SELECTING THE NUMBER OF CHANGE-POINTS IN SEGMENTED LINE REGRESSION

Hyune-Ju Kim\textsuperscript{1}, Binbing Yu\textsuperscript{2}, and Eric J. Feuer\textsuperscript{3}

\textsuperscript{1}Syracuse University, \textsuperscript{2}National Institute of Aging, and \textsuperscript{3}National Cancer Institute

Supplementary Material

This note contains proofs for Theorems 3.2.1 and 3.2.2, and this is the version revised in 2012.

Appendix A: Proof of Theorem 3.2.1

**Lemma A.1.** Suppose that conditions (A1) and (A2) in Assumption 3.2.1 are satisfied. Then, for $\alpha$ fixed and $j > i$, there exists $c = c_n = c_n(i, j; \alpha) = o(1)$ that asymptotically achieves the level $\alpha$.

**Lemma A.2.** Suppose that the assumptions in Lemma A.1 are satisfied and $c_n = o(1)$. 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, $c_n = o(1)$, and $\frac{n}{(\ln n)^2 c_n} \to \infty$ as $n \to \infty$. 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 (3.1) that

\[
P(\hat{\kappa} < k^* \mid \kappa = k^*) = \sum_{j=0}^{k^*-1} P(\hat{\kappa} = j \mid \kappa = k^*)
\]

\[
\leq \sum_{j=0}^{k^*-1} \sum_{k_0=0}^{j} d_{k_0} P(A_{k_0,k^*\alpha} | \kappa = k^*)
\]

\[
\leq \left( \sum_{j=0}^{k^*-1} \sum_{k_0=0}^{j} d_{k_0} \right) \max_{i=0,\ldots,k^*-1} P(A_{i,k^*\alpha} | \kappa = k^*)
\]

\[
= g_i(k^*, M) \max_{i=0,\ldots,k^*-1} P(A_{i,k^*\alpha} | \kappa = k^*),
\]

where $g_i(k^*, M)$ is a function of $k^*$ and $M$.
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 \( n \to \infty \). Since \( P(\hat{\kappa} > k^* | \kappa = k^* ) \leq \alpha_0 \) by the design of the permutation procedure, in general, we obtain that \( \lim_{n \to \infty} P(\hat{\kappa} = k^* | \kappa = k^* ) \geq 1 - \alpha_0 \).

If \( c = c_n = o(1) \) is chosen such that \( \frac{n}{(\ln n)^2} c_n \to \infty \), then we achieve the desired result by Lemma A.3.

**Proof of Lemma A.1.** Since, for \( j > i(= \kappa) \), \( 0 < \hat{\sigma}^2_j - \hat{\sigma}^2_i = O_p((\ln n)^2/n) \) and \( \hat{\sigma}^2_j \) converges to \( \sigma_0^2 \) in probability from Lemma 5.4 of Liu et al. (1997), where \( \hat{\sigma}^2_j = \text{RSS}(i)/n \) as in Liu et al., there exist \( B_\alpha \) and \( N_\alpha \) such that \( P(\sigma^2_j - \hat{\sigma}^2_j \geq B_\alpha (\ln n)^2/n | \kappa = i ) \leq \alpha \) for all \( n > N_\alpha \). Thus for \( n > N_\alpha \), there exists \( c = c_n \leq B_\alpha (\ln n)^2/n \) such that

\[
\alpha = P(\text{RSS}(i) \geq (1 + c) \text{RSS}(j) | \kappa = i) = P\left( \frac{\hat{\sigma}^2_i - \hat{\sigma}^2_j}{\hat{\sigma}^2_j} \geq c | \kappa = i \right).
\]

**Proof of Lemma A.2.** For \( i < k^* \),

\[
P(A_{i,k^*}; \kappa | \kappa = k^*) = P(\hat{\sigma}^2_i > (1 + c_n) \hat{\sigma}^2_{k^*}; \kappa = k^*) \\
= P_k(\hat{\sigma}^2_i > \sigma_0^2 + C, \hat{\sigma}^2_i < (1 + c_n) \hat{\sigma}^2_{k^*}) + P_k(\hat{\sigma}^2_i \leq \sigma_0^2 + C, \hat{\sigma}^2_i < (1 + c_n) \hat{\sigma}^2_{k^*}) \\
= P_1 + P_2,
\]

where \( C \) is a positive constant in Lemma 5.4 of Liu et al. (1997) for which \( P_k(\hat{\sigma}^2_i > \sigma_0^2 + C) \to 1 \) as \( n \to \infty \). Since \( \hat{\sigma}^2_{k^*} - \sigma_0^2 = o_p(1) \), \( c_n = o(1) \) and \( C > 0 \), we get for \( \kappa = k^* \),

\[
P_1 = P_k(\hat{\sigma}^2_i > \sigma_0^2 + C, \hat{\sigma}^2_i < (1 + c_n) \hat{\sigma}^2_{k^*}) \leq P_k(\hat{\sigma}^2_i - \sigma_0^2 > C - c_n \hat{\sigma}^2_{k^*})
\]

which converges to zero. Also,

\[
P_2 = P_k(\hat{\sigma}^2_i \leq \sigma_0^2 + C, \hat{\sigma}^2_i < (1 + c_n) \hat{\sigma}^2_{k^*}) \leq P_k(\hat{\sigma}^2_i \leq \sigma_0^2 + C),
\]

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

**Proof of Lemma A.3.** Note that

\[
P(\text{R}_{k^*; i, n} | \kappa = k^*) = P(\hat{\sigma}^2_{k^*} \geq (1 + c_n) \hat{\sigma}^2_j; \kappa = k^*) = P_k(\hat{\sigma}^2_{k^*} - \hat{\sigma}^2_j \geq c_n \hat{\sigma}^2_j).
\]
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 \( \hat{\sigma}_j^2 = \sigma_0^2 + o_p(1) \). If \( c_n = o(1) \) is chosen such that \( \frac{n}{(\ln n)^2} c_n \to \infty \),

\[
P_{k^*}(\hat{\sigma}_k^2 - \hat{\sigma}_j^2 \geq c_n \hat{\sigma}_j^2) = P_{k^*}\left(\frac{\hat{\sigma}_k^2 - \hat{\sigma}_j^2}{\hat{\sigma}_j^2} \geq c_n \frac{n}{(\ln n)^2}\right) \to 0 \quad \text{as} \quad n \to \infty.
\]

**Appendix B: Proof of Theorem 3.2.2**

Note that in this revision, the conditions (C1) and (C2) in Assumption 3.2.2 are replaced by (A1) and (A2) of Assumption 3.2.1.

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

(i) \( \eta_i \) is a decreasing function of \( i \).

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

**Lemma B.2.** Suppose that the assumptions in Lemma B.1 are satisfied. For \( \alpha_0 \) fixed and \( j > i \), there exists \( c = c_n = c_n(i,j; \alpha_0/M_n) \) that asymptotically achieves the level \( \alpha_0/M_n \), where \( M_n/\sqrt{\eta^*} \to 0 \) as \( n \to \infty \).

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

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

\[
P(A_{i,k^*+\alpha} | \kappa = k^*) \leq P\left(Z_{i,n} + \frac{y^T(B_1 + B_2 + B_3)\epsilon}{2\sigma_0\sqrt{\eta_i}} > \frac{\sqrt{\eta_i}}{2\sigma_0}\right),
\]

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

\[
Z_{i,n} = \frac{-2\mu^T(I - H_i(\tau_{k^*})\epsilon}{2\sigma_0\sqrt{\eta_i}},
\]

for \( \epsilon = y - E(y|x, \kappa = k^*) \).
Lemma B.5. Suppose that the assumptions in Lemma B.1 are satisfied. For \( i < k^* \),
\[ V_{i,n} = y^T (B_1 + B_2 + B_3) y / (2\sigma_0 \sqrt{\eta_i}) = O_p(\sqrt{\ln n}) + h_{i,n}, \]
where \( \sqrt{\eta_i} = O(1) \) and \( h_{i,n} \leq \gamma_{i,n} \sqrt{\eta_i} / (2\sigma_0) \) for \( \gamma_{i,n} \) such that \( 0 < \lim_{n \to \infty} (1 - \gamma_{i,n}) \leq 1 \).

Proof of Theorem 3.2.2.

We first show that \( P(\hat{k} < k^* | \kappa = k^*) \to 0 \) as \( n \to \infty \). Note that for \( V_{i,n} = y^T (B_1 + B_2 + B_3) y / (2\sigma_0 \sqrt{\eta_i}) \) \( (i < k^*) \),
\[
P(A_{i,k^*:n} | \kappa = k^*) \leq P(Z_{i,n} + V_{i,n} - h_{i,n} \geq (1 - \gamma_{i,n}) \sqrt{\eta_i} / (2\sigma_0))
\leq P(e^{\hat{Z}_{i,n} + \hat{V}_{i,n}} \geq e^{\sqrt{\eta_i} / (2\sigma_0)})
\leq E(e^{\hat{Z}_{i,n} + \hat{V}_{i,n}}) / e^{\sqrt{\eta_i} / (2\sigma_0)},
\]
where \( \hat{Z}_{i,n} = Z_{i,n}/((1 - \gamma_{i,n}) \ln n), \hat{V}_{i,n} = (V_{i,n} - h_{i,n})/((1 - \gamma_{i,n}) \ln n), \) and \( \sqrt{\eta_i} = \sqrt{\eta_i} / \ln n \),
and the last inequality is obtained by Markov’s inequality. Then,
\[
P(\hat{k} < k^* | \kappa = k^*) = \sum_{j=0}^{k^*-1} P(\hat{k} = j | \kappa = k^*)
\leq \sum_{j=0}^{k^*-1} \sum_{k_0=0}^{j} d_{k_0} P(A_{k_0,k^*:n} | \kappa = k^*)
\leq \left( \sum_{j=0}^{k^*-1} \sum_{k_0=0}^{j} d_{k_0} \right) \left( \max_{i=0,\ldots,k^*-1} E(e^{\hat{Z}_{i,n} + \hat{V}_{i,n}}) \right)
\leq g_2(k^*) \left( \max_{j=0,\ldots,k^*-1} \left( \frac{M^j}{j!} \right) \right) \left( \max_{i=0,\ldots,k^*-1} \frac{E(e^{\hat{Z}_{i,n} + \hat{V}_{i,n}})}{e^{\sqrt{\eta_i} / (2\sigma_0)}} \right)
\leq g_2(k^*) M^{k^*-1} \max_{i=0,\ldots,k^*-1} E(e^{\hat{Z}_{i,n} + \hat{V}_{i,n}})
\leq g_2(k^*) M^{k^*-1} \frac{e^{\sqrt{\eta_i} / (2\sigma_0)}}{e^{\sqrt{\eta_i} / (2\sigma_0)}} \max_{i=0,\ldots,k^*-1} E(e^{\hat{Z}_{i,n} + \hat{V}_{i,n}})
\leq \tilde{g}_2(k^*) \left( \frac{M}{\sqrt{\eta}} \right)^{k^*-1} \left( \frac{(\ln n)^2}{\sqrt{\eta}} \right)^{k^*-1} \max_{i=0,\ldots,k^*-1} E(e^{\hat{Z}_{i,n} + \hat{V}_{i,n}}),
\]
where \( \tilde{g}_2(k^*) \) is a positive function of \( k^* \). Since \( \hat{Z}_{i,n} + \hat{V}_{i,n} = o_p(1) \) 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, by using \( P(\hat{k} > k^* | \kappa = k^*) \leq \alpha_0 \), we can show that \( \lim_{n \to \infty} P(\hat{k} = k^* | \kappa = k^*) \geq 1 - \alpha_0 \). Similarly as in Theorem 3.2.1, by choosing \( c = c_n \) such that \( \sqrt{n}c_n = O(1) \) and the corresponding \( \alpha_0 \) approaches to zero, we can achieve the desired result.

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

\[
(I - H_i(t)) - (I - H_{i+1}(t)) = (I - H_i(t)) \left[ \frac{x_{i+1}(t)x^T_{i+1}(t)}{a_{i+1}^2} \right] (I - H_i(t)) > 0,
\]

where \( a_{i+1}^2 = x^T_{i+1}(t)(I - H_i(t))x_{i+1}(t) \).

Thus, for \( X_{k^*-1} = X_{k^*-1}(\tau_{k^*}) \), \( x_{k^*} = x_{k^*}(\tau_{k^*}) \), \( \mu^* = \mu(\tau_{k^*}) \) and \( H_i = H_i(\tau_{k^*}) \),

\[
\eta^* = \min_{i < k^*} \eta_i = \eta_{k^*-1} = (\mu^*)^T(I - H_{k^*-1})\mu^*
\]

\[
= (\mu^*)^T \left( I - H_{k^*} + (I - H_{k^*-1}) \left[ \frac{x_{k^*}x^T_{k^*}}{a_{k^*}^2} \right] (I - H_{k^*-1}) \right) \mu^*
\]

\[
= \beta^T \left( X_{k^*-1} \; x_{k^*} \right)^T (I - H_{k^*-1}) \left[ \frac{x_{k^*}x^T_{k^*}}{a_{k^*}^2} \right] (I - H_{k^*-1}) (X_{k^*-1} \; x_{k^*}) \beta
\]

\[
= \delta_{k^*}a_{k^*}^2 \delta_{k^*}
\]

\[
= \delta_{k^*}^2 \left[ x^T_{k^*}(I - H_{k^*-1})x_{k^*} \right]
\]

\[
= \delta_{k^*}^2 \sum_{m=k^*+1}^n \left\{ \sum_{j=k^*+1}^n (x_j - \tau_{k^*})b_{mj} \right\} (x_m - \tau_{k^*}),
\]

where \( (x_{k^*+1}, \ldots, x_n) \) are the observations in \( [\tau_{k^*}, 1] \) and \( I - H_{k^*-1} = (b_{mj}) \). Under (C1), it can be shown that for large \( n \), \( \eta^* \geq D_1n / \ln n \), where \( D_1 \) is a positive constant.

**Proofs of Lemma B.2. and Lemma B.3.**

The proof of Lemma B.3, which is based on lengthy and straightforward matrix algebra, is omitted, and the proof of Lemma B.2. is sketched below.

Suppose that for some \( a_n > 0 \) such that \( a_n \to \infty \) as \( n \to \infty \), \( Z_n = a_n \frac{\hat{\sigma}_2 - \hat{\sigma}_2^*}{\hat{\sigma}_2^*} \), under the null hypothesis of \( \kappa = i \), converges in distribution to \( Z \) with a cumulative distribution
function $F(\cdot)$ and the probability density function $f(\cdot)$. We then see that for $j > i$,

$$\frac{\alpha_0}{M_n} = P(RSS(i) \geq (1 + c_n)RSS(j)|\kappa = i) = P(Z_n \geq \hat{c}_n) \approx 1 - F(\hat{c}_n),$$

where $\hat{c}_n = a_n c_n$. Since $\frac{d}{dn} \frac{1}{M_n}$ is proportional to $-f(\hat{c}_n) \frac{d}{dn} \hat{c}_n$ and $\frac{d}{dn} g_n$ is proportional to $-\frac{\sqrt{n}}{n\sqrt{n}}$, where $1/\sqrt{n^2} \leq \sqrt{\frac{n}{B_n}} = g_n$, a slowly increasing function of $n$, $\hat{c}_n$, such that $\frac{\sqrt{n}}{n\sqrt{n}}/f(\hat{c}_n) \frac{d}{dn} \hat{c}_n \to 0$ as $n \to \infty$ satisfies the condition of $M = M_n$ such that $M/\sqrt{\eta^2} \to 0$ as $n \to \infty$. Using that $Z_n/a_n = O_p \left( \frac{M_n(\ln n)^2}{n} \right)$, it can also be shown that for appropriately chosen $c_n$, $\sqrt{n} c_n = O(1)$ since $\sqrt{n} c_n = \frac{\hat{c}_n}{a_n/\sqrt{n}}$ where $\hat{c}_n$ is slowly increasing and $a_n/\sqrt{n} \to \infty$ at least as fast as $\sqrt{n}/\{M_n(\ln n)^2\}$ does as $n \to \infty$. For example, if $f$ is a chi-square density with finite degrees of freedom, then $c_n$ such that $\hat{c}_n = a_n c_n = D_2 \ln n$ for $0 < D_2 < 1$ can be used.

**Proof of Lemma B.4.**

$$P(A_{i,k^*} | \kappa = k^*) = P_{k^*} \left[ y^T(I - H_i(\hat{\tau}_i))y < (1 + c) y^T(I - H_{k^*}(\hat{\tau}_{k^*}))y \right]$$

$$= P_{k^*} \left[ y^T(I - H_i(\tau_{k^*}))y + y^T(H_i(\tau_{k^*}) - H_i(\hat{\tau}_i))y < (1 + c) \{ y^T(I - H_{k^*}(\hat{\tau}_{k^*}))y \} \right].$$

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

$$P_{k^*} \left[ 2\mu^* y^T(I - H_i(\tau_{k^*})) \epsilon < -\mu^* y^T(I - H_i(\tau_{k^*}))\mu^* - \epsilon^T(H_{k^*}(\tau_{k^*}) - H_i(\tau_{k^*}))\epsilon \right. \left. y^T(H_{k^*}(\tau_{k^*}) - H_{k^*}(\hat{\tau}_{k^*}))y + c y^T(I - H_{k^*}(\hat{\tau}_{k^*}))y + y^T(H_i(\hat{\tau}_i) - H_i(\tau_{k^*}))y \right].$$

Since $\epsilon^T(H_{k^*}(\tau_{k^*}) - H_i(\tau_{k^*}))\epsilon > 0$ by Lemma B.3,

$$P(A_{i,k^*} | \kappa = k^*) \leq P \left( -2\mu^* y^T(I - H_i(\tau_{k^*})) \epsilon + y^T(B_1 + B_2 + B_3)y \epsilon^T(I - H_i(\tau_{k^*}))\epsilon \right)$$

$$= P \left( Z_{i,n} + \frac{y^T(B_1 + B_2 + B_3)y}{2\sigma_0^2/\sqrt{\eta_i}} > \frac{\sqrt{\eta_i}}{2\sigma_0} \right).$$
Proof of Lemma B.5.

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

(ii) \( y^T B_2 y / (2\sigma_0 \sqrt{\eta_i}) = c y^T (I - H_{k^*}(\hat{\tau}_{k^*})) y / (2\sigma_0 \sqrt{\eta_i}) = O_p(\sqrt{\ln n}) \) for a choice of \( c = c_n \) such that \( c\sqrt{n} = O(1) \). This can be shown because \( \sqrt{n / \eta_i} = O(\sqrt{\ln n}) \) and \( \hat{\sigma}_{k^*}^2 \) is a consistent estimator of \( \sigma_0^2 \).

(iii)
\[
y^T B_3 y / (2\sigma_0 \sqrt{\eta_i}) = \frac{y^T (I - H_i(\tau_{k^*})) y}{2\sigma_0 \sqrt{\eta_i}} - \frac{y^T (I - H_i(\hat{\tau}_i)) y}{2\sigma_0 \sqrt{\eta_i}} = \sqrt{\frac{n\sigma_0^2}{2\eta_i}} (Z_{1,n} - Z_{2,n}) + \frac{E_{k^*}[Q_1] - E_{k^*}[Q_2]}{2\sqrt{\eta_i}/\sigma_0},
\]
where \( Q_1 = y^T (I - H_i(\tau_{k^*})) y / \sigma_0^2 \), \( Q_2 = y^T (I - H_i(\hat{\tau}_i)) y / \sigma_0^2 \), \( Z_{1,n} = (Q_1 - E_{k^*}[Q_1]) / \sqrt{2n} \), and \( Z_{2,n} = (Q_2 - E_{k^*}[Q_2]) / \sqrt{2n} \). Matrix algebra shows that \( (E_{k^*}[Q_1] - E_{k^*}[Q_2]) / (2\sqrt{\eta_i}/\sigma_0) = h_{i,n} + O(\sqrt{\ln n}) \), where \( h_{i,n} \leq \gamma_{i,n} \sqrt{\eta_i} / (2\sigma_0) \) for \( \gamma_{i,n} \) such that \( 0 < \lim_{n \to \infty} (1 - \gamma_{i,n}) \leq 1 \). Since \( Z_{1,n} - Z_{2,n} = O_p(1) \) and \( \sqrt{n / \eta_i} = O(\sqrt{\ln n}) \), \( y^T B_3 y / (2\sigma_0 \sqrt{\eta_i}) = O_p(\sqrt{\ln n}) + h_{i,n} \).

Combining (i), (ii) and (iii), we obtain that \( V_{i,n} = O_p(\sqrt{\ln n}) + h_{i,n} \).