SELECTING THE NUMBER OF CHANGE-POINTS IN SEGMENTED LINE REGRESSION

Hyune-Ju Kim[1,], Binbing Yu[2], and Eric J. Feuer[3]

[1]Syracuse University, [2]National Institute of Aging, and [3]National Cancer Institute

Supplementary Material

This note contains proofs for Theorems 3.2.1 and 3.2.2.

Appendix A: Proof of Theorem 3.2.1

Lemma A.1. Suppose that conditions (A1) and (A2) in Assumption 3.2.1 are satisfied and that $k^* \leq M$ for a positive fixed constant M. Then $0 < c = c_n(i, j; \alpha) = O(1/\sqrt{n})$ when $j - i = O(1)$.

Lemma A.2. Suppose that the assumptions in Lemma A.1 are satisfied. Then, for $i < k^*$, $P(A_{i,k*;\alpha}|\kappa = k^*)$ converges to zero as $n \to \infty$.

Lemma A.3. Suppose that the assumptions in Lemma A.1 are satisfied. Then, for $j > k^*$, $P(R_{k*,j;\alpha}|\kappa = k^*)$ converges to zero as $n \to \infty$.

Proof of Theorem 3.2.1. First, note from (2) that

$$
\begin{aligned}
P(\hat{\kappa} < k^* \mid \kappa = k^*) &= \sum_{j=0}^{k^*-1} P(\hat{\kappa} = j \mid \kappa = k^*) \\
&\leq \sum_{j=0}^{k^*-1} \sum_{k_0=0}^{j} d_{k_0} P(A_{k_0,k*;\alpha} \,|\kappa = k^*) \\
&\leq \left(\sum_{j=0}^{k^*-1} \sum_{k_0=0}^{j} d_{k_0} \right) \max_{i=0,\ldots,k^*-1} P(A_{i,k*;\alpha}|\kappa = k^*) \\
&= g_1(k^*, M) \max_{i=0,\ldots,k^*-1} P(A_{i,k*;\alpha}|\kappa = k^*),
\end{aligned}
$$

where $g_1(k^*, M)$ is a positive function of k^* and M. Lemma A.2 then provides the result that the under-fitting probability converges to zero $n \to \infty$..

Now, based on (3), we see that

$$\begin{aligned} P(\hat{\kappa} > k^* \mid \kappa = k^*) &= \sum_{j=k^*+1}^{M} P(\hat{\kappa} = j \mid \kappa = k^*) \\ &\le \sum_{j=k^*+1}^{M} \sum_{k_1=j}^{M} d_{k_1} P(R_{k*,k_1;\alpha} \mid \kappa = k^*) \\ &\le \left(\sum_{j=k^*+1}^{M} \sum_{k_1=j}^{M} d_{k_1} \right) \max_{j=k^*+1,\ldots,M} P(R_{k*,k_1;\alpha} | \kappa = k^*) \\ &= g_2(k^*, M) \max_{j=k^*+1,\ldots,M} P(R_{k*,k_1;\alpha} | \kappa = k^*), \end{aligned}$$

where $g_2(k^*, M)$ is a positive function of k^* and M. Lemma A.3 then provides the result that the over-fitting probability also converges to zero as $n \to \infty$.

Proof of Lemma A.1. Note that in testing $H_0 : \kappa = i$ against $H_1 : \kappa = j$ for $i < j$,

$$\begin{aligned} \alpha &= P(RSS(i) \ge (1+c)RSS(j) \mid \kappa = i) \\ &= P_i \left(\frac{Z_{1,n} + Z_{2,n} + R_n}{\hat{\sigma}_j^2 / \sigma_0^2} \ge \frac{n-2-2j}{\sqrt{2(n-2-2i)}} \, c \right), \end{aligned}$$

where $Z_{1,n} = \frac{RSS(i)/\sigma_0^2 - (n-2-2i)}{\sqrt{2(n-2-2i)}}$, $Z_{2,n} = \frac{\sum_{l=1}^n (\epsilon_l^2 - \sigma_0^2)}{n\sigma_0^2} \, \frac{n-2-2j}{\sqrt{2(n-2-2i)}}$, and

$$R_n = \sqrt{\frac{n-2-2i}{2}} \left(1 - \frac{n-2-2j}{n-2-2i} \, \frac{\hat{\sigma}_j^2 - \sum_{l=1}^n \epsilon_l^2/n - \sigma_0^2}{\sigma_0^2} \right).$$

Since $\hat{\sigma}_j^2 - \sum_{l=1}^n \epsilon_l^2/n = O_p\left((\ln n)^2/n\right)$ for $j > i$ from Lemma 5.4 of Liu et al., $R_n = O_p\left((\ln n)^2/\sqrt{n}\right)$. Since the $\hat{\tau}$'s are consistent under the null model of $\kappa = i$ by Proposition 5.1 of Liu et al. and $\hat{\sigma}_j^2$ converges to σ_0^2 in probability, we see that $(Z_{1,n}+Z_{2,n}+R_n)/(\hat{\sigma}_j^2/\sigma_0^2)$ converges in distribution to a normal distribution with mean zero and finite variance. Thus for α fixed and i and j fixed, $c = O(1/\sqrt{n})$.

Proof of Lemma A.2. Note that for $\hat{\sigma}_i^2 = RSS(i)/(n-2-2i)$ and $0 < b_n = (1+c)(n-2-2k^*)/(n-2-2i) - 1 = O(1/\sqrt{n})$,

$$P(A_{i,k*;\alpha} | \kappa = k^*) = P(\hat{\sigma}_i^2 \le (1+b_n) \; \hat{\sigma}_{k^*}^2 | \kappa = k^*)$$

$$= \quad P_{k^*}(\hat\sigma_i^2 > \sigma_0^2 + C,\ \hat\sigma_i^2 \le (1+b_n)\ \hat\sigma_{k^*}^2) + P_{k^*}(\hat\sigma_i^2 \le \sigma_0^2 + C,\ \hat\sigma_i^2 \le (1+b_n)\ \hat\sigma_{k^*}^2)$$
$$= \quad P_1 + P_2,$$

where C is a positive constant in Lemma 5.4 of Liu et al. (1997) for which $P_{k^*}(\hat\sigma_i^2 > \sigma_0^2 + C) \to 1$ as $n \to \infty$. Since $\hat\sigma_{k^*}^2 - \sigma_0^2 = o_p(1)$, $b_n = O(1/\sqrt{n})$ and $C > 0$, we get for $\kappa = k^*$,

$$P_1 = P_{k^*}(\hat\sigma_i^2 > \sigma_0^2 + C,\ \hat\sigma_i^2 \le (1+b_n)\ \hat\sigma_{k^*}^2) \le P_{k^*}(\hat\sigma_{k^*}^2 - \sigma_0^2 > C - b_n\hat\sigma_{k^*}^2)$$

which converges to zero. Also, for $i < k^*$,

$$P_2 = P_{k^*}(\hat\sigma_i^2 < \sigma_0^2 + C,\ \hat\sigma_i^2 \le (1+b_n)\ \hat\sigma_{k^*}^2) \le P_{k^*}(\hat\sigma_i^2 < \sigma_0^2 + C),$$

and thus P_2 converges to zero by Lemma 5.4 of Liu et al.

Proof of Lemma A.3. Note that for b_n as in the proof of Lemma A.2,

$$P(R_{k*,j;\alpha}|\kappa = k^*) = P(\hat\sigma_{k^*}^2 > (1+b_n)\ \hat\sigma_j^2|\kappa = k^*) = P_{k^*}(\hat\sigma_{k^*}^2 - \hat\sigma_j^2 > b_n\ \hat\sigma_j^2).$$

From Lemma 5.4 of Liu et al. (1997), for $j > k^*$, $0 < \hat\sigma_{k^*}^2 - \hat\sigma_j^2 = O_p((\ln n)^2/n)$ and $\hat\sigma_j^2 = \sigma_0^2 + o_p(1)$. Since $0 < b_n = O(1/\sqrt{n})$,

$$P_{k^*}(\hat\sigma_{k^*}^2 - \hat\sigma_j^2 > b_n\ \hat\sigma_j^2) \to 0 \quad \text{as} \quad n \to 0.$$

Appendix B: Proof of Theorem 3.2.2

Lemma B.1. Suppose that conditions (C1), (C2) and (C3) in Assumption 3.2.2 are satisfied. Then the $\eta_i = \boldsymbol{\mu}^{*T}(I - H_i(\boldsymbol{\tau}_{k^*}))\boldsymbol{\mu}^*$ satisfy the followings:

(i) η_i is a decreasing function of i.

(ii) $1/\eta_{k^*-1} = O(\ln n/n)$.

Lemma B.2. Suppose that the assumptions in Lemma B.1 are satisfied. Then $c = c_n$ can be determined such that $c = o(1)$, $\sqrt{n}c_n = O(\sqrt{\ln n})$ and $\alpha_0/M_n = 1 - \Phi(\sqrt{n}\, c) + o(1/M_n)$, where Φ is the standard normal distribution function.

Lemma B.3. Suppose that the assumptions in Lemma B.1 are satisfied. For $i < k^*$, $H_{k^*}(\boldsymbol{\tau}_{k^*}) - H_i(\boldsymbol{\tau}_{k^*})$ is idempotent.

Lemma B.4. Suppose that the assumptions in Lemma B.1 are satisfied. For $i < k^*$,

$$P(A_{i,k*;\alpha}|\kappa = k^*) \leq P\left(Z_i + \frac{\mathbf{y}^T(B_1 + B_2 + B_3)\mathbf{y}}{2\sigma_0\sqrt{\eta_i}} > \frac{\sqrt{\eta_i}}{2\sigma_0}\right),$$

where $B_1 = H_{k^*}(\boldsymbol{\tau}_{k^*}) - H_{k^*}(\hat{\boldsymbol{\tau}}_{k^*})$, $B_2 = c(I - H_{k^*}(\hat{\boldsymbol{\tau}}_{k^*}))$, $B_3 = H_i(\hat{\boldsymbol{\tau}}_i) - H_i(\boldsymbol{\tau}_{k^*})$, and

$$Z_i = \frac{-2\boldsymbol{\mu}^{*T}(I - H_i(\boldsymbol{\tau}_{k^*}))\boldsymbol{\epsilon}}{2\sigma_0\sqrt{\eta_i}},$$

for $\boldsymbol{\epsilon} = \boldsymbol{y} - E(\boldsymbol{y}|\boldsymbol{x}, \kappa = k^*)$.

Lemma B.5. Suppose that the assumptions in Lemma B.1 are satisfied. For $i < k^*$, $V_{i,n} = \mathbf{y}^T(B_1 + B_2 + B_3)\mathbf{y}/(2\sigma_0\sqrt{\eta_i}) = O_p(\ln n) - d_{i,n}$, where $d_{i,n}$ is a positive constant.

Lemma B.6. For $j > k^*$,

$$P(R_{k*,j;\alpha}|\kappa = k^*) \leq P\left(Z_j^R + \frac{\mathbf{y}^T(B_1^R + B_2^R + B_3^R)\mathbf{y}}{2\sigma_0\sqrt{\eta_j}} > \frac{\sqrt{\eta_j}}{2\sigma_0}\right),$$

where $B_1^R = H_{k^*}(\boldsymbol{\tau}_{k^*}) - H_{k^*}(\hat{\boldsymbol{\tau}}_{k^*})$, $B_2^R = -c(I - H_{k^*}(\hat{\boldsymbol{\tau}}_{k^*}))$, $B_3^R = H_j(\hat{\boldsymbol{\tau}}_j) - H_j(\boldsymbol{\tau}_{k^*})$, and

$$Z_j^R = \frac{-2\boldsymbol{\mu}^{*T}(I - H_j(\boldsymbol{\tau}_{k^*}))\boldsymbol{\epsilon}}{2\sigma_0\sqrt{\eta_j}},$$

for $\boldsymbol{\epsilon} = \boldsymbol{y} - E(\boldsymbol{y}|\boldsymbol{x}, \kappa = k^*)$. Also $V_{j,n}^R = \mathbf{y}^T(B_1^R + B_2^R + B_3^R)\mathbf{y}/(2\sigma_0\sqrt{\eta_j}) = O_p(\ln n) - d_{j,n}^R$, where $d_{j,n}^R$ is a positive constant.

Proof of Theorem 3.2.2.

We first show that $P(\hat{\kappa} < k^*|\kappa = k^*) \to 0$ as $n \to \infty$. Note that for $V_{i,n} = \mathbf{y}^T(B_1 + B_2 + B_3)\mathbf{y}/(2\sigma_0\sqrt{\eta_i})$ $(i < k^*)$,

$$P(A_{i,k*;\alpha}\,|\kappa = k^*) \quad \leq \quad P\left(Z_i + V_{i,n} + d_{i,n} \geq \sqrt{\eta_i}/(2\sigma_0) + d_{i,n}\right)$$

$$\begin{aligned} &= P(e^{\tilde{Z}_i+\tilde{V}_{i,n}} \geq e^{\sqrt{\tilde{\eta}_i}/(2\sigma_0)}) \\ &\leq E(e^{\tilde{Z}_i+\tilde{V}_{i,n}})/e^{\sqrt{\tilde{\eta}_i}/(2\sigma_0)}, \end{aligned}$$

where $\tilde{Z}_i = Z_i/\ln n$, $\tilde{V}_{i,n} = (V_{i,n} + d_{i,n})/\ln n$, and $\sqrt{\tilde{\eta}_i} = \sqrt{\eta_i}/\ln n$, and the last inequality is obtained by Markov's inequality. Then,

$$\begin{aligned} P(\hat{\kappa} < k^* \mid \kappa = k^*) &= \sum_{j=0}^{k^*-1} P(\hat{\kappa} = j \mid \kappa = k^*) \\ &\leq \sum_{j=0}^{k^*-1} \sum_{k_0=0}^{j} d_{k_0} P(A_{k_0,k*;\alpha} \mid \kappa = k^*) \\ &\leq \left(\sum_{j=0}^{k^*-1} \sum_{k_0=0}^{j} d_{k_0}\right) \left(\max_{i=0,\ldots,k^*-1} \frac{E(e^{\tilde{Z}_i+\tilde{V}_{i,n}})}{e^{\sqrt{\tilde{\eta}_i}/(2\sigma_0)}}\right) \\ &\leq k^{*k^*} \left(\max_{j=0,\ldots,k^*-1} \binom{M}{j}\right) \left(\max_{i=0,\ldots,k^*-1} \frac{E(e^{\tilde{Z}_i+\tilde{V}_{i,n}})}{e^{\sqrt{\tilde{\eta}_i}/(2\sigma_0)}}\right) \\ &\leq k^{*k^*} \; M^{k^*-1} \, \frac{\max_{i=0,\ldots,k^*-1} E(e^{\tilde{Z}_i+\tilde{V}_{i,n}})}{\min_{i=0,\ldots,k^*-1} e^{\sqrt{\tilde{\eta}_i}/(2\sigma_0)}} \\ &\leq k^{*k^*} \, \frac{M^{k^*-1}}{e^{\sqrt{\tilde{\eta}^*}/(2\sigma_0)}} \max_{i=0,\ldots,k^*-1} E(e^{\tilde{Z}_i+\tilde{V}_{i,n}}) \\ &\leq g(k^*) \left(\frac{M}{\sqrt{\eta^*}}\right)^{k^*-1} \left(\frac{(\ln n)^2}{\sqrt{\eta^*}}\right)^{k^*-1} \max_{i=0,\ldots,k^*-1} E(e^{\tilde{Z}_i+\tilde{V}_{i,n}}), \end{aligned}$$

where $g(k^*)$ is a positive function of k^*. Since Z_i converges to a standard normal distribution and $\tilde{V}_{i,n} = O_p(1)$, and $\frac{(\ln n)^2}{\sqrt{\eta^*}} = o(1)$, the upper bound will converge to zero under a mild condition on M such as the one described in Assumption 3.2.2 (C3).

Then, using Lemma B.6, we can show that the over-fitting probability also converges to zero as $n \to \infty$.

Proof of Lemma B.1. Let $X_{i+1}(\boldsymbol{t}) = (X_i(\boldsymbol{t})\ \boldsymbol{x}_{i+1}(\boldsymbol{t}))$, where $\boldsymbol{x}_{i+1}(\boldsymbol{t}) = ((x_1 - t_{i+1})^+, \ldots, (x_n - t_{i+1})^+)^T$. Note that $\eta_i = \boldsymbol{\mu}^{*T}(I - H_i(\boldsymbol{\tau}_{k^*}))\boldsymbol{\mu}^*$ is a decreasing function of i, which can be proved by showing that

$$(I - H_i(\boldsymbol{t})) - (I - H_{i+1}(\boldsymbol{t})) = (I - H_i(\boldsymbol{t})) \left[\frac{\boldsymbol{x}_{i+1}(\boldsymbol{t})\boldsymbol{x}_{i+1}^T(\boldsymbol{t})}{a_{i+1}^{22}}\right] (I - H_i(\boldsymbol{t})) > 0,$$

where $a_{i+1}^{22} = \boldsymbol{x}_{i+1}^T(\boldsymbol{t})(I - H_i(\boldsymbol{t}))\boldsymbol{x}_{i+1}(\boldsymbol{t})$.

Thus, for $X_{k^*-1} = X_{k^*-1}(\boldsymbol{\tau}_{k^*})$, $\boldsymbol{x}_{k^*} = \boldsymbol{x}_{k^*}(\boldsymbol{\tau}_{k^*})$, $\boldsymbol{\mu}^* = \boldsymbol{\mu}(\boldsymbol{\tau}_{k^*})$ and $H_i = H_i(\boldsymbol{\tau}_{k^*})$,

$$
\begin{aligned}
\eta^* &= \min_{i<k^*} \eta_i = \eta_{k^*-1} = (\boldsymbol{\mu}^*)^T (I - H_{k^*-1}) \boldsymbol{\mu}^* \\
&= (\boldsymbol{\mu}^*)^T \left(I - H_{k^*} + (I - H_{k^*-1}) \left[\frac{\boldsymbol{x}_{k^*} \boldsymbol{x}_{k^*}^T}{a_{k^*}^{22}} \right] (I - H_{k^*-1}) \right) \boldsymbol{\mu}^* \\
&= \boldsymbol{\beta}^T \left(X_{k^*-1} \ \boldsymbol{x}_{k^*} \right)^T (I - H_{k^*-1}) \left[\frac{\boldsymbol{x}_{k^*} \boldsymbol{x}_{k^*}^T}{a_{k^*}^{22}} \right] (I - H_{k^*-1}) \left(X_{k^*-1} \ \boldsymbol{x}_{k^*} \right) \boldsymbol{\beta} \\
&= \delta_{k^*} a_{k^*}^{22} \delta_{k^*} \\
&= \delta_{k^*}^2 \left[\boldsymbol{x}_{k^*}^T (I - H_{k^*-1}) \boldsymbol{x}_{k^*} \right] \\
&= \delta_{k^*}^2 \sum_{m=l_{k^*+1}}^{n} \left\{ \sum_{j=l_{k^*+1}}^{n} (x_j - \tau_{k^*}) b_{mj} \right\} (x_m - \tau_{k^*}),
\end{aligned}
$$

where $(x_{l_{k^*}+1}, \ldots, x_n)$ are the observations in $[\tau_{k^*}, 1]$ and $I - H_{k^*-1} = (b_{mj})$. If we assume that there are at least $n/\ln n$ many observations in each segment of $[\tau_i, \tau_{i+1})$ for $i = 0, \ldots, k^*$, which was motivated by Corollary 3.22 of Feder (1975), then we see that $\eta^* \geq D_1 n/\ln n$, for some positive constant $D_1 > 1$.

Proof of Lemma B.2. Recall that the test proposed in Kim et al. (2000) rejects $H_0 : \kappa = i$ in favor of $H_1 : \kappa = k^*$ at level α if $T = RSS(i)/RSS(k^*) \geq (1 + c)$ for some $c = c_n(i, k^*; \alpha(i, k^*)) > 0$, where $RSS(i) = \mathbf{y}^T (I - H_i(\hat{\boldsymbol{\tau}}_i))\mathbf{y}$ and $RSS(k^*) = \mathbf{y}^T (I - H_{k^*}(\hat{\boldsymbol{\tau}}_{k^*}))\mathbf{y}$. Also, recall that $A_{i,k^*;\alpha}$ is the event that $H_0 : \kappa = i$ is not rejected at level α. Following the argument in the proof of Lemma A.1 and that $\hat{\sigma}_{k^*}^2 - \sigma_0^2 = O_p(1/\sqrt{n})$ in Feder (1975), we see that

$$
\frac{\alpha_0}{M_n} = P(RSS(i) \geq (1 + c) RSS(k^*) | \kappa = i) = P\left(Z + o_p(1) > \sqrt{n} c \right)
$$

for a stable distribution Z. If $\sqrt{n}\, c = D_2 \sqrt{\ln n}$ for some positive constant D_2, $0 < D_2 < 1$, we obtain that $\frac{d}{dn} \frac{1}{M_n}$ is proportional to $-1/(n^{1+D_2^2/2} \sqrt{\ln n})$. If we let $\eta^* = D_1\ (n/\ln n)$ for some constant $D_1 > 1$, then we see that $\frac{d}{dn} \frac{1}{\sqrt{\eta^*}}$ is proportional to $-\sqrt{\ln n}/(n\sqrt{n})$. This implies that such a choice of c satisfies the condition of $M = M_n$ such that $M/\sqrt{\eta^*} \to 0$ as $n \to \infty$.

The proof of Lemma B.3, which is based on lengthy and straightforward matrix algebra, is omitted.

Proof of Lemma B.4.

$$\begin{aligned} P(A_{i,k*;\alpha}|\kappa = k^*) &= P_{k^*}\Big[\ \mathbf{y}^T(I - H_i(\hat{\boldsymbol{\tau}}_i))\mathbf{y} < (1+c)\ \mathbf{y}^T(I - H_{k^*}(\hat{\boldsymbol{\tau}}_{k^*}))\mathbf{y}\ \Big] \\ &= P_{k^*}\Big[\ \mathbf{y}^T(I - H_i(\boldsymbol{\tau}_{k^*}))\mathbf{y} + \mathbf{y}^T(H_i(\boldsymbol{\tau}_{k^*}) - H_i(\hat{\boldsymbol{\tau}}_i))\mathbf{y} \\ &\qquad < (1+c)\ \Big\{\mathbf{y}^T(I - H_{k^*}(\hat{\boldsymbol{\tau}}_{k^*}))\mathbf{y})\Big\}\Big]. \end{aligned}$$

Noting that $\mathbf{y} = \boldsymbol{\mu}^* + \boldsymbol{\epsilon}$ when $\kappa = k^*$ and $(I - H_{k^*}(\boldsymbol{\tau}_{k^*}))\boldsymbol{\mu}^* = 0$, the right hand side is equivalent to

$$\begin{aligned} &P_{k^*}\Big[\ 2\boldsymbol{\mu}^{*T}(I - H_i(\boldsymbol{\tau}_{k^*}))\boldsymbol{\epsilon} < -\boldsymbol{\mu}^{*T}(I - H_i(\boldsymbol{\tau}_{k^*}))\boldsymbol{\mu}^* - \boldsymbol{\epsilon}^T(H_{k^*}(\boldsymbol{\tau}_{k^*}) - H_i(\boldsymbol{\tau}_{k^*}))\boldsymbol{\epsilon} \\ &\quad \mathbf{y}^T(H_{k^*}(\boldsymbol{\tau}_{k^*}) - H_{k^*}(\hat{\boldsymbol{\tau}}_{k^*}))\mathbf{y} + c\ \mathbf{y}^T(I - H_{k^*}(\hat{\boldsymbol{\tau}}_{k^*}))\mathbf{y} + \mathbf{y}^T(H_i(\hat{\boldsymbol{\tau}}_i) - H_i(\boldsymbol{\tau}_{k^*}))\mathbf{y}\ \Big]. \end{aligned}$$

Since $\boldsymbol{\epsilon}^T(H_{k^*}(\boldsymbol{\tau}_{k^*}) - H_i(\boldsymbol{\tau}_{k^*}))\boldsymbol{\epsilon} > 0$ by Lemma A.3,

$$\begin{aligned} P(A_{i,k*;\alpha}|\kappa = k^*) &\leq P\left(-2\boldsymbol{\mu}^{*T}(I - H_i(\boldsymbol{\tau}_{k^*}))\boldsymbol{\epsilon} + \mathbf{y}^T(B_1 + B_2 + B_3)\mathbf{y} > \boldsymbol{\mu}^{*T}(I - H_i(\boldsymbol{\tau}_{k^*}))\boldsymbol{\mu}^*\right) \\ &= P\left(Z_i + \frac{\mathbf{y}^T(B_1 + B_2 + B_3)\mathbf{y}}{2\sigma_0\sqrt{\eta_i}} > \frac{\sqrt{\eta_i}}{2\sigma_0}\right). \end{aligned}$$

Proof of Lemma B.5.

(i) $\mathbf{y}^T B_1\mathbf{y}/(2\sigma_0\sqrt{\eta_i}) = \mathbf{y}^T(H_{k^*}(\boldsymbol{\tau}_{k^*}) - H_{k^*}(\hat{\boldsymbol{\tau}}_{k^*}))\mathbf{y}/(2\sigma_0\sqrt{\eta_i}) = O_p(\sqrt{\ln n})$. This can be obtained by using $\hat{\sigma}^2_{k^*} - \sigma^2_0 = O_p(1/\sqrt{n})$ and $1/\sqrt{\eta_i} \leq 1/\sqrt{\eta^*} = O(\sqrt{\ln n/n})$.

(ii) $\mathbf{y}^T B_2\mathbf{y}/(2\sigma_0\sqrt{\eta_i}) = c\ \mathbf{y}^T(I - H_{k^*}(\hat{\boldsymbol{\tau}}_{k^*}))\mathbf{y}/(2\sigma_0\sqrt{\eta_i}) = O_p(\ln n)$ for a choice of c in Lemma B.2. This can be shown because $c\sqrt{n} = O(\sqrt{\ln n})$, $\sqrt{n/\eta_i} = O(\sqrt{\ln n})$, and $\hat{\sigma}^2_{k^*}$ is a consistent estimator of σ^2_0.

(iii)

$$\begin{aligned} \mathbf{y}^T B_3\mathbf{y}/(2\sigma_0\sqrt{\eta_i}) &= \frac{\mathbf{y}^T(I - H_i(\boldsymbol{\tau}_{k^*}))\mathbf{y}}{2\sigma_0\sqrt{\eta_i}} - \frac{\mathbf{y}^T(I - H_i(\hat{\boldsymbol{\tau}}_i))\mathbf{y}}{2\sigma_0\sqrt{\eta_i}} \\ &= \sqrt{\frac{n\sigma_0^2}{2\eta_i}}\,(Z_{1,n} + Z_{2,n}) - \frac{E_{k^*}[Q_2] - E_{k^*}[Q_1]}{2\sqrt{\eta_i}/\sigma_0}, \end{aligned}$$

where $Q_1 = \mathbf{y}^T(I - H_i(\boldsymbol{\tau}_{k^*}))\mathbf{y}/\sigma_0^2$, $Q_2 = \mathbf{y}^T(I - H_i(\hat{\boldsymbol{\tau}}_i))\mathbf{y}/\sigma_0^2$, $Z_{1,n} = (Q_1 - E_{k^*}[Q_1])/\sqrt{2n}$, and $Z_{2,n} = (Q_2 - E_{k^*}[Q_2])/\sqrt{2n}$. Matrix algebra shows that $(E_{k^*}[Q_2] - E_{k^*}[Q_1])/(2\sqrt{\eta_i}/\sigma_0) = d_{i,n} + O(\sqrt{\ln n/n})$, where $d_{i,n} > 0$. Since each of $Z_{1,n}$ and $Z_{2,n}$ converges to a standard normal distribution and $\sqrt{n/\eta_i} = O(\sqrt{\ln n})$, $\mathbf{y}^T B_3 \mathbf{y}/(2\sigma_0\sqrt{\eta_i}) = O_p(\sqrt{\ln n}) - d_{i,n}$.

Combining (i), (ii) and (iii), we obtain that $V_{i,n} = O_p(\ln n) - d_{i,n}$, where $d_{i,n} > 0$.

Similar arguments used in the proofs of Lemma B.4 and Lemma B.5 would lead to Lemma B.6.