We use the following notation in the proofs:

\[ Q_n(\beta) = n^{-1} \sum_{i<j} b_{ij} |e_i - e_j| + n \sum_{j=1}^d p'_\lambda(|\beta^0_j|)|\beta_j| \]
\[ D_n(\beta) = n^{-1} \sum_{i<j} b_{ij} |e_i - e_j| \]
\[ S_n(\beta) = n^{-1} \sum_{i<j} b_{ij} (x_i - x_j) \text{sgn}((Y_i - Y_j) - (x_i - x_j)'\beta) \]
\[ A_n(\beta) = (2\sqrt{3}\tau)^{-1}(\beta - \beta^0)'X'WX(\beta - \beta^0) - (\beta - \beta^0)'S_n(\beta_0) + D_n(\beta_0), \]

where \( \text{sgn}(x) \) stands for the sign of \( x \).

We first present and prove two useful lemmas about the unpenalized weighted Wilcoxon estimator under possible local contamination. These results will be useful later to establish the asymptotic properties of the penalized Wilcoxon estimator. In the proof of the two lemmas, we frequently refer to the book of Hettmansperger and McKean (1998), abbreviated as HM in the sequel.

**Lemma 0.1** Assume the regularity conditions in Section 3.1, then \( \forall \epsilon > 0, \forall c > 0, \)

\[
\left[ \sup_{\sqrt{n}\|\beta - \beta^0\| \leq c} |D_n(\beta) - A_n(\beta)| \geq \epsilon \right] \overset{p}{\rightarrow} 0
\]

under either \( H \) or \( H^*_n \).

**Proof.** The result under \( H \) was given in Sievers (1983), see also Section 5.2 of HM. To prove it under \( H^*_n \), let \( U_n(t) = n^{-1/2}[S_n(\beta_0 + t/\sqrt{n}) - S_n(\beta^0_0)]. \) Then \( U_n(t) - (\sqrt{3}\tau)^{-1}Ct = o_p(1) \) under \( H \). Since \( H^*_n \) is contiguous with respect to \( H \), by Le Cam’s first lemma (see,
for example, Chapter 6 of van der Vaart, 1998),

\[ U_n(t) - (\sqrt{3}\tau)^{-1}Ct \overset{p}{\rightarrow} 0 \quad \text{under } H^*_n. \]  

(2)

Let \( D_n(t) = D_n(\beta_0 + t/\sqrt{n}) \) and \( A_n(t) = A_n(\beta_0 + t/\sqrt{n}) \), then (2) implies that

\[ \nabla(D_n(t) - A_n(t)) \overset{p}{\rightarrow} 0 \quad \text{under } H^*_n. \]

Using the diagonal subsequencing argument (see the proof of Theorem A.3.7 of HM), we can show that \( D_n(t) - A_n(t) \overset{a.s.}{\rightarrow} 0 \) under \( H^*_n \) for all rational \( t \) and \( n \in \tilde{N} \), an infinite set of positive integers defined on page 414 of HM. Let \( J_n(t) = D_n(t) - D_n(0) + \frac{tn}{\sqrt{n}} - \frac{1}{2} S_n(\beta_0) \), then

\[ \{ J_n(t) - (2\sqrt{3}\tau)^{-1}t'(n^{-1}X'WX)t \} \overset{a.s.}{\rightarrow} 0 \quad \text{under } H^*_n \] uniformly on each compact subset of \( \mathbb{R}^d \). By the way \( \tilde{N} \) is constructed, Theorem 4 of Tucker (1967, page 103) implies that \( D_n(t) - A_n(t) \overset{p}{\rightarrow} 0 \) under \( H^*_n \) uniformly on each compact subset of \( \mathbb{R}^d \). Equation (1) follows by considering the compact subset \( \{ t : |t| \leq c \} \). \( \square \)

**Lemma 0.2** Assume the regularity conditions in Section 3.1, then \( n^{-1/2}S_n(\beta_0) \overset{d}{\rightarrow} N(0, V/3) \) under \( H \) and \( n^{-1/2}S_n(\beta_0) \overset{d}{\rightarrow} N(\eta, V/3) \) under \( H^*_n \), where \( \eta \) is defined in Theorem 2.

**Proof.** The result under \( H \) was given in Sievers (1983), see also Section 5.2 of HM. Note that \( S_n(\beta_0) = n^{-1}\sum_{i<j}b_{ij}(x_i - x_j)\text{sgn}(\epsilon_i - \epsilon_j) \) and it’s straightforward to check that the projection of \( S_n(\beta_0) \) under \( H \) is \( T_n(\beta_0) = n^{-1}\sum_{i=1}^n \sum_{k=1}^n b_{ki}(x_k - x_i)[2F(\epsilon_k) - 1] \).

Since \( n^{-1/2}[S_n(\beta_0) - T_n(\beta_0)] \overset{p}{\rightarrow} 0 \) under \( H \), applying Le Cam’s first lemma, we obtain

\[ n^{-1/2}[S_n(\beta_0) - T_n(\beta_0)] \overset{p}{\rightarrow} 0 \quad \text{under } H^*_n. \]
Thus it's sufficient to derive the asymptotic distribution of $n^{-1/2}T_n(\beta_0)$ under $H^*_n$.

\[
E_{H^*_n}\left[n^{-1/2}T_n(\beta_0)\right] = n^{-3/2} \sum_{i=1}^{n} \sum_{k=1}^{n} \iint b(x_1, x_2)(x_1 - x_2)[2F(y_1 - x'_1\beta_0) - 1]dH^*_n(x_1, y_1)dH^*_n(x_2, y_2)
\]

\[
= n^{-3/2} \left(1 - \frac{\epsilon}{\sqrt{n}}\right)^2 \iint b(x_1, x_2)(x_1 - x_2)[2F(y_1 - x'_1\beta_0) - 1]dH(x_1, y_1)dH(x_2, y_2)
\]

\[
+ n^{-3/2} \left(1 - \frac{\epsilon}{\sqrt{n}}\right) \frac{\epsilon}{\sqrt{n}} \sum_{i=1}^{n} \sum_{k=1}^{n} \iint b(x_1, x_2)(x_1 - x_2)[2F(y_1 - x'_1\beta_0) - 1]dH(x_1, y_1)dH(x_2, y_2)
\]

\[
+ n^{-3/2} \left(1 - \frac{\epsilon}{\sqrt{n}}\right) \frac{\epsilon}{\sqrt{n}} \sum_{i=1}^{n} \sum_{k=1}^{n} \iint b(x_1, x_2)(x_1 - x_2)[2F(y_1 - x'_1\beta_0) - 1]d\Delta(x'_1, y'_1)(x_1, y_1)dH(x_2, y_2)
\]

\[
+ n^{-5/2} \epsilon^2 \sum_{i=1}^{n} \sum_{k=1}^{n} \iint b(x_1, x_2)(x_1 - x_2)[2F(y_1 - x'_1\beta_0) - 1]d\Delta(x'_1, y'_1)(x_1, y_1)d\Delta(x'_2, y'_2)(x_2, y_2)
\]

\[
= \epsilon [2F(y^* - x^'\beta_0) - 1] \int b(x^*, x)(x^* - x)dM(x) + o_p(1).
\]

And

\[
Var_{H^*_n}\left[n^{-1/2}T_n(\beta_0)\right] = n^{-3} \sum_{k=1}^{n} E_{H^*_n}\left\{ \sum_{i=1}^{n} b_{ki}(x_k - x_i)[2F(\epsilon_k) - 1] \right\}^2
\]

\[
- n^{-3} \left\{ E_{H^*_n}\left[ \sum_{i=1}^{n} \sum_{k=1}^{n} b_{ki}(x_k - x_i)[2F(\epsilon_k) - 1] \right] \right\}^2
\]

\[
= n^{-3} \sum_{k=1}^{n} E_H\left\{ \sum_{i=1}^{n} b_{ki}(x_k - x_i)[2F(\epsilon_k) - 1] \right\}^2 + o(1)
\]

\[
= n^{-3} \sum_{k=1}^{n} E_H\left[ [2F(\epsilon_k) - 1]^2 \sum_{i=1}^{n} \sum_{j=1}^{n} b_{ki}b_{kj}(x_k - x_i)'(x_k - x_j) \right] + o(1)
\]

\[
= (3n)^{-1} E_H\left\{ \sum_{k=1}^{n} \sum_{i=1}^{n} \sum_{j=1}^{n} w_{ki}w_{kj}(x_k - x_i)'(x_k - x_j) \right\} + o(1)
\]

\[
\rightarrow V/3.
\]
Finally, to prove the asymptotic normality, note that we can write

\[ n^{-1/2} T_n(\beta_0) = \sum_{k=1}^{n} \left[ n^{-3/2} \sum_{i=1}^{n} b_{ki}(x_k - x_i) \right] [2F(\epsilon_k) - 1]. \]

Conditional on \( X \) first, this is a sum of independent but not identically distributed random variables. We can establish the conditional normality via the Lindeberg-Feller central limit theorem. The unconditional normality follows by Slutsky’s theorem. □

We next derive the asymptotic properties of the WW-SCAD. The proof is similar to that of Fan and Li (2001) but is more challenging since the function \( D_n \) is nonsmooth. Lemma 0.3 below shows that the WW-SCAD estimator is \( \sqrt{n} \)-consistent. Lemma 0.4 below suggests that the WW-SCAD estimator must possess the sparsity property. These two lemmas prepare us for the proof of Theorem 1 and Theorem 2.

**Lemma 0.3** Assume the regularity conditions in Section 3.1. If \( \lambda_n \to 0 \), then the WW-SCAD estimator \( \hat{\beta} \) satisfies \( ||\hat{\beta} - \beta_0|| = O_p(n^{-1/2}) \).

**Proof.** We will show that \( \forall \epsilon > 0 \), there exists a large constant \( C \) such that

\[
P \left( \inf_{||u|| = C} Q_n(\beta_0 + n^{-1/2}u) > Q_n(\beta_0) \right) \geq 1 - \epsilon,
\]

where \( u = (u_1, \ldots, u_d)' \). Since \( Q_n(\beta) \) is convex in \( \beta \), this implies that with probability at least \( 1 - \epsilon \) that the WW-SCAD estimator lies in the ball \( \{ \beta_0 + n^{-1/2}u : ||u|| \leq C \} \).

Let \( G_n(u) = Q_n(\beta_0 + n^{-1/2}u) - Q_n(\beta_0) \) and

\[
H_n = A_n(\beta_0 + n^{-1/2}u) - A_n(\beta_0) + n \sum_{j=1}^{d} p'_{\lambda_n}(|\beta_0^j|)(|\beta_{j0} + n^{-1/2}u_j| - |\beta_{j0}|).
\]

Then by Lemma 0.1, \( G_n(u) - H_n(u) \mathop{\to}^{p} 0 \) uniformly on \( \{ u : ||u|| \leq C \} \). It is sufficient to show that with probability approaching one, \( H_n(u) \) is positive for sufficiently large
\[ H_n(u) = (2\sqrt{3})^{-1}u'[n^{-1}X'WX]u - u'n^{-1/2}S_n(\beta_0) + n \sum_{j=1}^{d} p'_{\lambda_n}(|\beta_j^0|)(|\beta_j + n^{-1/2}u_j| - |\beta_j|) \]

\[ \geq (2\sqrt{3})^{-1}u'[n^{-1}X'WX]u - u'n^{-1/2}S_n(\beta_0) - \sqrt{n} \sum_{j=1}^{s} p'_{\lambda_n}(|\beta_j^0|)|u_j| \]  

(3)

Note that \( n^{-1}X'WX \overset{p}{\rightarrow} C \), a positive definite matrix, and \( n^{-1/2}S_n(\beta_0) = O_p(1) \) by Lemma 0.2. Furthermore, \( p'_{\lambda_n}(|\beta_j^0|) = p'_{\lambda_n}(|\beta_j^0|)I(|\beta_j^0| \leq a\lambda_n) \). Thus for any \( \epsilon > 0 \), \( P(\sqrt{n}p'_{\lambda_n}(|\beta_j^0|) > \epsilon) \leq P(|\beta_j^0| \leq a\lambda_n) \rightarrow 0 \) by the fact \( |\beta_j^0| - a\lambda_n \overset{p}{\rightarrow} |\beta_{0j}| > 0 \) for \( j = 1, \ldots, s \). This implies that \( \sqrt{n}p'_{\lambda_n}(\beta_j^0) = o_p(1) \). Therefore, for \( n \) sufficiently large, the first term on the right-hand side of (3) asymptotically dominates, which can be made positive and large with sufficiently large \( C \).

\[ \Box \]

**Lemma 0.4** Assume the regularity conditions in Section 3.1. If \( \lambda_n \rightarrow 0 \) and \( \sqrt{n}\lambda_n \rightarrow \infty \) as \( n \rightarrow \infty \), then with probability tending to one, for any \( \beta_1 \) satisfying \( ||\beta_1 - \beta_{10}|| = O_p(n^{-1/2}) \) and any constant \( C \),

\[ Q \left\{ \begin{pmatrix} \beta_1 \\ 0 \end{pmatrix} \right\} = \max_{||\beta_2|| \leq Cn^{-1/2}} Q \left\{ \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix} \right\} . \]

**Proof.** Since \( Q_n(\beta) \) is a convex, piecewise linear and almost everywhere differentiable function of \( \beta \), it suffices to show that with probability tending to one, for any \( \beta_1 \) satisfying \( ||\beta_1 - \beta_{10}|| = O_p(n^{-1/2}) \) and for any small \( \epsilon_n = Cn^{-1/2} \),

\[ \frac{\partial Q_n(\beta)}{\partial \beta_j} \begin{cases} > 0, & 0 < \beta_j < \epsilon_n \\ < 0, & -\epsilon_n < \beta_j < 0, \end{cases} \]
at any differentiable point $\beta$, for $j = s + 1, \ldots, d$. Note that

$$n^{-1/2} \frac{\partial Q_n(\beta)}{\partial \beta_k} = -n^{-3/2} \sum_{i < j} b_{ij}(x_{ik} - x_{jk}) \text{sgn}((Y_i - Y_j) - (X_i - X_j)'\beta) + n^{1/2} p'_\lambda(|\beta_k|) \text{sgn}(\beta_k).$$

By Lemma 0.2 and the regularity conditions, the first term on the right side is $O_p(1)$. Furthermore,

$$p'_\lambda_n(|\beta_0_k|) \lambda_n = p'_\lambda_n(|\beta_0_k|) \lambda_n I(|\beta_0_k| \leq \lambda_n) + p'_\lambda_n(|\beta_0_k|) \lambda_n I(|\beta_0_k| > \lambda_n) = 1 + p'_\lambda_n(|\beta_0_k|) \lambda_n I(|\beta_0_k| > \lambda_n).$$

Thus for any $\epsilon > 0$, $P(|p'_\lambda_n(|\beta_0_k|)/\lambda_n - 1| > \epsilon) \leq P(|\beta_0_k| > \lambda_n) = P(\sqrt{n}|\beta_0_k| > \sqrt{n}\lambda_n) \rightarrow 0$ by the fact that $\sqrt{n}|\beta_0_k|$ is bounded in probability (because $\sqrt{n}(|\beta_0_k| - \beta_{k0})$ is asymptotically normal) and $\sqrt{n}\lambda_n \rightarrow \infty$. This implies that $p'_\lambda_n(|\beta_0_k|)/\lambda_n \xrightarrow{p} 1$ and thus $n^{1/2} p'_\lambda_n(|\beta_0_k|) = (n^{1/2}\lambda_n)(p'_\lambda_n(|\beta_0_k|)/\lambda_n) \xrightarrow{p} \infty$ as $n \rightarrow \infty$. Therefore, the sign of the derivative is completely determined by that of $\beta_k$. This completes the proof. □

**Proof of Theorem 1.** It follows from Lemma 0.4 that part (i) holds. Below, we prove part (ii). Let $\tilde{\beta} = (\tilde{\beta}_1, \tilde{\beta}_2)'$ be the minimizer of

$$Q_n^*(\beta) = A_n(\beta) + n \sum_{j=1}^d p'_\lambda(|\beta_j^0|)|\beta_j|. \quad (4)$$

Similarly as in the proof of Lemma 0.3 and Lemma 0.4, we can show that $\tilde{\beta}$ is $\sqrt{n}$-consistent and $P(\tilde{\beta}_2 = 0) \rightarrow 1$ as $n \rightarrow \infty$. We next prove the asymptotic normality of $\tilde{\beta}_1$. With probability approaching one,

$$\lim_{n \rightarrow \infty} \frac{\partial Q_n^*(\beta)}{\partial \beta} \bigg|_{\beta = (\tilde{\beta}_1, 0)'} = 0.$$
Consider the first $s$-dimension of the above derivative, we obtain

$$(\sqrt{3}\tau)^{-1}(\hat{\beta}_1 - \beta_{10})'(X'WX)_{11} - S_{n1}(\beta_0) + n(p'_{\lambda_n}(|\beta_1^0|)sgn(\beta_1), \ldots, p'_{\lambda_n}(|\beta_s^0|)sgn(\beta_s))' = 0_s,$$

where $(X'WX)_{11}$ denotes the $s \times s$ submatrix in the upper-left corner of $X'WX$, and $S_{n1}(\beta_0)$ is the first $d$-dimension of $S_n(\beta_0)$. Thus

$$\sqrt{n}(\hat{\beta}_1 - \beta_{10}) = \sqrt{3}\tau(X'WX)^{-1}_{11}[n^{-1/2}S_{n1}(\beta_0) + \sqrt{n}(p'_{\lambda_n}(|\beta_1^0|)sgn(\beta_1), \ldots, p'_{\lambda_n}(|\beta_s^0|)sgn(\beta_s))'] \
\xrightarrow{d} N_s(0_s, \tau^2 C^{-1}_{11}V_{11}C_{11}).$$

We will finish the proof by showing $\sqrt{n}(\hat{\beta} - \tilde{\beta}) \overset{p}{\rightarrow} 0$, which implies that $\sqrt{n}(\hat{\beta}_1 - \tilde{\beta}_1) \overset{p}{\rightarrow} 0$. This is done using a convexity argument due to Jaeckel (1972), see also the proof of A.3.9. of HM, which we outline below. Choose $\epsilon > 0$ and $\delta > 0$. Since $\sqrt{n}(\hat{\beta} - \beta_0) = O_p(1)$, there exists a $C_0$ such that

$$P(||\beta - \tilde{\beta}|| \geq C_0n^{-1/2}) < \delta/2, \quad (5)$$

for $n$ sufficiently large. Let

$$T = \min\{Q_n^*(\beta) : ||\beta - \tilde{\beta}|| = \epsilon n^{-1/2}\} - Q_n^*(\tilde{\beta}). \quad (6)$$

Since $\tilde{\beta}$ minimizes $Q_n^*(\beta)$, $T > 0$. Hence by Lemma 0.1,

$$P\left[\max_{||\beta - \beta_0|| < (C_0 + \epsilon)n^{-1/2}}|Q_n(\beta) - Q_n^*(\beta)| \geq T/2\right] \leq \delta/2, \quad (7)$$

for sufficiently large $n$. By (5) and (6), with probability greater than $1 - \delta$, $Q_n^*(\tilde{\beta}) < Q_n(\tilde{\beta}) + T/2$ and $||\tilde{\beta} - \beta_0|| < C_0n^{-1/2}$ for sufficiently large $n$. Next, consider $\beta$ such that $||\beta - \tilde{\beta}|| = \epsilon n^{-1/2}$. For $||\tilde{\beta} - \beta_0|| < C_0n^{-1/2}$, it follows that $||\beta - \beta_0|| \leq (C_0 + \epsilon)n^{-1/2}$. 

7
Arguing as above, we have with probability greater than $1 - \delta$ that $Q_n(\beta) > Q_n^*(\beta) - T/2$ for sufficiently large $n$. From this, (6) and (7), we obtain that

$$Q_n(\beta) > Q_n^*(\beta) - T/2$$

$$\geq \min\{Q_n^*(\beta) : ||\beta - \tilde{\beta}|| = \epsilon n^{-1/2}\} - T/2$$

$$= T + Q_n^*(\tilde{\beta}) - T/2 = T/2 + Q_n^*(\tilde{\beta}) > Q_n(\tilde{\beta}).$$

Thus $Q_n(\beta) > Q_n(\tilde{\beta})$ for $||\beta - \tilde{\beta}|| = \epsilon n^{-1/2}$. Since $Q_n$ is convex, it follows that $Q_n(\beta) > Q_n(\tilde{\beta})$ for $||\beta - \tilde{\beta}|| > \epsilon n^{-1/2}$. But $Q_n(\beta) > Q_n(\tilde{\beta})$ since $\tilde{\beta}$ minimizes $Q_n$. Hence $\tilde{\beta}$ must lie inside the disk $||\beta - \tilde{\beta}|| = \epsilon n^{-1/2}$ with probability at least $1 - 2\delta$. That is $P(||\beta - \tilde{\beta}|| < \epsilon n^{-1/2}) > 1 - 2\delta$. This yields the results.

**Proof of Theorem 2.** First note that the conclusions of Lemmas 0.3 and 0.4 also hold under $H_n^*$. This is because all the $O_p$ and $o_p$ terms in the proofs of the two lemmas remain their orders under $H_n^*$. Next we mimic the proof of Theorem 1. It’s clear that part (i) holds. To prove part (ii), let $\tilde{\beta} = (\tilde{\beta}_1, \tilde{\beta}_2)'$ be the minimizer of $Q_n^*(\beta)$ in (4).

With probability approaching one (under $H_n^*$),

$$\frac{\partial Q_n^*(\beta)}{\partial \beta} \bigg|_{\beta = (\tilde{\beta}_1, 0)'} = 0.$$

Consider the first $s$-dimension of the above derivative, we obtain

$$(\sqrt{3}\tau)^{-1}(\tilde{\beta}_1 - \beta_{10})'(X'WX)_{11} - S_{n1}(\beta_0) + n(p_1^1(|\beta_1^0|)sgn(\beta_1), \ldots, p_s^1(|\beta_s^0|)sgn(\beta_s))' = 0_s.$$

By Lemma 0.2, we have under $H_n^*$,

$$\sqrt{n}(\tilde{\beta}_1 - \beta_{10}) = \sqrt{3}\tau(X'WX)^{-1}_{11} \left[ n^{-1/2} S_{n1}(\beta_0) + \sqrt{n}(p_1^1(|\beta_1^0|)sgn(\beta_1), \ldots, p_s^1(|\beta_s^0|)sgn(\beta_s))' \right]$$

$$\xrightarrow{d} N_d(\eta, \tau^2C_{11}^{-1}V_{11}C_{11}).$$
Finally, the same convexity arguments yields that $\sqrt{n}(\bar{\beta}_1 - \tilde{\beta}_1) = o_p(1)$ under $H_n^*$. 

**Proof of Theorem 3.** The proof proceeds as in Wang, Li and Tsai (2001). First, similarly as in their Lemma 3, $P(BIC_{\lambda_n} = BIC_{S_T}) \to 1$, which implies $BIC_{\lambda_n} \overset{p}{\to} \log(L_{S_T})$. Next we verify that $P(\inf_{\lambda \in \Omega_- \cup \Omega_+} BIC_{\lambda} > BIC_{\lambda_n}) \to 1$. This is done by considering two separate cases.

**Case 1: Underfitted model**, i.e., the model misses at least one covariate in the true model. For any $\lambda \in \Omega_-$, we have

$$BIC_{\lambda} = \log\left(n^{-2} \sum_{i<j} b_{ij}|(Y_i - X'_i\hat{\beta}_{\lambda}) - (Y_j - X'_j\hat{\beta}_{\lambda})|\right) + df_{\lambda} \log(n)/n \geq \log\left(n^{-2} \sum_{i<j} b_{ij}|(Y_i - X'_i\hat{\beta}_{S_{\lambda}}) - (Y_j - X'_j\hat{\beta}_{S_{\lambda}})|\right) \geq \inf_{S \neq S_T} \log\left(n^{-2} \sum_{i<j} b_{ij}|(Y_i - X'_i\hat{\beta}_S) - (Y_j - X'_j\hat{\beta}_S)|\right) \overset{p}{\to} \inf_{S \neq S_T} \log(L_n^S) > \log(L_{S_T})$$

in probability, where in the third step $\hat{\beta}_{S_{\lambda}}$ is the unpenalized weighted Wilcoxon estimator for model $S_{\lambda}$.

**Case 2: Overfitted model**, i.e., the model contains all the covariates in the true model and at least one covariate that does not belong to the true model. For any $\lambda \in \Omega_+$, we have $\frac{\sqrt{T}}{\tau}[D_n(\beta_{S_T}) - D_n(\beta_{S_{\lambda}})] \to \sum_{i=1}^q \gamma_i^2(1)$, where the $\gamma_i$'s are positive constants and $q$ is a positive integer, and $\chi^2(1), \ldots, \chi^2(1)$ are i.i.d. $\chi^2$ random variables each with one degree of freedom (Theorem 5.2.12 of HM, 1998). Thus for any overfitted model $S$,
\[ D_n(\hat{\beta}_{S_T}) - D_n(\hat{\beta}_S) = O_p(1) \]. With probability approaching one,

\[
n(BIC_\lambda - BIC_{\lambda_n}) = n \log \left( \frac{D_n(\hat{\beta}_\lambda)}{D_n(\hat{\beta}_{\lambda_n})} \right) + (df_\lambda - df_{\lambda_n}) \log n
\]

\[
= \{ [n^{-1} D_n(\hat{\beta}_{\lambda_n})]^{-1} (D_n(\hat{\beta}_\lambda) - D_n(\hat{\beta}_{\lambda_n})) + o_p(1) \} + (df_\lambda - s + o_p(1)) \log n
\]

\[
\geq (\log(L^{S_T}))^{-1} (D_n(\hat{\beta}_{S_T}) - D_n(\hat{\beta}_{S_T})) + o_p(1) + (1 + o_p(1)) \log n.
\]

With probability approaching one,

\[
\inf_{\lambda \in \Omega_+} n(BIC_\lambda - BIC_{\lambda_n}) \geq (\log(L^{S_T}))^{-1} \min_{S \supseteq S_T} (D_n(\hat{\beta}_S) - D_n(\hat{\beta}_{S_T})) + o_p(1) + (1 + o_p(1)) \log n.
\]

The first term on the right-hand side of the above is \( O_p(1) \) and the last term diverges to \(+\infty\) as \( n \to \infty \), which implies that \( P(\inf_{\lambda \in \Omega_+} n(BIC_\lambda - BIC_{\lambda_n}) > 0) \to 1 \). □

**Additional References**
