Online Supplementary Materials to
“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

PROOF OF THEOREM 1

Proof. For any \( \theta = (\tau, \alpha_0^+, \alpha_0^-, \beta_0^+, \beta_0^-) \in \Theta \), without loss of generality, we assume that \( \tau_0 \leq 1/2 \).
Now consider \( \theta_\Delta = (\tau + \Delta, \alpha_0^+, \beta_0^+, \beta_0^-) \) for sufficiently small \( \Delta > 0 \). For sufficiently small \( \Delta > 0 \), \( \theta_\Delta \in \Theta \).
As will become clearer later, \( f_{\theta_\Delta} \) corresponding to this particular form of \( \theta_\Delta \) does not necessarily give the closest approximation to \( f_\theta \) over the restricted parameter space while fixing the break-point location at \( \tau_0 + \Delta \). 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 \( \theta \) and \( \theta_\Delta \) by \( P_\theta^n \) and \( P^n_{\theta_\Delta} \), respectively. Then the Total Variation (TV) distance between \( P_\theta^n \) and \( P^n_{\theta_\Delta} \) is bounded above by

\[
\| P^n_\theta \circ P^n_{\theta_\Delta} \|_{TV} \leq 2 - 2 \prod_{i=1}^{n} \left[ 1 - d^2_{\text{Hell}} \left\{ N \left( f_\theta(i/(n+1)), \sigma^2 \right), N \left( f_{\theta_\Delta}(i/(n+1)), \sigma^2 \right) \right\} \right]
\]

\[
= 2 - 2 \exp \left[ - \frac{1}{8\sigma^2} \sum_{i:i/(n+1) \in (\tau_0, \tau_0 + \Delta)} \left\{ f_{\theta_\Delta}(i/(n+1)) - f_\theta(i/(n+1)) \right\}^2 \right]
\]

where \( d_{\text{Hell}} \) is the Hellinger distance, and recalling that \( d^2_{\text{Hell}} \left\{ N(\mu_1, \sigma), N(\mu_2, \sigma) \right\} = 1 - \exp\left\{ - (\mu_1 - \mu_2)^2 / (8\sigma^2) \right\} \). We also used the fact that the values of \( f_\theta(x) \) and \( f_{\theta_\Delta}(x) \) are only different over the interval \( (\tau_0, \tau_0 + \Delta) \).

If \( \theta \in \Theta \setminus \Theta^K \) (i.e. with a jump), then for any small \( \Delta > 0 \), say with \( \Delta < 0.5|\alpha_0^+ - \alpha_0^-| / |\beta_0^+ - \beta_0^-| \),

\[
\| P^n_\theta \circ P^n_{\theta_\Delta} \|_{TV} \leq 2 - 2 \exp \left\{ - \frac{1}{8\sigma^2} \left( \frac{\alpha_0^+ - \alpha_0^-}{2} \right)^2 \left( \frac{\Delta n}{2} \right) \right\},
\]

Since \( \theta \) is fixed, so is \(|\alpha_0^+ - \alpha_0^-| \), which is strictly positive because \( \theta \in \Theta \setminus \Theta^K \). By taking \( \Delta = cn^{-1} \) for sufficiently small \( c > 0 \), we have \( \| P^n_\theta \circ P^n_{\theta_\Delta} \|_{TV} \leq 1/4 \). Now it follows from Le Cam’s two-point method (cf. Le Cam (1986) or Yu (1997)) that for any \( \epsilon > 0 \) and large \( n \),

\[
\inf_{\theta} \sup_{\theta' \in \Theta: ||\theta' - \theta|| \leq \epsilon} R(\theta', \hat{\theta}) \geq \frac{\Delta}{2} \left( 1 - || P^n_\theta \circ P^n_{\theta_\Delta} \|_{TV} \right) \geq \frac{c}{4} n^{-1}.
\]

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

\[
\| P^n_\theta \circ P^n_{\theta_\Delta} \|_{TV} \leq 2 - 2 \exp \left\{ - \frac{1}{8\sigma^2} \left( \frac{\Delta}{2} \right)^2 \left( \frac{\Delta n}{2} \right) \left( \frac{\beta_0^+ - \beta_0^-}{2} \right)^2 \right\} \leq 2 - 2 \exp \left\{ - \frac{\beta_0^+ - \beta_0^-}{8\sigma^2} \left( \frac{\Delta}{2} \right)^2 \left( \frac{\Delta n}{2} \right) \right\}.
\]

© 2018 Biometrika Trust
By taking $\Delta = cn^{-1/3}$ for sufficiently small $c > 0$, we still have $\|P^n_\theta - P^n_{\theta_\Delta}\|_TV \leq 1/4$. Hence, Le Cam’s two-point method yields

$$\inf_\theta \sup_{\theta' \in \Theta : \|\theta' - \theta\| \leq \epsilon} R(\theta', \hat{\theta}) \geq \frac{\Delta}{2} \left(1 - \|P^n_\theta - P^n_{\theta_\Delta}\|_TV\right) \geq \frac{c}{4} n^{-1/3}.$$

The proof is completed by simple rearrangements of the terms. \qed

**Proof of Theorem 2**

**Proof.** This proof is similar to that of Theorem 1. Here only the differences are highlighted. For any $\theta = (\tau_\theta, \alpha_\theta, \beta_\theta, \beta_\theta' ) \in \Theta^K$, without loss of generality, we again assume that $\tau_\theta \leq 1/2$ 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,

$$\|P^n_\theta - P^n_{\theta_\Delta}\|_TV \leq 2 - 2 \exp \left\{- \frac{1}{8\sigma^2} \sum_{i/(n+1) \in [\tau_\theta + \Delta, n]} \left\{f_{\theta_\Delta}(i/(n+1)) - f_\theta(i/(n+1))\right\}^2\right\},$$

$$\leq 2 - 2 \exp \left\{- \frac{1}{8\sigma^2} (\beta_\theta^2 - \beta_\theta' )^2 \Delta^2 \left\{0.5 - \Delta\right\} n\right\} \leq 2 - 2 \exp \left( - \frac{2}{32\sigma^2} \Delta^2 n \right)$$

for small $\Delta$, say $< 0.25$. By taking $\Delta = cn^{-1/2}$ for sufficiently small $c > 0$, we have that $\|P^n_\theta - P^n_{\theta_\Delta}\|_TV \leq 1/4$. The result then follows as previously. \qed

**Proof of Theorem 3**

**Proof.** We assume that $\theta$ is fixed, and shall use $C, c > 0$ as generic constants. For notational convenience, we now write $f_\theta = (f_\theta(1/(n+1)), \ldots, f_\theta(n/(n+1)))$ (where the dependence on $n$ is suppressed), let $(x, y) = \sum_{i=1}^n x_i y_i$ 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 $\|x\|^2_n = \|x\|^2/n$. It follows from the definition of $\hat{\theta} := \hat{\theta}_{LS}$ (cf. Lemma 4.7 of van de Geer (2000)) that

$$\|f_\theta - f_{\hat{\theta}}\| \leq 2\sigma \left\langle \varepsilon_n, \frac{f_{\hat{\theta}} - f_\theta}{\|f_{\hat{\theta}} - f_\theta\|} \right\rangle$$

(5)

where $\varepsilon_n = (\varepsilon_{n1}, \ldots, \varepsilon_{nn})$. Here $f_{\hat{\theta}} - f_\theta$ 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 $\xi(s, e) = (\xi(s, e)(1), \ldots, \xi(s, e)(n))$ and $\gamma(s, e) = (\gamma(s, e)(1), \ldots, \gamma(s, e)(n))$, with

$$\xi(s, e)(i) = \begin{cases} \frac{s}{e-s} & i = s + 1, \ldots, e \\ 0 & \text{otherwise} \end{cases}, \quad \gamma(s, e)(i) = \begin{cases} \frac{i-(e-s+1)/2}{\sqrt{(e-s)(e-s+1)/2}} & i = s + 1, \ldots, e \\ 0 & \text{otherwise} \end{cases}.$$  

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,

$$\left\langle \varepsilon_n, \frac{f_{\hat{\theta}} - f_\theta}{\|f_{\hat{\theta}} - f_\theta\|} \right\rangle \leq \sup_{s, e} \left(\|\varepsilon_n, \xi(0,s)\| + \|\varepsilon_n, \xi(s,e)\| + \|\varepsilon_n, \xi(e,n)\| + \|\varepsilon_n, \gamma(0,s)\| + \|\varepsilon_n, \gamma(s,e)\| + \|\varepsilon_n, \gamma(e,n)\|\right).$$

(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
\[
P(B_n) \leq \frac{\sqrt{3}}{\sqrt{\pi} \log n} n^{-1} < n^{-1},
\]
where the event \( B_n = \left\{ 2 \sup_{\theta'} \left\langle \varepsilon_n, \frac{f_{\theta'} - f_0}{\|f_{\theta'} - f_0\|} \right\rangle \leq 12 \sqrt{6 \log n} \right\} \).

(7)

Plugging (7) back to (5), we have that when \( B_n \) holds, there exists some \( C > 0 \) such that
\[
\|f_{\bar{\theta}} - f_0\|_n^2 \leq C \log n/n.
\]

(8)

For any \( \theta \in \Theta \setminus \Theta^K \) (i.e. with a jump), \( \|f_{\bar{\theta}} - f_0\|_n^2 \to 0 \) as \( n \to \infty \) implies consistency of \( \hat{\theta} \). Furthermore, due to the presence of the jump, there exists \( C, c > 0 \) (independent of \( n \), but might depend on \( \theta \)) such that \( \|f_{\theta'} - f_0\|_n^2 \geq C(\tau_{\theta'} - \tau_0) \) for any \( \theta' \) with \( |\tau_{\theta'} - \tau_0| > c/n \). Comparing this with (8), we have
\[
|\tau_{\bar{\theta}} - \tau_0| \leq C \log n/n.
\]
Similarly, we obtain \( |(\alpha_\theta^+ - \alpha_{\bar{\theta}}) - (\alpha_\theta^- - \alpha_{\bar{\theta}})| \leq C \sqrt{\log n/n}, \) which yields
\[
|\alpha_\theta^+ - \alpha_{\bar{\theta}}| \geq |\alpha_\theta^- - \alpha_{\bar{\theta}}| - C \sqrt{\log n/n} > n^{-1/6}
\]
for sufficiently large \( n \). This implies \( \hat{\theta}^K = \hat{\theta} \) 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
\[
\mathcal{F}_n^\Theta(R) = \{ f_{\theta'} | \theta' \in \Theta, \|f_{\theta'} - f_0\|_n \leq R, |\tau_{\theta'} - \tau_0| \leq \epsilon' \}
\]
for some fixed \( \epsilon' \in \left(0, \min(\tau_0, 1 - \tau_0)/2\right) \). Then, for any \( f_{\theta'} \in \mathcal{F}_n^\Theta(R) \), we also have that
\[
\sup_{x \in [0, 1]} |f_{\theta'}(x) - f_0(x)| \leq \sup_{x \in [0, \tau_0], x \in [\tau_0, \tau_0]} |f_{\theta'}(x) - f_0(x)| \leq CR.
\]
Thus, \( \mathcal{F}_n^\Theta(R) \) (for \( L_2(Q_n) \)-metric) can be bounded above as
\[
H_2(2u, \mathcal{F}_n^\Theta(R), Q_n) \leq C \log \left( \frac{CR + u}{n} \right).
\]

Here one could derive this by construction, using the fact that for every \( f_{\theta'} \in \mathcal{F}_n^\Theta(R) \), \( f_{\theta'} - f_0 \) 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 \( \hat{\theta} \) be the least squares estimator defined like \( \hat{\theta} \) in (3) in Section 4 but being optimised over \( [\tau_0 - \epsilon', \tau_0 + \epsilon'] \times \mathbb{R}^d \) instead, then it follows from Theorem 9.1 and Example 9.3.1 of van de Geer (2000) that \( E_\theta\|f_{\theta'} - f_0\|_n^2 \leq C n^{-1} \). Note that \( \hat{\theta}^K = \hat{\theta} \) under \( B_n \). Also recall that \( \|f_{\theta'} - f_0\|_n^2 \geq C |\tau_{\theta'} - \tau_0| \) for any \( \theta' \) with \( |\tau_{\theta'} - \tau_0| > cn^{-1}. \) Simple manipulation entails
\[
E_\theta(n|\tau_{\bar{\theta}} - \tau_0|) = E_\theta(n|\tau_{\bar{\theta}} - \tau_0|) - E_\theta(n|\tau_{\bar{\theta}} - \tau_0|,B_n^c) P(B_n^c) + E_\theta(n|\tau_{\bar{\theta}} - \tau_0| \mid B_n^c) P(B_n^c) \leq C,
\]
which completes our proof for the case of \( \theta \in \Theta \setminus \Theta^K \).

Now for any \( \theta \in \Theta^K \) (i.e. with a kink), there exists some \( \epsilon' > 0 \) such that \( \|f_{\theta'} - f_0\|_n^2 \geq C |\tau_{\theta'} - \tau_0|^3 \) for any \( \theta' \in \Theta \) with \( |\tau_{\theta'} - \tau_0| \in (c/n, \epsilon'] \). See for instance, Lemma 7 of Baranowski et al. (2019). Moreover, comparing that with (8) leads to \( |(\alpha_\theta^+ - \alpha_{\bar{\theta}}^+) - (\alpha_\theta^- - \alpha_{\bar{\theta}}^-)| \leq C(\log n/n)^{1/3} \). So under \( B_n \), \( \hat{\theta}^K = \hat{\theta}^K \) for sufficiently large \( n \). Our next ingredient is the following statement: there exists some \( \epsilon' > 0 \) such that \( \|f_{\theta'} - f_0\|_n^2 \geq C |\tau_{\theta'} - \tau_0|^2 \) for any \( \theta' \in \Theta^K \) satisfying \( |\tau_{\theta'} - \tau_0| \in (c/n, \epsilon'] \). See for example, Lemma 8 of Baranowski et al. (2019). Since \( \mathcal{F}_{\hat{\theta}}(R) \subset \mathcal{F}_{\hat{\theta}}(R) \), we could again use Theorem 9.1 and Example 9.3.1 of van de Geer (2000) to see that \( E_\theta\|f_{\bar{\theta}} - f_0\|_n^2 \leq C n^{-1}. \) Consequently,
\[
E_\theta(\sqrt{n}|\tau_{\bar{\theta}} - \tau_0|) \leq \sqrt{E_\theta(n|\tau_{\bar{\theta}} - \tau_0|^2) \leq C \sqrt{E_\theta(n\|f_{\bar{\theta}} - f_0\|_n^2) \leq C},
\]
where \( C > 0 \) are generic constants. Finally, the proof is complete by noting that
\[
E_\theta(\sqrt{n}|\tau_{\bar{\theta}} - \tau_0|) = E_\theta(\sqrt{n}|\tau_{\bar{\theta}} - \tau_0|) - E_\theta(\sqrt{n}|\tau_{\bar{\theta}} - \tau_0| \mid B_n^c) P(B_n^c) + E_\theta(\sqrt{n}|\tau_{\bar{\theta}} - \tau_0| \mid B_n^c) P(B_n^c) \leq C + \sqrt{n}(n^{-1} + \sqrt{n}(n^{-1}) \leq C.
\]
ANCILLARY LEMMAS

Some ancillary results are required for the proof of Theorem 4. They concern the approximation of \( f_{\theta_{\Delta}} \) for \( \theta_{\Delta} \in \Theta \cap \Theta^K \) using \( f_{\theta'} \) with \( \theta' \in \Theta^K \), and the distance between \( \theta_{\Delta} \) and \( \theta' \) 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 \( \Theta^K = \{ \theta \in \Theta^K, \tau_0 \in [1/2 + \Delta - \Delta', 1/2 + \Delta + \Delta'] \} \). Then there exists \( c > 0 \) such that

\[
\inf_{\theta' \in \Theta^K} \| f_{\theta'} - f_{\theta_{\Delta}} \|^2_n \geq cn^{-2\gamma}.
\]

for every sufficiently large \( n \).

**Proof.** This proof can be divided into three parts.

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

\[
\inf_{\theta' \in \Theta^K} \| f_{\theta'} - f_{\theta_{\Delta}} \|^2_n \leq \| f_{\theta_{\Delta}}' - f_{\theta_{\Delta}} \|^2_n \leq \Delta^2 = n^{-2\gamma} \to 0,
\]

as \( n \to \infty \), where we have taken \( \theta_{\Delta}' = (1/2 + \Delta, 0, 0, -1, 1) \).

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

\[
\lim_{n \to \infty} \inf_{\theta' \in \Theta^K \cap B_\varepsilon(\theta)} \| f_{\theta'} - f_{\theta_{\Delta}} \|^2_n > 0
\]

for any fixed \( \varepsilon > 0 \), it now follows that for any sufficiently large \( n \),

\[
\inf_{\theta' \in \Theta^K} \| f_{\theta'} - f_{\theta_{\Delta}} \|^2_n = \inf_{\theta' \in \Theta^K \cap B_\varepsilon(\theta)} \| f_{\theta'} - f_{\theta_{\Delta}} \|^2_n.
\]

This also implies that

\[
| f_{\theta'}(1/2 + \Delta - \Delta') - f_{\theta'}(1/2 + \Delta + \Delta') | \leq 2(1 + \varepsilon)\Delta'.
\]

for any \( \theta' \in \Theta^K \cap B_\varepsilon(\theta) \). Thanks to the fact that

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

we have that

\[
\inf_{\theta' \in \Theta^K \cap B_\varepsilon(\theta)} \max \left\{ \left| f_{\theta_{\Delta}}(1/2 + \Delta - \Delta') - f_{\theta'}(1/2 + \Delta - \Delta') \right|, \left| f_{\theta_{\Delta}}(1/2 + \Delta + \Delta') - f_{\theta'}(1/2 + \Delta + \Delta') \right| \right\} 
\geq (2\Delta - 2(1 + \varepsilon)\Delta')/2 \geq \Delta/2
\]

for any sufficiently large \( n \).

Finally, note that for any \( \theta' \in \Theta^K \cap B_\varepsilon(\theta) \), both \( f_{\theta'} \) and \( f_{\theta_{\Delta}} \) 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,

\[
\inf_{\theta' \in \Theta^K} \| f_{\theta'} - f_{\theta_{\Delta}} \|^2_n = \inf_{\theta' \in \Theta^K \cap B_\varepsilon(\theta)} \| f_{\theta'} - f_{\theta_{\Delta}} \|^2_n \geq \frac{1}{3}(\Delta/2)^2 \times \frac{1}{4} \geq cn^{-2\gamma}.
\]

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

\[
\| f_{\theta_{\Delta}} - f_{\theta'} \|^2_n \leq C \log n/n,
\]

then
for all sufficiently large \( n \), then there exists some \( C' > 0 \) such that
\[
\max \left( \left| \tau_{\theta'} - \theta_{\Delta} \right|, \left| \alpha_{\theta'} - \alpha_{\theta_{\Delta}} \right|, \left| \alpha_{\theta'}^+ - \alpha_{\theta_{\Delta}}^+ \right| \right) \leq C'(\log n/n)^{1/3}
\]
for every sufficiently large \( n \).

**Proof.** We divide the proof into four steps.

First, it follows from \( \|f_{\theta_{\Delta}} - f_{\theta'}\|_n^2 \to 0 \) that \( \|\theta' - \theta_{\Delta}\| \to 0 \) as \( n \to \infty \). Let \( \theta = (1/2, 0, 0, -1, 1) \). Thus, \( \theta' \to \theta \), i.e. \( \|f_{\theta_{\Delta}} - f_{\theta'}\|_n^2 \leq C\log n/n \) 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 \( \theta_{\Delta} \in B_{\epsilon}(\theta) \).

Second, without loss of generality, consider the case of \( \tau_{\theta'} \geq \tau_{\theta_{\Delta}} \). Note that over the interval \([\tau_{\theta_{\Delta}}, \tau_{\theta'}]\), \( f_{\theta_{\Delta}} \) 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 \([\tau_{\theta_{\Delta}}, \tau_{\theta'}]\), we have that the absolute difference between \( f_{\theta'} \) and \( f_{\theta_{\Delta}} \) is at least \((\tau_{\theta'} - \tau_{\theta_{\Delta}})(2 - \epsilon)/4\) over a subinterval of length \((\tau_{\theta'} - \tau_{\theta_{\Delta}})/4\). Consequently, there exists a \( c > 0 \) such that given \( \tau_{\theta'} - \tau_{\theta_{\Delta}} > cn^{-1} \),
\[
C \log n/n \geq \|f_{\theta_{\Delta}} - f_{\theta'}\|_n^2 \geq (\tau_{\theta'} - \tau_{\theta_{\Delta}})^3/2.
\]

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

Third, with regard to \( |\alpha_{\theta'} - \alpha_{\theta_{\Delta}}| \), we again consider two scenarios. If \( \tau_{\theta'} \geq \tau_{\theta_{\Delta}} \), we shall use the following inequality:
\[
|\alpha_{\theta'} - \alpha_{\theta_{\Delta}}| = |f_{\theta'}(\tau_{\theta'}) - f_{\theta_{\Delta}}(\tau_{\theta_{\Delta}})| \leq |f_{\theta'}(\tau_{\theta_{\Delta}}) - f_{\theta_{\Delta}}(\tau_{\theta_{\Delta}})| + |f_{\theta'}(\tau_{\theta_{\Delta}}) - f_{\theta'}(\tau_{\theta'})|.
\]
To bound \( |f_{\theta'}(\tau_{\theta_{\Delta}}) - f_{\theta_{\Delta}}(\tau_{\theta_{\Delta}})| \), note that both \( f_{\theta'} \) and \( f_{\theta_{\Delta}} \) are linear over \([0, \tau_{\Delta}]\), thus for sufficiently large \( n \),
\[
C \log n/n \geq \|f_{\theta_{\Delta}} - f_{\theta'}\|_n^2 \geq 2(1/2 - \epsilon)/4,
\]
which implies that \( |f_{\theta'}(\tau_{\theta_{\Delta}}) - f_{\theta_{\Delta}}(\tau_{\theta_{\Delta}})| \leq C'(\log n/n)^{1/2} \) for some generic constant \( C' > 0 \). Moreover, to bound \( |f_{\theta'}(\tau_{\theta_{\Delta}}) - f_{\theta'}(\tau_{\theta'})| \), we have that
\[
|f_{\theta'}(\tau_{\theta_{\Delta}}) - f_{\theta'}(\tau_{\theta'})| \leq (1 + \epsilon)|\tau_{\theta'} - \tau_{\theta_{\Delta}}| \leq C'(\log n/n)^{1/3}
\]
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
\[
|\alpha_{\theta'} - \alpha_{\theta_{\Delta}}| = |f_{\theta'}(\tau_{\theta'}) - f_{\theta_{\Delta}}(\tau_{\theta_{\Delta}})| \leq |f_{\theta'}(\tau_{\theta'}) - f_{\theta_{\Delta}}(\tau_{\theta_{\Delta}})| + |f_{\theta_{\Delta}}(\tau_{\theta'}) - f_{\theta_{\Delta}}(\tau_{\theta_{\Delta}})|
\]
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
\[
\max \left( |\tau_{\theta'} - \tau_{\theta_{\Delta}}|, |\alpha_{\theta'} - \alpha_{\theta_{\Delta}}|, |\alpha_{\theta'}^+ - \alpha_{\theta_{\Delta}}^+| \right) \leq C'(\log n/n)^{1/3}
\]
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 \( \theta \in \Theta \setminus \Theta^K \). 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
\[
P(\tilde{B}_n) < n^{-1}, \quad \text{where the event } \tilde{B}_n = \left\{ 2 \sup_{\theta} \sup_{\theta'} \langle \varepsilon_n, f_{\theta'} - f_\theta \rangle / \|f_{\theta'} - f_\theta\| \leq 12 \sqrt{6 \log n} \right\}
\]
and, when $\hat{B}_n$ holds,
$$\sup_{\theta'} \|f_\theta - f_{\theta'}\|_n^2 \leq C \log n/n,$$
where $\hat{\theta} := \hat{\theta}_{LS}$ is the least squares estimator defined in (3). Moreover, for any fixed $\theta' \in \Theta \setminus \Theta^K$ and every $\theta''$ 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 $\theta$, it only contains $\theta''$ with a jump of significant size. Combining these facts entails consistency of $\hat{\theta}$ for estimating $\theta'$, and that $\hat{\theta}^3 = \hat{\theta}$ over $\hat{B}_n$ for every $\theta''$, where the true parameter is taken as $\theta' \in \Theta$ that lies in a small neighbourhood of $\theta$. 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 $\theta$, would go through. In particular, by taking a sufficiently small $\epsilon$, we have that
$$\sup_{\theta' \in \Theta, \|\theta' - \theta\| \leq \epsilon} E_{\theta'} \|f_\theta - f_{\theta'}\|_n^2 \leq C n^{-1}.$$ Consequently,
$$\limsup_{n \to \infty} \sup_{\theta' \in \Theta, \|\theta' - \theta\| \leq \epsilon} nR(\theta', \hat{\theta}^S) < \infty.$$ 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 $\theta_\Delta$, $\hat{\theta}^S = \hat{\theta}^K$ with arbitrarily high probability even though $\theta_\Delta \notin \Theta^K$. To see this, recall that $\|f_\theta - f_{\theta_\Delta}\|_n^2 \leq C \log n/n$ under $\hat{B}_n$. Therefore, it follows from Lemma 2 that
$$\max \left( |\alpha_{\theta}^- - \alpha_{\theta_\Delta}^-|, |\alpha_{\theta}^+ - \alpha_{\theta_\Delta}^+| \right) \leq C' (\log n/n)^{1/3}.$$ Since $|\alpha_{\theta_\Delta}^+ - \alpha_{\theta_\Delta}^-| = 2\Delta = 2n^{-1/4}$, we have that $|\alpha_{\theta}^+ - \alpha_{\theta}^-| \leq 3n^{-1/4} < n^{-1/6}$ for sufficiently large $n$. Therefore, by construction, $\hat{\theta}^S = \hat{\theta}^K$. Next, let’s fix $\gamma' \in (1/4, 1/3)$. Our goal is to show that $\hat{\theta}^K \notin [\theta_\Delta - n^{-\gamma'}, \theta_\Delta + n^{-\gamma'}]$ under $\hat{B}_n$ for any sufficiently large $n$. We prove this by contradiction. Suppose that $\hat{\theta}^K \in [\theta_\Delta - n^{-\gamma'}, \theta_\Delta + n^{-\gamma'}]$, then
$$\sum_{i=1}^{n} \{Y_{ni} - f_{\hat{\theta}^K}(X_{ni})\}^2 \leq \sum_{i=1}^{n} \{\sigma \varepsilon_{ni} + f_{\theta_\Delta}(X_{ni}) - f_{\hat{\theta}^K}(X_{ni})\}^2$$
$$= \sigma^2 \varepsilon_{ni}^2 + \|f_{\hat{\theta}^K} - f_{\theta_\Delta}\|^2 - 2\sigma \langle \varepsilon_{ni}, f_{\hat{\theta}^K} - f_{\theta_\Delta} \rangle$$
$$\geq \sigma^2 \varepsilon_{ni}^2 + \|f_{\hat{\theta}^K} - f_{\theta_\Delta}\|^2 - 12\sigma \sqrt{6 \log n} \|f_{\hat{\theta}^K} - f_{\theta_\Delta}\|$$
$$\geq \sigma^2 \varepsilon_{ni}^2 + \|f_{\hat{\theta}^K} - f_{\theta_\Delta}\|^2 / 2$$
$$\geq \sigma^2 \varepsilon_{ni}^2 + cn^{1/2}$$
for some $c > 0$. Here the third last inequality holds under $\hat{B}_n$, the second inequality is due to the fact that $\|f_{\hat{\theta}^K} - f_{\theta_\Delta}\| \geq 24\sigma \sqrt{6 \log n}$ 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 $\theta_\Delta$,
$$\sum_{i=1}^{n} \{Y_{ni} - f_{\hat{\theta}^K}(X_{ni})\}^2 \leq \sum_{i=1}^{n} \{Y_{ni} - f_{\theta}(X_{ni})\}^2 \leq \sigma^2 \varepsilon_{ni}^2 + \|f_{\theta} - f_{\theta_\Delta}\|^2 + 2\sigma^2 \langle \varepsilon_{ni}, f_{\theta} - f_{\theta_\Delta} \rangle$$
$$\leq \sigma^2 \varepsilon_{ni}^2 + \|f_{\theta} - f_{\theta_\Delta}\|^2 + 12\sigma \sqrt{6 \log n} \|f_{\theta} - f_{\theta_\Delta}\|$$
$$\leq \sigma^2 \varepsilon_{ni}^2 + cn^{1/4}$$
for some $c > 0$, where the second last inequality holds under $\hat{B}_n$, while the last inequality is derived from straightforward computation as illustrated below for sufficiently large $n$:
$$\|f_{\theta} - f_{\theta_\Delta}\|^2 \leq n \|f_{\theta} - f_{\theta_\Delta}\|_n^2 \leq nC \Delta^3 \leq Cn^{1-3\gamma} = Cn^{1/4}.$$
Putting things together, we see that

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

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

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

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 $\gamma$ and $\gamma'$ such that $1/6 < \gamma < \gamma' < 1/3$. In particular, we could take both of them to be sufficiently close to $1/6$ (so $\Delta$ 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 $\hat{\theta}^S$ satisfies

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

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

REFERENCES


