see that $V(t) \to \infty$ as $|t| \to \infty$ a.s., note that $\sum_{j=1}^{n} (\gamma - U_j)/n \to \gamma - \frac{1}{2} > 0$ and $\nu^-(-t) \to \infty$, a.s. So, we get $V(t) \to \infty$ as $t \to \infty$ a.s. Also, choose $\epsilon > 0$ and $\eta_\epsilon > 0$ such that $\gamma + \epsilon < 1$ and $E\Phi[\eta_\epsilon + Z_1] = \Phi[\eta_\epsilon/\sqrt{2}] = \gamma + \epsilon$. Then by the SLLN, $\sum_{j=1}^{n} (\Phi[\eta_\epsilon + Z_j] - \gamma)/n \to \epsilon$ a.s. As $S_j \to \infty$ and $\nu^+(t) \to \infty$ a.s., we get $\liminf_{t\to\infty} \{V(t)/\nu^+(t)\} \geq \epsilon$ a.s. Thus $V(t) \to \infty$ as $|t| \to \infty$ w.p. 1. Hence, by applying *Theorem 2* we get the desired result. □

A.3 Proof of Proposition 1

The proof of Proposition 1 follows along the same lines as that of Theorem 3. Here, we briefly justify the form of the limiting distribution. By calculations analogous to those used for simplifying (20), it can be shown that for $t > 0$, $E(\exp(\imath c V_n(t))) = \left(1 + \bar{\xi}_n/n\right)^n$, where

$$
\begin{aligned}
\bar{\xi}_n &= \int_0^t \left[\left\{ E\left(\exp\left(\imath c \left\{ \Phi\left(\frac{\sqrt{m}\mu(d_0 + u/n)}{\sigma_0} + Z_{1n} \right) - \gamma \right\} \right) \right) - 1 \right\} \right. \\
&\quad \left. \times f\left(d_0 + \frac{u}{n}\right) \right] du \\
&\to f(d_0) \int_0^t \left[\exp\left(\imath c \{1 - \gamma\} \right) - 1 \right] du = f(d_0) \{\exp(\imath c(1 - \gamma)) - 1\} t.
\end{aligned}
$$

The above convergence uses the fact $\sqrt{m}\mu(d_0 + u/n)/\sigma_0 \to \infty$ for $u > 0$, which can be justified through a k-th order Taylor expansion of μ around d_0. The limit here is precisely the characteristic function of $\bar{V}(t)$. Hence, the one-dimensional marginals of V_n converge to that of $\bar{V}$ on the positive half line. The remainder of the proof is almost identical to that for Theorem 3.

A.4 Proof of Proposition 2

For proving Proposition 2, we first prove the following lemma to justify imputing σ_0 in place of $\hat{\sigma}$ in the local processes.

Lemma 3. *Consider the case when $\beta < 2k$. Let $H_n(t) = n^{\beta/(2k)}\left\{\mathbb{M}_{m,n}\left(d_0 + t/n^{\beta/(2k)}, \sigma_0\right) - \mathbb{M}_{m,n}(d_0, \sigma_0)\right\}$. Then, for any $L > 0$,*

$$\sup_{t\in[-L,L]} |\hat{H}_n(t) - H_n(t)| \xrightarrow{P} 0.$$

Proof. For $t \in [-L, L]$,

$$\begin{aligned}
&|\hat{H}_n(t) - H_n(t)| \\
&= n^{\beta/(2k)-1}\left|\sum_{i=1}^n \left\{\Phi\left(\frac{\sqrt{m}\bar{Y}_i}{\hat{\sigma}}\right) - \Phi\left(\frac{\sqrt{m}\bar{Y}_i}{\sigma_0}\right)\right\}\left(1\left(X_i \le d_0 + \frac{t}{n^{\beta/(2k)}}\right) - 1\left(X_i \le d_0\right)\right)\right| \\
&\le n^{\beta/(2k)-1}\sup_{y\in\mathbb{R}}\left|\Phi\left(\frac{\sqrt{m}y}{\hat{\sigma}}\right) - \Phi\left(\frac{\sqrt{m}y}{\sigma_0}\right)\right|\sum_{i=1}^n 1\left(X_i \in \left[d_0 - \frac{L}{n^{\beta/(2k)}}, d_0 + \frac{L}{n^{\beta/(2k)}}\right]\right) \\
&\le \sup_{u\in\mathbb{R}}\left|\Phi(u) - \Phi\left(\frac{\hat{\sigma}}{\sigma_0}u\right)\right|\left\{n^{\beta/(2k)-1}\sum_{i=1}^n 1\left(X_i \in \left[d_0 - \frac{L}{n^{\beta/(2k)}}, d_0 + \frac{L}{n^{\beta/(2k)}}\right]\right)\right\}.
\end{aligned}$$

As in the proof of *Lemma 2*, the first term goes in probability to zero. As for the second term,

$$\begin{aligned}
&Var\left[n^{\beta/(2k)-1}\sum_{i=1}^n 1\left(X_i \in \left[d_0 - \frac{L}{n^{\beta/(2k)}}, d_0 + \frac{L}{n^{\beta/(2k)}}\right]\right)\right] \\
&= n^{2(\beta/(2k)-1)} n \; Var\left[1\left(X_i \in \left[d_0 - \frac{L}{n^{\beta/(2k)}}, d_0 + \frac{L}{n^{\beta/(2k)}}\right]\right)\right] \\
&= n^{2(\beta/(2k)-1)} n \; O(n^{-\frac{\beta}{k}}) = O(n^{-1}) \to 0,
\end{aligned}$$

and

$$\begin{aligned}
&E\left[n^{\beta/(2k)-1}\sum_{i=1}^n 1\left(X_i \in \left[d_0 - \frac{L}{n^{\beta/(2k)}}, d_0 + \frac{L}{n^{\beta/(2k)}}\right]\right)\right] \\
&= n^{(\beta/(2k)-1)} n \; E\left[1\left(X_i \in \left[d_0 - \frac{L}{n^{\beta/(2k)}}, d_0 + \frac{L}{n^{\beta/(2k)}}\right]\right)\right] \\
&= n^{(\beta/(2k)-1)} n \; O(n^{-\beta/(2k)}) = O(1).
\end{aligned}$$

Thus the second term is $O(1) + o_P(1)$. Hence, we get the result. □

We use a version of the Arzela-Ascoli theorem in several proofs 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.*

We now continue with the proof of Proposition 2. Using *Lemma 3*, for proving (13), it would suffice to show that

$$\sup_{t\in[-L,L]} |H_n(t) - c(t)| \xrightarrow{P} 0. \tag{22}$$

Let

$$c_n(t) = E\{H_n(t)\} = n^{\beta/(2k)} \int_{d_0}^{d_0+tn^{-\beta/(2k)}} E\left[\left\{ \Phi\left(\frac{\sqrt{m}\bar{Y}}{\sigma_0}\right) - \gamma \right\} \middle| X = x \right] f(x)dx.$$

For $x < 0$ and given $X = x$, $\Phi\left(\frac{\sqrt{m}\bar{Y}}{\sigma_0}\right) \xrightarrow{d} U(0,1)$. Hence, by the dominated convergence theorem (DCT), $c_n(t) \to \left(\frac{1}{2} - \gamma\right) f(d_0)t$, for $t \le 0$. For $t > 0$, we have:

$$\begin{aligned}
c_n(t) &= n^{\beta/(2k)} \int_{d_0}^{d_0+tn^{-\beta/(2k)}} E\left[\left\{ \Phi\left(\frac{\sqrt{m}\bar{Y}}{\sigma_0}\right) - \gamma \right\} \middle| X = x \right] f(x)dx \\
&= n^{\beta/(2k)} \int_{d_0}^{d_0+tn^{-\beta/(2k)}} \left\{ \Phi\left(\frac{\sqrt{m}\mu(x)}{\sqrt{2}\sigma_0}\right) - \gamma \right\} f(x)dx \\
&= \int_0^t \left\{ \Phi\left(\frac{\sqrt{m}\mu(d_0 + u/n^{\beta/(2k)})}{\sqrt{2}\sigma_0}\right) - \gamma \right\} f\left(d_0 + \frac{u}{n^{\beta/(2k)}}\right) du \\
&= \int_0^t \left\{ \Phi\left(\frac{\sqrt{m_0}\ \mu^{(k)}(d_0+)}{\sqrt{2}k!\sigma_0} u^k + o(1)\right) - \gamma \right\} f\left(d_0 + \frac{u}{n^{\beta/(2k)}}\right) du \\
&\to f(d_0) \int_0^t \left\{ \Phi\left(\frac{\sqrt{m_0}\ \mu^{(k)}(d_0+)}{\sqrt{2}k!\sigma_0} u^k\right) - \gamma \right\} du, \text{ by DCT.}
\end{aligned}$$

Hence, $c_n(t) \to c(t)$. In fact, this convergence is uniform on any compact set. To see this, note that $|c_n(t) - c_n(s)| \le \|f\|_\infty |t - s|$. So, c_ns are equicontinuous and thus by Arzela-Ascoli, the convergence is uniform on $[-L, L]$ for every $L > 0$. Further, let $\tilde{H}_n(t) = n^{1/2-\beta/(4k)}(H_n(t) - c_n(t))$. Then, for $t_1 < t < t_2$,

$$\begin{aligned}
&E|\tilde{H}_n(t) - \tilde{H}_n(t_1)|^2 |\tilde{H}_n(t_2) - \tilde{H}_n(t)|^2 = E\ |\tilde{H}_n(t) - \tilde{H}_n(t_1)|^2\ E\ |\tilde{H}_n(t_2) - \tilde{H}_n(t)|^2 \\
&= Var\left[n^{\beta/(4k)} \left\{ \Phi\left(\frac{\sqrt{m}\bar{Y}}{\sigma_0}\right) - \gamma \right\} 1\left(X_1 \in \left(d_0 + \frac{t_1}{n^{\beta/(2k)}}, d_0 + \frac{t}{n^{\beta/(2k)}} \right] \right) \right] \times \\
&\quad Var\left[n^{\beta/(4k)} \left\{ \Phi\left(\frac{\sqrt{m}\bar{Y}}{\sigma_0}\right) - \gamma \right\} 1\left(X_1 \in \left(d_0 + \frac{t}{n^{\beta/(2k)}}, d_0 + \frac{t_2}{n^{\beta/(2k)}} \right] \right) \right] \\
&\le n^{\beta/(2k)} E\left[\left\{ \Phi\left(\frac{\sqrt{m}\bar{Y}}{\sigma_0}\right) - \gamma \right\} 1\left(X_1 \in \left(d_0 + \frac{t_1}{n^{\beta/(2k)}}, d_0 + \frac{t}{n^{\beta/(2k)}} \right] \right) \right]^2 \times \\
&\quad n^{\beta/(2k)} E\left[\left\{ \Phi\left(\frac{\sqrt{m}\bar{Y}}{\sigma_0}\right) - \gamma \right\} 1\left(X_1 \in \left(d_0 + \frac{t}{n^{\beta/(2k)}}, d_0 + \frac{t_2}{n^{\beta/(2k)}} \right] \right) \right]^2 \\
&\le \|f\|_\infty^2 (t - t_1)(t_2 - t) \le \|f\|_\infty^2 (t_2 - t_1)^2.
\end{aligned}$$

So, by *Theorem 15.6* in Billingsley (1968), pp. 128, $\tilde{H}$ is tight in $D(\mathbb{R})$. As $\beta < 2k$, $(H_n(t) - c_n(t)) \xrightarrow{d} 0$ and hence $H_n(t) \xrightarrow{d} c(t)$ as processes in $D(\mathbb{R})$. As the limiting process in degenerate and $x(\cdot) \mapsto \sup_{t\in[-L,L]} |x(t)|$ is continuous, we get (22).

Moreover the limit process, $c(t)$, is continuous and has a unique minimum. Also, $n^{\beta/(2k)}(\hat{d}_{m,n} - d_0)$ is $O_P(1)$. Thus, by the argmin continuous mapping, we obtain the desired result. □

A.5 Proof of Proposition 4

We first show that

$$\frac{1}{nh_n^{2k+1}} g_1(\hat{d}_{m,n}, h_n) = \frac{f(d_0)}{2k+1} + o_P(1),$$

where

$$g_1(d,h) = \sum_{i=1}^{n} (X_i - d)^{2k} 1(X_i \in (d, d+h]).$$

Note that $h_n^{-(2k+1)}(\hat{d}_{m,n} - d_0) = o_P(1)$. Fix $\delta > 0$. Then $P[|\hat{d}_{m,n} - d_0| < \delta\ h_n^{2k+1}] \to 1$. On the set $[|\hat{d}_{m,n} - d_0| < \delta\ h_n^{2k+1}]$,

$$g_1(d_0 - \delta h_n^{2k+1}, h_n + 2\delta h_n^{2k+1}) \geq g_1(\hat{d}_{m,n}, h_n) \geq g_1(d_0 + \delta h_n^{2k+1}, h_n - 2\delta h_n^{2k+1}). \tag{23}$$

So, it suffices to show that the above two bounds converge in probability to $f(d_0)/(2k+1)$. Note that

$$\begin{aligned}
& E\left[\frac{1}{nh_n^{2k+1}} g_1(d_0 + \delta h_n^{2k+1}, h_n - 2\delta h_n^{2k+1})\right] \\
= & \frac{n}{nh_n^{2k+1}} \int_{d_0+\delta h_n^{2k+1}}^{d_0+h_n-\delta h_n^{2k+1}} (x - d_0 - \delta h_n^{2k+1})^{2k} f(x)\ dx \\
= & \frac{1}{h_n^{2k+1}} \int_0^{1-\delta h_n^{2k}} (uh_n)^{2k} f(d_0 + \delta h_n^{2k+1} + uh_n) h_n du \\
= & f(d_0) \int_0^1 u^{2k} du + o(1) = \frac{f(d_0)}{2k+1} + o(1),
\end{aligned}$$

and

$$\begin{aligned}
& Var\left[\frac{1}{nh_n^{2k+1}} g_1(d_0 + \delta h_n^{2k+1}, h_n - 2\delta h_n^{2k+1})\right] \\
= & \frac{n}{(nh_n^{2k+1})^2} Var\left[(X_1 - d_0 - \delta h_n^{2k+1})^{2k} 1(X_1 \in (d_0 + \delta h_n^{2k+1}, d_0 - \delta h_n^{2k+1} + h_n])\right] \\
\leq & \frac{n}{(nh_n^{2k+1})^2} E\left[(X_1 - d_0 - \delta h_n^{2k+1})^{4k} 1(X_1 \in (d_0 + \delta h_n^{2k+1}, d_0 - \delta h_n^{2k+1} + h_n])\right] \\
\leq & \frac{n}{(nh_n^{2k+1})^2} (h_n - 2\delta h_n^{2k+1})^{4k+1} (f(d_0) + o(1)) = \frac{O(1)}{nh_n} \to 0.
\end{aligned}$$

Thus, $\frac{1}{nh_n^{2k+1}} g_1(d_0 + \delta h_n^{2k+1}, h_n - 2\delta h_n^{2k+1}) = \frac{f(d_0)}{2k+1} + o_P(1)$. The treatment of the upper bound in (23) is similar. Next, let $g_2(d,h) = \sum \bar{Y}_i (X_i - d)^k 1(X_i \in (d, d+h])$. As the k-th derivative of μ is bounded in $(d_0, d_0 + \zeta)$ for sufficiently small ζ, we have

$$\begin{aligned}
& E\left[\frac{1}{nh_n^{2k+1}} g_2(d_0, h_n)\right] = \frac{n}{nh_n^{2k+1}} \int_{d_0}^{d_0+h_n} \mu(x)(x - d_0)^k f(x) dx \\
= & \frac{1}{h_n^{2k+1}} \int_0^1 \mu(d_0 + uh_n)(uh_n)^k f(d_0 + \delta h_n^{2k+1} + uh_n) h_n du \\
= & \frac{1}{h_n^{2k+1}} \int_0^1 (\xi(uh_n)^k + o((uh_n)^k))(uh_n)^k f(d_0 + \delta h_n^{2k+1} + uh_n) h_n du \\
= & \xi \frac{f(d_0)}{2k+1} + o(1), \text{by DCT.}
\end{aligned}$$

Also, by similar calculations,

$$
\begin{aligned}
& Var\left[\frac{1}{nh_n^{2k+1}} g_2(d_0, h_n)\right] \\
\leq \quad & \frac{n}{(nh_n^{2k+1})^2} E\left[\bar{Y}_1 (X_1 - d_0)^k 1(X_1 \in (d_0, d_0 + h_n])\right]^2 \\
= \quad & \frac{n}{(nh_n^{2k+1})^2} E\left[\left\{\mu^2(X_1) + \sigma_0^2/(m_0 n^{2k})\right\}(X_1 - d_0)^{2k} 1(X_1 \in (d_0, d_0 + h_n])\right] \\
\leq \quad & \frac{O(1)}{nh_n} + \frac{n\sigma_0^2}{c(nh_n^{2k+1})^2 n^{2k}} O(h_n^{2k+1}) \to 0.
\end{aligned}
$$

So, $\frac{1}{nh_n^{2k+1}} g_2(d_0, h_n) = \xi \frac{f(d_0)}{2k+1} + o_P(1)$. To conclude the final result, we need to show that

$$
\frac{1}{nh_n^{2k+1}} \left\{g_2(\hat{d}_{m,n}, h_n) - g_2(d_0, h_n)\right\} = o_P(1).
$$

Let $M_0 = \sup_{d \in (d_0, d_0+\zeta)} \mu(d)$, which is finite for sufficiently small ζ. On the set $[|\hat{d}_{m,n} - d_0| < \delta h_n^{2k+1}]$, and for large n,

$$
\begin{aligned}
& \left|g_2(\hat{d}_{m,n}, h_n) - g_2(d_0, h_n)\right| \\
\leq \quad & \sup_{|d-d_0|<\delta h_n^{2k+1}} \left|\frac{1}{nh_n^{2k+1}} \{g_2(d, h_n) - g_2(d_0, h_n)\}\right| \\
\leq \quad & \sup_{|d-d_0|<\delta h_n^{2k+1}} \sum_{i=1}^n \left[|\bar{Y}_i| (X_i - d)^k - (X_i - d_0)^k | 1(X_i \in (d_0, d_0 + h_n] \cap (d, d + h_n]) \right. \\
& \quad \left. + \ |\bar{Y}_i| (X_i - d \wedge d_0)^k 1(X_i \in (d_0, d_0 + h_n] \Delta (d, d + h_n])\right] \\
\leq \quad & \sup_{|d-d_0|<\delta h_n^{2k+1}} \sum_{i=1}^n \left[|\bar{Y}_i| k (X_i - d_0 + \delta h_n^{2k+1})^{k-1} |d - d_0| 1(X_i \in (d_0, d_0 + h_n]) \right. \\
& \quad \left. + |\bar{Y}_i| (X_i - d_0 - \delta h_n^{2k+1})^k \{1(|X_i - d_0| \leq \delta h_n^{2k+1}) + 1(|X_i - d_0 - h_n| \leq \delta h_n^{2k+1})\}\right] \\
= \quad & O(h_n^{3k}) \sum_{i=1}^n |\bar{Y}_i| 1(X_i \in (d_0, d_0 + h_n]) \\
& \quad + O(h_n^k) \sum_{i=1}^n |\bar{Y}_i| \{1(|X_i - d_0| \leq \delta h_n^{2k+1}) + 1(|X_i - d_0 - h_n| \leq \delta h_n^{2k+1})\} \\
\leq \quad & O(h_n^{3k}) \sum_{i=1}^n (M_0 + |\bar{\epsilon}_i|) 1(X_i \in (d_0, d_0 + h_n]) \\
& \quad + O(h_n^k) \sum_{i=1}^n (M_0 + |\bar{\epsilon}_i|) \{1(|X_i - d_0| \leq \delta h_n^{2k+1}) + 1(|X_i - d_0 - h_n| \leq \delta h_n^{2k+1})\} \\
\leq \quad & O(h_n^{3k}) O_P(nh_n) + O(h_n^k) O_P(nh_n^{2k+1}) = o_P(nh_n^{2k+1}).
\end{aligned}
$$

The last inequality follows from the fact that $\frac{1}{nh_n} \sum_{i=1}^n (M_0 + |\bar{\epsilon}_i|) 1(X_i \in (d_0, d_0 + h_n]) \stackrel{P}{\to} M_0 f(d_0)$, which can be justified by computing the limiting means and variances. This completes the proof. $\square$

A.6 Proof of Proposition 6

For $\epsilon > 0$ and $x \in \mathbb{R}$, let

$$U_n(x) = U_n(x,\beta) = \frac{1}{N_n}\sum_{j=1}^{N_n} 1\left[q_n(\hat{d}_{n,q_n,j} - d_0) \le x\right]$$

and $E_n = [q_n|\hat{d}_n - d_0| \le \epsilon]$. As $q_n/n \to 0$ and $n(\hat{d}_n - d_0) = O_P(1)$, $P(E_n) \to 1$. Moreover, on the set E_n,

$$U_n(x-\epsilon) \le L_{n,q}(x) \le U_n(x+\epsilon).$$

Hence, to show pointwise convergence (in probability) of $L_{n,q}(\cdot,\beta)$ to $G_\beta(\cdot)$, it suffices to show that $U_n(x,\beta) \xrightarrow{P} G_\beta(x)$. Note that $E\,[U_n(x)] = G_{n,\beta}(x) \to G_\beta(x)$. So, it suffices to show that $Var(U_n(x)) \to 0$. To this end, let $s_n = \lfloor n/q_n \rfloor$. For $j = 0,\ldots,(s_n-1)$, let $R_{n,q_n,j}$ be the statistic $\hat{d}_{q_n}$ computed from the data set $(X_{q_nj+1}, Y_{(q_nj+1)\ 1},\ldots,Y_{(q_nj+1)\ l_n};\ldots;X_{q_nj+q_n},Y_{(q_nj+q_n)\ 1},\ldots,\ Y_{(q_nj+q_n)\ l_n})$ and

$$\bar{U}_n(x) = \frac{1}{s_n}\sum_{j=1}^{s_n} 1\left[q_n(R_{n,q_n,j} - d_0) \le x\right].$$

$\bar{U}_n(x)$ has the same expectation as $U_n(x)$, but its summands are independent. Also each summand lies between 0 and 1, and hence has a variance bounded above by $1/4$. Let $X_{(i)}$s denote the ordered X_is and $Y_{[i](j)}$s be their ordered concomitants, i.e.,$Y_{[i](j)}$s are the replications at $X_{(i)}$s and $Y_{[i](j)} \le Y_{[i](j+1)}, j = 1,\ldots(m-1)$. It can be seen that

$$U_n(x) = E\left[\bar{U}_n(x)|X_{(i)}, Y_{[i](j)}, 1 \le i \le n, 1 \le j \le m\right].$$

So, by the Rao-Blackwell theorem, $Var(U_n(x)) \le Var(\bar{U}_n(x)) \le 1/(4s_n) \to 0$ as $s_n = \lfloor n/q_n \rfloor \to \infty$ and thus $U_n(x,\beta) \xrightarrow{P} G_\beta(x)$ for $x \in \mathbb{R}$. The uniform convergence in probability and (ii) follow from arguments for Theorem 15.7.1 in Lehmann and Romano (2005), given the pointwise convergence shown above. □

A.7 Proof of Proposition 7

We first justify that the rate of convergence of $\hat{d}_{m,n}$ remains unchanged when we impute a $\sqrt{mn}$-consistent estimator of τ. Recall that

$$\mathbb{M}_{m,n}(d,\sigma,\tilde{\tau}) = \mathbb{P}_n\left[\left\{\Phi\left(\frac{\sqrt{m}(\bar{Y}-\tilde{\tau})}{\sigma}\right) - \gamma\right\}1(X \le d)\right].$$

As in the proof of Theorem 1, we need to bound the expression

$$E^* \sup_{\substack{\rho_n(d,d_{m,n})<\delta \\ (\sigma,\tilde{\tau})\in V_{n,\epsilon}}} \sqrt{n}|\{\mathbb{M}_{m,n}(d,\sigma,\tilde{\tau}) - M_{m,n}(d,\sigma_0,\tau)\} - \{\mathbb{M}_{m,n}(d_{m,n},\sigma,\tilde{\tau}) - M_{m,n}(d_{m,n},\sigma_0,\tau)\}|,$$

where $M_{m,n}(d,\sigma,\tilde{\tau}) = E[\mathbb{M}_{m,n}(d,\sigma,\tilde{\tau})]$, and $V_{n,\epsilon} = [\sigma_0 - L_\epsilon/\sqrt{mn}, \sigma_0 + L_\epsilon/\sqrt{mn}] \times [\tau - L_\epsilon/\sqrt{mn}, \tau + L_\epsilon/\sqrt{mn}]$ is a set with L_ϵ chosen in such a way that $P[(\hat{\sigma},\tilde{\tau}) \in V_{n,\epsilon}] > 1-\epsilon$, for $\epsilon > 0$. Following the proof of Theorem 1, the above display can be bounded by

$$E^* \sup_{\substack{\rho_n(d,d_{m,n})<\delta \\ (\sigma,\tilde{\tau})\in V_{n,\epsilon}}} \left| \mathbb{G}_n \left[\left\{ \Phi\left(\frac{\sqrt{m}(\bar{Y}-\tilde{\tau})}{\sigma}\right) - \gamma \right\} \{1(X \le d) - 1(X \le d_{m,n})\} \right] \right|$$

$$+\sqrt{n} \sup_{\substack{\rho_n(d,d_{m,n})<\delta \\ \sqrt{mn}|\sigma-\sigma_0|<L_\epsilon}} \left| P_n \left[\left\{ \Phi\left(\frac{\sqrt{m}(\bar{Y}-\tau)}{\sigma}\right) - \Phi\left(\frac{\sqrt{m}(\bar{Y}-\tau)}{\sigma_0}\right) \right\} \{1(X \le d) - 1(X \le d_{m,n})\} \right] \right|$$

$$+\sqrt{n} \sup_{\substack{\rho_n(d,d_{m,n})<\delta \\ (\sigma,\tilde{\tau})\in V_{n,\epsilon}}} \left| P_n \left[\left\{ \Phi\left(\frac{\sqrt{m}(\bar{Y}-\tilde{\tau})}{\sigma}\right) - \Phi\left(\frac{\sqrt{m}(\bar{Y}-\tau)}{\sigma}\right) \right\} \{1(X \le d) - 1(X \le d_{m,n})\} \right] \right|.$$

The first term involves empirical process acting on a class of functions with VC-index at most 3 while the second term appears in the proof of Theorem 1. These two terms can be dealt in the same manner as in that proof. For the third term, note that $|\Phi\left(\sqrt{m}(\bar{Y}-\tilde{\tau})/\sigma\right) - \Phi\left(\sqrt{m}(\bar{Y}-\tau)/\sigma\right)| \le \sup_u |\Phi\left(u + \sqrt{m}(\tau-\tilde{\tau})/\sigma\right) - \Phi(u)|$ which equals $|\Phi\left(\sqrt{m}(\tau-\tilde{\tau})/2\sigma\right) - \Phi\left(-\sqrt{m}(\tau-\tilde{\tau})/2\sigma\right)|$. As Φ is Lipschitz of order 1, this is further bounded above by $\sqrt{m}|\tau-\tilde{\tau}|/\sigma$. Hence, for sufficiently large n, the third term in the above display is bounded by

$$2(L_\epsilon/\sigma_0) \sup_{\rho_n(d,d_{m,n})<\delta} P_n|1(X \le d) - 1(X \le d_{m,n})|.$$

Hence, this term has the same order as $\phi_n(\cdot)$ appearing in (15), in the proof of Theorem 1. The rest of the argument is identical to the proof for the known τ case and thus, we end up with the same rate of convergence.

To justify that the limiting distributions also stay the same, note that $n(\hat{d}_{m,n} - d_0)$ is a minimizer of the process $n\{\mathbb{M}_n(d_0 + t/n, \hat{\sigma}, \hat{\tau}) - \mathbb{M}_n(d_0, \hat{\sigma}, \tau)\}$, $t \in \mathbb{R}$. But by arguments analogous to the proof of Lemma 2, the difference $\sup_{t\in[-L,L]} n|\mathbb{M}_n(d_0+t/n, \hat{\sigma}, \hat{\tau}) - \mathbb{M}_n(d_0 + t/n, \hat{\sigma}, \tau)|$ is $\sqrt{m}(\hat{\tau}-\tau)/\hat{\sigma} \times O_P(1)$, which goes in probability to zero for any $L > 0$. Hence, the limiting distribution is not affected as long as we have a $\sqrt{mn}$-consistent estimate of τ. □

A.8 Proof of Proposition 8

For notational convenience, we denote $\mathbb{M}_{m,n}^{FD}(d)$ by $\mathbb{M}_n^{FD}(d)$ (as m is a function of n). Let Φ_n be as defined in Section 3.1 and

$$M_n^{FD}(d) = E\left[\mathbb{M}_n^{FD}(d)\right] = \frac{1}{n}\sum_{i=1}^n \left\{ \Phi_n\left(\frac{\sqrt{m}\ \mu(i/n)}{\sqrt{1+\sigma_0^2}}\right) - \gamma \right\} 1\left(i/n \le d\right).$$

The expression on the right side follows from calculations almost identical to (8). Let $d_n^{FD} = \text{sargmin}_{d\in[0,1]} M_n^{FD}(d)$. To prove Proposition 8, 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 $d \mapsto \rho_n(d, d_n^{FD})$ such that

$$M_n^{FD}(d) - M_n^{FD}(d_n^{FD}) \gtrsim \rho_n^2(d, d_n^{FD}).$$

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

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

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 4. *Let $\eta > 0$. Let $d \mapsto \rho_n(d, d_n^{FD})$ be a map from $(0,1)$ to $[0,\infty)$ such that*

$$\begin{aligned}\rho_n^2(d, d_n^{FD}) &= (1/n)\{|\lfloor nd \rfloor - \lfloor nd_0 \rfloor| 1(d \le d_0) \\ &\quad + |\lfloor nd \rfloor - \left\lfloor n(d_n^{FD} + \eta m^{-1/(2k)}) \right\rfloor| 1(d > d_n^{FD} + \eta m^{-1/(2k)})\right\}.\end{aligned}$$

Then η and $\kappa > 0$ can be chosen such that for sufficiently large n and $\rho_n(d, d_n^{FD}) < \kappa$, we have

$$M_n(d) - M_n(d_n^{FD}) \gtrsim \rho_n^2(d, d_n^{FD}).$$

Also, $(d_n^{FD} - d_0) = O(m^{-1/(2k)})$.

We first provide the proof of Proposition 8 using Lemma 4. Using the above lemma, there exists $A < \infty$ such that for sufficiently large n and any $\delta > 0$, $\{\rho_n(d, d_n^{FD}) < \delta\} \subset \{|d - d_n^{FD}| < A(\delta^2 + n^{-\alpha})\}$. Consider the case $d > d_n^{FD}$ and let

$$U(i,d) = \left\{\Phi\left(\sqrt{m}\bar{Y}_i\right) - \Phi_n\left(\frac{\sqrt{m}\mu(i/n)}{\sqrt{1+\sigma_0^2}}\right)\right\} 1\left(d_n^{FD} < i/n \le d\right).$$

Note that $E\{U(i,d)\} = 0$ and for $1 \le i \ne j \le n$, $U(i,d)$ and $U(j,d)$ are independent. Also, $S(i,d) := (\mathbb{M}_n^{FD} - M_n^{FD})(d) - (\mathbb{M}_n^{FD} - M_n^{FD})(d_n^{FD}) = (1/n)\sum_i U(i,d)$, a normalized sum of $(\lfloor nd \rfloor - \lfloor nd_n^{FD} \rfloor)$ non-zero independent terms, is a martingale in d, $d \ge d_n^{FD}$, 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

$$\begin{aligned} &E\left[\sup_{0 \le d - d_n^{FD} < A(\delta^2 + n^{-\alpha})} \sqrt{n}|S(i,d)|\right] \\ &\quad \le \sqrt{n}\left\{ES^2(i, d_n^{FD} + A(\delta^2 + n^{-\alpha}))\right\}^{1/2} \\ &\quad = \frac{1}{\sqrt{n}}\left[\sum_{i \le n} E\{U^2(i, d_n^{FD} + A(\delta^2 + n^{-\alpha}))\}\right]^{1/2} \\ &\quad \lesssim (\delta^2 + n^{-\alpha})^{1/2}. \end{aligned}$$

A similar bound can be established for the case $d \le d_n^{FD}$. Hence, we get

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

where $\phi_n(\delta) = (\delta^2 + n^{-\alpha})^{1/2}$. The function $\phi_n(\cdot)$ and $\rho_n(\cdot, d_n^{FD})$ satisfy the conditions of Theorem 3.2.5 of van der Vaart and Wellner (1996). Hence, the rate of convergence, say r_n, satisfies

$$r_n^2 \phi_n\left(\frac{1}{r_n}\right) \lesssim \sqrt{n} \ \Rightarrow (r_n^2 + r_n^4 n^{-\alpha}) \lesssim n.$$

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

Proof of Lemma 4. Since $\mu(x) = 0$ for $x \leq d_0$, note that $d_n^{FD} > d_0$ for sufficiently large n. As $\Phi_n(0)$ converges to $1/2$, it can be seen that for large n and $d \leq d_0$,

$$\begin{aligned} M_n(d) - M_n(d_n^{FD}) &\geq M_n(d) - M_n(d_0) \\ &= \sum_{i=1}^{n} \{\gamma - \Phi_n(0)\} 1 \left(d < \frac{i}{n} \leq d_0 \right) \\ &\geq \frac{1}{2} \left(\gamma - \frac{1}{2} \right) \{ \lfloor nd \rfloor - \lfloor nd_0 \rfloor \} / n. \end{aligned} \tag{24}$$

Next, we show that

$$\Phi_n \left(\frac{\sqrt{m}\mu(d_n^{FD} + \eta/n^{\alpha})}{\sqrt{1+\sigma_0^2}} \right) - \gamma > K_0, \tag{25}$$

for sufficiently large n and some $K_0 > 0$. It can be shown that d_n^{FD} converges to d_0. Hence, d_n^{FD} is not a boundary point of the interval $[1/n, 1]$ for large n; it corresponds to a local minimum of M_n, i.e,

$$\Phi_n \left(\sqrt{m}\mu(d_n^{FD}) / \sqrt{1+\sigma_0^2} \right) \leq \gamma < \Phi_n \left(\sqrt{m}\mu(d_n^{FD} + 1/n) / \sqrt{1+\sigma_0^2} \right).$$

Thus, $\Phi_n(\sqrt{m}\mu(d_n^{FD})/\sqrt{1+\sigma_0^2})$ converges to γ and consequently, $\sqrt{m}\mu(d_n^{FD})/\sqrt{1+\sigma_0^2}$ and $m^{-1/(2k)}(d_n - d_0)$ are $O(1)$. Thus, it suffices to show that $\sqrt{m}(\mu(d_n^{FD} + \eta/\nu_n) - \mu(d_n^{FD}))$ is bounded away from zero to justify (25). This can be shown in an identical manner as in the proof of Lemma 1.

Choose $\kappa > 0$ such that μ is non-decreasing in $(d_0, d_0 + \kappa)$. For sufficiently large n, $d_n^{FD} + \eta m^{-1/(2k)} + 1/n < d_0 + \kappa$ and hence,

$$\begin{aligned} M_n(d) - M_n(d_n^{FD}) &\geq M_n(d) - M_n(d_0 + \eta m^{-1/(2k)}) \\ &\geq \sum_{d_0 + \eta m^{-1/(2k)} \leq i/n \leq d} \left\{ \Phi_n \left(\sqrt{m}\mu(i/n) / \sqrt{1+\sigma_0^2} \right) - \gamma \right\} \\ &\geq K_0 (\lfloor nd \rfloor - \left\lfloor n(d_n^{FD} + \eta m^{-1/(2k)}) \right\rfloor)/n. \end{aligned} \tag{26}$$

Using (24) and (26), we get the result. □

References

Bambos, N. and Michailidis, G. (2004). Queueing and scheduling in random environments. *Adv. in Appl. Probab.*, 36(1):293–317.

Banerjee, M. and McKeague, I. W. (2007). Confidence sets for split points in decision trees. *Ann. Statist.*, 35(2):543–574.

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

Chen, Y.-I. and Chang, Y.-M. (2007). Identification of the minimum effective dose for right-censored survival data. *Comput. Statist. Data Anal.*, 51(6):3213–3222.

Cox, C. (1987). Threshold dose-response models in toxicology. *Biometrics*, 43(3):511–523.

Dümbgen, L. (1991). The asymptotic behavior of some nonparametric change-point estimators. *Ann. Statist.*, 19(3):1471–1495.

Gasser, T. and Müller, H.-G. (1984). Estimating regression functions and their derivatives by the kernel method. *Scand. J. Statist.*, 11(3):171–185.

Härdle, W. and Gasser, T. (1985). On robust kernel estimation of derivatives of regression functions. *Scand. J. Statist.*, 12(3):233–240.

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

Kim, J. and Pollard, D. (1990). Cube root asymptotics. *Ann. Statist.*, 18(1):191–219.

Korostelëv, A. P. (1987). Minimax estimation of a discontinuous signal. *Teor. Veroyatnost. i Primenen.*, 32(4):796–799.

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

Kosorok, M. R. (2008). *Introduction to empirical processes and semiparametric inference*. Springer Series in Statistics. Springer, New York.

Koul, H. L., Qian, L., and Surgailis, D. (2003). Asymptotics of M-estimators in two-phase linear regression models. *Stochastic Process. Appl.*, 103(1):123–154.

Lan, Y., Banerjee, M., and Michailidis, G. (2007). Change-point estimation under adaptive sampling. Technical report, http://www.stat.lsa.umich.edu/∼moulib/cp-techrep.pdf.

Lan, Y., Banerjee, M., and Michailidis, G. (2009). Change-point estimation under adaptive sampling. *Ann. Statist.*, 37(4):1752–1791.

Lehmann, E. L. and Romano, J. P. (2005). *Testing statistical hypotheses*. Springer Texts in Statistics. Springer, New York, third edition.

Lindvall, T. (1973). Weak convergence of probability measures and random functions in the function space $D(0, \infty)$. *J. Appl. Probability*, 10:109–121.

Loader, C. R. (1996). Change point estimation using nonparametric regression. *Ann. Statist.*, 24(4):1667–1678.

Mallik, A., Banerjee, M., and Sen, B. (2013). Asymptotics for p-value based threshold estimation in regression settings. submitted, available at *http://www.stat.lsa.umich.edu/∼moulib/th-estn-reg-v2.pdf*.

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(4):887–900.

Müller, H.-G. (1992). Change-points in nonparametric regression analysis. *Ann. Statist.*, 20(2):737–761.

Müller, H.-G. and Song, K.-S. (1997). Two-stage change-point estimators in smooth regression models. *Statist. Probab. Lett.*, 34(4):323–335.

Pons, O. (2003). Estimation in a Cox regression model with a change-point according to a threshold in a covariate. *Ann. Statist.*, 31(2):442–463. Dedicated to the memory of Herbert E. Robbins.

Tamhane, A. C. and Logan, B. R. (2002). Multiple test procedures for identifying the minimum effective and maximum safe doses of a drug. *J. Amer. Statist. Assoc.*, 97(457):293–301.

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.