Large dimensional empirical likelihood

Binbin Chen[1], Guangming Pan[1], Qing Yang[1], Wang Zhou[2]

[1]*Nanyang Technological University, Singapore, 637371*

[2]*National University of Singapore, Singapore, 117546*

Supplementary Material

In this supplementary document, we first provide detailed proofs for Step 1 to Step 4 required in proving Theorem 2.1. At the end of this document we prove Theorem 2.1, which is Step 5 in the main paper.

Lemma S1. Under the assumptions of Theorem 2.1,

$$\lambda'\mathbf{u} = o_p(s^{-1}). \tag{S.1}$$

Proof: Note that

$$\mathbf{y}_{n+1} - \beta = -s\mathbf{u} \quad \text{and} \quad \mathbf{y}_{n+2} - \beta = -2\beta + s\mathbf{u} = (2r+s)\mathbf{u}. \tag{S.2}$$

From the constraint in (A.1), we have

$$0 = \sum_{i=1}^{n+2} \omega_i(\mathbf{y}_i - \beta) = \sum_{i=1}^{n} \omega_i(\mathbf{y}_i - \beta) + \omega_{n+1}(-s\mathbf{u}) + \omega_{n+2}(2r+s)\mathbf{u}.$$

As $\mathbf{u}$ is the direction of $-\beta$ hence the unit vector, multiplying both sides by $\mathbf{u}'$ we obtain

$$s(\omega_{n+1} - \omega_{n+2}) = \sum_{i=1}^{n} \omega_i \mathbf{u}'(\mathbf{y}_i - \beta) + 2r\omega_{n+2} \triangleq I_1 + I_2.$$

Consider I_1 first. By the Hölder inequality and the fact that $\sum_{i=1}^n \omega_i^2 \le \sum_{i=1}^n \omega_i \le 1$,

$$\begin{aligned} |I_1| &\le \Big(\sum_{i=1}^{n} \omega_i^2\Big)^{1/2} \Big(\sum_{i=1}^{n} \mathbf{u}'(\mathbf{y}_i - \beta)(\mathbf{y}_i - \beta)'\mathbf{u}\Big)^{1/2} \\ &\le \sqrt{n}\Big(\mathbf{u}'\mathbf{S}_1\mathbf{u}\Big)^{1/2} = O_p(\sqrt{n}), \end{aligned}$$

where

$$\mathbf{S}_1 = \frac{1}{n}\sum_{i=1}^{n}(\mathbf{y}_i-\beta)(\mathbf{y}_i-\beta)' = \mathbf{A}^{-1}\frac{1}{n}\sum_{i=1}^{n}(\mathbf{x}_i-\mu)(\mathbf{x}_i-\mu)'(\mathbf{A}^{-1})' = \mathbf{A}^{-1}\Gamma\Big(\frac{1}{n}\sum_{i=1}^{n}\mathbf{z}_i\mathbf{z}_i'\Big)\Gamma'(\mathbf{A}^{-1})'.$$

The last equality follows from the assumption 2 in Theorem 2.1 and the facts that all the eigenvalues of $\mathbf{A}$ (Jiang (2004), Xiao and Zhou (2010)) and $\frac{1}{n}\sum_{i=1}^{n}\mathbf{z}_i\mathbf{z}_i'$ (Bai and Yin (1993)) are bounded from above and from below by positive constants in probability, hence $c_0 \le \|\mathbf{S}_1\| \le C_0$ in probability for some c_0 and C_0.

Consider I_2. Let

$$r^2 \triangleq \frac{T^2}{n} = (\bar{\mathbf{x}}-\mu)'\mathbf{S}^{-1}(\bar{\mathbf{x}}-\mu) = \bar{\mathbf{z}}'\Big(\frac{1}{n-1}\sum_{i=1}^{n}(\mathbf{z}_i-\bar{\mathbf{z}})(\mathbf{z}_i-\bar{\mathbf{z}})'\Big)^{-1}\bar{\mathbf{z}}. \tag{S.3}$$

By Theorem 1 of Pan and Zhou (2011),

$$r^2 = c_n(1-c_n)^{-1} + O_p(1/\sqrt{n}). \tag{S.4}$$

Thus $I_2 = O_p(1)$. Therefore,

$$s(\omega_{n+1}-\omega_{n+2}) = O_p(\sqrt{n}). \tag{S.5}$$

It follows from (2.4), (A.2), (A.6) and (A.9) that

$$\frac{1}{1-s\lambda'\mathbf{u}} - \frac{1}{1+(2r+s)\lambda'\mathbf{u}} = \frac{n+2}{s}\cdot s(\omega_{n+1}-\omega_{n+2}) = O_p\Big(\frac{n\sqrt{n}}{s}\Big) \xrightarrow{i.p.} 0. \tag{S.6}$$

For any $\epsilon > 0$, if $s\lambda'\mathbf{u} > \epsilon$, then

$$\frac{1}{1-s\lambda'\mathbf{u}} - \frac{1}{1+(2r+s)\lambda'\mathbf{u}} \ge \frac{1}{1-\epsilon} - \frac{1}{1+\epsilon} > 0,$$

which would contradict (S.6). Similarly, $s\lambda'\mathbf{u} < -\epsilon$ would also lead to a contradiction. Hence, $|s\lambda'\mathbf{u}| \le \epsilon$ in probability for any $\epsilon > 0$ which implies

$$\lambda'\mathbf{u} = o_p(s^{-1}).$$

Lemma S2. Under the assumptions of Theorem 2.1,

$$\|\lambda\| = o_p(s^{-1/2}), \qquad \max_{i\le n}|\lambda'(\mathbf{y}_i-\beta)| = o_p\Big(\sqrt{\frac{n}{s}}\Big). \tag{S.7}$$

Proof: Let $\lambda = \rho\theta$, where $\rho = \|\lambda\|$. From the model assumption (2.3) in Theorem 2.1, we have almost surely,

$$\begin{aligned}\max_{i\le n}|\theta'(\mathbf{y}_i-\beta)|^2 &= \max_{i\le n}|\theta'\mathbf{A}^{-1}\Gamma\mathbf{z}_i|^2 \le |\theta'\mathbf{A}^{-1}\Gamma\Gamma'(\mathbf{A}^{-1})'\theta|\max_{i\le n}|\mathbf{z}_i'\mathbf{z}_i| \\ &\le K\cdot p\max_{i\le n}\left|p^{-1}\sum_{j=1}^{p}\left(z_{ij}^2-1\right)\right| + Kp \qquad \text{(S.8)} \\ &= O_p(n).\end{aligned}$$

Here (and in what follows) K denotes a constant which may change from line to line and z_{ij} are the i.i.d components of $\mathbf{z}_i$. In the last step, we apply Lemma 5.2 in the appendix of the main paper with $X_{ij} = z_{ij}^2$ and $\alpha = \beta = 1$. By equation (S.2) and Lemma S1, we also have

$$|\lambda'(\mathbf{y}_{n+1}-\beta)| = |s\lambda'\mathbf{u}| = o_p(1), \qquad |\lambda'(\mathbf{y}_{n+2}-\beta)| = |(2r+s)\lambda'\mathbf{u}| = o_p(1). \qquad \text{(S.9)}$$

Recalling the identity (A.3) $\sum_{i=1}^{n+2}\frac{\mathbf{y}_i-\beta}{1+\lambda'(\mathbf{y}_i-\beta)} = 0$, by the formula $\frac{1}{1+x} = 1-\frac{x}{1+x}$ and the fact that $\sum_{i=1}^{n+2}\mathbf{y}_i = \sum_{i=1}^{n}\mathbf{y}_i = \mathbf{0}$, we have

$$\begin{aligned}0 = \sum_{i=1}^{n+2}\frac{\lambda'(\mathbf{y}_i-\beta)}{1+\lambda'(\mathbf{y}_i-\beta)} &= \sum_{i=1}^{n+2}\lambda'(\mathbf{y}_i-\beta) - \rho^2\sum_{i=1}^{n+2}\frac{\theta'(\mathbf{y}_i-\beta)(\mathbf{y}_i-\beta)'\theta}{1+\lambda'(\mathbf{y}_i-\beta)} \\ &= (n+2)r\lambda'\mathbf{u} - \rho^2\sum_{i=1}^{n+2}\frac{\theta'(\mathbf{y}_i-\beta)(\mathbf{y}_i-\beta)'\theta}{1+\lambda'(\mathbf{y}_i-\beta)},\end{aligned}$$

then, via (S.9)

$$\begin{aligned}\left|\frac{n+2}{n}r\lambda'\mathbf{u}\right| &= \frac{\rho^2}{n}\sum_{i=1}^{n+2}\frac{\theta'(\mathbf{y}_i-\beta)(\mathbf{y}_i-\beta)'\theta}{1+\lambda'(\mathbf{y}_i-\beta)} \\ &\ge \frac{\rho^2}{n}\frac{\sum_{i=1}^{n}\theta'(\mathbf{y}_i-\beta)(\mathbf{y}_i-\beta)'\theta + s^2(\theta'\mathbf{u})^2 + (2r+s)^2(\theta'\mathbf{u})^2}{1+\rho\max_{i\le n}|\theta'(\mathbf{y}_i-\beta)| + |s\lambda'\mathbf{u}| + |(2r+s)\lambda'\mathbf{u}|} \\ &\ge \frac{\rho^2\theta'\mathbf{S}_1\theta}{1+o_p(1)+\rho\max_{i\le n}|\theta'(\mathbf{y}_i-\beta)|}.\end{aligned}$$

It follows that

$$\rho^2\theta'\mathbf{S}_1\theta - \rho(1+o(1))|r\lambda'\mathbf{u}|\max_{i\le n}|\theta'(\mathbf{y}_i-\beta)| \le (1+o_p(1))|r\lambda'\mathbf{u}|. \qquad \text{(S.10)}$$

We claim that $\rho = o_p(s^{-1/2})$. If not, suppose that $\liminf_{n\to\infty}\rho\sqrt{s} > 0$. Then, $|r\lambda'\mathbf{u}|\max_{i\le n}|\theta'(\mathbf{y}_i-\beta)|/\rho = o_p(1)$ due to (S.1), (S.8) and the condition (2.4) $\frac{n\sqrt{n}}{s}\to 0$ in Theorem 2.1. So we have $\frac{1}{2}\rho^2\theta'\mathbf{S}_1\theta \le |r\lambda'\mathbf{u}|$ from (S.10), which results in $\rho = o_p(s^{-1/2})$ since $\theta'\mathbf{S}_1\theta > c_0$ in probability and $\lambda'\mathbf{u} = o_p(s^{-1})$ from Lemma S1. This leads to a contradiction. Therefore

$$\|\lambda\| = \rho = o_p(s^{-1/2}). \tag{S.11}$$

Combining (S.8) with (S.11), we have

$$\max_{i\le n} |\lambda'(\mathbf{y}_i - \beta)| = \|\lambda\| \cdot \max_{i\le n} |\theta'(\mathbf{y}_i - \beta)| = o_p\Big(\sqrt{\frac{n}{s}}\Big).$$

Lemma S3. Under the assumptions of Theorem 2.1, we can improve the estimate of λ to

$$\|\lambda\| = o_p(s^{-1}). \tag{S.12}$$

Proof: Let $\mathbf{y}_i - \beta = k_i\mathbf{u} + \mathbf{r}_i$, where $k_i = (\mathbf{y}_i - \beta)'\mathbf{u}$ and $\mathbf{r}_i = (\mathbf{y}_i - \beta) - k_i\mathbf{u}$, for $i = 1, 2, \ldots, n+2$. Thus $\mathbf{u}'\mathbf{r}_i = \mathbf{0}$. By (S.2), we note that $\mathbf{r}_{n+1} = \mathbf{r}_{n+2} = \mathbf{0}$. Since the matrix $\mathbf{S}_1$ is of full rank with probability one due to $p/n \to c < 1$, $span\{\mathbf{y}_i - \beta, i = 1, 2, \ldots, n\} = \mathbb{R}^p$ with probability one. Hence there exist $a_1, \ldots, a_n$ with probability one such that

$$\theta = a_1(\mathbf{y}_1 - \beta) + a_2(\mathbf{y}_2 - \beta) + \cdots + a_n(\mathbf{y}_n - \beta). \tag{S.13}$$

Substituting $\mathbf{y}_i - \beta = k_i\mathbf{u} + \mathbf{r}_i$ into (S.13), we have

$$\theta = \Big(\sum_{i=1}^n a_i k_i\Big)\mathbf{u} + a_1\mathbf{r}_1 + \cdots + a_n\mathbf{r}_n. \tag{S.14}$$

Multiplying (S.14) by $\mathbf{u}'$ and θ', respectively, we obtain

$$\begin{cases} \mathbf{u}'\theta = \sum_{i=1}^n a_i k_i, \\ \quad 1 = \Big(\sum_{i=1}^n a_i k_i\Big)\theta'\mathbf{u} + \sum_{i=1}^n a_i \theta'\mathbf{r}_i. \end{cases}$$

Thus,

$$\Big|1 - \Big(\sum_{i=1}^n a_i k_i\Big)^2\Big| = \Big|\sum_{i=1}^n a_i\theta'\mathbf{r}_i\Big| \le \Big|\sum_{i=1}^n a_i^2 \cdot \sum_{i=1}^n (\theta'\mathbf{r}_i)^2\Big|^{1/2}.$$

Suppose that the following two relations are true,

$$\sum_{i=1}^n a_i^2 = O_p(1/n). \tag{S.15}$$

$$\sum_{i=1}^{n}(\theta'\mathbf{r}_i)^2 = o_p(n^2/s^2). \tag{S.16}$$

Then under the condition (2.4) in Theorem 2.1, we obtain

$$|\mathbf{u}'\theta| = |\sum_{i=1}^{n} a_i k_i| = o_p(\sqrt{n}/s) \xrightarrow{i.p.} 1. \tag{S.17}$$

From Lemma S1, we have $|\lambda'\mathbf{u}| = \|\lambda\| \cdot |\theta'\mathbf{u}| = o_p(s^{-1})$ and hence $\|\lambda\| = o_p(s^{-1})$ via (S.17). It remains to prove (S.15) and (S.16).

We first prove (S.15). Recalling $\mathbf{y}_i - \beta = \mathbf{A}^{-1}\Gamma\mathbf{z}_i$, from (S.13) we have

$$\mathbf{A}\theta = a_1\Gamma\mathbf{z}_1 + a_2\Gamma\mathbf{z}_2 + \cdots + a_n\Gamma\mathbf{z}_n.$$

Hence,

$$\begin{aligned}
\theta'\mathbf{A}'\mathbf{A}\theta &= (\sum_{i=1}^{n} a_i\Gamma\mathbf{z}_i)'(\sum_{i=1}^{n} a_i\Gamma\mathbf{z}_i) \\
&= tr\Gamma'\Gamma\sum_{i=1}^{n} a_i^2 + \sum_{i=1}^{n} a_i^2(\mathbf{z}_i'\Gamma'\Gamma\mathbf{z}_i - tr\Gamma'\Gamma) + \sum_{i\neq j} a_i a_j \mathbf{z}_i'\Gamma'\Gamma\mathbf{z}_j \\
&\triangleq J_1 + J_2 + J_3.
\end{aligned} \tag{S.18}$$

It's easy to see $J_1 \neq 0$, otherwise all a_i's will be zero, which would imply that the unit vector θ is zero from expression (S.13), a contradiction. We next show that J_1 is the dominant term. Let $\mathbf{z}_i = (z_{i1}, \ldots, z_{in})'$ and $\Gamma'\Gamma = (\vartheta_{ij})$.

$$\begin{aligned}
Var(J_2) = EJ_2^2 &= \sum_{i=1}^{n} a_i^4 E(\sum_{k=1}^{n}(z_{ik}^2 - 1)\vartheta_{kk} + \sum_{k\neq t} z_{ik}z_{it}\vartheta_{kt})^2 \\
&= \sum_{i=1}^{n} a_i^4 E(\sum_{k=1}^{n}(z_{ik}^2 - 1)^2\vartheta_{kk}^2 + \sum_{k\neq t} z_{ik}^2 z_{it}^2\vartheta_{kt}\vartheta_{tk}) \\
&= \sum_{i=1}^{n} a_i^4((\mu_4 - 3)\sum_{k=1}^{n}\vartheta_{kk}^2 + 2tr\Gamma'\Gamma\Gamma'\Gamma) \leq (\mu_4 - 1)tr(\Gamma'\Gamma\Gamma'\Gamma)\sum_{i=1}^{n} a_i^4,
\end{aligned}$$

and

$$\begin{aligned}
Var(J_3) = EJ_3^2 &= E\sum_{i\neq j}\sum_{s\neq t} a_i a_j a_s a_t \mathbf{z}_i'\Gamma'\Gamma\mathbf{z}_j\mathbf{z}_s'\Gamma'\Gamma\mathbf{z}_t \\
&= 2\sum_{i\neq j} a_i^2 a_j^2 tr(\Gamma'\Gamma\Gamma'\Gamma).
\end{aligned}$$

Hence

$$P\Big(\Big|\frac{J_2}{J_1}\Big| > \epsilon\Big) \le \frac{Var(J_2)}{J_1^2\epsilon^2} \le \frac{\mu_4 - 1}{\epsilon^2} \cdot \frac{tr\Gamma'\Gamma\Gamma'\Gamma}{(tr\Gamma'\Gamma)^2} \cdot \frac{\sum_{i=1}^n a_i^4}{(\sum_{i=1}^n a_i^2)^2} \le \frac{K}{n} \to 0,$$

and

$$P\Big(\Big|\frac{J_3}{J_1}\Big| > \epsilon\Big) \le \frac{Var(J_3)}{J_1^2\epsilon^2} = \frac{2}{\epsilon^2} \cdot \frac{tr\Gamma'\Gamma\Gamma'\Gamma}{(tr\Gamma'\Gamma)^2} \cdot \frac{\sum_{i\neq j} a_i^2 a_j^2}{(\sum_{i=1}^n a_i^2)^2} \le \frac{K}{n} \to 0.$$

Therefore, $J_2/J_1 \xrightarrow{i.p.} 0, J_3/J_1 \xrightarrow{i.p.} 0$, as $n \to \infty$. By (S.18), we have $\theta'\mathbf{A}'\mathbf{A}\theta = tr\Gamma'\Gamma\sum_{i=1}^n a_i^2(1 + o_p(1))$. Since θ is a unit vector, $\|\mathbf{A}'\mathbf{A}\|$ (Jiang (2004)) is bounded from above in probability and $tr\Gamma'\Gamma\Gamma'\Gamma \ge cp$ for some positive constant c,

$$\sum_{i=1}^n a_i^2 = O_p(1/n).$$

Let us turn to (S.16). By the definition of k_i, $\sum_{i=1}^n \mathbf{y}_i = \mathbf{0}$ and $\beta = -r\mathbf{u}$, we first have

$$\sum_{i=1}^n \mathbf{r}_i = \sum_{i=1}^{n+2}(\mathbf{y}_i - \beta) - \sum_{i=1}^{n+2} k_i\mathbf{u} = (n+2)r\mathbf{u} - (n+2)r\mathbf{u}\mathbf{u}'\mathbf{u} = \mathbf{0}. \tag{S.19}$$

Also, note that $k_i = (\mathbf{y}_i - \beta)'\mathbf{u}$ and $\mathbf{y}_i$ are standardized so that

$$\begin{aligned}\sum_{i=1}^n k_i\mathbf{r}_i &= \sum_{i=1}^n k_i\Big((\mathbf{y}_i - \beta) - k_i\mathbf{u}\Big) \\ &= \sum_{i=1}^n(\mathbf{y}_i - \beta)(\mathbf{y}_i - \beta)'\mathbf{u} - \sum_{i=1}^n k_i^2\mathbf{u} \\ &= \Big((n-1)\mathbf{I}_p + n\beta\beta'\Big)\mathbf{u} - \Big(\sum_{i=1}^n k_i^2\Big)\mathbf{u} \\ &= \Big((n-1) + nr^2 - \sum_{i=1}^n k_i^2\Big)\mathbf{u}.\end{aligned}$$

Since $\mathbf{u}$ and $\mathbf{r}_i$ $i = 1, \ldots, n$ are orthogonal, we have

$$\sum_{i=1}^n k_i\mathbf{r}_i = 0. \tag{S.20}$$

Rewriting the constraint (A.3) on the Lagrange multiplier λ as

$$\mathbf{0} = \sum_{i=1}^{n+2} \frac{\mathbf{y}_i - \beta}{1 + \lambda'(\mathbf{y}_i - \beta)} = \sum_{i=1}^{n+2} \frac{k_i\mathbf{u}}{1 + \lambda'(\mathbf{y}_i - \beta)} + \sum_{i=1}^{n} \frac{\mathbf{r}_i}{1 + \lambda'(\mathbf{y}_i - \beta)},$$

since $\mathbf{u}$ and $\mathbf{r}_i, i = 1, \ldots, n$ are orthogonal, we also have

$$\sum_{i=1}^{n} \frac{\mathbf{r}_i}{1 + \lambda'(\mathbf{y}_i - \beta)} = 0. \tag{S.21}$$

By (S.19), (S.20), (S.21) and applying the equality $\frac{1}{1+x} = 1 - \frac{x}{1+x}$ in the following second and fourth equalities, we have

$$\begin{aligned}
0 &= \sum_{i=1}^{n} \frac{\theta'\mathbf{r}_i}{1 + \lambda'(\mathbf{y}_i - \beta)} = \sum_{i=1}^{n} \theta'\mathbf{r}_i - \sum_{i=1}^{n} \frac{\theta'\mathbf{r}_i\lambda'(\mathbf{y}_i - \beta)}{1 + \lambda'(\mathbf{y}_i - \beta)} \\
&= -\sum_{i=1}^{n} \frac{k_i\theta'\mathbf{r}_i\lambda'\mathbf{u}}{1 + \lambda'(\mathbf{y}_i - \beta)} - \sum_{i=1}^{n} \frac{\theta'\mathbf{r}_i\lambda'\mathbf{r}_i}{1 + \lambda'(\mathbf{y}_i - \beta)} \\
&= -\sum_{i=1}^{n} k_i\theta'\mathbf{r}_i\lambda'\mathbf{u} + \sum_{i=1}^{n} \frac{k_i\theta'\mathbf{r}_i\lambda'\mathbf{u}\lambda'(\mathbf{y}_i - \beta)}{1 + \lambda'(\mathbf{y}_i - \beta)} \\
&\quad - \sum_{i=1}^{n} \theta'\mathbf{r}_i\lambda'\mathbf{r}_i + \sum_{i=1}^{n} \frac{\theta'\mathbf{r}_i\lambda'\mathbf{r}_i\lambda'(\mathbf{y}_i - \beta)}{1 + \lambda'(\mathbf{y}_i - \beta)} \\
&= \lambda'\mathbf{u}\rho \sum_{i=1}^{n} \frac{k_i\theta'\mathbf{r}_i\theta'(\mathbf{y}_i - \beta)}{1 + \lambda'(\mathbf{y}_i - \beta)} - \rho\sum_{i=1}^{n}(\theta'\mathbf{r}_i)^2 + \rho\sum_{i=1}^{n} \frac{(\theta'\mathbf{r}_i)^2\lambda'(\mathbf{y}_i - \beta)}{1 + \lambda'(\mathbf{y}_i - \beta)}.
\end{aligned}$$

If $\rho = \|\lambda\| = 0$, then Lemma S3 is obviously true. For $\rho \neq 0$, dividing both sides by ρ and by Hölder's inequality, Lemma S1, Lemma S2 and (S.8) we have

$$\begin{aligned}
\sum_{i=1}^{n}(\theta'\mathbf{r}_i)^2 &= \lambda'\mathbf{u}\sum_{i=1}^{n} \frac{k_i\theta'\mathbf{r}_i\theta'(\mathbf{y}_i - \beta)}{1 + \lambda'(\mathbf{y}_i - \beta)} + \sum_{i=1}^{n} \frac{(\theta'\mathbf{r}_i)^2\lambda'(\mathbf{y}_i - \vartheta)}{1 + \lambda'(\mathbf{y}_i - \beta)} \\
&\leq \lambda'\mathbf{u}\max_{i\leq n}|\theta'(\mathbf{y}_i - \beta)| \sum_{i=1}^{n} \frac{|k_i\theta'\mathbf{r}_i|}{1 - \max_{i\leq n}|\lambda'(\mathbf{y}_i - \beta)|} \\
&\quad + \rho\max_{i\leq n}|\theta'(\mathbf{y}_i - \beta)| \sum_{i=1}^{n} \frac{(\theta'\mathbf{r}_i)^2}{1 - \max_{i\leq n}|\lambda'(\mathbf{y}_i - \beta)|} \\
&\leq O_p(1)o_p(\sqrt{n}/s)\Big(\sum_{i=1}^{n}(\theta'\mathbf{r}_i)^2 \cdot \sum_{i=1}^{n} k_i^2\Big)^{1/2} + O_p(1)\sum_{i=1}^{n}(\theta'\mathbf{r}_i)^2 \cdot o_p(\sqrt{n/s}).
\end{aligned}$$

Together with the observation that $\sum_{i=1}^{n} k_i^2 = n\mathbf{u}'\mathbf{S}_1\mathbf{u} = O_p(n)$, we get

$$(1 + o_p(\sqrt{n/s}))\sum_{i=1}^{n}(\theta'\mathbf{r}_i)^2 \leq o_p\Big(\frac{n}{s}\Big)\Big(\sum_{i=1}^{n}(\theta'\mathbf{r}_i)^2\Big)^{1/2}.$$

Hence, $\Big[\sum_{i=1}^{n}(\theta'\mathbf{r}_i)^2\Big]^{\frac{1}{2}} = o_p(n/s)$, which concludes the proof of (S.16).

Lemma S4. Under the assumptions of Theorem 2.1, we have

$$s^2\lambda'\mathbf{u} = (n+2)r/2 + o_p(n^2/s) + o_p(1), \qquad \rho = \|\lambda\| = o_p(n/s^2). \tag{S.22}$$

Proof: By (S.2) and applying the identity $\frac{1}{1+x} = 1 - x + \frac{x^2}{1+x}$ to the constraint equality (A.3), we have

$$\begin{aligned}
0 &= \sum_{i=1}^{n+2} \frac{\mathbf{u}'(\mathbf{y}_i - \beta)}{1+\lambda'(\mathbf{y}_i-\beta)} = \sum_{i=1}^{n+2} \mathbf{u}'(\mathbf{y}_i-\beta) - \sum_{i=1}^{n+2} \mathbf{u}'(\mathbf{y}_i-\beta)(\mathbf{y}_i-\beta)'\lambda + \sum_{i=1}^{n+2} \frac{\mathbf{u}'(\mathbf{y}_i-\beta)\Big((\mathbf{y}_i-\beta)'\lambda\Big)^2}{1+\lambda'(\mathbf{y}_i-\beta)} \\
&= (n+2)r - \Big(n\mathbf{u}'\mathbf{S}_1\lambda + s^2\lambda'\mathbf{u} + (2r+s)^2\lambda'\mathbf{u}\Big) \\
&\quad + \sum_{i=1}^{n} \frac{\mathbf{u}'(\mathbf{y}_i-\beta)\Big((\mathbf{y}_i-\beta)'\lambda\Big)^2}{1+\lambda'(\mathbf{y}_i-\beta)} - \frac{s^3(\lambda'\mathbf{u})^2}{1-s\lambda'\mathbf{u}} + \frac{(s+2r)^3(\lambda'\mathbf{u})^2}{1-(2r+s)\lambda'\mathbf{u}} \\
&= (n+2)r - \Big(n\mathbf{u}'\mathbf{S}_1\lambda + 2s^2\lambda'\mathbf{u} + (4sr+4r^2)\lambda'\mathbf{u}\Big) \\
&\quad + \sum_{i=1}^{n} \frac{\mathbf{u}'(\mathbf{y}_i-\beta)\Big((\mathbf{y}_i-\beta)'\lambda\Big)^2}{1+\lambda'(\mathbf{y}_i-\beta)} - s^3(\lambda'\mathbf{u})^2\Big[(n+2)(\omega_{n+2}-\omega_{n+1})\Big] + \frac{(6s^2r+12sr^2+8r^3)(\lambda'\mathbf{u})^2}{1-(2r+s)\lambda'\mathbf{u}}.
\end{aligned}$$

By Lemma S1, Lemma S3 and (S.5), we have the following estimates

$$\begin{aligned}
&n\mathbf{u}'\mathbf{S}_1\lambda = n\rho\mathbf{u}'\mathbf{S}_1\theta = o_p(1), \quad s\lambda'\mathbf{u} = o_p(1), \quad \frac{(6s^2r+12sr^2+8r^3)(\lambda'\mathbf{u})^2}{1-(2r+s)\lambda'\mathbf{u}} = o_p(1), \\
&s^3(\lambda'\mathbf{u})^2\Big[(n+2)(\omega_{n+2}-\omega_{n+1})\Big] = s^2(\lambda'\mathbf{u})o_p(s\lambda'\mathbf{u}), \\
&\Big|\sum_{i=1}^{n} \frac{\mathbf{u}'(\mathbf{y}_i-\beta)\Big((\mathbf{y}_i-\beta)'\lambda\Big)^2}{1+\lambda'(\mathbf{y}_i-\beta)}\Big| \le \frac{\max_{i\le n}|\mathbf{u}'(\mathbf{y}_i-\beta)|}{1-\rho\max_{i\le n}|\theta'(\mathbf{y}_i-\beta)|} \cdot n\rho^2\theta'\mathbf{S}_1\theta = o_p(n\sqrt{n}/s^2) = o_p(1).
\end{aligned}$$

It follows from (S.4) and the above inequalities that

$$(n+2)r - 2s^2\lambda'\mathbf{u} + s^2\lambda'\mathbf{u}\cdot o_p(s\lambda'\mathbf{u}) + o_p(1) = 0. \tag{S.23}$$

Since $o_p(s\lambda'\mathbf{u}) = o_p(1)$ from Lemma S1, we first obtain $\lambda'\mathbf{u} = O_p(n/s^2)$. Hence via (S.17), $\rho = \|\lambda\| = \lambda'\mathbf{u}/|\mathbf{u}'\theta| = O_p(n/s^2)$, which implies the second bound in (S.22). Furthermore $s^2\lambda'\mathbf{u}\cdot o_p(s\lambda'\mathbf{u}) = o_p(n^2/s)$, thus from (S.23), we have $s^2\lambda'\mathbf{u} = \frac{1}{2}(n+2)r + o_p(n^2/s) + o_p(1)$. The Lemma is proved.

Proof of Theorem 2.1. $\frac{2s^2 W(\mu)}{(n+2)^2} - \frac{T^2}{n} = o_p(n/s) + o_p(1/n), \quad \text{as} \quad n \to \infty.$

Proof: By (S.7) and (S.9), $\max\limits_{i \leq n+2} |\lambda'(\mathbf{y}_i - \beta)| = o_p(1)$. So we can use Taylor's expansion,

$$-\log{(n+2)\omega_i} = \log{(1+\lambda'(\mathbf{y}_i-\beta))} = \lambda'(\mathbf{y}_i-\beta) - \frac{1}{2}\Big(\lambda'(\mathbf{y}_i-\beta)\Big)^2 + \frac{1}{3}\Big(\lambda'(\mathbf{y}_i-\beta)\Big)^3 - \eta_i, \tag{S.24}$$

where $\eta_i = \frac{1}{4}\Big(\frac{\lambda'(\mathbf{y}_i-\beta)}{1+\xi_i}\Big)^4$ and $|\xi_i| \leq |\lambda'(\mathbf{y}_i-\beta)|$. Then by (S.7), (S.9) and Lemma S4,

$$\begin{aligned}
\sum_{i=1}^{n+2}\eta_i &= \sum_{i=1}^{n}\eta_i + \frac{1}{4}\Big(\frac{s\lambda'\mathbf{u}}{1+\xi_{n+1}}\Big)^4 + \frac{1}{4}\Big(\frac{(2r+s)\lambda'\mathbf{u}}{1+\xi_{n+2}}\Big)^4 \\
&\leq K\max_{i\leq n}|\lambda'(\mathbf{y}_i-\beta)|^2 \cdot \frac{\sum_{i=1}^n \lambda'(\mathbf{y}_i-\beta)(\mathbf{y}_i-\beta)'\lambda}{1-\max\limits_{i\leq n}|\lambda'(\mathbf{y}_i-\beta)|} + K\Big(\frac{s\lambda'\mathbf{u}}{1+o_p(1)}\Big)^4 + K\Big(\frac{(s+2r)\lambda'\mathbf{u}}{1+o_p(1)}\Big)^4 \\
&\leq O_p(1)\rho^4\max_{i\leq n}|\theta'(\mathbf{y}_i-\beta)|^2\cdot n\cdot\theta'\mathbf{S}_1\theta + O_p(1)\cdot(s\lambda'\mathbf{u})^4 \\
&\leq O_p(n^4/s^4).
\end{aligned}$$

By the formula $\breve{W}(\beta) = -2\sum_{i=1}^{n+2}\log\Big((n+2)\omega_i\Big) = 2\sum_{i=1}^{n+2}\log{(1+\lambda'(\mathbf{y}_i-\beta))}$, we have

$$\begin{aligned}
\breve{W}(\beta) &= 2\Big[\sum_{i=1}^{n+2}\lambda'(\mathbf{y}_i-\beta) - \frac{1}{2}\sum_{i=1}^{n+2}\Big(\lambda'(\mathbf{y}_i-\beta)\Big)^2 + \frac{1}{3}\sum_{i=1}^{n+2}\Big(\lambda'(\mathbf{y}_i-\beta)\Big)^3 - \sum_{i=1}^{n+2}\eta_i\Big] \\
&= 2\Big[(n+2)r\lambda'\mathbf{u} - \frac{1}{2}\Big(\sum_{i=1}^{n}\Big(\lambda'(\mathbf{y}_i-\beta)\Big)^2 + s^2(\lambda'\mathbf{u})^2 + (s+2r)^2(\lambda'\mathbf{u})^2\Big) \\
&\quad + \frac{1}{3}\Big(\sum_{i=1}^{n}\Big(\lambda'(\mathbf{y}_i-\beta)\Big)^3 + s^3(\lambda'\mathbf{u})^3 - (s+2r)^3(\lambda'\mathbf{u})^3\Big) - O_p(n^4/s^4)\Big] \\
&= 2\Big[(n+2)r\lambda'\mathbf{u} - \frac{1}{2}\Big(\sum_{i=1}^{n}\Big(\lambda'(\mathbf{y}_i-\beta)\Big)^2 + 2s^2(\lambda'\mathbf{u})^2 + (4rs+4r^2)(\lambda'\mathbf{u})^2\Big) \\
&\quad + \frac{1}{3}\Big(\sum_{i=1}^{n}\Big(\lambda'(\mathbf{y}_i-\beta)\Big)^3 - (6s^2r+12sr^2+8r^3)(\lambda'\mathbf{u})^3\Big) - O_p(n^4/s^4)\Big] \\
&= 2\Big[(n+2)r\lambda'\mathbf{u} - s^2(\lambda'\mathbf{u})^2 + O_p(n^4/s^4)\Big],
\end{aligned} \tag{S.25}$$

where by Lemma S4 and condition (2.4) the last equality follows from the fact that

$$\sum_{i=1}^{n}\Big(\lambda'(\mathbf{y}_i-\beta)\Big)^2 = n\rho^2\theta'\mathbf{S}_1\theta = O_p(n^3/s^4), \quad (4rs+4r^2)(\lambda'\mathbf{u})^2 = O_p(n^2/s^3),$$
$$\Big|\sum_{i=1}^{n}\Big(\lambda'(\mathbf{y}_i-\beta)\Big)^3\Big| \le n\rho^3\max_{i\le n}|\theta'(\mathbf{y}_i-\beta)|\cdot\theta'\mathbf{S}_1\theta = o_p(n^3\sqrt{n}/(s^4\sqrt{s})),$$
$$(6s^2r+12sr^2+8r^3)(\lambda'\mathbf{u})^3 = O_p(n^3/s^4).$$

By Lemma S4, multiplying (S.25) by s^2 and using $s^2\lambda'\mathbf{u} = (n+2)r/2+o_p(n^2/s)+o_p(1)$, we have

$$\begin{aligned} s^2\breve{W}(\beta) &= 2\Big[(n+2)rs^2\lambda'\mathbf{u} - (s^2\lambda'\mathbf{u})^2 + O_p(n^4/s^2)\Big] \\ &= \frac{1}{2}(n+2)^2r^2 + o_p(n^3/s) + O_p(n^4/s^2) + o_p(n^2/s) + o_p(1) + o_p(n), \end{aligned} \tag{S.26}$$

It follows from (S.26), (S.3) and condition (2.4) that

$$\frac{2s^2\breve{W}(\beta)}{(n+2)^2} - \frac{T^2}{n} = o_p(n/s) + o_p(1/n), \quad \text{as} \quad n\to\infty.$$

The proof of Theorem 2.1 is completed.