# Supplementary Materials: Higher Order Bias Correcting Moment Equation for M-Estimation and Its Higher Order Efficiency

**Kyoo il Kim**

## Appendix A. Technical Lemmas and Proofs

This Appendix provides a series of lemmas and their proofs that are useful to derive the main results presented in the paper.

### *Appendix A.1. Some Preliminary Lemmas*

**Lemma A.1.** *(Uniform Weak Convergence Theorem with Compactness) Suppose (i)* $\{z_i : i = 1, \ldots, n\}$ *are iid; (ii)* $m(z, \theta)$ *is continuous at each* $\theta \in \Theta$ *for all* $z \in \mathcal{Z}$ *with probability one; (iii)* $E\left[\sup_{\theta \in \Theta} \|m(z_i, \theta)\|\right] < \infty$*; (iv)* $\Theta$ *is compact. Then,* $E\left[\|m(z_i, \theta)\|\right]$ *is continuous for all* $\theta \in \Theta$ *and* $\sup_{\theta \in \Theta} \left\| \frac{1}{n} \sum_{i=1}^{n} m(z_i, \theta) - E\left[m(z_i, \theta)\right] \right\| = o_p(1)$*.*

**Proof.** This result is implied by Lemma 1 of Tauchen (1985) [1] or can be verified by showing the stochastic equicontinuity of $\left\{ \frac{1}{n} \sum_{i=1}^{n} \left(m(z_i, \theta) - E\left[m(z_i, \theta)\right]\right) : n \geq 1 \right\}$ for $\theta \in \Theta$ as in Newey (1991) [2] observing that $E\left[\sup_{\theta \in \Theta} \|m(z_i, \theta)\|\right] < \infty$ is stronger than the Lipschitz condition used in Newey (1991) [2]. The continuity of $E\left[\|m(z_i, \theta)\|\right]$ is obtained from the Dominated Convergence theorem with the dominating function $\sup_{\theta \in \Theta} \|m(z_i, \theta)\| < \infty$. Here we provide an alternative proof for the stochastic equicontinuity. We use the following definition of the stochastic equicontinuity:

**Definition A.1.** $\{M_n(\theta) | n \geq 1\}$ *is stochastically equicontinuous on* $\Theta$ *if* $\forall \varepsilon > 0$ $\exists \delta > 0$ *such that*

$$\overline{\lim_{n \to \infty}} P\left( \sup_{\theta \in \Theta} \sup_{\theta' \in B(\theta, \delta)} \|M_n(\theta') - M_n(\theta)\| > \varepsilon \right) < \varepsilon.$$

Now define $M_n(\theta) = \frac{1}{n} \sum_{i=1}^{n} m(z_i, \theta) - E\left[m(z_i, \theta)\right]$ and $Y_{i\delta} = \sup_{\theta \in \Theta} \sup_{\theta' \in B(\theta, \delta)} \|m(z_i, \theta') - m(z_i, \theta)\|$. Note $E\left[Y_{i\delta}\right] \leq 2E\left[\sup_{\theta \in \Theta} \|m(z_i, \theta)\|\right] < \infty$ by Condition (iii). We claim that $E\left[Y_{i\delta}\right] \to 0$ as $\delta \to 0$ by noting $Y_{i\delta} \to 0$ as $\delta \to 0$ with probability one, since Condition (ii) and (iv) implies uniform continuity. Furthermore, $Y_{i\delta} \leq 2 \sup_{\theta \in \Theta} \|m(z_i, \theta)\|$ $\forall \delta > 0$ and $E\left[\sup_{\theta \in \Theta} \|m(z_i, \theta)\|\right] < \infty$ by Condition (iii) and hence from the dominated convergence theorem, the claim follows. Now let $\varepsilon > 0$, then

$$\begin{aligned} &\overline{\lim}_{n \to \infty} P\left( \sup_{\theta \in \Theta} \sup_{\theta' \in B(\theta, \delta)} \|M_n(\theta') - M_n(\theta)\| > \varepsilon \right) \leq \overline{\lim}_{n \to \infty} P\left( \tfrac{1}{n} \textstyle\sum_{i=1}^{n} \left(Y_{i\delta} + E[Y_{i\delta}]\right) > \varepsilon \right) \\ &\leq \overline{\lim}_{n \to \infty} E\left[ \tfrac{1}{n} \textstyle\sum_{i=1}^{n} \left(Y_{i\delta} + E[Y_{i\delta}]\right) \right] / \varepsilon = 2E[Y_{i\delta}] / \varepsilon \to 0 \text{ as } \delta \to 0, \end{aligned}$$

where the first inequality follows by Triangle inequality, the second holds by Markov inequality, and the last equality holds by $E\left[Y_{i\delta}\right] \to 0$ as $\delta \to 0$. This proves $M_n(\theta)$ is stochastically equicontinuous and the uniform convergence follows noting Condition (iii) is sufficient for the pointwise weak convergence. This is proved when $\Theta$ is bounded (not necessarily compact) in the proof of Lemma A.2. □

**Lemma A.2.** *(Uniform Weak Convergence Theorem without Compactness) Suppose (i)* $\{z_i : i = 1, \ldots, n\}$ *are iid; (ii)* $m(z_i, \theta)$ *satisfies the Lipschitz condition in* $\theta$ *as* $\|m(z_i, \theta_1) - m(z_i, \theta_2)\| \leq B(z_i) \|\theta_1 - \theta_2\|$, $\forall \theta_1, \theta_2 \in \Theta$ *for some function* $B(\cdot) \colon \mathcal{Z} \to R$ *and* $E\left[B(\cdot)^{2+\delta}\right] < \infty$*; (iii)* $E\left[\sup_{\theta \in \Theta} \|m(z_i, \theta)\|^{2+\delta}\right] < \infty$ *for some*

*$\delta > 0$; (iv) $\Theta$ is bounded. Then, $\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(m\left(z_i,\theta\right)-E[m\left(z_i,\theta\right)]\right)$ is stochastically equicontinuous and thus $\sup_{\theta\in\Theta}\left\|\frac{1}{n}\sum_{i=1}^{n}m\left(z_i,\theta\right)-E\left[m\left(z_i,\theta\right)\right]\right\| = o_p(1)$.*

**Proof.** From condition (ii), we note that $m(\cdot,\cdot)$ belongs to Type II class in Andrews (1994) [3] with envelopes given by $\max(\sup_{\theta\in\Theta}\|m(\cdot,\theta)\|, B(\cdot))$ and hence satisfies Pollard's entropy condition by Theorem 2 in Andrews (1994) [3], which is Assumption A of Theorem 1 in Andrews (1994) [3]. Condition (iii) implies Assumption B of Theorem 1 in Andrews (1994) [3]. Condition (i) is stronger than Assumption C for Theorem 1 in Andrews (1994) [3] and hence stochastic equicontinuity follows. Now noting Condition (iii) is sufficient for pointwise weak convergences of $\frac{1}{n}\sum_{i=1}^{n}m\left(z_i,\theta\right)$ to $E\left[m\left(z_i,\theta\right)\right]$ for all $\theta\in\Theta$ and combining this with the stochastic equicontinuity result, we have the uniform convergence as assuming $\Theta$ is bounded. To be more precise, first, note that the stochastic equicontinuity of $\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(m\left(z_i,\theta\right)-E[m\left(z_i,\theta\right)]\right)$ implies the stochastic equicontinuity of $\frac{1}{n}\sum_{i=1}^{n}\left(m\left(z_i,\theta\right)-E[m\left(z_i,\theta\right)]\right)$. Now define $v_n(\theta)=\frac{1}{n}\sum_{i=1}^{n}\left(m\left(z_i,\theta\right)-E[m\left(z_i,\theta\right)]\right)$ and let $\varepsilon>0$ and take a $\delta$ such that $\overline{\lim}_{n\to\infty}P\left(\sup_{\theta\in\Theta}\sup_{\theta'\in B(\theta,\delta)}\|v_n(\theta')-v_n(\theta)\|>\varepsilon\right)<\varepsilon$. Such $\delta$ exists by the definition of the stochastic equicontinuity. Now note that from the boundedness of $\Theta$, we can construct a finite cover of $\Theta$ as $\{B(\theta_j,\delta): j=1,\ldots,J\}$. Then it follows

$$
\begin{aligned}
&\overline{\lim}_{n\to\infty}P\left(\sup_{\theta'\in\Theta}\|v_n(\theta')\|>2\varepsilon\right)\\
&\leq \overline{\lim}_{n\to\infty}P\left(\max_{j\leq J}\left(\sup_{\theta'\in B(\theta_j,\delta)}\|v_n(\theta')-v_n(\theta_j)\|+\|v_n(\theta_j)\|\right)>2\varepsilon\right)\\
&\leq \overline{\lim}_{n\to\infty}P\left(\max_{j\leq J}\sup_{\theta'\in B(\theta_j,\delta)}\|v_n(\theta')-v_n(\theta_j)\|>\varepsilon\right)+\overline{\lim}_{n\to\infty}P\left(\max_{j\leq J}\|v_n(\theta_j)\|>\varepsilon\right)\\
&\leq \overline{\lim}_{n\to\infty}P\left(\sup_{\theta\in\Theta}\sup_{\theta'\in B(\theta,\delta)}\|v_n(\theta')-v_n(\theta)\|>\varepsilon\right)+\overline{\lim}_{n\to\infty}P\left(\max_{j\leq J}\|v_n(\theta_j)\|>\varepsilon\right)\leq\varepsilon,
\end{aligned}
$$

where the first inequality is from Triangle Inequality and by the construction of $\{B(\theta_j,\delta): j=1,\ldots,J\}$. The last inequality comes from the stochastic equicontinuity of $v_n(\theta)$ and the pointwise weak convergence of $v_n(\theta)$ and hence the uniform convergence result follows. □

In addition to the assumption of $\widehat{\theta}$ being consistent, we provides two alternative primitive conditions that satisfy the higher level conditions used in Lemma 1. The first possible set of primitive conditions is

**Assumption A.1.** *(i) $\{z_i\}_{i=1}^{n}$ are iid; (ii) $s(z,\theta)$ is $\kappa$-times continuously differentiable in a neighborhood of $\theta_0$, denoted by $\Theta_0$ for all $z\in\mathcal{Z}$, $\kappa\geq 3$ with probability one; (iii) $E\left[\sup_{\theta\in\Theta_0}\|\nabla^{v}s(z,\theta)\|\right]<\infty$, $v=\{0,1,2,\ldots\kappa\}$, $\kappa\geq 3$; (iv) $\Theta$ is compact; (v) $\theta_0$ is in the interior of $\Theta$ and is the only $\theta$ satisfying (1).*

**Assumption A.2.** *$E\left[\|\nabla^{v}s(z,\theta_0)\|^2\right]<\infty$, $v=\{0,1,2,\ldots\kappa\}$, $\kappa\geq 3$.*

**Assumption A.3.** *$E\left[\nabla s(z,\theta_0)\right]$ is nonsingular.*

Instead of Assumption A.1, alternatively we may assume

**Assumption A.4.** *(i) $\{z_i\}_{i=1}^{n}$ are iid; (ii) $\nabla^{v}s(z,\theta)$ satisfies the Lipschitz condition in $\theta$ as*

$$\|\nabla^{v}s(z,\theta_1)-\nabla^{v}s(z,\theta_2)\|\leq B_v(z)\|\theta_1-\theta_2\|\quad\forall\theta_1,\theta_2\in\Theta_0$$

*for some function $B_v(\cdot)\colon\mathcal{Z}\to R$ and $E\left[B_v(\cdot)^{2+\delta}\right]<\infty$, $v=\{0,1,2,\ldots\kappa\}$ in a neighborhood of $\theta_0$, denoted by $\Theta_0$ for all $z\in\mathcal{Z}$, $\kappa\geq 3$ with probability one; (iii) $E\left[\sup_{\theta\in\Theta_0}\|\nabla^{v}s(z,\theta)\|^{2+\delta}\right]<\infty$ for $\exists\ \delta>0$, $v=\{0,1,2,\ldots\kappa\}$, $\kappa\geq 3$; (iv) $\Theta$ is bounded; (v) $\theta_0$ is in the interior of $\Theta$ and is the only $\theta$ satisfying (1).*

**Lemma A.3.** *(Local Uniform Weak Convergence with Compactness)*
*Suppose Assumption A.1 holds, then we have* $\left\|\frac{1}{n}\sum_{i=1}^n \nabla^v s\left(z_i,\bar{\theta}\right) - E\left[\nabla^v s\left(z_i,\theta_0\right)\right]\right\| = o_p(1)$ *for* $\bar{\theta} = \theta_0 + o_p(1)$ *and* $v \in \{0,1,2,\ldots,\kappa\}$.

**Proof.** Consider

$$\begin{aligned}
&\left\|\tfrac{1}{n}\sum_{i=1}^n \nabla^v s\left(z_i,\bar{\theta}\right) - E\left[\nabla^v s\left(z_i,\theta_0\right)\right]\right\| \\
&\le \left\|\tfrac{1}{n}\sum_{i=1}^n \nabla^v s\left(z_i,\bar{\theta}\right) - E\left[\nabla^v s\left(z_i,\bar{\theta}\right)\right]\right\| + \left\|E\left[\nabla^v s\left(z_i,\bar{\theta}\right)\right] - E\left[\nabla^v s\left(z_i,\theta_0\right)\right]\right\| \\
&\le \sup_{\theta\in\Theta_0}\left\|\tfrac{1}{n}\sum_{i=1}^n \nabla^v s\left(z_i,\theta\right) - E\left[\nabla^v s\left(z_i,\theta\right)\right]\right\| + \left\|E\left[\nabla^v s\left(z_i,\bar{\theta}\right)\right] - E\left[\nabla^v s\left(z_i,\theta_0\right)\right]\right\|.
\end{aligned}$$

We have $\sup_{\theta\in\Theta_0}||\frac{1}{n}\sum_{i=1}^n \nabla^v s\left(z_i,\theta\right) - E\left[\nabla^v s\left(z_i,\theta\right)\right]|| = o_p(1)$ from Lemma A.1 by letting $m(z,\theta) = \nabla^v s\left(z,\theta\right)$ and noting Assumption A.1 satisfies all the conditions in Lemma A.1 for $\theta \in \Theta_0$. The continuity of $E\left[\nabla^v s\left(z_i,\theta\right)\right]$ at $\theta_0$ (by the Dominated Convergence theorem with the dominating function $\sup_{\theta\in\Theta_0}\|\nabla^v s(z_i,\theta)\|$) implies that $\left\|E\left[\nabla^v s\left(z_i,\bar{\theta}\right)\right] - E\left[\nabla^v s\left(z_i,\theta_0\right)\right]\right\| = o_p(1)$, since $\bar{\theta} = \theta_0 + o_p(1)$ and hence from this and the result above, it follows that $\left\|\frac{1}{n}\sum_{i=1}^n \nabla^v s\left(z_i,\bar{\theta}\right) - E\left[\nabla^v s\left(z_i,\theta_0\right)\right]\right\| = o_p(1)$. □

**Lemma A.4.** *(Local Uniform Weak Convergence without Compactness)*
*Under Assumption A.4, we have* $\left\|\frac{1}{n}\sum_{i=1}^n \nabla^v s\left(z_i,\bar{\theta}\right) - E\left[\nabla^v s\left(z_i,\theta_0\right)\right]\right\| = o_p(1)$ *for* $\bar{\theta} = \theta_0 + o_p(1)$ *and* $v \in \{0,1,2,\ldots,\kappa\}$.

**Proof.** Again noting Assumption A.4 satisfies all the conditions in Lemma A.2 for $\theta \in \Theta_0$, we have the uniform convergence and the dominated convergence theorem assures the continuity of $E\left[\nabla^v s\left(z_i,\theta\right)\right]$ for $\theta \in \Theta_0$ and hence the result follows. □

Now we show that conditions (i)-(viii) in Lemma 1 are satisfied under Assumption A.1-A.3 or Assumption A.4, A.2-A.3. Condition (i) and (iia) are directly assumed. Condition (iib) is by the dominated convergence theorem with the dominating function given by $\sup_{\theta\in\Theta_0}\left\|\nabla^3 s(z,\theta)\right\|$ under Condition (i), (iia), and $E[\sup_{\theta\in\Theta_0}\left\|\nabla^3 s(z,\theta)\right\|] < \infty$. Condition (iii) holds from Lemma A.3 or A.4.

Condition (iv) holds by the stochastic equicontinuity of $\frac{1}{\sqrt{n}}\sum_{i=1}^n\left(\nabla^2 s\left(z_i,\theta\right) - E[\nabla^2 s\left(z_i,\theta\right)]\right)$ for $\theta \in \Theta_0$ as discussed in A.2 with $m(z,\theta) = \nabla^2 s\left(z,\theta\right)$. Condition (iv) is used to show that

$$\left(\frac{1}{n}\sum_{i=1}^n \nabla^2 s\left(z_i,\tilde{\theta}\right) - \frac{1}{n}\sum_{i=1}^n \nabla^2 s\left(z_i,\theta_0\right)\right)\left(\left(\hat{\theta}-\theta_0\right)\otimes\left(\left(\hat{\theta}-\theta_0\right)\right)\right) = O_p(n^{-3/2})$$

in the proof of Lemma 1. Alternatively, it can be shown as

$$\begin{aligned}
&\left\|\left(\tfrac{1}{n}\sum_{i=1}^n \nabla^2 s\left(z_i,\tilde{\theta}\right) - \tfrac{1}{n}\sum_{i=1}^n \nabla^2 s\left(z_i,\theta_0\right)\right)\left(\left(\hat{\theta}-\theta_0\right)\otimes\left(\left(\hat{\theta}-\theta_0\right)\right)\right)\right\| \\
&\le \left\|\tfrac{1}{n}\sum_{i=1}^n \nabla^3 s\left(z_i,\tilde{\tilde{\theta}}\right)\right\| \left\|\tilde{\theta}-\theta_0\right\| \left\|\hat{\theta}-\theta_0\right\|^2 \\
&= \left\|E\left[\nabla^3 s\left(z_i,\theta_0\right)\right] + o_p(1)\right\| \left\|\tilde{\theta}-\theta_0\right\| \left\|\hat{\theta}-\theta_0\right\|^2 = O_p(n^{-3/2}),
\end{aligned}$$

where $\tilde{\tilde{\theta}}$ $(\tilde{\theta})$ lies between $\tilde{\theta}$ $(\hat{\theta})$ and $\theta_0$ noting $\hat{\theta}-\theta_0 = O_p\left(\frac{1}{\sqrt{n}}\right)$. The second last equality is obtained from Lemma A.3 under Assumption A.1. This implies that Condition (iv) can be replaced with another local uniform convergence condition $\left\|\frac{1}{n}\sum_{i=1}^n \nabla^3 s\left(z_i,\bar{\theta}\right) - E\left[\nabla^3 s\left(z_i,\theta_0\right)\right]\right\| = o_p(1)$ for $\bar{\theta} = \theta_0 + o_p(1)$ under Assumption A.1. Condition (v) is assumed in Assumption A.3. Condition (vi)-(viii) are by CLT provided that $E\left[\left\|\nabla^v s(z,\theta_0)\right\|^2\right] < \infty$, $v = \{0,1,2\}$ respectively, which are satisfied under Assumption A.2.

Now to establish additional preliminary lemmas, we need a stronger set of conditions as Assumption 1-2 or 3-2. Note that Assumption 1-2 implies Assumption A.1-A.3 and Assumption A.4 is weaker than Assumption 3. First, under Assumption 1 or 3, we have the uniform weak convergences (U-WCON) for the normalized sums of functions in $\nabla^v s(z,\theta), v = \{0,1,2,\dots\kappa\}$ up to the second order as in a neighborhood of $\theta_0$, denoted by $\Theta_0$ and hence it is not difficult to show that

**Lemma A.5.** *Under Assumption 1 or 3, we have*

$$\sup_{\theta\in\Theta}\left|\frac{1}{n}\sum_{i=1}^n \|\nabla^{v_1}s(z_i,\theta)\|\,\|\nabla^{v_2}s(z_i,\theta)\| - E\left[\|\nabla^{v_1}s(z_i,\theta)\|\,\|\nabla^{v_2}s(z_i,\theta)\|\right]\right| = o_p(1), \tag{A1}$$

*for* $v_1, v_2 \in \{0,1,2,\dots,\kappa\}, \kappa \geq 4$.

**Proof.** Provided Assumption 1 holds, (A1) is obtained by applying the Uniform Convergence theorem of Lemma A.1 by letting $m(z,\theta) = \|\nabla^{v_1}s(z_i,\theta)\|\,\|\nabla^{v_2}s(z_i,\theta)\|$. Noting

$$\begin{aligned} &E\left[\sup_{\theta\in\Theta_0}\|\nabla^{v_1}s(z_i,\theta)\|\,\|\nabla^{v_2}s(z_i,\theta)\|\right] \leq E\left[\sup_{\theta\in\Theta_0}\frac{\|\nabla^{v_1}s(z_i,\theta)\|^2+\|\nabla^{v_2}s(z_i,\theta)\|^2}{2}\right] \\ &\leq E\left[\sup_{\theta\in\Theta_0}\|\nabla^{v_1}s(z_i,\theta)\|^2\right]/2 + E\left[\sup_{\theta\in\Theta_0}\|\nabla^{v_2}s(z_i,\theta)\|^2\right]/2 < \infty, \end{aligned} \tag{A2}$$

which is satisfied by Assumption 1 (iii), all the conditions for Lemma A.1 are trivially satisfied.

Alternatively under Assumption 3, we obtain (A1) directly from Theorem 1-3 in Andrews (1994) [3], which is a quite general result and hence we rather provide a simple proof for our specific purpose. Noting other conditions for Lemma A.2 are trivially satisfied under Assumption 3, the uniform convergence result of (A1) is obtained upon verifying the Lipschitz condition for $\forall\theta_1,\theta_2 \in \Theta$ as

$$\begin{aligned} &\left|\|\nabla^{v_1}s(z,\theta_1)\|\,\|\nabla^{v_2}s(z,\theta_1)\| - \|\nabla^{v_1}s(z,\theta_2)\|\,\|\nabla^{v_2}s(z,\theta_2)\|\right| \\ &\leq \left|\|\nabla^{v_1}s(z,\theta_1)\|\,\|\nabla^{v_2}s(z,\theta_1)\| - \|\nabla^{v_1}s(z,\theta_1)\|\,\|\nabla^{v_2}s(z,\theta_2)\|\right| \\ &\quad + \left|\|\nabla^{v_1}s(z,\theta_1)\|\,\|\nabla^{v_2}s(z,\theta_2)\| - \|\nabla^{v_1}s(z,\theta_2)\|\,\|\nabla^{v_2}s(z,\theta_2)\|\right| \\ &= \|\nabla^{v_1}s(z,\theta_1)\|\left|\|\nabla^{v_2}s(z,\theta_1)\| - \|\nabla^{v_2}s(z,\theta_2)\|\right| + \|\nabla^{v_2}s(z,\theta_2)\|\left|\|\nabla^{v_1}s(z,\theta_1)\| - \|\nabla^{v_1}s(z,\theta_2)\|\right| \\ &\leq \|\nabla^{v_1}s(z,\theta_1)\|\,B_{v_2}(z)\,\|\theta_1-\theta_2\| + \|\nabla^{v_2}s(z,\theta_2)\|\,B_{v_1}(z)\,\|\theta_1-\theta_2\| \\ &\leq \sup_{\theta\in\Theta}\|\nabla^{v_1}s(z,\theta)\|\,B_{v_2}(z)\,\|\theta_1-\theta_2\| + \sup_{\theta\in\Theta}\|\nabla^{v_2}s(z,\theta)\|\,B_{v_1}(z)\,\|\theta_1-\theta_2\| \\ &= \left(\sup_{\theta\in\Theta}\|\nabla^{v_1}s(z,\theta)\|\,B_{v_2}(z) + \sup_{\theta\in\Theta}\|\nabla^{v_2}s(z,\theta)\|\,B_{v_1}(z)\right)\|\theta_1-\theta_2\| \equiv M(z)\,\|\theta_1-\theta_2\|, \end{aligned}$$

where the first inequality is by Triangle Inequality and the second inequality is obtained by the Lipschitz conditions for $\nabla^{v_1}s(z,\theta)$ and $\nabla^{v_2}s(z,\theta)$, since for $v = v_1, v_2$, $\left|\|\nabla^v s(z,\theta_1)\| - \|\nabla^v s(z,\theta_2)\|\right| \leq \|\nabla^v s(z,\theta_1) - \nabla^v s(z,\theta_2)\|$ by Triangle Inequality. Now we need to verify that $E\left[M(z)^{2+\delta}\right] < \infty$, which is true, since

$$E\left[\sup_{\theta\in\Theta}\|\nabla^{v_a}s(z,\theta)\|^{2+\delta}B_{v_b}(z)^{2+\delta}\right] \leq E\left[\left(\sup_{\theta\in\Theta}\|\nabla^{v_a}s(z,\theta)\|^{4+2\delta} + B_{v_b}(z)^{4+2\delta}\right)/2\right] < \infty \tag{A3}$$

for $(v_a, v_b) \in \{(v_1,v_2),(v_2,v_1)\}$ under $E\left[\sup_{\theta\in\Theta}\|\nabla^v s(z,\theta)\|^{4+\delta'}\right] < \infty$ for $v = v_1, v_2$, $E\left[B_{v_1}(z)^{4+\delta'}\right] < \infty$, and $E\left[B_{v_2}(z)^{4+\delta'}\right] < \infty$ with $\delta' = 2\delta$. □

**Lemma A.6.** *(Consistency of $\theta^*$) Suppose $\theta_0$ is the unique solution of (1) and $\theta^*$ solves (7) and further suppose $\sup_{\theta\in\Theta}\left\|\frac{1}{n}\sum_{i=1}^n s(z_i,\theta) - E[s(z,\theta)]\right\| = o_p(1)$ and $\sup_{\theta\in\Theta}\|\widehat{c}(\theta)\| = O_p(1)$, then $\theta^*$ is a consistent estimator of $\theta_0$.*

**Proof.** Let $\varepsilon > 0$. Then, there exists $\delta > 0$ such that whenever $\theta \in \Theta \backslash B(\theta_0, \varepsilon)$, we have $\|E[s(z_i, \theta)]\| > \delta$ provided that $\theta_0$ is the unique solution of (1). This implies

$$\begin{aligned}
&\Pr\left(\left\|\theta^* - \theta_0\right\| > \varepsilon\right) \leq \Pr\left(\left\|E[s(z_i, \theta^*)\right\| > \delta\right) = \Pr\left(\left\|E[s(z_i, \theta^*)] - \tfrac{1}{n}\textstyle\sum_{i=1}^n s(z_i, \theta^*) - \tfrac{1}{n}\widehat{c}(\theta^*)\right\| > \delta\right) \\
&\leq \Pr\left(\left\|E[s(z_i, \theta^*)] - \tfrac{1}{n}\textstyle\sum_{i=1}^n s(z_i, \theta^*)\right\| + \left\|\tfrac{1}{n}\widehat{c}(\theta^*)\right\| > \delta\right) \\
&\leq \Pr\left(\sup_{\theta\in\Theta}\left\|E[s(z_i, \theta)] - \tfrac{1}{n}\textstyle\sum_{i=1}^n s(z_i, \theta)\right\| + \tfrac{1}{n}\sup_{\theta\in\Theta}\|\widehat{c}(\theta)\| > \delta\right) \\
&= \Pr\left(o_p(1) > \delta\right) \to 0,
\end{aligned}$$

where the second inequality is by Triangle Inequality and the last equality is obtained provided that the uniform convergence of $\frac{1}{n}\sum_{i=1}^n s(z_i, \theta)$ to $E[s(z, \theta)]$ over $\theta \in \Theta$ and $\sup_{\theta\in\Theta}\|\widehat{c}(\theta)\| = O_p(1)$. The uniform convergence holds by Lemma A.1 or Lemma A.2 with $m(z, \theta) = s(z, \theta)$ provided that all the conditions in Lemma A.1 or Lemma A.2 are satisfied. The second necessary condition $\sup_{\theta\in\Theta}\|\widehat{c}(\theta)\| = O_p(1)$ is satisfied assuming conditions in Assumption 1-2 or Assumption 3-2 hold for the whole parameter space $\Theta$ instead of $\Theta_0$ similarly with Lemma A.7. □

**Lemma A.7.** *Under Assumption 1-2 or 3-2 (a) we have* $\widehat{c}(\theta) = c(\theta) + o_p(1)$ *uniformly over* $\theta \in \Theta_0 \subset \Theta$ *and (b) moreover, we have* $\widehat{c}(\theta_0) = c(\theta_0) + O_p\left(\frac{1}{\sqrt{n}}\right)$.

**Proof. Lemma A.7 (a)**

First we note that $c(\theta)$ is bounded uniformly over $\theta \in \Theta_0$ under Assumption 1 (ii)-(iii) and Assumption 2. This is obvious, since we can bound $\sup_{\theta\in\Theta_0}\|c(\theta)\|$ by sums and products of $\sup_{\theta\in\Theta_0}\|Q(\theta)\|$, $\sup_{\theta\in\Theta_0}\|\nabla s(z_i, \theta)\|^2$, and $\sup_{\theta\in\Theta_0}\|s(z_i, \theta)\|^2$ using Triangle Inequality, Cauchy-Schwarz Inequality, and the Dominated Convergence theorem. In what follows, we bound each term uniformly over $\theta \in \Theta_0$ and suppress the sup-norm over $\theta \in \Theta_0$ otherwise it is noted. Now for any $\theta \in \Theta_0$, note

$$\begin{aligned}
&\widehat{c}(\theta) - c(\theta) \\
= \;& \frac{1}{2}\left(\widehat{H}_2(\theta)(\frac{1}{n}\sum_{i=1}^n[\widehat{Q}(\theta)s(z_i, \theta) \otimes \widehat{Q}(\theta)s(z_i, \theta)]) - H_2(\theta)(E[Q(\theta)s(z_i, \theta) \otimes Q(\theta)s(z_i, \theta)])\right) && \text{(A4)} \\
& + \left(\frac{1}{n}\textstyle\sum_{i=1}^n\left[\nabla s(z_i, \theta)\widehat{Q}(\theta)s(z_i, \theta)\right] - E\left[\nabla s(z_i, \theta)Q(\theta)s(z_i, \theta)\right]\right) && \text{(A5)}
\end{aligned}$$

Now rewrite (A5) as

$$\begin{aligned}
&\frac{1}{n}\textstyle\sum_{i=1}^n\left[\nabla s(z_i, \theta)\widehat{Q}(\theta)s(z_i, \theta)\right] - E\left[\nabla s(z_i, \theta)Q(\theta)s(z_i, \theta)\right] \\
= \;& \frac{1}{n}\textstyle\sum_{i=1}^n\left[\nabla s(z_i, \theta)\widehat{Q}(\theta)s(z_i, \theta) - \nabla s(z_i, \theta)Q(\theta)s(z_i, \theta)\right] && \text{(A6)} \\
& + \frac{1}{n}\textstyle\sum_{i=1}^n\left[\nabla s(z_i, \theta)Q(\theta)s(z_i, \theta) - E\left[\nabla s(z_i, \theta)Q(\theta)s(z_i, \theta)\right]\right]. && \text{(A7)}
\end{aligned}$$

Then we have for (A6),

$$\left\|\frac{1}{n}\textstyle\sum_{i=1}^n\left[\nabla s(z_i, \theta)\left(\widehat{Q}(\theta) - Q(\theta)\right)s(z_i, \theta)\right]\right\| \leq \frac{1}{n}\textstyle\sum_{i=1}^n\|\nabla s(z_i, \theta)\|\,\|s(z_i, \theta)\|\left\|\widehat{Q}(\theta) - Q(\theta)\right\| \quad \text{(A8)}$$

by Triangle Inequality and Cauchy-Schwarz Inequality. In what follows, again we treat $\widehat{H}_1(\theta)\left(= -\widehat{Q}(\theta)^{-1}\right)$ as nonsingular for $\theta \in \Theta_0$. This is innocuous, since by Lemma A.1 or A.2 with

$m(z,\theta) = \nabla s(z,\theta)$ and Assumption 2, with probability approaching to one, $\widehat{H}_1(\theta)$ is nonsingular for $\theta \in \Theta_0$. Now note

$$\begin{aligned} & \left\| \widehat{Q}(\theta) - Q(\theta) \right\| \\ = \; & \left\| Q(\theta) \left( \widehat{Q}(\theta)^{-1} - Q(\theta)^{-1} \right) \widehat{Q}(\theta) \right\| \leq \|Q(\theta)\| \left\| \widehat{Q}(\theta) \right\| \left\| \widehat{Q}(\theta)^{-1} - Q(\theta)^{-1} \right\| = o_p(1) \end{aligned} \tag{A9}$$

by the uniform convergence of $\widehat{Q}(\theta)^{-1}$ to $Q(\theta)^{-1}$ and Assumption 2 applying the Slutsky theorem. We have $\frac{1}{n}\sum_{i=1}^{n} \|\nabla s(z_i,\theta)\| \, \|s(z_i,\theta)\| = O_p(1)$ by (A2) and Lemma A.5 with $v_1 = 1$ and $v_2 = 0$. Together with (A9), this implies (A6) is $o_p(1)$. Now note we have

$$\begin{aligned} & E\left[\sup_{\theta\in\Theta_0} \|\nabla s(z_i,\theta)\, Q(\theta) s(z_i,\theta)\|\right] \\ \leq \; & E\left[\sup_{\theta\in\Theta_0} \|s(z_i,\theta)\| \, \|\nabla s(z_i,\theta)\| \, \|Q(\theta)\|\right] \leq CE\left[\sup_{\theta\in\Theta_0} \|s(z_i,\theta)\| \, \|s(z_i,\theta)\|\right] < \infty, \end{aligned}$$

from (A2) and $\sup_{\theta\in\Theta_0} \|Q(\theta)\| < \infty$ or Lipschitz condition as

$$\begin{aligned} & \|\nabla s(z,\theta_1)\, Q(\theta_1) s(z,\theta_1) - \nabla s(z,\theta_2)\, Q(\theta_2) s(z,\theta_2)\| \\ \leq \; & \sup_{\theta\in\Theta_0} \|Q(\theta) s(z,\theta)\| \, \|\nabla s(z,\theta_1) - \nabla s(z,\theta_2)\| \\ & + \sup_{\theta\in\Theta_0} \|Q(\theta)\nabla s(z,\theta)\| \, \|s(z,\theta_1) - s(z,\theta_2)\| + \sup_{\theta\in\Theta_0} \|s(z,\theta)\, \nabla s(z,\theta)\| \, \|Q(\theta_1) - Q(\theta_2)\| \\ \leq \; & \sup_{\theta\in\Theta_0} \|Q(\theta)\| \sup_{\theta\in\Theta_0} \|s(z,\theta)\| \, B_1(z) \, \|\theta_1 - \theta_2\| \\ & + \sup_{\theta\in\Theta_0} \|Q(\theta)\| \sup_{\theta\in\Theta_0} \|\nabla s(z,\theta)\| \, B_0(z) \, \|\theta_1 - \theta_2\| \\ & + \sup_{\theta\in\Theta_0} \|s(z,\theta)\| \sup_{\theta\in\Theta_0} \|\nabla s(z,\theta)\| \left(\sup_{\theta\in\Theta_0} \|Q(\theta)\|\right)^2 \sup_{\theta\in\Theta_0} \|\nabla H_1(\theta)\| \, \|\theta_1 - \theta_2\| \\ \equiv \; & \overline{M}(z) \, \|\theta_1 - \theta_2\|, \end{aligned}$$

where the first inequality is obtained by Triangle Inequality and Cauchy-Schwarz Inequality and the second inequality is obtained by Lipschitz conditions for $s(z,\theta)$ and $\nabla s(z,\theta)$ and since $\|Q(\theta_1) - Q(\theta_2)\| = \|Q(\theta_1)\| \left\|Q^{-1}(\theta_1) - Q^{-1}(\theta_2)\right\| \cdot \|Q(\theta_2)\|$. In the last equality, we set

$$\begin{aligned} \overline{M}(z) = \; & \sup_{\theta\in\Theta_0} \|Q(\theta)\| \sup_{\theta\in\Theta_0} \|s(z,\theta)\| \, B_1(z) + \sup_{\theta\in\Theta_0} \|Q(\theta)\| \sup_{\theta\in\Theta_0} \|\nabla s(z,\theta)\| \, B_0(z) \\ & + \sup_{\theta\in\Theta_0} \|s(z,\theta)\| \sup_{\theta\in\Theta_0} \|\nabla s(z,\theta)\| \left(\sup_{\theta\in\Theta_0} \|Q(\theta)\|\right)^2 \sup_{\theta\in\Theta_0} \|\nabla H_1(\theta)\| \end{aligned}$$

and we have $E\left[\overline{M}(z)^{2+\delta}\right] < \infty$ by a similar argument with (A3) provided that $E\left[\sup_{\theta\in\Theta_0} \|s(z,\theta)\|^{4+\delta'}\right] < \infty$, $E\left[\sup_{\theta\in\Theta_0} \|\nabla s(z,\theta)\|^{4+\delta'}\right] < \infty$, $E\left[B_1(z)^{4+\delta'}\right] < \infty$, and $E\left[B_0(z)^{4+\delta'}\right] < \infty$ with $\delta' = 2\delta$, and also assuming $\sup_{\theta\in\Theta_0} \|\nabla H_1(\theta)\| < \infty$. Therefore, we can apply the Uniform Convergence theorem of Lemma A.1 or Lemma A.2 to (A7) and have

$$\sup_{\theta\in\Theta_0} \left\| \frac{1}{n} \sum_{i=1}^{n} \left[\nabla s(z_i,\theta)\, Q(\theta) s(z_i,\theta) - E\left[\nabla s(z_i,\theta)\, Q(\theta) s(z_i,\theta)\right]\right] \right\| = o_p(1). \tag{A10}$$

From (A6)=$o_p(1)$ and (A10), we conclude (A5) is $o_p(1)$ uniformly over $\theta \in \Theta_0$. Now consider

$$\sup_{\theta\in\Theta_0} \left\| \widehat{H}_2(\theta) - H_2(\theta) \right\| = o_p(1) \tag{A11}$$

by the uniform convergence from Lemma A.1 or A.2 with $m(z,\theta) = \nabla^2 s(z,\theta)$ and that

$$\left\| \frac{1}{n} \sum_{i=1}^{n} s(z_i,\theta) s(z_i,\theta)' - E\left[s(z_i,\theta) s(z_i,\theta)'\right] \right\| = o_p(1) \tag{A12}$$

uniformly over $\theta \in \Theta_0$ by the uniform convergence result of Lemma A.1 with $m(z,\theta) = s(z_i,\theta)s(z_i,\theta)'$ provided that $E\left[\sup_{\theta\in\Theta_0} \|s(z_i,\theta)\|^2\right] < \infty$ or Lemma A.2 by verifying the Lipschitz condition as

$$\begin{aligned} & \left\|s(z,\theta_1)s(z,\theta_1)' - s(z,\theta_2)\,s(z,\theta_2)'\right\| \\ \le\ & \left\|s(z,\theta_1)s(z,\theta_1)' - s(z,\theta_1)\,s(z,\theta_2)'\right\| + \left\|s(z,\theta_1)s(z,\theta_2)' - s(z,\theta_2)\,s(z,\theta_2)'\right\| \\ \le\ & 2\sup_{\theta\in\Theta} \|s(z,\theta)\|\,\|s(z,\theta_1) - s(z,\theta_2)\| \le 2\sup_{\theta\in\Theta}\|s(z,\theta)\|\,B_0(z)\,\|\theta_1-\theta_2\| \end{aligned} \tag{A13}$$

by Triangle Inequality and noting that $E\left[\sup_{\theta\in\Theta}\|s(z,\theta)\|^{2+\delta} B_0(z)^{2+\delta}\right] < \infty$ under $E\left[\sup_{\theta\in\Theta}\|s(z,\theta)\|^{4+\delta'}\right] < \infty$ and $E\left[B_0(z)^{4+\delta'}\right] < \infty$ with $\delta' = 2\delta$. From (A9), (A11), and (A12), it follows that

$$\begin{aligned} & \widehat{H}_2(\theta)\left(\tfrac{1}{n}\textstyle\sum_{i=1}^n\left[\widehat{Q}(\theta)s(z_i,\theta)\otimes\widehat{Q}(\theta)s(z_i,\theta)\right]\right) \\ &= \widehat{H}_2(\theta)\left(vec\left(\widehat{Q}(\theta)\left(\tfrac{1}{n}\textstyle\sum_{i=1}^n s(z_i,\theta)\,s(z_i,\theta)'\right)\widehat{Q}(\theta)'\right)\right) \\ &= \left(H_2(\theta)+o_p(1)\right)\left(vec\left((Q(\theta)+o_p(1))\left(E\left[s(z_i,\theta)\,s(z_i,\theta)'\right]+o_p(1)\right)(Q(\theta)+o_p(1))'\right)\right) \\ &= H_2(\theta)\left(vec\left(Q(\theta)\left(E\left[s(z_i,\theta)\,s(z_i,\theta)'\right]\right)Q(\theta)'\right)\right)+o_p(1) \\ &= H_2(\theta)\left(E\left[Q(\theta)s(z_i,\theta)\otimes Q(\theta)s(z_i,\theta)\right]\right)+o_p(1), \end{aligned} \tag{A14}$$

where the first and the last equality come from $vec(gg') = g\otimes g$ for a column vector $g$ and hence we bound (A4) as $o_p(1)$ uniformly over $\theta\in\Theta_0$. This concludes $\widehat{c}(\theta) = c(\theta)+o_p(1)$ uniformly over $\theta\in\Theta_0$. □

**Proof. Lemma A.7 (b)**

Note

$$\sqrt{n}\left(\widehat{Q}(\theta_0) - Q(\theta_0)\right) = O_p(1) \tag{A15}$$

by the Slutsky theorem and that

$$\sqrt{n}\left(\widehat{H}_2(\theta_0) - H_2(\theta_0)\right) = \frac{1}{\sqrt{n}}\sum_{i=1}^n\left(\nabla^2 s(z_i,\theta_0) - E\left[\nabla^2 s(z_i,\theta_0)\right]\right) = O_p(1) \tag{A16}$$

by the CLT under $E[\|\nabla^2 s(z_i,\theta_0)\|^2] < \infty$ and that by the CLT

$$\frac{1}{\sqrt{n}}\sum_{i=1}^n s(z_i,\theta_0)s(z_i,\theta_0)' - E\left[s(z_i,\theta_0)s(z_i,\theta_0)'\right] = O_p(1) \tag{A17}$$

under $E\left[\|s(z_i,\theta_0)s(z_i,\theta_0)'\|^2\right] = E\left[\|s(z_i,\theta_0)\|^4\right] < \infty$. We can also apply the CLT to

$$\frac{1}{\sqrt{n}}\sum_{i=1}^n \|\nabla s(z_i,\theta_0)\|\,\|s(z_i,\theta_0)\| = E\left[\|\nabla s(z_i,\theta_0)\|\,\|s(z_i,\theta_0)\|\right] + O_p(1) \tag{A18}$$

under (a) $E\left[\|\nabla s(z_i,\theta_0)\|^2\,\|s(z_i,\theta_0)\|^2\right] < \infty$ and to

$$\frac{1}{\sqrt{n}}\sum_{i=1}^n\left[\nabla s(z_i,\theta_0)\,Q(\theta_0)s(z_i,\theta_0) - E\left[\nabla s(z_i,\theta_0)\,Q(\theta_0)s(z_i,\theta_0)\right]\right] = O_p(1) \tag{A19}$$

under (b) $E\left[\|\nabla s(z_i,\theta_0)Q(\theta_0)s(z_i,\theta_0)\|^2\right]<\infty$. Both (a) and (b) are satisfied provided that $E\left[\|s(z_i,\theta_0)\|^4\right]<\infty$ and $E\left[\|\nabla s(z_i,\theta_0)\|^4\right]<\infty$, since

$$\begin{aligned}&E\left[\|\nabla s(z_i,\theta_0)\,Q(\theta_0)s(z_i,\theta_0)\|^2\right]\le\|Q(\theta_0)\|^2\,E\left[\|s(z_i,\theta_0)\|^2\,\|\nabla s(z_i,\theta_0)\|^2\right]\\ \le\;&C\left(E\left[\left(\|s(z_i,\theta_0)\|^4+\|\nabla s(z_i,\theta_0)\|^4\right)/2\right]\right)<\infty\end{aligned}$$

by Cauchy-Schwarz Inequality. Applying the results of (A15), (A16), and (A17) to (A14), we can show that (A4)=$O_p\left(1/\sqrt{n}\right)$ for $\theta=\theta_0$. Similarly, plugging the results of (A15), (A18), and (A19) into (A7) and (A8), we obtain (A5)=$O_p\left(1/\sqrt{n}\right)$ for $\theta=\theta_0$ and hence we conclude that $\hat{c}(\theta_0)=c(\theta_0)+O_p\left(1/\sqrt{n}\right)$. $\square$

To characterize $\nabla\hat{c}(\theta)$, we introduce some matrix differentiation results consistent with our notation. We denote a $m\times n$ matrix $D$ as $(d_{ij})_n^m$, where $d_{ij}$ is the $i$-th row and the $j$-th column element of $D$. Also we denote a $m\times k^n$ matrix $E$ as $[e_{ij}]_n^m$, where $e_{ij}$ is a $1\times k$ vector such that

$$E=[e_{ij}]_n^m=\begin{pmatrix}e_{11}&\cdots&e_{1k^{n-1}}\\ \vdots&\ddots&\vdots\\ e_{m1}&\cdots&e_{mk^{n-1}}\end{pmatrix}$$

and hence $e_{ij}=(E_{i,(j-1)k+1},E_{i,(j-1)k+2},\ldots,E_{i,j\cdot k})$ by defining $E_{u,v}$ as the $u$-th row and the $v$-th column element of $E$.

**Remark A.1.** *For $k\times k$ matrices $A$ and $B$, we have $\nabla(AB)=A\nabla B+B'\nabla(A')$.*

**Proof.** Let $C=AB$. Then, we have $c_{ij}=\sum_{l=1}^k a_{il}b_{lj}$ and hence

$$\begin{aligned}\nabla C&=\left[\nabla c_{ij}\right]_2^k=\left[\nabla(\textstyle\sum_{l=1}^k a_{il}b_{lj})\right]_2^k=\left[\textstyle\sum_{l=1}^k a_{il}\nabla b_{lj}\right]_2^k+\left[\textstyle\sum_{l=1}^k b_{lj}\nabla a_{il}\right]_2^k\\ &=(a_{ij})_k^k\left[\nabla b_{ij}\right]_2^k+(b_{ji})_k^k\left[\nabla a_{ji}\right]_2^k=A\nabla B+B'\nabla(A')\end{aligned}$$

$\square$

**Remark A.2.** *For a $k^m\times k^n$ matrix $A$ and a $k^n\times 1$ vector $b$ with $m,n=0,1,2,\ldots$, we have*

$$\nabla(Ab)=A\nabla b+vec^*\left(b'\nabla(A')\right),$$

*where $vec^*((a_1,a_2,\ldots,a_k))=\begin{pmatrix}a_1\\ \vdots\\ a_k\end{pmatrix}$ and $a_j$ is a $1\times k$ vector for $j=1,\ldots,k$. For completeness, we let $vec^*(c)=c$ for a scalar $c$.*

**Proof.** Let $c=Ab$ and note $\nabla c_i=\sum_{l=1}^{k^n}a_{il}\nabla b_l+\sum_{l=1}^{k^n}b_l\nabla a_{il}$. This implies

$$\begin{aligned}\left[\nabla c_i\right]_1^{k^m}&=\left[\textstyle\sum_{l=1}^{k^n}a_{il}\nabla b_l\right]_1^{k^m}+\left[\textstyle\sum_{l=1}^{k^n}b_l\nabla a_{il}\right]_1^{k^m}\\ &=(a_{ij})_{k^n}^{k^m}\nabla b+\begin{pmatrix}b'[\nabla a_{j1}]_1^{k^n}\\ \vdots\\ b'[\nabla a_{jk^m}]_1^{k^n}\end{pmatrix}=A\nabla b+vec^*\left(b'\nabla(A')\right).\end{aligned}$$

$\square$

**Remark A.3.** *Moreover, we have* $\nabla\left(vec^{*}(\cdot)\right) = vec^{*}(\nabla(\cdot))$ *by definition of* $vec^{*}$.

**Proof.** For a $1 \times k^m$ vector $c = (c_1, \ldots, c_m)$ with $c_i$ to be a $1 \times k$ vector and $i = 1, \ldots, m$, consider

$$\begin{aligned}
&\nabla\left(vec^{*}(c)\right) \\
&= \nabla\left(vec^{*}((c_1, \ldots, c_m))\right) = \nabla\begin{pmatrix} c_1 \\ \vdots \\ c_m \end{pmatrix} = \begin{pmatrix} \nabla c_1 \\ \vdots \\ \nabla c_m \end{pmatrix} = vec^{*}\left((\nabla c_1, \ldots, \nabla c_m)\right) = vec^{*}(\nabla c).
\end{aligned}$$

□

**Remark A.4.** *For matrices (including column and row vectors) A and B, we have*

$$\nabla(A \otimes B) = (A \otimes \nabla B) + \left(\nabla A \otimes^{*} B\right),$$

*where we define* $\otimes^{*}$ *for matrices* $D$ $(m \times k^n)$ *and* $E$ $(p \times q)$ *as*

$$\begin{aligned}
&D \otimes^{*} E \\
&\equiv \begin{pmatrix}
d_{11}e_{11} & \cdots & d_{11}e_{1q} & & d_{1k^{n-1}}e_{11} & \cdots & d_{1k^{n-1}}e_{1q} \\
\vdots & \ddots & \vdots & \cdots & \vdots & \ddots & \vdots \\
d_{11}e_{p1} & \cdots & d_{11}e_{pq} & & d_{1k^{n-1}}e_{p1} & \cdots & d_{1k^{n-1}}e_{pq} \\
 & \vdots & & \ddots & & \vdots & \\
d_{m1}e_{11} & \cdots & d_{m1}e_{1q} & & d_{mk^{n-1}}e_{11} & \cdots & d_{mk^{n-1}}e_{1q} \\
\vdots & \ddots & \vdots & \cdots & \vdots & \ddots & \vdots \\
d_{m1}e_{p1} & \cdots & d_{m1}e_{pq} & & d_{mk^{n-1}}e_{p1} & \cdots & d_{mk^{n-1}}e_{pq}
\end{pmatrix} \\
&= \begin{pmatrix}
E \otimes d_{11} & \cdots & E \otimes d_{1k^{n-1}} \\
\vdots & \ddots & \vdots \\
E \otimes d_{m1} & \cdots & E \otimes d_{mk^{n-1}}
\end{pmatrix}
\end{aligned}$$

*for* $1 \times k$ *vector* $d_{ij}$, $i = 1, \ldots, m$ *and* $j = 1, \ldots, n$ *and* $e_{uv}$ *is the u-th row and the v-th column element of E.*

**Proof.** Consider

$$\begin{aligned}
\nabla(A \otimes B) &= \begin{pmatrix}
\nabla(a_{11}B) & \cdots & \nabla(a_{1k^{n-1}}B) \\
\vdots & \ddots & \vdots \\
\nabla(a_{m1}B) & \cdots & \nabla(a_{mk^{n-1}}B)
\end{pmatrix} \\
&= \begin{pmatrix}
a_{11}\nabla B & \cdots & a_{1k^{n-1}}\nabla B \\
\vdots & \ddots & \vdots \\
a_{m1}\nabla B & \cdots & a_{mk^{n-1}}\nabla B
\end{pmatrix} + \begin{pmatrix}
B \otimes \nabla a_{11} & \cdots & B \otimes \nabla a_{1k^{n-1}} \\
\vdots & \ddots & \vdots \\
B \otimes \nabla a_{m1} & \cdots & B \otimes \nabla a_{mk^{n-1}}
\end{pmatrix} \\
&= (A \otimes \nabla B) + \left(\nabla A \otimes^{*} B\right).
\end{aligned}$$

□

**Remark A.5.** *For an invertible matrix A* $(k \times k)$, *we have* $\nabla\left(A^{-1}\right) = -A^{-1}\left(A'\right)^{-1}\nabla(A') = -(A'A)^{-1}\nabla(A')$.

**Proof.** From $(A')^{-1}A' = I$, we have $\nabla\left((A')^{-1}A'\right) = \nabla I = 0$ and hence from Remark A.1, $(A')^{-1}\nabla(A') + A\nabla\left(A^{-1}\right) = 0$. Multiplying $A^{-1}$ each side, we have

$$A^{-1}\left((A')^{-1}\nabla(A') + A\nabla\left(A^{-1}\right)\right) = 0,\ \ (A'A)^{-1}\nabla(A') + \nabla\left(A^{-1}\right) = 0,$$

which gives $\nabla\left(A^{-1}\right) = -A^{-1}(A')^{-1}\nabla(A')$. $\square$

**Lemma A.8.** *Under Assumption 1-2 or 3-2,we have (a)* $\left\|\nabla^v\left(\widehat{Q}(\theta)'\right)\right\| = \left\|\nabla^v\widehat{Q}(\theta)\right\| = O_p(1)$ *and (b)* $\nabla^{v-1}\left(\widehat{Q}(\theta_0)'\right) - \nabla^{v-1}\left(Q(\theta_0)'\right) = O_p\left(1/\sqrt{n}\right)$ *for* $\theta \in \Theta_0$ *and* $v = \{1,2,3\}$

**Proof.** For $\theta \in \Theta_0$, note Remark A.5 implies (noting $\widehat{Q}(\theta)^{-1} = -\widehat{H}_1(\theta) = -\frac{1}{n}\sum_{i=1}^n \nabla s(z_i,\theta)$ and $\widehat{H}_2(\theta) = \nabla\widehat{H}_1(\theta)$ by definition)

$$\nabla\left(\widehat{Q}(\theta)'\right) = -\nabla\left(\left(\widehat{H}_1(\theta)'\right)^{-1}\right) = \left(\widehat{H}_1(\theta)'\right)^{-1}\widehat{H}_1(\theta)^{-1}\nabla\widehat{H}_1(\theta) = \widehat{Q}(\theta)'\widehat{Q}(\theta)\widehat{H}_2(\theta) \tag{A20}$$

and hence $||\nabla(\widehat{Q}(\theta)')|| \leq ||\widehat{Q}(\theta)||^2||\widehat{H}_2(\theta)|| = O_p(1)$ by (A9) and (A11). Now consider

$$\begin{aligned}
&\left\|\nabla^2\left(\widehat{Q}(\theta)'\right)\right\| = \left\|\nabla\left(\widehat{Q}(\theta)'\widehat{Q}(\theta)\widehat{H}_2(\theta)\right)\right\| \\
&\leq \left\|\nabla\left(\widehat{Q}(\theta)'\widehat{Q}(\theta)\right)\right\|\left\|\widehat{H}_2(\theta)\right\| + \left\|\widehat{Q}(\theta)'\widehat{Q}(\theta)\right\|\left\|\nabla\widehat{H}_2(\theta)\right\| \\
&\leq 2\left\|\widehat{Q}(\theta)'\right\|\left\|\nabla\widehat{Q}(\theta)\right\|\left\|\widehat{H}_2(\theta)\right\| + \left\|\widehat{Q}(\theta)'\widehat{Q}(\theta)\right\|\left\|\nabla\widehat{H}_2(\theta)\right\| \\
&\leq 2\left\|\widehat{Q}(\theta)\right\|\left\|\nabla\widehat{Q}(\theta)\right\|\left\|\widehat{H}_2(\theta)\right\| + \left\|\widehat{Q}(\theta)\right\|^2\left\|\nabla\widehat{H}_2(\theta)\right\| = O_p(1),
\end{aligned} \tag{A21}$$

noting $\left\|\nabla\left(\widehat{Q}(\theta)'\right)\right\| = \left\|\nabla\widehat{Q}(\theta)\right\|$ and since

$$\left\|\nabla\widehat{H}_2(\theta)\right\| = \left\|\frac{1}{n}\sum_{i=1}^n \nabla^3 s(z_i,\theta)\right\| = \left\|E\left[\nabla^3 s(z_i,\theta)\right]\right\| + o_p(1) = O_p(1) \tag{A22}$$

applying the Uniform Convergence theorem of Lemma A.1 or A.2 with $m(z,\theta) = \nabla^3 s(z,\theta)$. Similarly we can show that

$$\begin{aligned}
&\left\|\nabla^3\left(\widehat{Q}(\theta)'\right)\right\| = \left\|\nabla^2\left(\widehat{Q}(\theta)'\widehat{Q}(\theta)\widehat{H}_2(\theta)\right)\right\| \\
&\leq \left\|\nabla^2\left(\widehat{Q}(\theta)'\widehat{Q}(\theta)\right)\right\|\left\|\widehat{H}_2(\theta)\right\| + \left\|\nabla\left(\widehat{Q}(\theta)'\widehat{Q}(\theta)\right)\right\|\left\|\nabla\widehat{H}_2(\theta)\right\| \\
&\quad + \left\|\nabla\left(\widehat{Q}(\theta)'\widehat{Q}(\theta)\right)\right\|\left\|\nabla\widehat{H}_2(\theta)\right\| + \left\|\widehat{Q}(\theta)'\widehat{Q}(\theta)\right\|\left\|\nabla^2\widehat{H}_2(\theta)\right\| \\
&= \left\|\nabla^2\left(\widehat{Q}(\theta)'\widehat{Q}(\theta)\right)\right\|\left\|\widehat{H}_2(\theta)\right\| + O_p(1) + O_p(1) + O_p(1) = O_p(1).
\end{aligned}$$

from (A9), (A20), (A21), and by the uniform convergence of $\nabla^2\widehat{H}_2(\theta) = \frac{1}{n}\sum_{i=1}^n \nabla^4 s(z_i,\theta)$ to $E\left[\nabla^4 s(z,\theta)\right]$ from Lemma A.1 or A.2 with $m(z,\theta) = \nabla^4 s(z,\theta)$. The last equality is obtained noting we have

$$\left\|\nabla^2\left(\widehat{Q}(\theta)'\widehat{Q}(\theta)\right)\right\| \leq 2\left\|\nabla\widehat{Q}(\theta)\right\|\left\|\nabla\widehat{Q}(\theta)\right\| + 2\left\|\widehat{Q}(\theta)'\right\|\left\|\nabla^2\widehat{Q}(\theta)\right\| = O_p(1)$$

from (A9), (A20), and (A21). Now to show the second result, first note that we can rewrite

$$\widehat{Q}(\theta_0) = \left(-H_1 - V/\sqrt{n}\right)^{-1} = Q + O_p\left(1/\sqrt{n}\right) \tag{A23}$$

by the Slutsky theorem and $V = O_p(1)$ by CLT and also we rewrite

$$\widehat{H}_2(\theta_0) = H_2 + W/\sqrt{n} = H_2 + O_p\left(1/\sqrt{n}\right), \tag{A24}$$

since $W = O_p(1)$ by CLT. From (A20), consider

$$\begin{aligned}\nabla\left(\widehat{Q}(\theta_0)'\right) &= \widehat{Q}(\theta_0)'\widehat{Q}(\theta_0)\widehat{H}_2(\theta_0) \\ &= \left(Q + O_p\left(1/\sqrt{n}\right)\right)'\left(Q + O_p\left(1/\sqrt{n}\right)\right)\left(H_2 + O_p\left(1/\sqrt{n}\right)\right) \\ &= Q'QH_2 + O_p\left(1/\sqrt{n}\right) = \nabla\left(Q(\theta_0)'\right) + O_p\left(1/\sqrt{n}\right)\end{aligned}$$

using (A23) and (A24). Similarly from (A21), we have $\nabla^2\left(\widehat{Q}(\theta_0)'\right) = \nabla^2\left(Q(\theta_0)'\right) + O_p\left(1/\sqrt{n}\right)$ from (A23) and (A24)and noting $\nabla\widehat{H}_2(\theta_0) = \nabla H_2 + \nabla\left(W/\sqrt{n}\right)$ and $\nabla\left(W/\sqrt{n}\right) = \nabla W/\sqrt{n} = W_3/\sqrt{n} = O_p\left(1/\sqrt{n}\right)$. □

In the following proof, we will apply Triangle Inequality and Cauchy-Schwarz Inequality whenever they are necessary without noting them.

**Lemma A.9.** *Under Assumption 1-2 or 3-2, Condition 1-3 are satisfied.*

**Proof. Condition 1**

$\widehat{c}(\theta_0) = O_p(1)$ is obvious from Lemma A.7. □

**Proof. Condition 2**

Again we bound each term uniformly over $\theta \in \Theta_0$ and suppress the sup-norm otherwise it is noted. Now consider for $\theta \in \Theta_0$

$$\begin{aligned}&\nabla\widehat{c}(\theta) \\ &= \nabla\left(\tfrac{1}{2}\widehat{H}_2(\theta)\left(\tfrac{1}{n}\textstyle\sum_{i=1}^n\left[\widehat{Q}(\theta)s(z_i,\theta)\otimes\widehat{Q}(\theta)s(z_i,\theta)\right]\right) + \tfrac{1}{n}\textstyle\sum_{i=1}^n\left[\nabla s(z_i,\theta)\widehat{Q}(\theta)s(z_i,\theta)\right]\right) \\ &= \nabla\left(\tfrac{1}{2}\widehat{H}_2(\theta)\left(\tfrac{1}{n}\textstyle\sum_{i=1}^n\left[\widehat{Q}(\theta)s(z_i,\theta)\otimes\widehat{Q}(\theta)s(z_i,\theta)\right]\right)\right) + \nabla\left(\tfrac{1}{n}\textstyle\sum_{i=1}^n\left[\nabla s(z_i,\theta)\widehat{Q}(\theta)s(z_i,\theta)\right]\right) \\ &= \tfrac{1}{2}vec^*\left(\left(\tfrac{1}{n}\textstyle\sum_{i=1}^n\left[\widehat{Q}(\theta)s(z_i,\theta)\otimes\widehat{Q}(\theta)s(z_i,\theta)\right]\right)'\left(\nabla\left(\widehat{H}_2(\theta)'\right)\right)\right) \\ &\quad + \tfrac{1}{2}\widehat{H}_2(\theta)\nabla\left(\tfrac{1}{n}\textstyle\sum_{i=1}^n\left[\widehat{Q}(\theta)s(z_i,\theta)\otimes\widehat{Q}(\theta)s(z_i,\theta)\right]\right) + \nabla\left(\tfrac{1}{n}\textstyle\sum_{i=1}^n\left[\nabla s(z_i,\theta)\widehat{Q}(\theta)s(z_i,\theta)\right]\right),\end{aligned} \tag{A25}$$

using Remark A.2. For the first RHS term of the last equality in (A25), note

$$\begin{aligned}&\left\|\nabla\left(\widehat{H}_2(\theta)'\right)\right\| = \left\|\tfrac{1}{n}\textstyle\sum_{i=1}^n\nabla\left(\left(\nabla^2 s(z_i,\theta)\right)'\right)\right\| \\ &\leq \tfrac{1}{n}\textstyle\sum_{i=1}^n\left\|\nabla\left(\left(\nabla^2 s(z_i,\theta)\right)'\right)\right\| = \tfrac{1}{n}\textstyle\sum_{i=1}^n\left\|\nabla^3 s(z_i,\theta)\right\| = E\left[\left\|\nabla^3 s(z_i,\theta)\right\|\right] + o_p(1) = O_p(1)\end{aligned} \tag{A26}$$

uniformly over $\theta \in \Theta_0$ applying the Uniform Convergence theorem of Lemma A.1 or Lemma A.2 with $m(z,\theta) = \nabla^3 s(z,\theta)$. We have shown that

$$\left\|\frac{1}{n}\sum_{i=1}^n\left[\widehat{Q}(\theta)s(z_i,\theta)(\theta)\otimes\widehat{Q}(\theta)s(z_i,\theta)(\theta)\right] - Q(\theta)\,E\left[s(z_i,\theta)s(z_i,\theta)'\right]Q(\theta)'\right\| = o_p(1) \tag{A27}$$

uniformly over $\theta \in \Theta_0$ in (A14) and from this result with (A26), we bound the first RHS term of the last equality in (A25) as

$$\begin{aligned}
&\left\| \tfrac{1}{2} vec^* \left( \left( \tfrac{1}{n} \textstyle\sum_{i=1}^n \left[ \widehat{Q}(\theta) s(z_i,\theta) \otimes \widehat{Q}(\theta) s(z_i,\theta) \right] \right)' \left( \nabla \left( \widehat{H}_2(\theta)' \right) \right) \right\| \\
&= \tfrac{1}{2} \left\| \left( \tfrac{1}{n} \textstyle\sum_{i=1}^n \left[ \widehat{Q}(\theta) s(z_i,\theta)(\theta) \otimes \widehat{Q}(\theta) s(z_i,\theta) \right] \right)' \left( \nabla \left( \widehat{H}_2(\theta)' \right) \right) \right\| \\
&\le \left\| \tfrac{1}{n} \textstyle\sum_{i=1}^n \left[ \widehat{Q}(\theta) s(z_i,\theta) \otimes \widehat{Q}(\theta) s(z_i,\theta) \right] \right\| \left\| \left( \nabla \left( \widehat{H}_2(\theta)' \right) \right) \right\| = O_p(1).
\end{aligned}$$

Now consider

$$\begin{aligned}
& \nabla \left( \frac{1}{n} \textstyle\sum_{i=1}^n \left[ \widehat{Q}(\theta) s(z_i,\theta) \otimes \widehat{Q}(\theta) s(z_i,\theta) \right] \right) && \text{(A28)} \\
=\; & \frac{1}{n} \textstyle\sum_{i=1}^n \nabla \left[ \widehat{Q}(\theta) s(z_i,\theta) \otimes \widehat{Q}(\theta) s(z_i,\theta) \right] \\
=\; & \frac{1}{n} \textstyle\sum_{i=1}^n \left( \nabla \left( \widehat{Q}(\theta) s(z_i,\theta) \right) \otimes^* \widehat{Q}(\theta) s(z_i,\theta) \right) && \text{(A29)} \\
& + \frac{1}{n} \textstyle\sum_{i=1}^n \left( \widehat{Q}(\theta) s(z_i,\theta) \otimes \nabla \left( \widehat{Q}(\theta) s(z_i,\theta) \right) \right) && \text{(A30)}
\end{aligned}$$

from Remark A.4. Noting $\nabla \left( \widehat{Q}(\theta) s(z_i,\theta) \right) = \widehat{Q}(\theta) \nabla s(z_i,\theta) + vec^* \left( s(z_i,\theta)' \nabla \left( \widehat{Q}(\theta)' \right) \right)$ from Remark A.2, we rewrite (A29) as

$$\begin{aligned}
& \frac{1}{n} \textstyle\sum_{i=1}^n \left( \widehat{Q}(\theta) \nabla s(z_i,\theta) + vec^* \left( s(z_i,\theta)' \nabla \left( \widehat{Q}(\theta)' \right) \right) \otimes^* \widehat{Q}(\theta) s(z_i,\theta) \right) && \text{(A31)} \\
=\; & \frac{1}{n} \textstyle\sum_{i=1}^n \left( \widehat{Q}(\theta) \nabla s(z_i,\theta) \otimes^* \widehat{Q}(\theta) s(z_i,\theta) \right) && \text{(A32)} \\
& + \frac{1}{n} \textstyle\sum_{i=1}^n \left( vec^* \left( s(z_i,\theta)' \nabla \left( \widehat{Q}(\theta)' \right) \right) \otimes^* \widehat{Q}(\theta) s(z_i,\theta) \right).
\end{aligned}$$

Now note that $\left\| A \otimes^* B \right\| = \|A \otimes B\| = \|A\| \|B\|$ for matrices $A$ and $B$ including column or row vectors. This implies for (A32)

$$\begin{aligned}
& \left\| \tfrac{1}{n} \textstyle\sum_{i=1}^n \left( \widehat{Q}(\theta) \nabla s(z_i,\theta) \otimes^* \widehat{Q}(\theta) s(z_i,\theta) \right) \right\| \le \tfrac{1}{n} \textstyle\sum_{i=1}^n \left\| \left( \widehat{Q}(\theta) s(z_i,\theta) \otimes^* \widehat{Q}(\theta) s(z_i,\theta) \right) \right\| \\
&= \tfrac{1}{n} \textstyle\sum_{i=1}^n \left\| \widehat{Q}(\theta) \nabla s(z_i,\theta) \right\| \left\| \widehat{Q}(\theta) s(z_i,\theta) \right\| \le \tfrac{1}{n} \textstyle\sum_{i=1}^n \|\nabla s(z_i,\theta)\| \|s(z_i,\theta)\| \left\| \widehat{Q}(\theta) \right\|^2 = O_p(1)
\end{aligned} \qquad \text{(A33)}$$

uniformly over $\theta \in \Theta_0$ by Lemma A.5 for $(v_1, v_2) = (1, 0)$ under $E\left[ \sup_{\theta \in \Theta_0} \|\nabla^v s(z_i,\theta)\|^2 \right] < \infty, v = 1, 0$ and by (A9). This gives

$$\begin{aligned}
& \left\| \tfrac{1}{n} \textstyle\sum_{i=1}^n \left( vec^* \left( s(z_i,\theta)' \nabla \left( \widehat{Q}(\theta)' \right) \right) \otimes^* \widehat{Q}(\theta) s(z_i,\theta) \right) \right\| \\
\le\; & \tfrac{1}{n} \textstyle\sum_{i=1}^n \left\| vec^* \left( s(z_i,\theta)' \nabla \left( \widehat{Q}(\theta)' \right) \right) \right\| \left\| \widehat{Q}(\theta) s(z_i,\theta) \right\| \\
\le\; & \tfrac{1}{n} \textstyle\sum_{i=1}^n \left\| s(z_i,\theta)' \nabla \left( \widehat{Q}(\theta)' \right) \right\| \left\| \widehat{Q}(\theta) s(z_i,\theta) \right\| \\
\le\; & \tfrac{1}{n} \textstyle\sum_{i=1}^n \|s(z_i,\theta)\|^2 \left\| \widehat{Q}(\theta) \right\| \left\| \nabla \left( \widehat{Q}(\theta)' \right) \right\| = O_p(1)
\end{aligned} \qquad \text{(A34)}$$

by the uniform convergence of $\frac{1}{n} \sum_{i=1}^n \|s(z_i,\theta)\|^2$ to $E\left[ \|s(z,\theta)\|^2 \right] < \infty$, (A9), and Lemma A.8 noting $\|vec^*(\cdot)\| = \|\cdot\|$ and hence we show that (A29) is $O_p(1)$ uniformly over $\theta \in \Theta_0$ from (A33) and (A34). Similarly we can show that (A30) is $O_p(1)$ around the neighborhood of $\theta_0$ and hence we have

$$\nabla \left( \frac{1}{n} \textstyle\sum_{i=1}^n \left[ \widehat{Q}(\theta) s(z_i,\theta) \otimes \widehat{Q}(\theta) s(z_i,\theta) \right] \right) = O_p(1). \qquad \text{(A35)}$$

Together with (A11), this shows the second RHS term of the last equality in (A25) is $O_p(1)$. Now consider for the third RHS term of the last equality in (A25),

$$\begin{aligned}
&\nabla\left(\tfrac{1}{n}\sum_{i=1}^n\left[\nabla s(z_i,\theta)\widehat{Q}(\theta)s(z_i,\theta)\right]\right)\\
&=\tfrac{1}{n}\sum_{i=1}^n\nabla s(z_i,\theta)\nabla\left(\widehat{Q}(\theta)s(z_i,\theta)\right)+\tfrac{1}{n}\sum_{i=1}^n vec^*\left(\left(\widehat{Q}(\theta)s(z_i,\theta)\right)'\nabla\left((\nabla s(z_i,\theta))'\right)\right)\\
&=\tfrac{1}{n}\sum_{i=1}^n\nabla s(z_i,\theta)\left(\widehat{Q}(\theta)\nabla s(z_i,\theta)+vec^*\left(s(z_i,\theta)'\nabla\left(\widehat{Q}(\theta)'\right)\right)\right)\\
&\quad+\tfrac{1}{n}\sum_{i=1}^n vec^*\left(s(z_i,\theta)'\widehat{Q}(\theta)'\nabla\left((\nabla s(z_i,\theta))'\right)\right).
\end{aligned} \tag{A36}$$

This implies

$$\begin{aligned}
&\left\|\nabla\left(\tfrac{1}{n}\sum_{i=1}^n\left[\nabla s(z_i,\theta)\widehat{Q}(\theta)s(z_i,\theta)\right]\right)\right\|\leq\tfrac{1}{n}\sum_{i=1}^n\|\nabla s(z_i,\theta)\|^2\left\|\widehat{Q}(\theta)\right\|\\
&\quad+\tfrac{1}{n}\sum_{i=1}^n\|\nabla s(z_i,\theta)\|\left\|vec^*\left(s(z_i,\theta)'\nabla\left(\widehat{Q}(\theta)'\right)\right)\right\|\\
&+\tfrac{1}{n}\sum_{i=1}^n\left\|vec^*\left(s(z_i,\theta)'\widehat{Q}(\theta)'\nabla\left((\nabla s(z_i,\theta))'\right)\right)\right\|\\
&=\tfrac{1}{n}\sum_{i=1}^n\|\nabla s(z_i,\theta)\|^2\left\|\widehat{Q}(\theta)\right\|+\tfrac{1}{n}\sum_{i=1}^n\|\nabla s(z_i,\theta)\|\left\|s(z_i,\theta)'\nabla\left(\widehat{Q}(\theta)'\right)\right\|\\
&\quad+\tfrac{1}{n}\sum_{i=1}^n\left\|s(z_i,\theta)'\widehat{Q}(\theta)'\nabla\left((\nabla s(z_i,\theta))'\right)\right\|\\
&\leq\tfrac{1}{n}\sum_{i=1}^n\|\nabla s(z_i,\theta)\|^2\left\|\widehat{Q}(\theta)\right\|+\tfrac{1}{n}\sum_{i=1}^n\|\nabla s(z_i,\theta)\|\,\|s(z_i,\theta)\|\left\|\nabla\left(\widehat{Q}(\theta)'\right)\right\|\\
&\quad+\tfrac{1}{n}\sum_{i=1}^n\|s(z_i,\theta)\|\left\|\nabla^2 s(z_i,\theta)\right\|\left\|\widehat{Q}(\theta)\right\|.
\end{aligned} \tag{A37}$$

We have the first RHS term in the last inequality of (A37) equals to $O_p(1)$ uniformly over $\theta\in\Theta_0$ by (A9) and the uniform convergence of $\frac{1}{n}\sum_{i=1}^n\|\nabla s(z_i,\theta)\|^2$ to $E\left[\|\nabla s(z,\theta)\|^2\right]<\infty$ by applying Lemma A.1 under $\sup_{\theta\in\Theta_0}E\left[\|\nabla s(z,\theta)\|^2\right]<\infty$ or by applying Lemma A.2 (Lipschitz condition holds similarly with (A13) under $E\left[\sup_{\theta\in\Theta}\|\nabla s(z,\theta)\|B_1(z)\right]<\infty$) with $m(z,\theta)=\|\nabla s(z,\theta)\|^2$. Clearly the second RHS term of the last inequality is $O_p(1)$ uniformly over $\theta\in\Theta_0$ from Lemma A.5 and Lemma A.8. Finally, we obtain the last RHS term of the last inequality in (A37) equals to $O_p(1)$ uniformly over $\theta\in\Theta_0$ from (A9) and Lemma A.5 with $(v_1,v_2)=(0,2)$ and thus we bound the third RHS term of the last equality in (A25) to be $O_p(1)$ uniformly over $\theta\in\Theta_0$. This completes the proof. □

For later uses, here we summarize the differentiation results of $\nabla\hat{c}(\theta)$ and $\nabla c(\theta)$, respectively, as

$$\nabla\hat{c}(\theta)=\left\{\begin{array}{l}
\frac{1}{2}vec^*\left(\left(\frac{1}{n}\sum_{i=1}^n\left[\widehat{Q}(\theta)s(z_i,\theta)\otimes\widehat{Q}(\theta)s(z_i,\theta)\right]\right)'\left(\nabla\left(\widehat{H}_2(\theta)'\right)\right)\right)\\
+\frac{1}{2}\widehat{H}_2(\theta)\left(\begin{array}{c}\frac{1}{n}\sum_{i=1}^n\left(\widehat{Q}(\theta)\nabla s(z_i,\theta)\otimes^*\widehat{Q}(\theta)s(z_i,\theta)\right)\\+\frac{1}{n}\sum_{i=1}^n\left(vec^*\left(s(z_i,\theta)'\nabla\left(\widehat{Q}(\theta)'\right)\right)\otimes^*\widehat{Q}(\theta)s(z_i,\theta)\right)\end{array}\right)\\
+\frac{1}{2}\widehat{H}_2(\theta)\left(\begin{array}{c}\frac{1}{n}\sum_{i=1}^n\left(\widehat{Q}(\theta)s(z_i,\theta)\otimes\widehat{Q}(\theta)\nabla s(z_i,\theta)\right)\\+\frac{1}{n}\sum_{i=1}^n\left(\widehat{Q}(\theta)s(z_i,\theta)\right)\otimes vec^*\left(s(z_i,\theta)'\nabla\left(\widehat{Q}(\theta)'\right)\right)\end{array}\right)\\
+\frac{1}{n}\sum_{i=1}^n\nabla s(z_i,\theta)\left(\widehat{Q}(\theta)\nabla s(z_i,\theta)+vec^*\left(s(z_i,\theta)'\nabla\left(\widehat{Q}(\theta)'\right)\right)\right)\\
+\frac{1}{n}\sum_{i=1}^n vec^*\left(s(z_i,\theta)'\widehat{Q}(\theta)'\nabla\left((\nabla s(z_i,\theta))'\right)\right)
\end{array}\right\} \tag{A38}$$

$$\nabla c(\theta) = \tag{A39}$$
$$\begin{aligned}
&\tfrac{1}{2}vec^{*}\left(\left(E\left[Q(\theta)s(z_i,\theta)\otimes Q(\theta)s(z_i,\theta)\right]\right)'\left(\nabla\left(H_2(\theta)'\right)\right)\right)\\
&+\tfrac{1}{2}H_2(\theta)\left(E\left[Q(\theta)\nabla s(z_i,\theta)\otimes^{*}Q(\theta)s(z_i,\theta)\right]+E\left[vec^{*}\left(s(z_i,\theta)'\nabla\left(Q(\theta)'\right)\right)\otimes^{*}Q(\theta)s(z_i,\theta)\right]\right)\\
&+\tfrac{1}{2}H_2(\theta)\left(E\left[Q(\theta)s(z_i,\theta)\otimes Q(\theta)\nabla s(z_i,\theta)\right]+E\left[\left(Q(\theta)s(z_i,\theta)\right)\otimes vec^{*}\left(s(z_i,\theta)'\nabla\left(Q(\theta)'\right)\right)\right]\right)\\
&+E\left[\nabla s\left(z_i,\theta\right)\left(Q(\theta)\nabla s(z_i,\theta)+vec^{*}\left(s(z_i,\theta)'\nabla\left(Q(\theta)'\right)\right)\right)\right]\\
&+E\left[vec^{*}\left(s(z_i,\theta)'Q(\theta)'\nabla\left(\left(\nabla s\left(z_i,\theta\right)\right)'\right)\right)\right].
\end{aligned}$$

**Proof. Condition 3**

In what follows, we will apply Triangle Inequality and Cauchy-Schwarz Inequality whenever they are necessary without noting them. Again we bound each term uniformly over $\theta\in\Theta_0$ and suppress the sup-norm otherwise it is noted. From (A25), consider

$$\nabla^2\widehat{c}(\theta)=\left\{\begin{array}{l}
\nabla\left(\frac{1}{2}vec^{*}\left(\left(\frac{1}{n}\sum_{i=1}^{n}\left[\widehat{Q}(\theta)s(z_i,\theta)\otimes\widehat{Q}(\theta)s(z_i,\theta)\right]\right)'\left(\nabla\left(\widehat{H}_2(\theta)'\right)\right)\right)\right)\\
+\nabla\left(\frac{1}{2}\widehat{H}_2(\theta)\nabla\left(\frac{1}{n}\sum_{i=1}^{n}\left[\widehat{Q}(\theta)s(z_i,\theta)\otimes\widehat{Q}(\theta)s(z_i,\theta)\right]\right)\right)\\
+\nabla^2\left(\frac{1}{n}\sum_{i=1}^{n}\left[\nabla s(z_i,\theta)\widehat{Q}(\theta)s(z_i,\theta)\right]\right).
\end{array}\right\}\tag{A40}$$

Considering $||\nabla(\widehat{H}_2(\theta)')||=||\nabla\widehat{H}_2(\theta)||$ and $||\nabla^2(\widehat{H}_2(\theta)')||=||\nabla^2\widehat{H}_2(\theta)||$, for the first RHS term of (A40), we have

$$\begin{aligned}
&\left\|\nabla\left(\tfrac{1}{2}vec^{*}\left(\left(\tfrac{1}{n}\textstyle\sum_{i=1}^{n}\left[\widehat{Q}(\theta)s(z_i,\theta)\otimes\widehat{Q}(\theta)s(z_i,\theta)\right]\right)'\left(\nabla\left(\widehat{H}_2(\theta)'\right)\right)\right)\right)\right\|\\
&\leq\tfrac{1}{2}\left\|\nabla^2\widehat{H}_2(\theta)\right\|\left\|\tfrac{1}{n}\textstyle\sum_{i=1}^{n}\left[\widehat{Q}(\theta)s(z_i,\theta)\otimes\widehat{Q}(\theta)s(z_i,\theta)\right]\right\|\\
&+\tfrac{1}{2}\left\|\nabla\widehat{H}_2(\theta)\right\|\left\|\nabla\left(\tfrac{1}{n}\textstyle\sum_{i=1}^{n}\left[\widehat{Q}(\theta)s(z_i,\theta)\otimes\widehat{Q}(\theta)s(z_i,\theta)\right]\right)\right\|\\
&=O_p(1)O_p(1)+O_p(1)O_p(1)=O_p(1)
\end{aligned}$$

uniformly over $\theta\in\Theta_0$ from (A22), (A27), (A35), and since $\nabla^2\widehat{H}_2(\theta)=\frac{1}{n}\sum_{i=1}^{n}\nabla^4 s(z_i,\theta)=E\left[\nabla^4 s(z_i,\theta)\right]+o_p(1)=O_p(1)$ by the Uniform Convergence theorem of Lemma A.1 or Lemma A.2 with $m(z,\theta)=\nabla^4 s(z,\theta)$. Now we bound the second RHS term as

$$\begin{aligned}
&\left\|\nabla\left(\tfrac{1}{2}\widehat{H}_2(\theta)\nabla\left(\tfrac{1}{n}\sum_{i=1}^{n}\left[\widehat{Q}(\theta)s(z_i,\theta)\otimes\widehat{Q}(\theta)s(z_i,\theta)\right]\right)\right)\right\|\\
&\leq\tfrac{1}{2}\left\|\nabla\widehat{H}_2(\theta)\right\|\left\|\nabla\left(\tfrac{1}{n}\sum_{i=1}^{n}\left[\widehat{Q}(\theta)s(z_i,\theta)\otimes\widehat{Q}(\theta)s(z_i,\theta)\right]\right)\right\|\\
&+\tfrac{1}{2}\left\|\widehat{H}_2(\theta)\right\|\left\|\nabla^2\left(\tfrac{1}{n}\sum_{i=1}^{n}\left[\widehat{Q}(\theta)s(z_i,\theta)\otimes\widehat{Q}(\theta)s(z_i,\theta)\right]\right)\right\|
\end{aligned}\tag{A41}$$

Note, for the first RHS term in (A41)

$$\begin{aligned}
&\left\|\tfrac{1}{2}\nabla\widehat{H}_2(\theta)\nabla\left(\tfrac{1}{n}\textstyle\sum_{i=1}^{n}\left[\widehat{Q}(\theta)s(z_i,\theta)\otimes\widehat{Q}(\theta)s(z_i,\theta)\right]\right)\right\|\\
&\leq C\left\|\nabla\widehat{H}_2(\theta)\right\|\left\|\nabla\left(\tfrac{1}{n}\textstyle\sum_{i=1}^{n}\left[\widehat{Q}(\theta)s(z_i,\theta)\otimes\widehat{Q}(\theta)s(z_i,\theta)\right]\right)\right\|=O_p(1)
\end{aligned}$$

by (A22) and (A35). From (A28) and (A31), for the second RHS term in (A41), we have

$$\begin{aligned}
&\left\|\nabla^2\left(\tfrac{1}{n}\textstyle\sum_{i=1}^{n}\left[\widehat{Q}(\theta)s(z_i,\theta)\otimes\widehat{Q}(\theta)s(z_i,\theta)\right]\right)\right\|\\
&\leq\left\|\nabla\left(\tfrac{1}{n}\textstyle\sum_{i=1}^{n}\left(\widehat{Q}(\theta)\nabla s(z_i,\theta)\otimes^{*}\widehat{Q}(\theta)s(z_i,\theta)\right)\right)\right\|\\
&\quad+\left\|\nabla\left(\tfrac{1}{n}\textstyle\sum_{i=1}^{n}\left(vec^{*}\left(s(z_i,\theta)'\nabla\left(\widehat{Q}(\theta)'\right)\right)\otimes^{*}\widehat{Q}(\theta)s(z_i,\theta)\right)\right)\right\|\\
&\quad+\left\|\nabla\left(\tfrac{1}{n}\textstyle\sum_{i=1}^{n}\left(\widehat{Q}(\theta)s(z_i,\theta)\otimes\nabla\left(\widehat{Q}(\theta)s(z_i,\theta)\right)\right)\right)\right\|.
\end{aligned}\tag{A42}$$

First we bound the first RHS term of (A42) uniformly over $\theta \in \Theta_0$ as

$$
\begin{aligned}
&\left\|\nabla\left(\tfrac{1}{n}\textstyle\sum_{i=1}^{n}\left(\widehat{Q}(\theta)\nabla s(z_i,\theta)\otimes^{*}\widehat{Q}(\theta)s(z_i,\theta)\right)\right)\right\| \\
&\leq \tfrac{1}{n}\textstyle\sum_{i=1}^{n}\left\|\nabla\left(\widehat{Q}(\theta)\nabla s(z_i,\theta)\otimes^{*}\widehat{Q}(\theta)s(z_i,\theta)\right)\right\| \\
&\leq \tfrac{1}{n}\textstyle\sum_{i=1}^{n}\left\|\nabla\left(\widehat{Q}(\theta)\nabla s(z_i,\theta)\right)\right\|\left\|\widehat{Q}(\theta)s(z_i,\theta)\right\|+\tfrac{1}{n}\textstyle\sum_{i=1}^{n}\left\|\widehat{Q}(\theta)\nabla s(z_i,\theta)\right\|\left\|\nabla\left(\widehat{Q}(\theta)s(z_i,\theta)\right)\right\| \\
&= \tfrac{1}{n}\textstyle\sum_{i=1}^{n}\left\|\widehat{Q}(\theta)\nabla^2 s(z_i,\theta)+(\nabla s(z_i,\theta))'\,\nabla\left(\widehat{Q}(\theta)'\right)\right\|\left\|\widehat{Q}(\theta)s(z_i,\theta)\right\| \\
&\quad+\tfrac{1}{n}\textstyle\sum_{i=1}^{n}\left\|\widehat{Q}(\theta)\nabla s(z_i,\theta)\right\|\left\|\widehat{Q}(\theta)\nabla s(z_i,\theta)+vec^{*}\left(s(z_i,\theta)'\nabla\left(\widehat{Q}(\theta)'\right)\right)\right\| \\
&\leq \tfrac{1}{n}\textstyle\sum_{i=1}^{n}\|\nabla s(z_i,\theta)\|\,\|s(z_i,\theta)\|\left\|\widehat{Q}(\theta)\right\|\left\|\nabla\widehat{Q}(\theta)\right\|+\tfrac{1}{n}\textstyle\sum_{i=1}^{n}\|\nabla^2 s(z_i,\theta)\|\,\|s(z_i,\theta)\|\left\|\widehat{Q}(\theta)\right\|^2 \\
&\quad+\tfrac{1}{n}\textstyle\sum_{i=1}^{n}\|\nabla s(z_i,\theta)\|^2\left\|\widehat{Q}(\theta)\right\|^2+\tfrac{1}{n}\textstyle\sum_{i=1}^{n}\|\nabla s(z_i,\theta)\|\,\|s(z_i,\theta)\|\left\|\widehat{Q}(\theta)\right\|\left\|\nabla\widehat{Q}(\theta)\right\| \\
&= O_p(1)+O_p(1)+O_p(1)+O_p(1)=O_p(1),
\end{aligned}
$$

where the second inequality is from Remark A.4, the first equality is from Remark A.1 and Remark A.2, the third inequality is from Remark A.3. The second last equality comes from Lemma A.5, Lemma A.8 and (A9). For the second RHS term of (A42), from Remark A.2-A.4, it follows that

$$
\begin{aligned}
&\left\|\nabla\left(\tfrac{1}{n}\textstyle\sum_{i=1}^{n}\left(vec^{*}\left(s(z_i,\theta)'\nabla\left(\widehat{Q}(\theta)'\right)\right)\otimes^{*}\widehat{Q}(\theta)s(z_i,\theta)\right)\right)\right\| \\
&\leq \tfrac{1}{n}\textstyle\sum_{i=1}^{n}\left\|\nabla\left(vec^{*}\left(s(z_i,\theta)'\nabla\left(\widehat{Q}(\theta)'\right)\right)\right)\right\|\left\|\widehat{Q}(\theta)s(z_i,\theta)\right\| \\
&\quad+\tfrac{1}{n}\textstyle\sum_{i=1}^{n}\left\|vec^{*}\left(s(z_i,\theta)'\nabla\left(\widehat{Q}(\theta)'\right)\right)\right\|\left\|\nabla\left(\widehat{Q}(\theta)s(z_i,\theta)\right)\right\| \\
&\leq \tfrac{1}{n}\textstyle\sum_{i=1}^{n}\|\nabla s(z_i,\theta)\|\,\|s(z_i,\theta)\|\left\|\nabla\left(\widehat{Q}(\theta)'\right)\right\|\left\|\widehat{Q}(\theta)\right\|+\tfrac{1}{n}\textstyle\sum_{i=1}^{n}\|s(z_i,\theta)\|^2\left\|\nabla^2\left(\widehat{Q}(\theta)'\right)\right\|\left\|\widehat{Q}(\theta)\right\| \\
&\quad+\tfrac{1}{n}\textstyle\sum_{i=1}^{n}\|s(z_i,\theta)\|\,\|\nabla s(z_i,\theta)\|\left\|\nabla\left(\widehat{Q}(\theta)'\right)\right\|\left\|\widehat{Q}(\theta)\right\|+\tfrac{1}{n}\textstyle\sum_{i=1}^{n}\|s(z_i,\theta)\|^2\left\|\nabla\left(\widehat{Q}(\theta)'\right)\right\|^2 \\
&= O_p(1)+O_p(1)+O_p(1)+O_p(1)=O_p(1)
\end{aligned}
$$

uniformly over $\theta \in \Theta_0$. The last equality comes from Lemma A.5, Lemma A.8, and (A9). Similarly we can also bound the last RHS term of (A42) uniformly over $\theta \in \Theta_0$ as

$$
\begin{aligned}
&\left\|\nabla\left(\tfrac{1}{n}\textstyle\sum_{i=1}^{n}\left(\widehat{Q}(\theta)s(z_i,\theta)\otimes\nabla\left(\widehat{Q}(\theta)s(z_i,\theta)\right)\right)\right)\right\| \\
&\leq \tfrac{1}{n}\textstyle\sum_{i=1}^{n}\left\|\nabla\left(\widehat{Q}(\theta)s(z_i,\theta)\right)\otimes^{*}\nabla\left(\widehat{Q}(\theta)s(z_i,\theta)\right)\right\|+\tfrac{1}{n}\textstyle\sum_{i=1}^{n}\left\|\widehat{Q}(\theta)s(z_i,\theta)\otimes\nabla^2\left(\widehat{Q}(\theta)s(z_i,\theta)\right)\right\| \\
&\leq \tfrac{1}{n}\textstyle\sum_{i=1}^{n}\|s(z_i,\theta)\|^2\left\|\nabla\left(\widehat{Q}(\theta)'\right)\right\|^2+\tfrac{1}{n}\textstyle\sum_{i=1}^{n}\|\nabla s(z_i,\theta)\|^2\left\|\widehat{Q}(\theta)\right\|^2 \\
&\quad+2\tfrac{1}{n}\textstyle\sum_{i=1}^{n}\|\nabla s(z_i,\theta)\|\,\|s(z_i,\theta)\|\left\|\widehat{Q}(\theta)\right\|\left\|\nabla\left(\widehat{Q}(\theta)'\right)\right\| \\
&\quad+\tfrac{1}{n}\textstyle\sum_{i=1}^{n}\|s(z_i,\theta)\|^2\left\|\widehat{Q}(\theta)\right\|\left\|\nabla\left(\widehat{Q}(\theta)'\right)\right\|^2+\tfrac{1}{n}\textstyle\sum_{i=1}^{n}\|s(z_i,\theta)\|\,\|\nabla s(z_i,\theta)\|\left\|\widehat{Q}(\theta)\right\|\left\|\nabla\left(\widehat{Q}(\theta)'\right)\right\| \\
&\quad+\tfrac{1}{n}\textstyle\sum_{i=1}^{n}\|s(z_i,\theta)\|\,\|\nabla s(z_i,\theta)\|^2\left\|\widehat{Q}(\theta)\right\|^2 \\
&= O_p(1)+O_p(1)+O_p(1)+O_p(1)+O_p(1)+O_p(1)=O_p(1)
\end{aligned}
$$

using Remark A.2 and Remark A.4. The second last equality is obtained from Lemma A.5, Lemma A.8, (A9), and by the uniform convergence of $\frac{1}{n}\sum_{i=1}^{n}\|\nabla s(z_i,\theta)\|^2$ to $E\left[\|\nabla s(z,\theta)\|^2\right]$. The results above together bound the second RHS term of (A40) to be $O_p(1)$. Finally we rewrite the third RHS term of (A40) as

$$
\begin{aligned}
&\nabla^2\left(\frac{1}{n}\sum_{i=1}^n\left[\nabla s\left(z_i,\theta\right)\widehat{Q}(\theta)s(z_i,\theta)\right]\right)\\
=\;&\nabla\left(\begin{array}{c}\frac{1}{n}\sum_{i=1}^n\nabla s(z_i,\theta)\left(\widehat{Q}(\theta)\nabla s(z_i,\theta)+vec^*\left(s(z_i,\theta)'\nabla\left(\widehat{Q}(\theta)'\right)\right)\right)\\ +\frac{1}{n}\sum_{i=1}^n vec^*\left(s(z_i,\theta)'\widehat{Q}(\theta)'\nabla\left((\nabla s(z_i,\theta))'\right)\right)\end{array}\right)\\
=\;&\nabla\left(\frac{1}{n}\sum_{i=1}^n\nabla s(z_i,\theta)\left(\widehat{Q}(\theta)\nabla s(z_i,\theta)+vec^*\left(s(z_i,\theta)'\nabla\left(\widehat{Q}(\theta)'\right)\right)\right)\right) \qquad \text{(A43)}\\
&+\nabla\left(\frac{1}{n}\sum_{i=1}^n vec^*\left(s(z_i,\theta)'\widehat{Q}(\theta)'\nabla\left((\nabla s(z_i,\theta))'\right)\right)\right) \qquad \text{(A44)}
\end{aligned}
$$

from (A36). For (A43), note

$$
\begin{aligned}
&\nabla\left(\tfrac{1}{n}\sum_{i=1}^n\nabla s(z_i,\theta)\left(\widehat{Q}(\theta)\nabla s(z_i,\theta)+vec^*\left(s(z_i,\theta)'\nabla\left(\widehat{Q}(\theta)'\right)\right)\right)\right)\\
&=\tfrac{1}{n}\sum_{i=1}^n\left(\nabla s(z_i,\theta)\nabla\left(\widehat{Q}(\theta)\nabla s(z_i,\theta)\right)+(\nabla s(z_i,\theta))'\,\widehat{Q}(\theta)'\nabla\left((\nabla s(z_i,\theta))'\right)\right)\\
&\quad+\tfrac{1}{n}\sum_{i=1}^n\nabla s(z_i,\theta)vec^*\left(\nabla\left(s(z_i,\theta)'\nabla\left(\widehat{Q}(\theta)'\right)\right)\right)\\
&\quad+\tfrac{1}{n}\sum_{i=1}^n\left(vec^*\left(s(z_i,\theta)'\nabla\left(\widehat{Q}(\theta)'\right)\right)\right)'\nabla\left((\nabla s(z_i,\theta))'\right)\\
&=\tfrac{1}{n}\sum_{i=1}^n\left(\nabla s(z_i,\theta)\left(\widehat{Q}(\theta)\nabla^2 s(z_i,\theta)+(\nabla s(z_i,\theta))'\,\nabla(\widehat{Q}(\theta)')\right)+(\nabla s(z_i,\theta))'\,\widehat{Q}(\theta)'\nabla\left((\nabla s(z_i,\theta))'\right)\right)\\
&\quad+\tfrac{1}{n}\sum_{i=1}^n\nabla s(z_i,\theta)vec^*\left(\nabla\left(s(z_i,\theta)'\nabla\left(\widehat{Q}(\theta)'\right)\right)\right)\\
&\quad+\tfrac{1}{n}\sum_{i=1}^n\left(vec^*\left(s(z_i,\theta)'\nabla\left(\widehat{Q}(\theta)'\right)\right)\right)'\nabla\left((\nabla s(z_i,\theta))'\right)
\end{aligned}
$$

by Remark A.1 and Remark A.3 and hence

$$
\begin{aligned}
&\left\|\nabla\left(\tfrac{1}{n}\sum_{i=1}^n\nabla s(z_i,\theta)\left(\widehat{Q}(\theta)\nabla s(z_i,\theta)+vec^*\left(s(z_i,\theta)'\nabla\left(\widehat{Q}(\theta)'\right)\right)\right)\right)\right\|\\
&\le\tfrac{1}{n}\sum_{i=1}^n\|\nabla s(z_i,\theta)\|\,\|\nabla^2 s(z_i,\theta)\|\left\|\widehat{Q}(\theta)\right\|+\tfrac{1}{n}\sum_{i=1}^n\|\nabla s(z_i,\theta)\|^2\left\|\nabla(\widehat{Q}(\theta)')\right\|\\
&\quad+\tfrac{1}{n}\sum_{i=1}^n\|\nabla s(z_i,\theta)\|^2\left\|\nabla(\widehat{Q}(\theta)')\right\|+\tfrac{1}{n}\sum_{i=1}^n\|\nabla s(z_i,\theta)\|\,\|s(z_i,\theta)\|\left\|\nabla^2(\widehat{Q}(\theta)')\right\|\\
&\quad+\tfrac{1}{n}\sum_{i=1}^n\|s(z_i,\theta)\|\,\|\nabla^2 s(z_i,\theta)\|\left\|\nabla(\widehat{Q}(\theta)')\right\|=O_p(1)+O_p(1)+O_p(1)+O_p(1)+O_p(1)=O_p(1)
\end{aligned}
$$

from (A9), (A2), and Lemma A.8 and since i) $\frac{1}{n}\sum_{i=1}^n\|\nabla^v s\left(z_i,\theta\right)\|\left\|\nabla^2 s\left(z_i,\theta\right)\right\|=O_p(1)$ for $v\in\{0,1\}$ by Lemma A.5 and since ii)$\frac{1}{n}\sum_{i=1}^n\|\nabla s\left(z_i,\theta\right)\|^2=O_p(1)$ by the uniform convergence. Now consider for (A44)

$$
\begin{aligned}
&\left\|\nabla\left(\tfrac{1}{n}\sum_{i=1}^n vec^*\left(s(z_i,\theta)'\widehat{Q}(\theta)'\nabla\left((\nabla s(z_i,\theta))'\right)\right)\right)\right\|\\
&=\left\|\tfrac{1}{n}\sum_{i=1}^n vec^*\left(\nabla\left(s(z_i,\theta)'\widehat{Q}(\theta)'\nabla\left((\nabla s(z_i,\theta))'\right)\right)\right)\right\|\\
&\le\tfrac{1}{n}\sum_{i=1}^n\left\|\nabla\left(s(z_i,\theta)'\widehat{Q}(\theta)'\nabla\left((\nabla s(z_i,\theta))'\right)\right)\right\|\\
&\le\tfrac{1}{n}\sum_{i=1}^n\|\nabla s(z_i,\theta)\|\,\|\nabla^2 s(z_i,\theta)\|\left\|\widehat{Q}(\theta)\right\|+\tfrac{1}{n}\sum_{i=1}^n\|s(z_i,\theta)\|\,\|\nabla^3 s(z_i,\theta)\|\left\|\widehat{Q}(\theta)\right\|\\
&\quad+\tfrac{1}{n}\sum_{i=1}^n\|s(z_i,\theta)\|\,\|\nabla^2 s(z_i,\theta)\|\left\|\nabla\widehat{Q}(\theta)\right\|=O_p(1)+O_p(1)+O_p(1)
\end{aligned}
$$

by (A9), Lemma A.8, and Lemma A.5 and hence we have the third RHS term of (A40) equals to $O_p(1)$ uniformly over $\theta\in\Theta_0$. This completes the proof. □

*Appendix A.2. Additional Preliminary Lemmas for the Third Order Expansion*

First, note that Lemma A.3 and Lemma A.4 trivially hold under Assumption 1 and Assumption 3, respectively considering that Assumption 1 and Assumption 3 are stronger than Assumption A.1

and Assumption A.4, respectively. We establish conditions (i)-(ix) in Lemma 3 are satisfied under Assumption 1-2 or Assumption 3-2. Again Condition (i) and (iia) are directly assumed. Condition (iib) is by the dominated convergence theorem with the dominating function given by $\sup_{\theta\in\Theta_0}\|\nabla^4 s(z,\theta)\|$ under Condition (i), (iia), and $E[\sup_{\theta\in\Theta_0}\|\nabla^4 s(z,\theta)\|] < \infty$. Condition (iii) holds by the stochastic equicontinuity of $\frac{1}{\sqrt{n}}\sum_{i=1}^n\left(\nabla^3 s(z_i,\theta) - E[\nabla^3 s(z_i,\theta)]\right)$ for $\theta \in \Theta_0$ as discussed in Lemma A.2 with $m(z,\theta) = \nabla^3 s(z,\theta)$ under Assumption 3. Instead, under Assumption 1, Condition (iii) is replaced by another local uniform convergence condition as $\left\|\frac{1}{n}\sum_{i=1}^n \nabla^4 s\left(z_i,\overline{\theta}\right) - E\left[\nabla^4 s\left(z_i,\theta_0\right)\right]\right\| = o_p(1)$ for $\overline{\theta} = \theta_0 + o_p(1)$ similarly with our replacing Condition (iv) of Lemma 1 with $\left\|\frac{1}{n}\sum_{i=1}^n \nabla^3 s\left(z_i,\overline{\theta}\right) - E\left[\nabla^3 s\left(z_i,\theta_0\right)\right]\right\| = o_p(1)$ for $\overline{\theta} = \theta_0 + o_p(1)$. Condition (iv) is implied by Assumption 2. Condition (v) through (viii) holds by CLT provided that $E\left[\|\nabla^v s(z,\theta_0)\|^2\right] < \infty$, $v = \{0,1,2,3\}$ respectively, which are satisfied under Assumption 1 (iii) or 3 (iii). Condition (ix) is the result of Lemma 1. We also need to verify following lemmas.

**Lemma A.10.** *Under Assumption 1-2 or 3-2, Condition 7 (i): $\nabla\widehat{c}(\theta_0) = \nabla c(\theta_0) + O_p\left(1/\sqrt{n}\right)$ is satisfied.*

**Proof.** This can be proved similarly with Lemma A.7 (b). From (A38) and (A39), it follows that

$$
\begin{aligned}
&\|\nabla\widehat{c}(\theta_0) - \nabla c(\theta_0)\| \\
\leq &\left\| \begin{array}{c} \frac{1}{2}vec^*\left(\left(\frac{1}{n}\sum_{i=1}^n\left[\widehat{Q}()s(z_i,\theta_0)\otimes\widehat{Q}(\theta_0)s(z_i,\theta_0)\right]\right)'\left(\nabla\left(\widehat{H}_2(\theta_0)'\right)\right)\right) \\ -\frac{1}{2}vec^*\left(\left(E\left[Q(\theta_0)s(z_i,\theta_0)\otimes Q(\theta_0)s(z_i,\theta_0)\right]\right)'\left(\nabla\left(H_2(\theta_0)'\right)\right)\right) \end{array} \right\| &\text{(A45)}\\
+ &\begin{array}{c} \frac{1}{2}\widehat{H}_2(\theta_0)\left(\begin{array}{c} \frac{1}{n}\sum_{i=1}^n\left(\widehat{Q}(\theta_0)\nabla s(z_i,\theta_0)\otimes^*\widehat{Q}(\theta_0)s(z_i,\theta_0)\right) \\ +\frac{1}{n}\sum_{i=1}^n\left(vec^*\left(s(z_i,\theta_0)'\nabla\left(\widehat{Q}(\theta_0)'\right)\right)\otimes^*\widehat{Q}(\theta_0)s(z_i,\theta_0)\right)\end{array}\right) \\ -\frac{1}{2}H_2(\theta_0)\left(\begin{array}{c} E\left[Q(\theta_0)\nabla s(z_i,\theta_0)\otimes^* Q(\theta_0)s(z_i,\theta_0)\right] \\ +E\left[vec^*\left(s(z_i,\theta_0)'\nabla\left(Q(\theta_0)'\right)\right)\otimes^* Q(\theta_0)s(z_i,\theta_0)\right]\end{array}\right)\end{array} &\text{(A46)}\\
+ &\left\|\begin{array}{c} \frac{1}{2}\widehat{H}_2(\theta_0)\left(\begin{array}{c} \frac{1}{n}\sum_{i=1}^n\left(\widehat{Q}(\theta_0)s(z_i,\theta_0)\otimes\widehat{Q}(\theta_0)\nabla s(z_i,\theta_0)\right) \\ +\frac{1}{n}\sum_{i=1}^n\left(\widehat{Q}(\theta_0)s(z_i,\theta_0)\right)\otimes vec^*\left(s(z_i,\theta_0)'\nabla\left(\widehat{Q}(\theta_0)'\right)\right)\end{array}\right) \\ -\frac{1}{2}H_2(\theta_0)\left(\begin{array}{c} E\left[Q(\theta_0)s(z_i,\theta_0)\otimes Q(\theta_0)\nabla s(z_i,\theta_0)\right] \\ +E\left[\left(Q(\theta_0)s(z_i,\theta_0)\right)\otimes vec^*\left(s(z_i,\theta_0)'\nabla\left(Q(\theta_0)'\right)\right)\right]\end{array}\right)\end{array}\right\| &\text{(A47)}\\
+ &\left\|\begin{array}{c} \frac{1}{n}\sum_{i=1}^n\nabla s(z_i,\theta_0)\left(\widehat{Q}(\theta_0)\nabla s(z_i,\theta_0) + vec^*\left(s(z_i,\theta_0)'\nabla\left(\widehat{Q}(\theta_0)'\right)\right)\right) \\ -E\left[\nabla s(z_i,\theta_0)\left(Q(\theta_0)\nabla s(z_i,\theta_0) + vec^*\left(s(z_i,\theta_0)'\nabla\left(Q(\theta_0)'\right)\right)\right)\right]\end{array}\right\| &\text{(A48)}\\
+ &\left\|\begin{array}{c} \frac{1}{n}\sum_{i=1}^n vec^*\left(s(z_i,\theta_0)'\widehat{Q}(\theta_0)'\nabla\left(\left(\nabla s(z_i,\theta_0)\right)'\right)\right) \\ -E\left[vec^*\left(s(z_i,\theta_0)'Q(\theta_0)'\nabla\left(\left(\nabla s(z_i,\theta_0)\right)'\right)\right)\right]\end{array}\right\|. &\text{(A49)}
\end{aligned}
$$

We show (A45), (A46), (A47), (A48), and (A49) are $O_p\left(1/\sqrt{n}\right)$, respectively. First, observe that applying the CLT, we have $\frac{1}{n}\sum_{i=1}^n\left[s(z_i,\theta_0)s(z_i,\theta_0)'\right] = E\left[s(z_i,\theta_0)s(z_i,\theta_0)'\right] + O_p\left(1/\sqrt{n}\right)$ under $E\left[\|s(z_i,\theta_0)\|^4\right] < \infty$ and have $\nabla\left(\widehat{H}_2(\theta_0)'\right) = \nabla\left(H_2(\theta_0)'\right) + O_p\left(1/\sqrt{n}\right)$ under $E\left[\|\nabla^3 s(z_i,\theta_0)\|^2\right] < \infty$. Recalling (A23), this implies

$$
\begin{aligned}
&\tfrac{1}{2}vec^*\left(\left(\tfrac{1}{n}\sum_{i=1}^n\left[\widehat{Q}(\theta_0)s(z_i,\theta_0)\otimes\widehat{Q}(\theta_0)s(z_i,\theta_0)\right]\right)'\left(\nabla\left(\widehat{H}_2(\theta_0)'\right)\right)\right)\\
&= \tfrac{1}{2}vec^*\left(vec\left(\widehat{Q}(\theta_0)\tfrac{1}{n}\sum_{i=1}^n\left[s(z_i,\theta_0)s(z_i,\theta_0)'\right]\widehat{Q}(\theta_0)'\right)'\left(\nabla\left(\widehat{H}_2(\theta_0)'\right)\right)\right)\\
&= \tfrac{1}{2}vec^*\left(vec\left(Q(\theta_0)E\left[s(z_i,\theta_0)s(z_i,\theta_0)'\right]Q(\theta_0)'\right)'\left(\nabla\left(H_2(\theta_0)'\right)\right)\right) + O_p\left(1/\sqrt{n}\right)\\
&= \tfrac{1}{2}vec^*\left(\left(E\left[Q(\theta_0)s(z_i,\theta_0)\otimes Q(\theta_0)s(z_i,\theta_0)\right]\right)'\left(\nabla\left(H_2(\theta_0)'\right)\right)\right) + O_p\left(1/\sqrt{n}\right)
\end{aligned}
$$

and hence (A45) is $O_p\left(1/\sqrt{n}\right)$. Now for notational simplicity, define $\|A\|_1 = \|A\|$ and $\|A\|_0 = A$. Then, for $d_1, d_2, d_3 \in \{0,1\}$, we have

$$\begin{aligned} &\tfrac{1}{n}\textstyle\sum_{i=1}^{n}\left\|\left(Q(\theta_0)^{d_1}\nabla s(z_i,\theta_0)\otimes^{*} Q(\theta_0)^{d_2}s(z_i,\theta_0)\right)\right\|_{d_3} \\ &= E\left[\left\|Q(\theta_0)^{d_1}\nabla s(z_i,\theta_0)\otimes^{*} Q(\theta_0)^{d_2}s(z_i,\theta_0)\right\|_{d_3}\right] + O_p\left(1/\sqrt{n}\right)\end{aligned}$$

applying the CLT from

$$E\left[\left\|Q(\theta_0)^{d_1}\nabla s(z_i,\theta_0)\otimes^{*} Q(\theta_0)^{d_2}s(z_i,\theta_0)\right\|^2\right] \le \|Q(\theta_0)\|^{2(d_1+d_2)} E\left[\|\nabla s(z_i,\theta_0)\|^2\|\nabla s(z_i,\theta_0)\|^2\right] < \infty$$

under $\|Q(\theta_0)\| < \infty$, $E\left[\|\nabla s(z_i,\theta_0)\|^4\right] < \infty$, and $E\left[\|s(z_i,\theta_0)\|^4\right] < \infty$. Similarly, for $l_1, l_2, l_3 \in \{0,1\}$, we have

$$\begin{aligned} &\tfrac{1}{n}\textstyle\sum_{i=1}^{n}\left\|\left(vec^{*}\left(s(z_i,\theta_0)'\left(\nabla\left(Q(\theta_0)'\right)\right)^{l_1}\right)\otimes^{*} Q(\theta_0)^{l_2}s(z_i,\theta_0)\right)\right\|_{l_3} \\ &= E\left[\left\|vec^{*}\left(s(z_i,\theta_0)'\left(\nabla\left(Q(\theta_0)'\right)\right)^{l_1}\right)\otimes^{*} Q(\theta_0)^{l_2}s(z_i,\theta_0)\right\|_{l_3}\right] + O_p\left(1/\sqrt{n}\right)\end{aligned}$$

by the CLT under

$$E\left[\left\|s(z_i,\theta_0)'\left(\nabla\left(Q(\theta_0)'\right)\right)^{l_1}\otimes^{*} Q(\theta_0)^{l_2}s(z_i,\theta_0)\right\|^2\right] \le \|Q(\theta_0)\|^{2l_1}\left\|\nabla\left(Q(\theta_0)'\right)\right\|^{2l_2} E\left[\|\nabla s(z_i,\theta_0)\|^4\right] < \infty$$

recalling that $\|\nabla\left(Q(\theta_0)'\right)\| = \|Q(\theta_0)\|^2\|H_2(\theta_0)\| < \infty$. Applying these two results together with (A23), (A24), and Lemma A.8 (b), we have (A46) equals to $O_p\left(1/\sqrt{n}\right)$ by the Triangle inequality. Similarly we have (A47)=$O_p\left(1/\sqrt{n}\right)$.

For $t_1, t_2 \in \{0,1\}$, now consider we have

$$\frac{1}{n}\sum_{i=1}^{n}\left\|\nabla s(z_i,\theta_0)Q(\theta_0)^{t_1}\nabla s(z_i,\theta_0)\right\|_{t_2} = E\left[\left\|\nabla s(z_i,\theta_0)Q(\theta_0)^{t_1}\nabla s(z_i,\theta_0)\right\|_{t_2}\right] + O_p\left(1/\sqrt{n}\right)$$

by the CLT under $\|Q(\theta_0)\|^{2t_1} E\left[\|\nabla s(z_i,\theta_0)\|^2\right] < \infty$ and have

$$\begin{aligned} &\tfrac{1}{n}\textstyle\sum_{i=1}^{n}\left\|\nabla s(z_i,\theta_0)vec^{*}\left(s(z_i,\theta_0)'\left(\nabla\left(Q(\theta_0)'\right)\right)^{t_1}\right)\right\|_{t_2} \\ &= E\left[\left\|\nabla s(z_i,\theta_0)vec^{*}\left(s(z_i,\theta_0)'\left(\nabla\left(Q(\theta_0)'\right)\right)^{t_1}\right)\right\|_{t_2}\right] + O_p\left(1/\sqrt{n}\right)\end{aligned}$$

by applying the CLT provided that $\left\|\nabla\left(Q(\theta_0)'\right)\right\|^{2t_1} E\left[\|\nabla s(z_i,\theta_0)\|^2\|s(z_i,\theta_0)\|^2\right] < \infty$ that holds under $\|\nabla\left(Q(\theta_0)'\right)\| < \infty$, $E\left[\|\nabla s(z_i,\theta_0)\|^4\right] < \infty$, and $E\left[\|s(z_i,\theta_0)\|^4\right] < \infty$. Applying these two results together with (A23) and Lemma A.8 (b), we have (A48)=$O_p\left(1/\sqrt{n}\right)$. Finally, for $j_1, j_2 \in \{0,1\}$, note

$$\begin{aligned} &\tfrac{1}{n}\textstyle\sum_{i=1}^{n}\left\|vec^{*}\left(s(z_i,\theta_0)'Q(\theta_0)^{j_1\prime}\nabla\left(\left(\nabla s\left(z_i,\theta_0\right)\right)'\right)\right)\right\|_{j_2} \\ &= E\left[\left\|vec^{*}\left(s(z_i,\theta_0)'Q(\theta_0)^{j_1\prime}\nabla\left(\left(\nabla s\left(z_i,\theta_0\right)\right)'\right)\right)\right\|_{j_2}\right] + O_p\left(1/\sqrt{n}\right)\end{aligned}$$

by the CLT since $\|Q(\theta_0)\|^{2j_1} E\left[\|\nabla^2 s(z_i,\theta_0)\|^2\|s(z_i,\theta_0)\|^2\right] < \infty$ holds by $\|Q(\theta_0)\| < \infty$, $E\left[\|\nabla^2 s(z_i,\theta_0)\|^4\right] < \infty$, and $E\left[\|s(z_i,\theta_0)\|^4\right] < \infty$. It implies (A49)=$O_p\left(1/\sqrt{n}\right)$ together with (A23). This completes the proof. $\square$

**Lemma A.11.** *Under Assumption 1-2 or 3-2 with $\kappa \geq 5$, Condition 7 (ii):$\nabla^3 \widehat{c}(\theta) = O_p(1)$ in the neighborhood of $\theta_0$ is satisfied.*

**Proof.** This can be proved similarly with Lemma A.9 for Condition 3, which is straightforward but still demands many algebras. Here we provide a simple proof for Condition 1-3 and 7 when $\dim(\theta) = 1$ as an illustrational purpose. With $\dim(\theta) = 1$, we can rewrite the correction term (5) as

$$\begin{aligned} c(\theta) &= \tfrac{1}{2} H_2(\theta) Q(\theta)^2 E\left[s(z_i,\theta)^2\right] + Q(\theta) E\left[\nabla s(z_i,\theta)\, s(z_i,\theta)\right] \\ &= \frac{1}{2(E[\nabla s(z_i,\theta)])^2} E\left[\nabla^2 s(z_i,\theta)\right] E\left[s(z_i,\theta)^2\right] - \frac{1}{E[\nabla s(z_i,\theta)]} E\left[\nabla s(z_i,\theta)\, s(z_i,\theta)\right]. \end{aligned}$$

Now define $c(\theta) \equiv \tau(E[m(z_i,\theta)])$ where $\tau(t_1,t_2,t_3,t_4) \equiv \frac{1}{2t_1^2} t_2 t_3 - \frac{1}{t_1} t_4$, $t_1 = E[\nabla s(z_i,\theta)]$, $t_2 = E\left[\nabla^2 s(z_i,\theta)\right]$, $t_3 = E\left[s(z_i,\theta)^2\right]$, $t_4 = E\left[\nabla s(z_i,\theta)\, s(z_i,\theta)\right]$, and $m(z_i,\theta) \equiv (\nabla s(z_i,\theta), \nabla^2 s(z_i,\theta), s(z_i,\theta)^2, \nabla s(z_i,\theta)\, s(z_i,\theta))'$. The sample analogue of $c(\theta)$, $\widehat{c}(\theta)$ can be written as $\widehat{c}(\theta) \equiv \tau\left(\frac{1}{n}\sum_{i=1}^n m(z_i,\theta)\right)$ accordingly. Further define $\overline{m}(\theta) \equiv E[m(z_i,\theta)]$ ($\widehat{m}(\theta) \equiv \frac{1}{n}\sum_{i=1}^n m(z_i,\theta)$), $\tau_m(\theta) \equiv \frac{\partial \tau(\overline{m}(\theta))}{\partial m'}$ ($\widehat{\tau}_m(\theta) \equiv \frac{\partial \tau(\widehat{m}(\theta))}{\partial m'}$), $\tau_{mm}(\theta) \equiv \frac{\partial^2 \tau(\overline{m}(\theta))}{\partial m' \otimes \partial m'}$ ($\widehat{\tau}_{mm}(\theta) \equiv \frac{\partial^2 \tau(\widehat{m}(\theta))}{\partial m' \otimes \partial m'}$), and $\tau_{mmm}(\theta) \equiv \frac{\partial^3 \tau(\overline{m}(\theta))}{\partial m' \otimes \partial m' \otimes \partial m'}$ ($\widehat{\tau}_{mmm}(\theta) \equiv \frac{\partial^3 \tau(\widehat{m}(\theta))}{\partial m' \otimes \partial m' \otimes \partial m'}$) noting $\tau(\cdot)$ is a smooth function. Also define $\widehat{m}_\theta(\theta)$, $\widehat{m}_{\theta\theta}(\theta)$, and $\widehat{m}_{\theta\theta\theta}(\theta)$ are the first, the second, and the third derivative $\widehat{m}(\theta)$ with respect to $\theta$.

For $\theta \in \Theta_0$, now consider

$$\begin{aligned} \widehat{c}(\theta) &= \tau(\widehat{m}(\theta)),\ \nabla \widehat{c}(\theta) = \widehat{\tau}_m(\theta) \widehat{m}_\theta(\theta) \\ \nabla^2 \widehat{c}(\theta) &= \widehat{\tau}_{mm}(\theta)\left(\widehat{m}_\theta(\theta) \otimes \widehat{m}_\theta(\theta)\right) + \widehat{\tau}_m(\theta) \widehat{m}_{\theta\theta}(\theta) \\ \nabla^3 \widehat{c}(\theta) &= \widehat{\tau}_{mmm}(\theta)\left(\widehat{m}_\theta(\theta) \otimes \widehat{m}_\theta(\theta) \otimes \widehat{m}_\theta(\theta)\right) \\ &\quad + \widehat{\tau}_{mm}(\theta)\left(\widehat{m}_{\theta\theta}(\theta) \otimes \widehat{m}_\theta(\theta)\right) + \widehat{\tau}_{mm}(\theta)\left(\widehat{m}_\theta(\theta) \otimes \widehat{m}_{\theta\theta}(\theta)\right) + \widehat{\tau}_m(\theta) \widehat{m}_{\theta\theta\theta}(\theta). \end{aligned}$$

From the Slutsky theorem, it follows that

$$\widehat{\tau}_m(\theta) = \tau_m(\theta) + o_p(1), \widehat{\tau}_{mm}(\theta) = \tau_{mm}(\theta) + o_p(1), and \widehat{\tau}_{mmm}(\theta) = \tau_{mmm}(\theta) + o_p(1),$$

since $\widehat{m}(\theta) = \overline{m}(\theta) + o_p(1)$ by Lemma A.5 under $E\left[\sup_{\theta\in\Theta_0} \|\nabla s(z_i,\theta)\|^2\right] < \infty$, $E\left[\sup_{\theta\in\Theta_0} \|\nabla^2 s(z_i,\theta)\|\right] < \infty$, and $E\left[\sup_{\theta\in\Theta_0} \|s(z_i,\theta)\|^2\right] < \infty$, if we assume Assumption 1. Also it is clear that $\widehat{m}_\theta(\theta) = \overline{m}_\theta(\theta) + o_p(1)$, $\widehat{m}_{\theta\theta}(\theta) = \overline{m}_{\theta\theta}(\theta) + o_p(1)$, and $\widehat{m}_{\theta\theta\theta}(\theta) = \overline{m}_{\theta\theta\theta}(\theta) + o_p(1)$ by Lemma A.5 under $E[\sup_{\theta\in\Theta_0} \|\nabla^{\overline{v}} s(z_i,\theta)\|^2] < \infty$ for $\overline{v} = \{0,1,2,3,4\}$ and $E[\sup_{\theta\in\Theta_0} \|\nabla^5 s(z_i,\theta)\|]$ $< \infty$, if we assume Assumption 1. These imply that

$$\begin{aligned} \widehat{c}(\theta) &= \tau(\widehat{m}(\theta)) = \tau(\overline{m}(\theta)) + o_p(1) = c(\theta) + o_p(1) \\ \nabla \widehat{c}(\theta) &= \widehat{\tau}_m(\theta) \widehat{m}_\theta(\theta) = \overline{\tau}_m(\theta) \overline{m}_\theta(\theta) + o_p(1) = \nabla c(\theta) + o_p(1) \\ \nabla^2 \widehat{c}(\theta) &= \widehat{\tau}_{mm}(\theta)\left(\widehat{m}_\theta(\theta) \otimes \widehat{m}_\theta(\theta)\right) + \widehat{\tau}_m(\theta) \widehat{m}_{\theta\theta}(\theta) \\ &= \overline{\tau}_{mm}(\theta)\left(\overline{m}_\theta(\theta) \otimes \overline{m}_\theta(\theta)\right) + \overline{\tau}_m(\theta) \overline{m}_{\theta\theta}(\theta) + o_p(1) = \nabla^2 c(\theta) + o_p(1) \\ \nabla^3 \widehat{c}(\theta) &= \widehat{\tau}_{mmm}(\theta)\left(\widehat{m}_\theta(\theta) \otimes \widehat{m}_\theta(\theta) \otimes \widehat{m}_\theta(\theta)\right) \\ &\quad + \widehat{\tau}_{mm}(\theta)\left(\widehat{m}_{\theta\theta}(\theta) \otimes \widehat{m}_\theta(\theta)\right) + \widehat{\tau}_{mm}(\theta)\left(\widehat{m}_\theta(\theta) \otimes \widehat{m}_{\theta\theta}(\theta)\right) + \widehat{\tau}_m(\theta) \widehat{m}_{\theta\theta\theta}(\theta) \\ &= \overline{\tau}_{mmm}(\theta)\left(\overline{m}_\theta(\theta) \otimes \overline{m}_\theta(\theta) \otimes \overline{m}_\theta(\theta)\right) \\ &\quad + \overline{\tau}_{mm}(\theta)\left(\overline{m}_{\theta\theta}(\theta) \otimes \overline{m}_\theta(\theta)\right) + \overline{\tau}_{mm}(\theta)\left(\overline{m}_\theta(\theta) \otimes \overline{m}_{\theta\theta}(\theta)\right) + \overline{\tau}_m(\theta) \overline{m}_{\theta\theta\theta}(\theta) \\ &= \nabla^3 c(\theta) + o_p(1) \end{aligned}$$

uniformly over $\theta \in \Theta_0$, which imply Condition 1 (i), 2, 3, 7 (ii), respectively. Moreover, it is also clear that $\widehat{m}(\theta_0) = \overline{m}(\theta_0) + O_p\left(1/\sqrt{n}\right)$ by the CLT under $E[\|s(z_i,\theta_0)\|^4] < \infty$, $E[\|\nabla s(z_i,\theta_0)\|^4] < \infty$, and

$E[\left\|\nabla^2 s(z_i,\theta_0)\right\|^2] < \infty$ and that $\widehat{m}_\theta(\theta_0) = \overline{m}_\theta(\theta_0) + O_p\left(1/\sqrt{n}\right)$ by the CLT under $E[\left\|\nabla^{\overline{\overline{v}}} s(z_i,\theta_0)\right\|^4] < \infty$ for $\overline{\overline{v}} = \{0,1,2\}$ and $E[\left\|\nabla^3 s(z_i,\theta_0)\right\|^2] < \infty$. Also note that $\widehat{\tau}_m(\theta_0) = \overline{\tau}_m(\theta_0) + O_p\left(1/\sqrt{n}\right)$ by the Slutsky theorem and $\widehat{m}(\theta_0) = \overline{m}(\theta_0) + O_p\left(1/\sqrt{n}\right)$. These imply that $\widehat{c}(\theta_0) = c(\theta_0) + O_p\left(1/\sqrt{n}\right)$ and $\nabla\widehat{c}(\theta_0) = \nabla c(\theta_0) + O_p\left(1/\sqrt{n}\right)$, which are Condition 1 (ii) and 7 (i), respectively. □

**Lemma A.12.** *Under Assumption 1-2 or 3-2, Condition 4-6 are satisfied.*

**Proof.** **Condition 4:** Note

$$\begin{aligned}
&\widehat{B}(\theta_0) - B(\theta_0) = \widehat{Q}(\theta_0)\widehat{c}(\theta_0) - \widehat{Q}(\theta_0)c(\theta_0) + \widehat{Q}(\theta_0)c(\theta_0) - Q(\theta_0)c(\theta_0) \\
&= \widehat{Q}(\theta_0)\left(\widehat{c}(\theta_0) - c(\theta_0)\right) + \left(\widehat{Q}(\theta_0) - Q(\theta_0)\right)c(\theta_0) = O_p(1)\,O_p\left(1/\sqrt{n}\right) + O_p\left(1/\sqrt{n}\right),
\end{aligned}$$

since $\widehat{c}(\theta_0) - c(\theta_0) = O_p\left(1/\sqrt{n}\right)$ by Condition 1 (ii) and $\widehat{Q}(\theta_0) - Q(\theta_0) = O_p\left(1/\sqrt{n}\right)$ by Lemma A.8.

**Condition 5:** From Remark A.2, Condition 1 (ii), Condition 7 (i), $\widehat{Q}(\theta_0) = Q(\theta_0) + O_p\left(1/\sqrt{n}\right)$, and $\nabla\left(\widehat{Q}(\theta_0)'\right) = \nabla\left(Q(\theta_0)'\right) + O_p\left(1/\sqrt{n}\right)$ by Lemma A.8, we have

$$\begin{aligned}
&\nabla\widehat{B}(\theta_0) = \nabla\left(\widehat{Q}(\theta_0)\widehat{c}(\theta_0)\right) = \widehat{Q}(\theta_0)\nabla\widehat{c}(\theta_0) + vec^*\left(\widehat{c}(\theta_0)'\nabla\left(\widehat{Q}(\theta_0)'\right)\right) \\
&= Q(\theta_0)\nabla c(\theta_0) + vec^*\left(c(\theta_0)'\nabla\left(Q(\theta_0)'\right)\right) + O_p\left(1/\sqrt{n}\right) = \nabla B(\theta_0) + O_p\left(1/\sqrt{n}\right).
\end{aligned}$$

**Condition 6**: From Remark A.1-A.5, we have

$$\begin{aligned}
&\left\|\nabla^2\widehat{B}(\widetilde{\theta})\right\| = \left\|\nabla^2\left(\widehat{Q}(\widetilde{\theta})\widehat{c}(\widetilde{\theta})\right)\right\| = \left\|\nabla\left(\widehat{Q}(\theta_0)\nabla\widehat{c}(\theta_0)\right) + \nabla\left(vec^*\left(\widehat{c}(\theta_0)'\nabla\left(\widehat{Q}(\theta_0)'\right)\right)\right)\right\| \\
&\leq \left\|\widehat{Q}(\widetilde{\theta})\right\|\left\|\nabla^2\widehat{c}(\widetilde{\theta})\right\| + 2\left\|\nabla\widehat{Q}(\widetilde{\theta})\right\|\left\|\nabla\widehat{c}(\widetilde{\theta})\right\| + \left\|\nabla^2\widehat{Q}(\widetilde{\theta})\right\|\left\|\widehat{c}(\widetilde{\theta})\right\| = O_p(1),
\end{aligned}$$

from Remark A.2, Lemma A.8, $\left\|\widehat{Q}(\widetilde{\theta})\right\| = O_p(1)$, and Condition 2 and 3. □

## References


1. Tauchen, G.E. Diagnostic Testing and Evaluation of Maximum Likelihood Models. *J. Econom.* **1985**, *30*, 415–443.
2. Newey, W.K. Uniform Convergence in Probability and Stochastic Equicontinuity. *Econometrica* **1991**, *58*, 809–837.
3. Andrews, D.W.K. Empirical Process Methods in Econometrics. In *Handbook of Econometrics IV*; Engle, R.F., McFadden, D.L., Eds.; North Holland Publishing Co.: Amsterdam, The Netherlands, 1994; pp. 2247–2294.