Jump or Kink: Note on Super-efficiency in Segmented Linear Regression Break-point Estimation

By YINING CHEN

Department of Statistics, London School of Economics and Political Science
Houghton Street, London WC2A 2AE, U.K.
y.chen101@lse.ac.uk

SUMMARY

We consider the problem of segmented linear regression with a single break-point, with the focus on estimating of the location of the break-point. Let n be the sample size, we show that the global minimax convergence rate for this problem in terms of the mean absolute error is $O(n^{-1/3})$. On the other hand, we demonstrate the construction of a super-efficient estimator that achieves the pointwise convergence rate of either $O(n^{-1})$ or $O(n^{-1/2})$ for every fixed parameter value, depending on whether the structural change is a jump or a kink. The implications of this example and a potential remedy are discussed.

Key words: change-point; minimax rate; pointwise rate; structural break

1. INTRODUCTION

Asymptotic analysis is commonly used to facilitate comparison between different statistical estimators from a frequentist's perspective. Once the consistency of an estimator is established, the focus then naturally moves onto its rate of convergence. In general, statements concern the following two types of rates: the pointwise rate where the limit is taken when the unknown parameter is fixed, and the uniform rate where the limit is taken as the supremum over some or all of the parameter space. In addition, the convergence rate of the estimator that achieves the fastest uniform rate among all the estimators is known as the minimax rate. Often the (global) minimax rate is used to characterise the hardness of the problem.

In many settings, the pointwise rate, the uniform rate and the minimax rate are the same, in which case the corresponding estimator is usually regarded as rate-optimal. However, there are exceptions where caution must be exercised. A notable example arises from the phenomenon of super-efficiency, first documented by Joseph L. Hodges, Jr. in 1951. This topic was later treated comprehensively by Le Cam (1953) and Hájek (1972), among many others, in the settings of regular parametric models. See Stigler (2007) and Vovk (2009) for excellent reviews of the turbulent history of early studies. More recently, super-efficiency has also been investigated in other more complicated settings. For instance, Brown et al. (1997) studied it in nonparametric function estimation, Heinrich & Kahn (2018) studied it in mixture models, and Banerjee et al. (2019) studied it in the setting of isotonic regression.

Mathematically, let's denote the parameter space of interest by Θ, any estimator of $\theta \in \Theta$ by $\widehat{\theta}$, the loss function by $L(\theta, \widehat{\theta})$, and the corresponding risk function by $R(\theta, \widehat{\theta}) = E_\theta L(\theta, \widehat{\theta})$. For every $\theta \in \Theta$, suppose that there exists some $\gamma_\theta > 0$ such that

$$0 < \liminf_{\epsilon \to 0+} \liminf_{n \to \infty} \inf_{\widehat{\theta}} \sup_{\|\theta' - \theta\| \le \epsilon} n^{\gamma_\theta} R(\theta', \widehat{\theta}) \le \limsup_{\epsilon \to 0+} \limsup_{n \to \infty} \inf_{\widehat{\theta}} \sup_{\|\theta' - \theta\| \le \epsilon} n^{\gamma_\theta} R(\theta', \widehat{\theta}) < \infty,$$

then $n^{-\gamma_\theta}$ is known as the local minimax rate at θ. In this context, an estimator $\widehat{\theta}$ is super-efficient in its convergence rate if

$$\limsup_{n\to\infty} n^{\gamma_\theta}\ R(\theta,\widehat{\theta}) < \infty \ \text{ for every } \theta\in\Theta \quad \text{ and } \quad \limsup_{n\to\infty} n^{\gamma_\theta}\ R(\theta,\widehat{\theta}) = 0 \ \text{ for some } \theta\in\Theta.$$

The purpose of this short note is to demonstrate that super-efficiency can occur in the setting of segmented linear regression, even with only a single break-point. In spite of the popularity of segmented regression in statistics and econometrics literatures, to our knowledge, this phenomenon has not been widely understood in these contexts. In particular, since the class of segmented linear regression models is not regular (e.g. not differentiable in quadratic mean), existing results regarding super-efficiency in regular parametric models cannot be immediately applied. By focusing on estimating the location of the single break-point and taking the loss function to be the Euclidean distance between the true location and estimated location of the break-point, we show that the global minimax convergence rate of the risk is at least $O(n^{-1/3})$, i.e.

$$\liminf_{n\to\infty}\ \inf_{\widehat{\theta}} \sup_{\theta\in\Theta} n^{1/3}\ R(\theta,\widehat{\theta}) > 0.$$

We then illustrate super-efficiency by constructing an estimator $\widehat{\theta}^S$ that for every fixed $\theta\in\Theta$, depending on whether the break-point is a jump or a kink, satisfies either

$$n\, R(\theta,\widehat{\theta}^S) < \infty \quad \text{ or } \quad n^{1/2}\, R(\theta,\widehat{\theta}^S) < \infty.$$

These findings point to an interesting and dramatic scenario in which the break-point can be estimated at the rate of $O_p(n^{-1/3})$ in the minimax sense. However, as long as we are willing to assume that the truth does exist (i.e. the location, as a proportion of the data sequence, and the size of change do not vary with n), the break-point can then be estimated at a much faster rate of either $O_p(n^{-1})$ or $O_p(n^{-1/2})$. Consequently, for this particular break-point estimation problem, caution must be taken when one uses the global minimax rate to characterise its difficulty, and when one compares and interprets the convergence rates of different estimators.

Here our focus is on the segmented linear regression with a single break-point with the aim of illustrating super-efficiency. These results also hold in the setting of multiple break-points under suitable conditions. There has been a number of recent work on estimating the number and locations of unknown break-points in the settings of both continuous and discontinuous piecewise inear mean signals. See Bai and Perron (1998), Muggeo (2003), Das et al. (2016), Maidstone et al. (2019), Baranowski et al. (2019), to name but a few. In particular, with less stringent spacing conditions between consecutive break-points in the setting of a continuous piecewise linear mean signal, Maidstone et al. (2019) and Baranowski et al. (2019) proposed estimators that could achieve within a logarithmic factor of $O_p(n^{-1/3})$ in estimating the locations of all unknown break-points. The estimator's convergence rate was further improved to within a logarithmic factor of $O_p(n^{-1/2})$ in Baranowski et al. (2019) under more restrictive assumptions. See also Hansen (2017) for inference in the presence of a kink, Hidalgo et al. (2019) for a test of continuity at the break-point, and Dong (2018) for a related problem on treatment effect evaluation. There have also been work on kink location estimation in various univariate nonparametric regression settings. See Raimondo (1998), Goldenshluger et al. (2006), Cheng and Raimondo (2008), Wishart & Kulik (2010), Wishart (2011) and references therein. Finally, we mention the work of Korostelev & Lepski (2008) who investigated a version of jump location estimation problem with a growing dimension.

The rest of the manuscript is organised as follows. We formulate the break-point estimation problem mathematically in Section 2. The corresponding minimax rates are given in Section 3. Section 4 gives a super-efficient estimator, followed by a numerical experiment in Section 5. We discuss its implications and a potential remedy in Section 6. All proofs are deferred to the appendix.

2. MODEL SETUP: SEGMENTED LINEAR REGRESSION WITH A SINGLE BREAK-POINT

Suppose that we observe (X_{ni}, Y_{ni}) for $i = 1, \dots, n$. Consider the fixed design setting where

$$X_{ni} = i/(n+1)$$
$$Y_{ni} = f_\theta(X_{ni}) + \sigma\varepsilon_{ni}$$

for some $\sigma > 0$ and some function $f_\theta : [0,1] \to \mathbb{R}$. Here $\varepsilon_{n1}, \dots, \varepsilon_{nn}$ are independent and identically distributed $N(0,1)$ random variables. Furthermore, f_θ is a piecewise linear function indexed by $\theta = (\tau_\theta, \alpha_\theta^-, \alpha_\theta^+, \beta_\theta^-, \beta_\theta^+) \in \Theta \subset [0,1] \times \mathbb{R}^4$ of the form

$$f_\theta(x) = \begin{cases} \alpha_\theta^- + \beta_\theta^-(x - \tau_\theta) & \text{if } x \in [0, \tau_\theta] \\ \alpha_\theta^+ + \beta_\theta^+(x - \tau_\theta) & \text{if } x \in (\tau_\theta, 1] \end{cases}.$$

In other words, f_θ has a single break-point at τ_θ, with its linear part over $[0, \tau_\theta)$ determined by the slope β_θ^- and the intercept α_θ^- at $(\tau_\theta)-$, and its linear part over $(\tau_\theta, 1]$ determined by the slope β_θ^+ and the intercept α_θ^+ at $(\tau_\theta)+$. For simplicity, we have assumed that f_θ is left-continuous, so $f_\theta(\tau_\theta) = \alpha_\theta^-$. If $|\alpha_\theta^+ - \alpha_\theta^-| \neq 0$, then we refer to τ_θ as a jump. Otherwise, if $|\alpha_\theta^+ - \alpha_\theta^-| = 0$ but $|\beta_\theta^+ - \beta_\theta^-| \neq 0$, then we call τ_θ a kink.

To asymptotically analyse the break-point estimator based on $(X_{n1}, Y_{n1}), \dots (X_{nn}, Y_{nn})$, it is common to assume that the actual break-point does not occur too close to the boundary at $x = 0$ or $x = 1$, and the structural change is "noticeable", so at least one of the following two quantities, $|\alpha_\theta^+ - \alpha_\theta^-|$ and $|\beta_\theta^+ - \beta_\theta^-|$, is reasonably large. As such, it is natural to restrict ourselves to the parameter space of

$$\Theta = \left\{ \theta \in [0,1] \times \mathbb{R}^4 \;\middle|\; \tau_\theta \in [\delta, 1-\delta],\ \max\left(|\alpha_\theta^+ - \alpha_\theta^-|, |\beta_\theta^+ - \beta_\theta^-| \right) \geq \delta \right\}$$

for some fixed but perhaps unknown small $\delta > 0$. Here the dependence of Θ on δ is suppressed.

As mentioned previously, our main focus is on estimating the location of the break-point. In a sense, we treat $\alpha_\theta^-, \alpha_\theta^+, \beta_\theta^-$ and β_θ^+ as nuisance parameters. For any estimator $\widehat{\theta} = (\tau_{\widehat{\theta}}, \alpha_{\widehat{\theta}}^-, \alpha_{\widehat{\theta}}^+, \beta_{\widehat{\theta}}^-, \beta_{\widehat{\theta}}^+)$, we evaluate its performance based on the estimated break-point's absolute loss, namely, with the loss function $L(\theta, \widehat{\theta}) = |\tau_{\widehat{\theta}} - \tau_\theta|$ and the risk function $R(\theta, \widehat{\theta}) = E_\theta L(\theta, \widehat{\theta})$. Analogous conclusions could also be made under other losses, such as $L(\theta, \widehat{\theta}) = |\tau_{\widehat{\theta}} - \tau_\theta|^q$ for some $q > 1$.

Finally, we remark that this particular fix design is selected with the aim to better connect to the existing change-point detection literature.

3. MINIMAX RATE OF CONVERGENCE

First, we investigate the local minimax rate of convergence. We separate the parameter space into two disjoint sets Θ^K and $\Theta \backslash \Theta^K$, where Θ^K is the parameter space representing functions with a kink, i.e.

$$\Theta^K = \left\{ \theta \in \Theta \middle| \; \alpha_\theta^- = \alpha_\theta^+ \right\}.$$

THEOREM 1. *Under the setup mentioned in Section 2,*

$$\liminf_{\epsilon \to 0+} \liminf_{n \to \infty} \inf_{\widehat{\theta}} \sup_{\theta' \in \Theta : \|\theta' - \theta\| \leq \epsilon} n\, R(\theta', \widehat{\theta}) > 0 \quad \textit{for every } \theta \in \Theta \backslash \Theta^K$$

and

$$\liminf_{\epsilon \to 0+} \liminf_{n \to \infty} \inf_{\widehat{\theta}} \sup_{\theta' \in \Theta : \|\theta' - \theta\| \leq \epsilon} n^{1/3}\, R(\theta', \widehat{\theta}) > 0 \quad \textit{for every } \theta \in \Theta^K. \tag{1}$$

Theorem 1 implies that when τ_θ is a jump, the local minimax rate for estimating the location of the break-point in terms of the magnitude of $\tau_{\widehat{\theta}} - \tau_\theta$ is at least of $O_p(n^{-1})$. However, this rate slows down considerably to $O_p(n^{-1/3})$ when τ_θ is a kink.

The next corollary concerns the global minimax rate, which immediately follows from Theorem 1.

COROLLARY 1. *Under the setup mentioned in Section* 2,

$$\liminf_{n\to\infty} \inf_{\widehat{\theta}} \sup_{\theta\in\Theta} n^{1/3}\ R(\theta,\widehat{\theta}) > 0.$$

Although it is known that when the exact type of the break-point is given a priori, a jump could be estimated at $O_p(n^{-1})$ and a kink could be estimated at $O_p(n^{-1/2})$, we emphasize that these facts alone are far from implying the minimax rate of $O_p(n^{-1/3})$ for $\tau_{\widehat{\theta}} - \tau_\theta$ as shown above. It is also interesting to note that the $O_p(n^{-1/3})$ rate also appears in Raimondo (1998) who considers a related problem in the nonparametric setting where there is a jump in the first derivative of a continuous mean. However, the class of functions he considered in deriving this rate is different from ours.

On the other hand, if we further constrain the parameter space from Θ to Θ^K, then the following minimax result for kink location estimation in the setting of continuous segmented linear regression holds.

THEOREM 2. *Under the setup mentioned in Section* 2,

$$\liminf_{n\to\infty} \inf_{\widehat{\theta}} \sup_{\theta\in\Theta^K} n^{1/2}\ R(\theta,\widehat{\theta}) > 0. \tag{2}$$

Note that $O_p(n^{-1/2})$ implied by (2) is faster than $O_p(n^{-1/3})$ implied by (1). This seemingly counter-intuitive difference in the rates is due to the different choices of the parameter spaces in their derivations and can be explained by examining the proofs of Theorem 1 and Theorem 2 in the online supplementary materials. Our proofs follow from Le Cam's two-point method. See Le Cam (1986) or Yu (1997).

To give some intuitions, we first confine ourselves to $\theta_1 = (1/2, 0, 0, -1, 1)$ and $\theta_2 = (1/2 + \Delta, -\Delta, \Delta, -1, 1)$ for some small $\Delta > 0$, with $\theta_1, \theta_2 \in \Theta$. Denote the distribution of $(Y_{n1}, \ldots, Y_{nn})$ using the data generating process described in Section 2 with θ_1 as $\mathcal{P}^n_{\theta_1}$, and that with θ_2 as $\mathcal{P}^n_{\theta_2}$. Then break-point estimation could be viewed as the problem of differentiating between $\mathcal{P}^n_{\theta_1}$ and $\mathcal{P}^n_{\theta_2}$ based on the observations, whose hardness is dictated by the squared total variation distance between them. In the meantime, this squared total variation distance, denoted by $\|\mathcal{P}^n_{\theta_1} - \mathcal{P}^n_{\theta_2}\|^2_{\mathrm{TV}}$, can be bounded under suitable conditions as follows, with $\int_0^1 \left\{f_{\theta_1}(x) - f_{\theta_2}(x)\right\}^2 dx$ playing a crucial role in characterising the hardness of the original problem:

$$\begin{aligned}
\|\mathcal{P}^n_{\theta_1} - \mathcal{P}^n_{\theta_2}\|^2_{\mathrm{TV}} &\le 2 - 2\prod_{i=1}^n \left[1 - d^2_{\mathrm{hel}}\Big\{N\Big(f_{\theta_1}\big(i/(n+1)\big), \sigma^2\Big), N\Big(f_{\theta_2}\big(i/(n+1)\big), \sigma^2\Big)\Big\}\right] \\
&= 2 - 2\exp\left[-\frac{1}{8\sigma^2}\sum_{i=1}^n \Big\{f_{\theta_1}\big(i/(n+1)\big) - f_{\theta_2}\big(i/(n+1)\big)\Big\}^2\right] \\
&\to 2 - 2\exp\left[-\frac{n}{8\sigma^2}\int_0^1 \left\{f_{\theta_1}(x) - f_{\theta_2}(x)\right\}^2 dx\right].
\end{aligned}$$

Here d_{hel} is the Hellinger distance and $d^2_{\mathrm{hel}}\big\{N(\mu_1,\sigma), N(\mu_2,\sigma)\big\} = 1 - \exp\{-(\mu_1-\mu_2)^2/(8\sigma^2)\}$. For this particular pair of θ_1 and θ_2, $\int_0^1 \left\{f_{\theta_1}(x) - f_{\theta_2}(x)\right\}^2 dx = O(\Delta^3)$ for $\Delta \to 0$. See Fig. 1(a). Here $f_{\theta_1}(x)$ and $f_{\theta_2}(x)$ only differ over $x \in (1/2, 1/2+\Delta]$, meaning that the problem can be viewed as a *local* one as most pairs of the observations, namely, (X_{ni}, Y_{ni}) with $X_{ni} \notin (1/2, 1/2+\Delta]$, are irrelevant. In contrast, with the same value of θ_1, if we only consider parameters in Θ^K we are then unable to find a $\theta_2 \in \Theta^K$ with the corresponding break-point at $1/2+\Delta$ such that $f_{\theta_1}(x)$ and $f_{\theta_2}(x)$ only differ over a small neighbourhood. In fact, $f_{\theta_1}(x)$ and $f_{\theta_2}(x)$ will have to differ over a substantial interval that does not shrink as $\Delta \to 0$, so the problem of distinguishing between f_{θ_1} and f_{θ_2} appears more *global* than before. By taking, for example, $\theta_2 = (1/2+\Delta, -\Delta, -\Delta, -1, 1)$ (while keeping the same θ_1), we obtain $\int_0^1 \left\{f_{\theta_1}(x) - f_{\theta_2}(x)\right\}^2 dx = O(\Delta^2)$, as demonstrated in Fig. 1(b). In fact, we can further show that $\inf_{\theta_2\in\Theta^K} \int_0^1 \left\{f_{\theta_1}(x) - f_{\theta_2}(x)\right\}^2 dx = O(\Delta^2)$. This order difference of $\int_0^1 \left\{f_{\theta_1}(x) - f_{\theta_2}(x)\right\}^2 dx$

in Δ implies that the cases of $\theta \in \Theta$ and $\theta \in \Theta^K$ are fundamentally different! To give more details, in Le Cam's method, the minimax rate can be derived by picking Δ such that $\|\mathcal{P}_{\theta_1}^n - \mathcal{P}_{\theta_2}^n\|_{\mathrm{TV}} \leq C$ for some constant $C < 1$. In our settings, as derived as above, this roughly amounts to requiring $\int_0^1 \{f_{\theta_1}(x) - f_{\theta_2}(x)\}^2 dx = O(n^{-1})$. As such, Δ would be taken as $O(n^{-1/3})$ in Fig. 1(a), and $O(n^{-1/2})$ in Fig. 1(b), which are also the minimax convergence rates for break-point estimation under $\theta \in \Theta$ and $\theta \in \Theta^K$ respectively in terms of the expected absolute loss. Finally, for completeness, we also illustrate the case of a "noticeable" jump in Fig. 1(c), where $\theta_1 = (1/2, -1/2, 1/2, -1, 1)$ (i.e. with jump size of 1) and $\theta_2 = (1/2 + \Delta, -1/2 - \Delta, 1/2 + \Delta, -1, 1)$. Since here $\int_0^1 \{f_{\theta_1}(x) - f_{\theta_2}(x)\}^2 dx = O(\Delta)$, the minimax rate for estimating break-point of this type is $O(n^{-1})$.

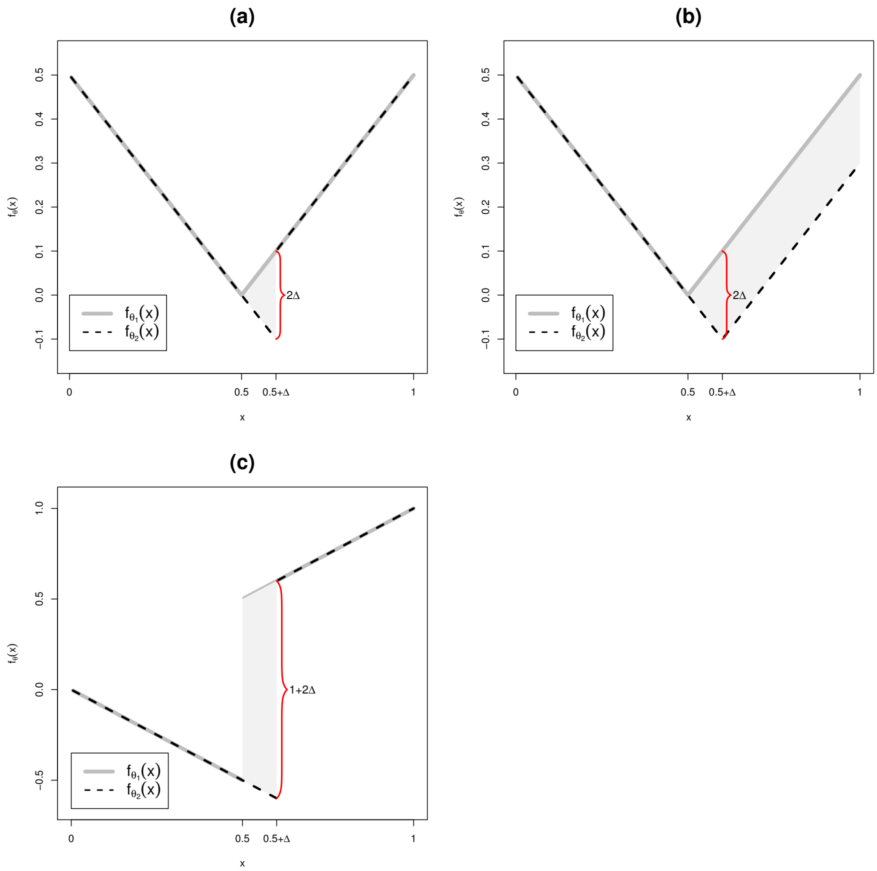

Fig. 1. Plots of f_{θ_1} and f_{θ_2} with their difference shaded in light grey. In (a), $\theta_1 = (1/2, 0, 0, -1, 1)$ and $\theta_2 = (1/2 + \Delta, -\Delta, \Delta, -1, 1)$. In (b), the continuity constraint is enforced with the same θ_1 but $\theta_2 = (1/2 + \Delta, -\Delta, -\Delta, -1, 1)$. Finally, (c) demonstrates the case of a non-vanishing jump with $\theta_1 = (1/2, -1/2, 1/2, -1, 1)$ and $\theta_2 = (1/2 + \Delta, -1/2 - \Delta, 1/2 + \Delta, -1, 1)$. In all the plots, the difference between f_{θ_1} and f_{θ_2} at $x = 1/2 + \Delta$ is highlighted using a curly bracket.

4. A SUPER-EFFICIENT ESTIMATOR

Write $\bar{\Theta} = [0,1] \times \mathbb{R}^4$ and $\bar{\Theta}^K = \{\theta \in \bar{\Theta} | \ \alpha_\theta^- = \alpha_\theta^+\}$. For notational convenience, in Section 4 and Section 5, $\widehat{\theta}$ is denoted as the least squares estimator satisfying

$$\widehat{\theta} := \widehat{\theta}_{LS} \in \operatorname{argmin}_{\theta \in \bar{\Theta}} \sum_{i=1}^{n} \{Y_{ni} - f_\theta(X_{ni})\}^2. \tag{3}$$

Here we minimise over $\bar{\Theta}$ instead of Θ, because δ of Θ is not always known a priori. Equivalently, we can write

$$\tau_{\widehat{\theta}} \in \operatorname{argmin}_{\tau \in [0,1]} \left\{ \min_{\alpha,\beta} \sum_{i=1}^{n} (Y_{ni} - \beta X_{ni} - \alpha)^2 \mathbf{1}_{\{X_{ni} \in [0,\tau]\}} + \min_{\alpha,\beta} \sum_{i=1}^{n} (Y_{ni} - \beta X_{ni} - \alpha)^2 \mathbf{1}_{\{X_{ni} \in (\tau,1]\}} \right\}.$$

Similarly, for kink estimation where we further restrict ourselves to Θ^K, we denote the corresponding least squares estimator by $\widehat{\theta}^K \in \operatorname{argmin}_{\theta \in \bar{\Theta}^K} \sum_{i=1}^n \{Y_{ni} - f_\theta(X_{ni})\}^2$, with

$$\tau_{\widehat{\theta}^K} \in \operatorname{argmin}_{\tau \in [0,1]} \left[\min_{\alpha,\beta^-,\beta^+} \sum_{i=1}^{n} \left\{ Y_{ni} - \alpha - \beta^-(X_{ni} - \tau)\mathbf{1}_{\{X_{ni} \in [0,\tau]\}} - \beta^+(X_{ni} - \tau)\mathbf{1}_{\{X_{ni} \in (\tau,1]\}} \right\}^2 \right].$$

For uniqueness, we take $\tau_{\widehat{\theta}}$ and $\tau_{\widehat{\theta}^K}$ to be the smallest element in the respective sets of minima.

As we shall see, $\tau_{\widehat{\theta}}$ achieves $O_p(n^{-1})$ for fixed $\theta \in \Theta \backslash \Theta^K$, but slows down to $O_p(n^{-1/3})$ when $\theta \in \Theta^K$. Meanwhile, $\tau_{\widehat{\theta}^K}$ achieves $O_p(n^{-1/2})$ for $\theta \in \Theta^K$. Therefore, making the correct extra assumption of continuity at the break-point and using the corresponding estimator could improve the convergence rate from $O_p(n^{-1/3})$ to $O_p(n^{-1/2})$ for $\theta \in \Theta^K$. This motivates us to shrink $\widehat{\theta}$ towards Θ^K in certain cases to improve the pointwise rate. In particular, we could estimate the break-point by $\tau_{\widehat{\theta}^S}$, where

$$\widehat{\theta}^S = \begin{cases} \widehat{\theta}^K & \text{if } |\alpha_{\widehat{\theta}}^+ - \alpha_{\widehat{\theta}}^-| \le n^{-1/6} \\ \widehat{\theta} & \text{if } |\alpha_{\widehat{\theta}}^+ - \alpha_{\widehat{\theta}}^-| > n^{-1/6} \end{cases}. \tag{4}$$

We are now in the position to discuss the pointwise and local uniform convergence rates of $\tau_{\widehat{\theta}^S}$.

THEOREM 3. *Under the setup mentioned in Section 2, $\tau_{\widehat{\theta}^S}$ is a super-efficient estimator for τ_θ. In particular, we have*

$$\limsup_{n\to\infty} \ nR(\theta, \widehat{\theta}^S) < \infty \quad \textit{for every } \theta \in \Theta\backslash\Theta^K$$

and

$$\limsup_{n\to\infty} \ n^{1/2}R(\theta, \widehat{\theta}^S) < \infty \quad \textit{for every } \theta \in \Theta^K.$$

It follows from Theorem 3 that $\sup_{\theta\in\Theta} \limsup_{n\to\infty} \ n^{1/2}R(\theta, \widehat{\theta}^S) < \infty$, i.e. the pointwise rate of break-point estimation via $\widehat{\theta}^S$ for every $\theta \in \Theta$ is faster than the global minimax rate of $O(n^{-1/3})$.

THEOREM 4. *Under the setup mentioned in Section 2, we have that*

$$\limsup_{\epsilon\to 0+} \limsup_{n\to\infty} \sup_{\theta'\in\Theta:\|\theta'-\theta\|\le\epsilon} nR(\theta', \widehat{\theta}^S) < \infty \quad \textit{for every } \theta \in \Theta\backslash\Theta^K$$

and

$$\liminf_{\epsilon\to 0+} \liminf_{n\to\infty} \sup_{\theta'\in\Theta:\|\theta'-\theta\|\le\epsilon} n^{1/3}R(\theta', \widehat{\theta}^S) = \infty \quad \textit{for every } \theta \in \Theta^K.$$

Theorem 4 implies that the global uniform rate of $\tau_{\widehat{\theta}^S}$ in terms of the absolute loss is worse than the global minimax rate of $O(n^{-1/3})$, which could actually be achieved by $\tau_{\widehat{\theta}}$. This type of behaviour is typical for super-efficient estimators that tend to achieve better pointwise convergence rates at the cost of worse uniform convergence rates.

In addition, the construction of the super-efficient estimator is by no means unique. In fact, here the threshold $n^{-1/6}$ in (4) can be replaced by $cn^{-\gamma}$ for any fixed $c > 0$ and $\gamma \in (0, 1/3)$. Alternatively, one could replace $|\alpha^+_{\widehat{\theta}} - \alpha^-_{\widehat{\theta}}|$ in (4) by the difference between residual sum of squares from fitting the model over either $\bar{\Theta}^K$ or $\bar{\Theta}$, and then choose the cut-off decision boundary accordingly.

5. Numerical experiment

We run a small simulation study to compare the behaviour of $\tau_{\widehat{\theta}}$ and $\tau_{\widehat{\theta}^S}$. Two different scenarios are considered under the settings of Section 2:

(a) $\theta_1 = (0.5, 0, 0, -1, 1) \in \Theta^K$, i.e. $f_{\theta_1}(x) = |x - 0.5|$;
(b) $\theta_2 = (0.5, 0, 0.5, -1, 1) \in \Theta \backslash \Theta^K$, i.e. $f_{\theta_2}(x) = |x - 0.5| + \mathbf{1}_{\{x>0.5\}}$.

Here we take $\sigma = 0.5$ and $n = 100, 200, 500, 1000, 2000$. All experiments are repeated 1000 times. The estimated values of $R(\theta, \widehat{\theta})$ and $R(\theta, \widehat{\theta}^S)$, also known as the mean absolute errors (MAEs) of $\tau_{\widehat{\theta}}$ and $\tau_{\widehat{\theta}^S}$, are reported in Fig. 2 on a log-log scale.

In Fig. 2(a), the super-efficiency phenomenon is visible, where the super-efficient estimator $\tau_{\widehat{\theta}^S}$ performs better than the least squares estimator $\tau_{\widehat{\theta}}$ in the presence of a kink, especially for large n. It is also evident from the plot that $\tau_{\widehat{\theta}^S}$ and $\tau_{\widehat{\theta}}$ have different pointwise convergence rates there, as indicated in Section 4. Meanwhile, Fig. 2(b) demonstrates that in the presence of a jump, $\tau_{\widehat{\theta}^S}$ and $\tau_{\widehat{\theta}}$ perform similarly. In particular, for large $n = 2000$, they are exactly the same in all the 1000 runs. Finally, Fig. 2 confirms that in terms of pointwise rates, estimating the location of a jump is easier than estimating that of a kink.

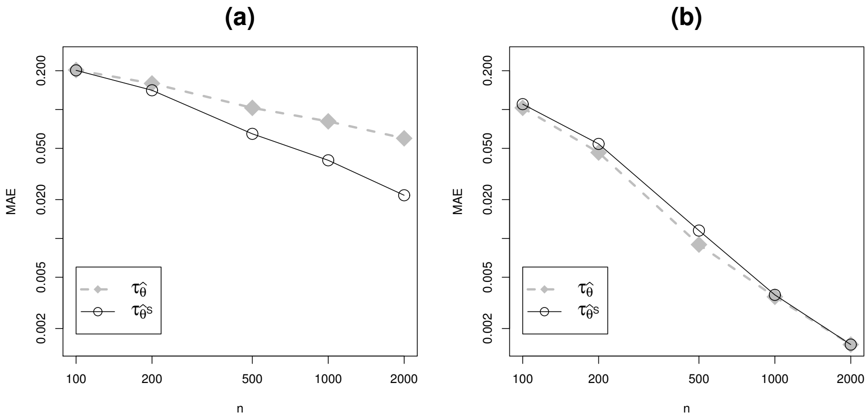

Fig. 2. Estimated MAEs of $\tau_{\widehat{\theta}}$ and $\tau_{\widehat{\theta}^S}$ for $n = 100, 200, 500, 1000, 2000$ on a log-log scale under different scenarios: (a) θ_1 with a kink; (b) θ_2 with a jump.

6. Discussion

Although the super-efficient estimator $\tau_{\widehat{\theta}^S}$ achieves a pointwise rate faster than the global minimax rate for every $\theta \in \Theta$, it is clear that in our example super-efficiency only occurs over Θ^K, which is a Lebesgue null set in comparison to Θ. However, if we were to focus solely on τ_θ by treating $\alpha^-_\theta, \alpha^+_\theta, \beta^-_\theta$ and β^+_θ as nuisance parameters, then super-efficiency could occur at every $\tau_\theta \in [\delta, 1 - \delta]$. Here $[\delta, 1 - \delta]$ is no longer a Lebesgue null set in comparison to $[0, 1]$.

On the other hand, whilst $\tau_{\widehat{\theta}^S}$ achieves a better pointwise convergence rate for locating the break-point, like Hodges' estimator, it is penalised at local neighbouring points. In particular, $\limsup_{n\to\infty} \sup_{\theta\in\Theta} n^{1/3} R(\theta, \widehat{\theta}^S) = \infty$, but $\limsup_{n\to\infty} \sup_{\theta\in\Theta} n^{1/3} R(\theta, \widehat{\theta}) < \infty$, where $\widehat{\theta}$ is taken as the least squares estimator defined in (3). Thus, in terms of the uniform convergence rate, $\tau_{\widehat{\theta}^S}$ actually performs worse than $\tau_{\widehat{\theta}}$.

However, it is useful to think about whether this perspective of uniformity is what we are really interested in. In this break-point estimation problem, the global minimax convergence rate is derived by considering alternatives with the jump size $\to 0$ as $n \to \infty$. It is entirely possible that these alternatives might violate the modeller's real brief in practice, whose intention dictates that whenever there is a jump or a change in slope (or both), the corresponding change size has to be significant. Mathematically, this plausibly more appropriate parameter space would be a restricted version of Θ, given by

$$\Theta^* = \left\{ \theta \in \Theta \,\middle|\, \min\left(\mathbf{1}_{\{\alpha_\theta^- = \alpha_\theta^+\}} + \left|\alpha_\theta^+ - \alpha_\theta^-\right| \mathbf{1}_{\{\alpha_\theta^- \neq \alpha_\theta^+\}},\ \mathbf{1}_{\{\beta_\theta^- = \beta_\theta^+\}} + \left|\beta_\theta^+ - \beta_\theta^-\right| \mathbf{1}_{\{\beta_\theta^- \neq \beta_\theta^+\}} \right) \geq \delta \right\}.$$

It can then be shown that for every $\theta \in \Theta^K$,

$$\liminf_{\epsilon \to 0+} \liminf_{n \to \infty} \inf_{\widehat{\theta}} \sup_{\theta' \in \Theta^* : \|\theta' - \theta\| \leq \epsilon} n^{1/2}\, R(\theta', \widehat{\theta}) > 0,$$

implying that the local minimax convergence rate for estimating the kink over Θ^* is $O_p(n^{-1/2})$, and thus $\widehat{\theta}^S$ is no longer super-efficient over Θ^*.

Besides, one interesting feature of this break-point estimation problem is that the local minimax convergence rates are different across the parameter space. In this circumstance, it might not be entirely adequate to summarise the hardness of the problem by a single global minimax rate. See also Donoho et al. (1995).

Our example can be generalised in more complex settings, such as segmented polynomial regression with higher order polynomials, as well as those with heterogeneous sub-Gaussian errors and multiple break-points. It shows the importance of choosing suitable parameter spaces for the calculation of uniform or minimax convergence rates. We hope that it also demonstrates the need of interpreting different types of convergence rates and the practical meaning of rate-optimality with care and caution.

APPENDIX: PROOFS

Proof of Theorem 1

Proof. For any $\theta = (\tau_\theta, \alpha_\theta^-, \alpha_\theta^+, \beta_\theta^-, \beta_\theta^+) \in \Theta$, without loss of generality, we assume that $\tau_\theta \leq 1/2$. Now consider $\theta_\Delta = (\tau_\theta + \Delta, \alpha_\theta^- + \beta_\theta^- \Delta, \alpha_\theta^+ + \beta_\theta^+ \Delta, \beta_\theta^-, \beta_\theta^+)$. For sufficiently small $\Delta > 0$, $\theta_\Delta \in \Theta$. As will become clearer later, f_{θ_Δ} corresponding to this particular form of θ_Δ does not necessarily give the closest approximation to f_θ over the restricted parameter space while fixing the break-point location as $\tau_\theta + \Delta$. Nevertheless, using it does allow us to establish the correct minimax convergence rate, which is sufficient for our purpose here.

Now denote the distribution of $(Y_{n1}, \ldots, Y_{nn})$ using the data generating processes with θ and θ_Δ by $\mathcal{P}_\theta^n$ and $\mathcal{P}_{\theta_\Delta}^n$, respectively. Then the Total Variation (TV) distance between $\mathcal{P}_\theta^n$ and $\mathcal{P}_{\theta_\Delta}^n$ is bounded above by

$$\begin{aligned} \|\mathcal{P}_\theta^n - \mathcal{P}_{\theta_\Delta}^n\|_{\mathrm{TV}}^2 &\leq 2 - 2\prod_{i=1}^n \left[1 - d_{\mathrm{hel}}^2 \left\{ N\Big(f_\theta\big(i/(n+1)\big), \sigma^2\Big), N\Big(f_{\theta_\Delta}\big(i/(n+1)\big), \sigma^2\Big) \right\} \right] \\ &= 2 - 2\exp\left[-\frac{1}{8\sigma^2} \sum_{i: i/(n+1) \in (\tau_\theta, \tau_\theta + \Delta]}^n \Big\{ f_{\theta_\Delta}(i/(n+1)) - f_\theta(i/(n+1)) \Big\}^2 \right] \end{aligned}$$

where d_{hel} is the Hellinger distance, and recalling that $d_{\mathrm{hel}}^2 \Big\{ N(\mu_1, \sigma), N(\mu_2, \sigma) \Big\} = 1 - \exp\{-(\mu_1 - \mu_2)^2/(8\sigma^2)\}$. We also used the fact that the values of $f_\theta(x)$ and $f_{\theta_\Delta}(x)$ are only different over the interval $(\tau_\theta, \tau_\theta + \Delta]$.

If $\theta \in \Theta \backslash \Theta^K$ (i.e. with a jump), then for any small $\Delta > 0$, say with $\Delta < 0.5|\alpha_\theta^+ - \alpha_\theta^-|/|\beta_\theta^+ - \beta_\theta^-|$,

$$\|\mathcal{P}_\theta^n - \mathcal{P}_{\theta_\Delta}^n\|_{\mathrm{TV}}^2 \le 2 - 2\exp\left\{-\frac{1}{8\sigma^2}\left(\frac{\alpha_\theta^+ - \alpha_\theta^-}{2}\right)^2 \left\lfloor \frac{\Delta n}{2} \right\rfloor\right\}.$$

Since θ is fixed, so is $|\alpha_\theta^+ - \alpha_\theta^-|$, which is strictly positive because $\theta \in \Theta \backslash \Theta^K$. By taking $\Delta = cn^{-1}$ for sufficiently small $c > 0$, we have $\|\mathcal{P}_\theta^n - \mathcal{P}_{\theta_\Delta}^n\|_{\mathrm{TV}}^2 \le 1/4$. Now it follows from Le Cam's two-point method (cf. Le Cam (1986) or Yu (1997)) that for any $\epsilon > 0$ and large n,

$$\inf_{\widehat{\theta}} \sup_{\theta' \in \Theta : \|\theta' - \theta\| \le \epsilon} R(\theta', \widehat{\theta}) \ge \frac{\Delta}{2}\Big(1 - \|\mathcal{P}_\theta^n - \mathcal{P}_{\theta_\Delta}^n\|_{\mathrm{TV}}\Big) \ge \frac{c}{4} n^{-1}.$$

On the other hand, if $\theta \in \Theta^K$ (i.e. with a kink), then

$$\|\mathcal{P}_\theta^n - \mathcal{P}_{\theta_\Delta}^n\|_{\mathrm{TV}}^2 \le 2 - 2\exp\left\{-\frac{1}{8\sigma^2}\left(\frac{\Delta}{2}\right)^2 \left\lfloor \frac{\Delta n}{2} \right\rfloor (\beta_\theta^+ - \beta_\theta^-)^2\right\} \le 2 - 2\exp\left\{-\frac{\delta^2}{8\sigma^2}\left(\frac{\Delta}{2}\right)^2 \left\lfloor \frac{\Delta n}{2} \right\rfloor\right\}.$$

By taking $\Delta = cn^{-1/3}$ for sufficiently small $c > 0$, we still have $\|\mathcal{P}_\theta^n - \mathcal{P}_{\theta_\Delta}^n\|_{\mathrm{TV}}^2 \le 1/4$. Hence, Le Cam's two-point method yields

$$\inf_{\widehat{\theta}} \sup_{\theta' \in \Theta : \|\theta' - \theta\| \le \epsilon} R(\theta', \widehat{\theta}) \ge \frac{\Delta}{2}\Big(1 - \|\mathcal{P}_\theta^n - \mathcal{P}_{\theta_\Delta}^n\|_{\mathrm{TV}}\Big) \ge \frac{c}{4} n^{-1/3}.$$

The proof is completed by simple rearrangements of the terms. □

Proof of Theorem 2

Proof. This proof is similar to that of Theorem 1. Here only the differences are highlighted.

For any $\theta = (\tau_\theta, \alpha_\theta, \alpha_\theta, \beta_\theta^-, \beta_\theta^+) \in \Theta^K$, without loss of generality, we again assume that $\tau_\theta \le 1/2$ and consider $\theta_\Delta = (\tau_\theta + \Delta, \alpha_\theta + \beta_\theta^- \Delta, \alpha_\theta + \beta_\theta^- \Delta, \beta_\theta^-, \beta_\theta^+)$. Here $\theta_\Delta \in \Theta^K$ for small $\Delta > 0$. Furthermore,

$$\begin{aligned}\|\mathcal{P}_\theta^n - \mathcal{P}_{\theta_\Delta}^n\|_{\mathrm{TV}}^2 &\le 2 - 2\exp\left[-\frac{1}{8\sigma^2}\sum_{i : i/(n+1) \in [\tau_\theta + \Delta, 1]}^{n} \Big\{f_{\theta_\Delta}(i/(n+1)) - f_\theta(i/(n+1))\Big\}^2\right] \\ &\le 2 - 2\exp\left\{-\frac{1}{8\sigma^2}(\beta_\theta^+ - \beta_\theta^-)^2 \Delta^2 \big\lfloor (0.5 - \Delta)n \big\rfloor\right\} \le 2 - 2\exp\left(-\frac{\delta^2}{32\sigma^2}\Delta^2 n\right)\end{aligned}$$

for small Δ, say < 0.25. By taking $\Delta = cn^{-1/2}$ for sufficiently small $c > 0$, we have that $\|\mathcal{P}_\theta^n - \mathcal{P}_{\theta_\Delta}^n\|_{\mathrm{TV}}^2 \le 1/4$. The result then follows as previously. □

Proof of Theorem 3

Proof. We assume that θ is fixed, and shall use $C, c > 0$ as generic constants. For notational convenience, we now write $f_\theta = \Big(f_\theta(1/(n+1)), \ldots, f_\theta(n/(n+1))\Big)$ (where the dependence on n is supressed), let $\langle x, y \rangle = \sum_{i=1}^n x_i y_i$ for any $x = (x_1, \ldots, x_n)$ and $y = (y_1, \ldots, y_n) \in \mathbb{R}^n$, $\|x\|^2 = \langle x, x \rangle$ and $\|x\|_n^2 = \|x\|^2/n$. It follows from the definition of $\widehat{\theta} := \widehat{\theta}_{LS}$ (cf. Lemma 4.7 of van de Geer (2000)) that

$$\|f_{\widehat{\theta}} - f_\theta\| \le 2\sigma \left\langle \varepsilon_n, \frac{f_{\widehat{\theta}} - f_\theta}{\|f_{\widehat{\theta}} - f_\theta\|} \right\rangle \tag{5}$$

where $\varepsilon_n = (\varepsilon_{n1}, \ldots, \varepsilon_{nn})$. Here $f_{\widehat{\theta}} - f_\theta$ can be viewed as a piecewise linear vector with at most two break-points (i.e. three components). For any $s \in \{0, \ldots, n-1\}$ and $e \in \{1, \ldots, n\}$ satisfying $s < e$ or $s + 1 < e$, respectively write $\xi_{(s,e]} = \Big(\xi_{(s,e]}(1), \ldots, \xi_{(s,e]}(n)\Big)$ and $\gamma_{(s,e]} = \Big(\gamma_{(s,e]}(1), \ldots, \gamma_{(s,e]}(n)\Big)$, with

$$\xi_{(s,e]}(i)=\begin{cases}\frac{1}{\sqrt{e-s}} & i=s+1,\ldots,e\\ 0 & \text{otherwise}\end{cases},\qquad \gamma_{(s,e]}(i)=\begin{cases}\frac{i-(e+s+1)/2}{\sqrt{(e-s-1)(e-s)(e-s+1)/12}} & i=s+1,\ldots,e\\ 0 & \text{otherwise}\end{cases}.$$

Once defined, it is easy to check that $\xi_{(0,s]}$, $\xi_{(s,e]}$, $\xi_{(e,n]}$, $\gamma_{(0,s]}$, $\gamma_{(s,e]}$ and $\gamma_{(e,n]}$ are mutually orthonormal in the Euclidean space. Hence,

$$\left|\left\langle \varepsilon_n, \frac{f_{\widehat{\theta}}-f_\theta}{\|f_{\widehat{\theta}}-f_\theta\|}\right\rangle\right| \le \sup_{s,e}\Big(|\langle\varepsilon_n,\xi_{(0,s]}\rangle| + |\langle\varepsilon_n,\xi_{(s,e]}\rangle| + |\langle\varepsilon_n,\xi_{(e,n]}\rangle| \\ + |\langle\varepsilon_n,\gamma_{(0,s]}\rangle| + |\langle\varepsilon_n,\gamma_{(s,e]}\rangle| + |\langle\varepsilon_n,\gamma_{(e,n]}\rangle|\Big). \tag{6}$$

Each of the six terms on the right hand side of (6) follows $|N(0,1)|$ for any pre-given $0 < s-1 < e < n+1$; for other pairs of s and e, some of these terms will simply degenerate to zero. Using a standard union bound argument (cf. proof of Theorem 1 of Baranowski et al. (2019)), we obtain

$$P(B_n^c) \le \frac{\sqrt{3}}{\sqrt{\pi\log n}}n^{-1} < n^{-1}, \quad \text{where the event } B_n = \left\{2\sup_{\theta'}\left\langle\varepsilon_n, \frac{f_{\theta'}-f_\theta}{\|f_{\theta'}-f_\theta\|}\right\rangle \le 12\sqrt{6\log n}\right\}. \tag{7}$$

Plugging (7) back to (5), we have that when B_n holds, there exists some $C>0$ such that

$$\|f_{\widehat{\theta}}-f_\theta\|_n^2 \le C\log n/n. \tag{8}$$

For any $\theta\in\Theta\backslash\Theta^K$ (i.e. with a jump), $\|f_{\widehat{\theta}}-f_\theta\|_n^2\to 0$ as $n\to\infty$ implies consistency of $\widehat{\theta}$. Furthermore, due to the presence of the jump, there exists $C,c>0$ (independent of n, but might depend on θ) such that $\|f_{\theta'}-f_\theta\|_n^2 \ge C|\tau_{\theta'}-\tau_\theta|$ for any θ' with $|\tau_{\theta'}-\tau_\theta|>c/n$. Comparing this with (8), we have $|\tau_{\widehat{\theta}}-\tau_\theta|\le C\log n/n$. Similarly, we obtain $\big|(\alpha^+_{\widehat{\theta}}-\alpha^-_{\widehat{\theta}})-(\alpha^+_\theta-\alpha^-_\theta)\big|\le C\sqrt{\log n/n}$, which yields

$$\big|\alpha^+_{\widehat{\theta}}-\alpha^-_{\widehat{\theta}}\big| \ge \big|\alpha^+_\theta-\alpha^-_\theta\big| - C\sqrt{\log n/n} > n^{-1/6}$$

for sufficiently large n. This implies $\widehat{\theta}^S=\widehat{\theta}$ under B_n. Next, to fine-tune the convergence rate (i.e. getting rid of the factor of $\log n$), we shall resort to empirical process theory. Let Q_n be the empirical distribution of $\{1/(n+1),\ldots,n/(n+1)\}$ and let

$$\mathcal{F}^n_{\bar{\Theta}}(R) = \{f_{\theta'} \mid \theta'\in\bar{\Theta}, \|f_{\theta'}-f_\theta\|_n\le R, |\tau_{\theta'}-\tau_\theta|\le\epsilon'\}$$

for some fixed $\epsilon'\in\Big(0,\min(\tau_\theta,1-\tau_\theta)/2\Big)$. Then, for any $f_{\theta'}\in\mathcal{F}^n_{\bar{\Theta}}(R)$, we also have that

$$\sup_{x\in[0,1]}|f_{\theta'}(x)-f_\theta(x)| \le \sup_{x\in\{0,(\tau_{\theta'})\pm,(\tau_\theta)\pm,1\}}|f_{\theta'}(x)-f_\theta(x)| \le CR.$$

Thus, $\mathcal{F}^n_{\bar{\Theta}}(R)$ (for $L_2(Q_n)$-metric) can be bounded above as

$$H_2(2u,\mathcal{F}^n_{\bar{\Theta}}(R),Q_n) \le C\log\Big(\frac{CR+u}{u}\Big).$$

Here one could derive this by construction, using the fact that for every $f_{\theta'}\in\mathcal{F}^n_{\bar{\Theta}}(R)$, $f_{\theta'}-f_\theta$ is piecewise linear, bounded by CR, and with at most two break-points, as well as Corollary 2.6 of van de Geer (2000). Let $\tilde{\theta}$ be the least squares estimator defined like $\widehat{\theta}$ in (3) in Section 4 but being optimised over $[\tau_\theta-\epsilon',\tau_\theta+\epsilon']\times\mathbb{R}^4$ instead, then it follows from Theorem 9.1 and Example 9.3.1 of van de Geer (2000) that $E_\theta\big\|f_{\tilde{\theta}}-f_\theta\|_n^2\le Cn^{-1}$. Note that $\widehat{\theta}^S=\widehat{\theta}=\tilde{\theta}$ under B_n. Also recall that $\|f_{\theta'}-f_\theta\|_n^2\ge C|\tau_{\theta'}-\tau_\theta|$ for any θ' with $|\tau_{\theta'}-\tau_\theta|>cn^{-1}$. Simple manipulation entails

$$E_\theta\big(n|\tau_{\widehat{\theta}^S}-\tau_\theta|\big) = E_\theta\big(n|\tau_{\tilde{\theta}}-\tau_\theta|\big) - E_\theta\big(n|\tau_{\tilde{\theta}}-\tau_\theta| \mid B_n^c\big)P(B_n^c) + E_\theta\big(n|\tau_{\widehat{\theta}^S}-\tau_\theta| \mid B_n^c\big)P(B_n^c) \le C,$$

which completes our proof for the case of $\theta\in\Theta\backslash\Theta^K$.

Now for any $\theta \in \Theta^K$ (i.e. with a kink), there exists some $\epsilon' > 0$ such that $\|f_{\theta'} - f_\theta\|_n^2 \geq C|\tau_{\theta'} - \tau_\theta|^3$ for any $\theta' \in \bar{\Theta}$ with $|\tau_{\theta'} - \tau_\theta| \in (c/n, \epsilon']$. See for instance, Lemma 7 of Baranowski et al. (2019). Moreover, comparing that with (8) leads to $|\alpha^+_{\widehat{\theta}} - \alpha^-_{\widehat{\theta}}| = \big|(\alpha^+_{\widehat{\theta}} - \alpha^-_{\widehat{\theta}}) - (\alpha^+_\theta - \alpha^-_\theta)\big| \leq C(\log n/n)^{1/3}$. So under B_n, $\widehat{\theta}^S = \widehat{\theta}^K$ for sufficiently large n. Our next ingredient is the following statement: there exists some $\epsilon' > 0$ such that $\|f_{\theta'} - f_\theta\|_n^2 \geq C|\tau_{\theta'} - \tau_\theta|^2$ for any $\theta' \in \bar{\Theta}^K$ satisfying $|\tau_{\theta'} - \tau_\theta| \in (c/n, \epsilon']$. See for example, Lemma 8 of Baranowski et al. (2019). Since $\mathcal{F}^n_{\bar{\Theta}^K}(R) \subset \mathcal{F}^n_{\bar{\Theta}}(R)$, we could again use Theorem 9.1 and Example 9.3.1 of van de Geer (2000) to see that $E_\theta\big\|f_{\widehat{\theta}^K} - f_\theta\|_n^2 \leq Cn^{-1}$. Consequently,

$$E_\theta\big(\sqrt{n}|\tau_{\widehat{\theta}^K} - \tau_\theta|\big) \leq \sqrt{E_\theta\big(n|\tau_{\widehat{\theta}^K} - \tau_\theta|^2\big)} \leq C\sqrt{E_\theta\big(n\big\|f_{\widehat{\theta}^K} - f_\theta\|_n^2\big)} \leq C,$$

where $C > 0$ are generic constants. Finally, the proof is complete by noting that

$$\begin{aligned} E_\theta\big(\sqrt{n}|\tau_{\widehat{\theta}^S} - \tau_\theta|\big) &= E_\theta\big(\sqrt{n}|\tau_{\widehat{\theta}^K} - \tau_\theta|\big) - E_\theta\big(\sqrt{n}|\tau_{\widehat{\theta}^K} - \tau_\theta| \;\big|\; B_n^c\big)P(B_n^c) + E_\theta\big(\sqrt{n}|\tau_{\widehat{\theta}^S} - \tau_\theta| \;\big|\; B_n^c\big)P(B_n^c) \\ &\leq C + \sqrt{n}(n^{-1}) + \sqrt{n}(n^{-1}) \leq C. \end{aligned}$$

Ancillary Lemmas

Some ancillary results are required for the proof of Theorem 4. They concern the approximation of f_{θ_Δ} for $\theta_\Delta \in \Theta\backslash\Theta^K$ using $f_{\theta'}$ with $\theta' \in \Theta^K$, and the distance between θ_Δ and θ' when $\|f_{\theta_\Delta} - f_{\theta'}\|_n$ is sufficiently small. In particular, for the reason that will become clear later in the proof of Theorem 4, in the remaining we shall focus solely on $\theta_\Delta = (1/2 + \Delta, -\Delta, \Delta, -1, 1)$ with $\Delta = n^{-\gamma}$ for some $\gamma \in (0, 1/3)$. All the lemmas presented below can be easily modified to handle more general cases.

LEMMA 1. *Let* $\theta_\Delta = (1/2 + \Delta, -\Delta, \Delta, -1, 1)$ *with* $\Delta = n^{-\gamma}$ *for some* $\gamma \in (0, 1/3)$. *In addition, let* $\Delta' = n^{-\gamma'}$ *with* $0 < \gamma < \gamma' < 1/3$ *and write* $\Theta^K_* = \Big\{\theta \mid \theta \in \bar{\Theta}^K, \tau_\theta \in [1/2 + \Delta - \Delta', 1/2 + \Delta + \Delta']\Big\}$. *Then there exists* $c > 0$ *such that*

$$\inf_{\theta' \in \Theta^K_*} \|f_{\theta'} - f_{\theta_\Delta}\|_n^2 \geq cn^{-2\gamma}.$$

for every sufficiently large n.

Proof. This proof can be divided into three parts.

First, we claim that $\inf_{\theta' \in \Theta^K_*} \|f_{\theta'} - f_{\theta_\Delta}\|_n^2 \to 0$ as $n \to \infty$. This is due to the fact that

$$\inf_{\theta' \in \Theta^K_*} \|f_{\theta'} - f_{\theta_\Delta}\|_n^2 \leq \|f_{\theta'_\Delta} - f_{\theta_\Delta}\|_n^2 \leq \Delta^2 = n^{-2\gamma} \to 0,$$

as $n \to \infty$, where we have taken $\theta'_\Delta = (1/2 + \Delta, 0, 0, -1, 1)$.

Second, let $B_\epsilon(\theta)$ denote the (closed) $\epsilon-$Ball around $\theta = (1/2, 0, 0, -1, 1)$, since

$$\liminf_{n \to \infty} \inf_{\theta' \in \Theta \backslash B_\epsilon(\theta)} \|f_{\theta'} - f_{\theta_\Delta}\|_n^2 > 0$$

for any fixed $\epsilon > 0$, it now follows that for any sufficiently large n,

$$\inf_{\theta' \in \Theta^K_*} \|f_{\theta'} - f_{\theta_\Delta}\|_n^2 = \inf_{\theta' \in \Theta^K_* \cap B_\epsilon(\theta)} \|f_{\theta'} - f_{\theta_\Delta}\|_n^2.$$

This also implies that

$$\Big|f_{\theta'}(1/2 + \Delta - \Delta') - f_{\theta'}(1/2 + \Delta + \Delta')\Big| \leq 2(1 + \epsilon)\Delta'.$$

for any $\theta' \in \Theta^K_* \cap B_\epsilon(\theta)$. Thanks to the fact that

$$\Big|f_{\theta_\Delta}(1/2 + \Delta - \Delta') - f_{\theta_\Delta}(1/2 + \Delta + \Delta')\Big| = 2\Delta,$$

we have that

$$\inf_{\theta' \in \Theta_*^K \cap B_\epsilon(\theta)} \max\left\{\left|f_{\theta_\triangle}(1/2+\Delta-\Delta') - f_{\theta'}(1/2+\Delta-\Delta')\right|, \left|f_{\theta_\triangle}(1/2+\Delta+\Delta') - f_{\theta'}(1/2+\Delta+\Delta')\right|\right\}$$
$$\geq \{2\Delta - 2(1+\epsilon)\Delta'\}/2 \geq \Delta/2$$

for any sufficiently large n.

Finally, note that for any $\theta' \in \Theta_*^K \cap B_\epsilon(\theta)$, both $f_{\theta'}$ and $f_{\theta_\triangle}$ are linear over $[0, 1/2+\Delta-\Delta']$ and $[1/2+\Delta+\Delta', 1]$, with both segments having non-vanishing width (i.e. $> 1/4$ for sufficiently large n). Consequently,

$$\inf_{\theta' \in \Theta_*^K} \|f_{\theta'} - f_{\theta_\triangle}\|_n^2 = \inf_{\theta' \in \Theta_*^K \cap B_\epsilon(\theta)} \|f_{\theta'} - f_{\theta_\triangle}\|_n^2 \geq \frac{1}{3}(\Delta/2)^2 \times \frac{1}{4} \geq cn^{-2\gamma}.$$

LEMMA 2. *Let $\theta_\triangle = (1/2+\Delta, -\Delta, \Delta, -1, 1)$ with $\Delta = n^{-\gamma}$ for some $\gamma \in (0,1)$, and let $C > 0$ be a constant. For any sequence of $\theta' \in \bar{\Theta}$ (N.B. here for notation convenience, the element-wise dependence of this sequence on n is suppressed), if θ' satisfies*

$$\|f_{\theta_\triangle} - f_{\theta'}\|_n^2 \leq C \log n/n,$$

for all sufficiently large n, then there exists some $C' > 0$ such that

$$\max\Big(|\tau_{\theta'} - \tau_{\theta_\triangle}|, |\alpha_{\theta'}^- - \alpha_{\theta_\triangle}^-|, |\alpha_{\theta'}^+ - \alpha_{\theta_\triangle}^+|\Big) \leq C'(\log n/n)^{1/3}$$

for every sufficiently large n.

Proof. We divide the proof into four steps.

First, it follows from $\|f_{\theta_\triangle} - f_{\theta'}\|_n^2 \to 0$ that $\|\theta' - \theta_\triangle\| \to 0$ as $n \to \infty$. Let $\theta = (1/2, 0, 0, -1, 1)$. Thus, $\theta' \to \theta$, i.e. $\|f_{\theta_\triangle} - f_{\theta'}\|_n^2 \leq C\log n/n$ implies that $\theta' \in B_\epsilon(\theta)$ for any small $\epsilon \in (0,1)$ for sufficiently large n. Note that by construction we also have $\theta_\triangle \in B_\epsilon(\theta)$.

Second, without loss of generality, consider the case of $\tau_{\theta'} \geq \tau_{\theta_\triangle}$. Note that over the interval $[\tau_{\theta_\triangle}, \tau_{\theta'}]$, $f_{\theta_\triangle}$ is linear with a slope of 1, $f_{\theta'}$ is linear with a slope between $[-1-\epsilon, -1+\epsilon]$. Since the difference between the slopes is non-vanishing, thanks to the linearity, by restricting ourselves to $[\tau_{\theta_\triangle}, \tau_{\theta'}]$, we have that the absolute difference between $f_{\theta'}$ and $f_{\theta_\triangle}$ is at least $(\tau_{\theta'} - \tau_{\theta_\triangle})(2-\epsilon)/4$ over a subinterval of length $(\tau_{\theta'} - \tau_{\theta_\triangle})/4$. Consequently, there exists a $c > 0$ such that given $\tau_{\theta'} - \tau_{\theta_\triangle} > cn^{-1}$,

$$C\log n/n \geq \|f_{\theta_\triangle} - f_{\theta'}\|_n^2 \geq (\tau_{\theta'} - \tau_{\theta_\triangle})^3 4^{-3}/2.$$

Note that the above claim is also true under the case where $\tau_{\theta'} < \tau_{\theta_\triangle}$, which can be established by an almost identical argument. As a result, we have that there exists a generic constant $C' > 0$ such that $|\tau_{\theta'} - \tau_{\theta_\triangle}| \leq C'(\log n/n)^{1/3}$ for any sufficiently large n.

Third, with regard to $|\alpha_{\theta'}^- - \alpha_{\theta_\triangle}^-|$, we again consider two scenarios. If $\tau_{\theta'} \geq \tau_{\theta_\triangle}$, we shall use the following inequality:

$$|\alpha_{\theta'}^- - \alpha_{\theta_\triangle}^-| = |f_{\theta'}(\tau_{\theta'}) - f_{\theta_\triangle}(\tau_{\theta_\triangle})| \leq |f_{\theta'}(\tau_{\theta_\triangle}) - f_{\theta_\triangle}(\tau_{\theta_\triangle})| + |f_{\theta'}(\tau_{\theta_\triangle}) - f_{\theta'}(\tau_{\theta'})|. \quad (9)$$

To bound $|f_{\theta'}(\tau_{\theta_\triangle}) - f_{\theta_\triangle}(\tau_{\theta_\triangle})|$, note that both $f_{\theta'}$ and $f_{\theta_\triangle}$ are linear over $[0, \theta_\triangle]$, thus for sufficiently large n,

$$C\log n/n \geq \|f_{\theta_\triangle} - f_{\theta'}\|_n^2 \geq \{|f_{\theta'}(\tau_{\theta_\triangle}) - f_{\theta_\triangle}(\tau_{\theta_\triangle})|/2\}^2(1/2-\epsilon)/4,$$

which implies that $|f_{\theta'}(\tau_{\theta_\triangle}) - f_{\theta_\triangle}(\tau_{\theta_\triangle})| \leq C'(\log n/n)^{1/2}$ for some generic constant $C' > 0$. Moreover, to bound $|f_{\theta'}(\tau_{\theta_\triangle}) - f_{\theta'}(\tau_{\theta'})|$, we have that

$$|f_{\theta'}(\tau_{\theta_\triangle}) - f_{\theta'}(\tau_{\theta'})| \leq (1+\epsilon)|\tau_{\theta'} - \tau_{\theta_\triangle}| \leq C'(\log n/n)^{1/3}$$

where the bound on $|\tau_{\theta'} - \tau_{\theta_\Delta}|$ follows from our result from Step 2 as above. By putting things together, we now have that $|\alpha^-_{\theta'} - \alpha^-_{\theta_\Delta}| \leq C'(\log n/n)^{1/3}$ if $\tau_{\theta'} \geq \tau_{\theta_\Delta}$. Otherwise, if $\tau_{\theta'} < \tau_{\theta_\Delta}$, we could use

$$|\alpha^-_{\theta'} - \alpha^-_{\theta_\Delta}| = |f_{\theta'}(\tau_{\theta'}) - f_{\theta_\Delta}(\tau_{\theta_\Delta})| \leq |f_{\theta'}(\tau_{\theta'}) - f_{\theta_\Delta}(\tau_{\theta'})| + |f_{\theta_\Delta}(\tau_{\theta'}) - f_{\theta_\Delta}(\tau_{\theta_\Delta})|$$

instead of (9) to derive the same conclusion.

Finally, using an argument similar to that presented in Step 3, we could also establish that $|\alpha^+_{\theta'} - \alpha^+_{\theta_\Delta}| \leq C'(\log n/n)^{1/3}$. Consequently, we have that

$$\max\Big(|\tau_{\theta'} - \tau_{\theta_\Delta}|, |\alpha^-_{\theta'} - \alpha^-_{\theta_\Delta}|, |\alpha^+_{\theta'} - \alpha^+_{\theta_\Delta}|\Big) \leq C'(\log n/n)^{1/3}$$

for all sufficiently large n.

Proof of Theorem 4

Proof. In this proof, we again use $C, c > 0$ as generic constants.

First, we prove the first part of the statement where we fix $\theta \in \Theta \backslash \Theta^K$. It builds upon the proof of Theorem 3. Note that here a closer inspection reveals that (7) and (8) in the proof of Theorem 3 hold uniformly for all $\theta \in \Theta$. More specifically, we have

$$P(\widetilde{B}_n^c) < n^{-1}, \qquad \text{where the event } \widetilde{B}_n = \left\{ 2 \sup_{\theta} \sup_{\theta'} \left\langle \varepsilon_n, \frac{f_{\theta'} - f_\theta}{\|f_{\theta'} - f_\theta\|} \right\rangle \leq 12\sqrt{6\log n} \right\}$$

and, when $\widetilde{B}_n$ holds,

$$\sup_{\theta} \|f_{\widehat{\theta}} - f_\theta\|_n^2 \leq C \log n/n,$$

where $\widehat{\theta} := \widehat{\theta}_{LS}$ is the least squares estimator defined in (3). Moreover, for any fixed $\theta \in \Theta\backslash\Theta^K$ and every θ' with $\|\theta' - \theta\| \leq \epsilon$, we always have that $|\alpha^+_{\theta'} - \alpha^-_{\theta'}| > c$ for some $c > 0$, as $\epsilon \to 0$. In other words, by picking a sufficiently small neighbourhood of θ, it only contains θ' with a jump of significant size. Combining these facts entails consistency of $\widehat{\theta}$ for estimating θ', and that $\widehat{\theta}^S = \widehat{\theta}$ over $\widetilde{B}_n$ for every θ', where the true parameter is taken as $\theta' \in \Theta$ that lies in a small neighbourhood of θ. Furthermore, it is straightforward to verify that the rest of the arguments in the proof of Theorem 3, when applied uniformly over a small neighbourhood of θ, would go through. In particular, by taking a sufficiently small ϵ, we have that

$$\sup_{\theta' \in \Theta : \|\theta' - \theta\| \leq \epsilon} E_{\theta'} \|f_{\widehat{\theta}} - f_{\theta'}\|_n^2 \leq C n^{-1}.$$

Consequently,

$$\limsup_{n \to \infty} \sup_{\theta' \in \Theta : \|\theta' - \theta\| \leq \epsilon} nR(\theta', \widehat{\theta}^S) < \infty.$$

For the proof of the second part, to simplify our arguments, we shall fix $\theta = (1/2, 0, 0, -1, 1)$ and take the truth parameter as $\theta_\Delta = (1/2 + \Delta, -\Delta, \Delta, -1, 1)$ with $\Delta = n^{-\gamma}$ and $\gamma = 1/4$. We first establish that under the truth θ_Δ, $\widehat{\theta}^S = \widehat{\theta}^K$ with arbitrarily high probability even though $\theta_\Delta \notin \Theta^K$. To see this, recall that $\|f_{\widehat{\theta}} - f_{\theta_\Delta}\|_n^2 \leq C \log n/n$ under $\widetilde{B}_n$. Therefore, it follows from Lemma 2 that

$$\max\Big(|\alpha^-_{\widehat{\theta}} - \alpha^-_{\theta_\Delta}|, |\alpha^+_{\widehat{\theta}} - \alpha^+_{\theta_\Delta}|\Big) \leq C'(\log n/n)^{1/3}.$$

Since $|\alpha^+_{\theta_\Delta} - \alpha^-_{\theta_\Delta}| = 2\Delta = 2n^{-1/4}$, we have that $|\alpha^+_{\widehat{\theta}} - \alpha^-_{\widehat{\theta}}| \leq 3n^{-1/4} < n^{-1/6}$ for sufficiently large n. Therefore, by construction, $\widehat{\theta}^S = \widehat{\theta}^K$. Next, let's fix $\gamma' \in (1/4, 1/3)$. Our goal is to show that $\widehat{\theta}^K \notin [\theta_\Delta - n^{-\gamma'}, \theta_\Delta + n^{-\gamma'}]$ under $\widetilde{B}_n$ for any sufficiently large n. We prove this by contradiction. Suppose that $\widehat{\theta}^K \in [\theta_\Delta - n^{-\gamma'}, \theta_\Delta + n^{-\gamma'}]$, then

$$\begin{aligned}\sum_{i=1}^n \big\{Y_{ni} - f_{\widehat{\theta}^K}(X_{ni})\big\}^2 &= \sum_{i=1}^n \big\{\sigma\varepsilon_{ni} + f_{\theta_\Delta}(X_{ni}) - f_{\widehat{\theta}^K}(X_{ni})\big\}^2 \\ &= \sigma^2\|\varepsilon_n\|^2 + \|f_{\widehat{\theta}^K} - f_{\theta_\Delta}\|^2 - 2\sigma\langle\varepsilon_n, f_{\widehat{\theta}^K} - f_{\theta_\Delta}\rangle \\ &\geq \sigma^2\|\varepsilon_n\|^2 + \|f_{\widehat{\theta}^K} - f_{\theta_\Delta}\|^2 - 12\sigma\sqrt{6\log n}\,\|f_{\widehat{\theta}^K} - f_{\theta_\Delta}\| \\ &\geq \sigma^2\|\varepsilon_n\|^2 + \|f_{\widehat{\theta}^K} - f_{\theta_\Delta}\|^2/2 \\ &\geq \sigma^2\|\varepsilon_n\|^2 + cn^{1/2}\end{aligned}$$

for some $c > 0$. Here the third last inequality holds under $\widetilde{B}_n$, the second inequality is due to the fact that $\|f_{\widehat{\theta}^K} - f_{\theta_\Delta}\| \geq 24\sigma\sqrt{6\log n}$ which follows from Lemma 1, and the final inequality follows from Lemma 1 (all with $\gamma = 1/4$). On the other hand, since $\theta := (1/2, 0, 0, -1, 1) \in \Theta^K$, under the truth θ_Δ,

$$\begin{aligned}\sum_{i=1}^n \big\{Y_{ni} - f_{\widehat{\theta}^K}(X_{ni})\big\}^2 \leq \sum_{i=1}^n \big\{Y_{ni} - f_{\theta}(X_{ni})\big\}^2 &= \sigma^2\|\varepsilon_n\|^2 + \|f_\theta - f_{\theta_\Delta}\|^2 + 2\sigma^2\langle\varepsilon_n, f_{\theta_\Delta} - f_\theta\rangle \\ &\leq \sigma^2\|\varepsilon_n\|^2 + \|f_\theta - f_{\theta_\Delta}\|^2 + 12\sigma^2\sqrt{6\log n}\,\|f_\theta - f_{\theta_\Delta}\| \\ &\leq \sigma^2\|\varepsilon_n\|^2 + Cn^{1/4}\end{aligned}$$

for some $c > 0$, where the second last inequality holds under $\widetilde{B}_n$, while the last inequality is derived from straightforward computation as illustrated below for sufficiently large n:

$$\|f_\theta - f_{\theta_\Delta}\|^2 = n\|f_\theta - f_{\theta_\Delta}\|_n^2 \leq nC\Delta^3 \leq Cn^{1-3\gamma} = Cn^{1/4}.$$

Putting things together, we see that

$$\sigma^2\|\varepsilon_n\|^2 + cn^{1/2} \leq \sum_{i=1}^n \big\{Y_{ni} - f_{\widehat{\theta}^K}(X_{ni})\big\}^2 \leq \sigma^2\|\varepsilon_n\|^2 + Cn^{1/4},$$

leading to a contradiction. This entails that $\tau_{\widehat{\theta}^K} \notin [\tau_{\theta_\Delta} - n^{-\gamma'}, \tau_{\theta_\Delta} + n^{-\gamma'}]$ under $\widetilde{B}_n$ for any sufficiently large n. Consequently, for any $\epsilon > 0$,

$$\sup_{\theta' \in \Theta : \|\theta' - \theta\| \leq \epsilon} n^{1/3} R(\theta', \widehat{\theta}^S) \geq n^{1/3} R(\theta_\Delta, \widehat{\theta}^S) \geq n^{1/3} n^{-\gamma'}(1 - n^{-1}) \to \infty$$

as $n \to \infty$, from which the second claim of this theorem follows immediately.

As a final remark, we note that the second part of the proof would also go through if we pick any γ and γ' such that $1/6 < \gamma < \gamma' < 1/3$. In particular, we could take both of them to be sufficiently close to $1/6$ (so Δ would be of order close to $n^{-1/6}$). This implies that for any fixed $\gamma \in (1/6, 1/3)$ and every $\theta \in \Theta^K$, our proposed super-efficient estimator $\widehat{\theta}^S$ satisfies

$$\liminf_{\epsilon \to 0+} \liminf_{n \to \infty} \sup_{\theta' \in \Theta : \|\theta' - \theta\| \leq \epsilon} n^\gamma R(\theta', \widehat{\theta}^S) = \infty,$$

i.e. its uniform convergence rate is worse than $O(n^{-\gamma})$ for any fixed $\gamma > 1/6$.

ACKNOWLEDGEMENTS

We thank Moulinath Banerjee, Haeran Cho and Piotr Fryzlewicz for their helpful comments on an earlier draft. We also thank the editors and three anonymous reviewers for their valuable suggestions.

REFERENCES

BAI, J. & PERRON, P. (1998). Estimating and testing linear models with multiple structural changes. *Econometrica*, **66**, 47–78.

BANERJEE, M., DUROT, C. & SEN, B. (2019). Divide and conquer in non-standard problems and the super-efficiency phenomenon. *Ann. Statist.* **47**, 720–757.

BARANOWSKI, R., CHEN, Y. & FRYZLEWICZ, P. (2019). Narrowest-over-threshold detection of multiple change-points and change-point-like features. *J. R. Statist. Soc.* B **81**, 649–672.

BROWN, L. D., LOW, M. G. & ZHAO, L. H. (1997). Superefficiency in nonparametric function estimation. *Ann. Statist.* **25**, 2607–2625.

CHENG, M-Y., & RAIMONDO, M. (2008). Kernel methods for optimal change-points estimation in derivatives. *J. Comput. Graph. Statist.* **17**, 56–75.

DAS, R., BANERJEE, M., NAN, B. & ZHENG, H. (2016). Fast estimation of regression parameters in a broken-stick model for longitudinal data. *J. Amer. Statist. Soc.* **111**, 1132–1143.

DONG, Y. (2018). Jump or kink? regression probability jump and kink design for treatment effect evaluation. *Unpublished Manuscript* Available at *http://www.yingyingdong.com/Research/RDwithoutDiscontinuity51.pdf*

DONOHO, D. L., JOHNSTONE, I. M., KERKYACHARIAN, G. & PICHARD, D. (1995). Wavelet shrinkage: asymptopia? (with discussion). *J. R. Statist. Soc.* B **57**, 301–369.

GOLDENSHLUGER, A., TSYBAKOV, A. & ZEEVI, A. (2006). Optimal change-point estimation from indirect observations. *Ann. Statist.* **34**, 350–372.

HÁJEK, J. (1972). Local asymptotic minimax and admissibility in estimation. *Proc. Sixth Berkeley Symp. Math. Statist. Probab.* **1**, 175–194.

HANSEN, B. E. (2017). Regression kink with an unknown threshold. *J. Bus. Econ. Statist.* **35**, 228–240.

HEINRICH, P. & KAHN, J. (2018). Strong identifiability and optimal minimax rates for finite mixture estimation. *Ann. Statist.* **46**, 2844–2870.

HIDALGO, J., LEE, J. & SEO, M. H. (2019). Robust inference for threshold regression models. *J. Econometrics* **210**, 291–309.

KOROSTELEV, A. & LEPSKI, O. (2008). On a multi-channel change-point. *Math. Methods Statist.* **17**, 187–197.

LE CAM, L. (1953). On some asymptotic properties of maximum likelihood estimates and related Bayes estimates. *University of California Publications in Statistics* **1**, 277–330.

LE CAM, L. (1986). *Asymptotic methods in statistical decision theory*. Springer.

MUGGEO, V. M. R. (2003). Estimating regression models with unknown break-points. *Statist. Med.* **22**, 3055–3071.

MAIDSTONE, R., FEARNHEAD, P. & LETCHFORD, A. (2019). Detecting changes in slope with an L_0 penalty. *J. Comput. Graph. Statist.* **28**, 265–275.

RAIMONDO, M. (1998). Minimax estimation of sharp change points. *Ann. Statist.* **26**, 1379–1397.

STIGLER, S. M. (2007). The epic story of maximum likelihood. *Statist. Sci.* **22**, 598–620.

VAN DE GEER, S. (2000). *Empirical processes in M-estimation*. Cambridge University Press.

VOVK, V. (2009). Superefficiency from the vantage point of computability. *Statist. Sci.* **24**, 73–86.

WISHART, J. R. (2011). Minimax lower bound for kink location estimators in a nonparametric regression model with long-range dependence. *Statist. Probab. Lett.* **81**, 1871–1875.

WISHART, J. R. & KULIK, R.(2010). Kink estimation in stochastic regression with dependent errors and predictors. *Electron. J. Statist* **4**, 875–913.

YU, B. (1997). Assouad, Fano, and Le Cam. *In* Festschrift for Lucien Le Cam (D.Pollard, E.Torgersen and G.L. Yang, eds.), 423–435, Springer.