SMU ECONOMICS & STATISTICS
WORKING PAPER SERIES

SINGAPORE MANAGEMENT UNIVERSITY

Higher Order Bias Correcting Moment Equation for M-Estimation and its Higher Order Efficiency

Kyoo il Kim

September 2006

Paper No. 17-2006

Higher Order Bias Correcting Moment Equation for M-Estimation and its Higher Order Efficiency

Kyoo il Kim*
School of Economics and Social Sciences
Singapore Management University

September 21, 2006

Abstract

This paper studies an alternative bias correction for the M-estimator, which is obtained by correcting the moment equation in the spirit of Firth (1993). In particular, this paper compares the stochastic expansions of the analytically bias-corrected estimator and the alternative estimator and finds that the third-order stochastic expansions of these two estimators are identical. This implies that at least in terms of the third order stochastic expansion, we cannot improve on the simple one-step bias correction by using the bias correction of moment equations. Though the result in this paper is for a fixed number of parameters, our intuition may extend to the analytical bias correction of the panel data models with individual specific effects. Noting the M-estimation can nest many kinds of estimators including IV, 2SLS, MLE, GMM, and GEL, our finding is a rather strong result.

Keywords: Third-order Stochastic Expansion, Bias Correction, M-estimation
JEL Classification: C10

1 Introduction

Asymptotic bias corrections are pursued to make estimators closer to the truth values. There are several ways of achieving this goal including analytical corrections, jackknife, and bootstrap methods. This variety of bias correction methods evokes the issue whether one method is preferable to the others at least on asymptotic efficiency grounds. Hahn, Kuersteiner, and Newey (2004) deal with this issue. For the maximum likelihood (ML) estimation, they show that a method of bias correction does not affect the higher-order efficiency of

*I am truly grateful to Jinyong Hahn for his advice and encouragement throughout my doctoral study in UCLA. I would also like to thank Joe Hotz, Moshe Buchinsky, Janet Currie, Duncan Thomas, and Patrik Guggenberger for their useful comments. I also thank proseminar participants in Applied Microeconomics and Econometrics. All remaining errors are mine. Correspondence: Kyoo il Kim, School of Economics and Social Sciences, Singapore Management University, 90 Stamford Road Singapore 178903, Tel. +65 6828-0876, Fax. +65 6828-0833, E-mail: kyookim@smu.edu.sg.

any estimator that is first-order efficient in parametric or semiparametric models. An ML estimator is a class of M-estimator and this paper extends their intuition to a general class of M-estimator.[1]

Specifically, this paper considers an alternative bias correction for the M-estimator, which is achieved by correcting the moment equation in the spirit of Firth (1993). In particular, we compare the stochastic expansions of the analytically bias-corrected estimator (which is referred to one-step bias correction) and the alternative estimator and find that the third-order stochastic expansions of these two estimators are identical. This is a stronger result, since it implies that these two estimators do not only have the same higher-order variances but also agree upon more properties in terms of their stochastic expansions.

In the literature (see Hahn and Newey (2004) and Ferández-Val (2004)), it has been discussed that removing the bias directly from the moment equations has the attractive features that it does not use pre-estimated parameters that are not bias corrected, though this alternative approach requires more intensive computations. This paper, however, illustrates that at least for the third order stochastic expansion, there is no benefit of using the bias correction of the moment equations over the simple one-step bias correction. Though our result is for the fixed number of parameters, we conjecture this is also true for the panel data models with individual specific parameters.

Obvious examples of the M-estimation include MLE, least squares and instrumental variable (IV) estimation. Many other popular estimators can also fit into the M-estimation framework with appropriate definition of the moment equations. It includes some cases of generalized method of moments (GMM, see examples in Rilstone, Srivastava, and Ullah (1996)), and two-step estimators (Newey (1984)). More interestingly the generalized empirical likelihood (GEL) also fits into this framework. This suggests that our approach can be an alternative to Newey and Smith (2004) when one is obtaining the higher-order bias and variance terms of GEL. From the finding of our paper, it follows that the stochastic expansions for the one-step bias corrected estimator and the bias corrected moment equation estimator of GEL will be identical, at least up to the third order.

Our paper is organized as follows. In Section 2 we derive the higher-order stochastic expansion of the M-estimator and consider the one-step bias correction. Section 3 introduces the bias corrected moment equations estimator and derives its higher-order stochastic expansion. Section 4 discusses the higher-order efficiency properties of several analytically bias-corrected estimators. We conclude in Section 5. Primitive conditions for the validity of the higher-order stochastic expansions and mathematical details are discussed in Appendix.

2 Higher Order Expansion for M-Estimator

Consider a moment condition

$$E\left[s\left(z_i,\theta_0\right)\right]=0 \tag{1}$$

[1] This possible extension is noted in Hahn and Newey (2004).

where $s(z_i, \theta)$ is a known $k \times 1$ vector-valued function of the data and a parameter vector $\theta \in \Theta \subset R^k$ and z_i may include both endogenous and exogenous variables. The M-estimator is obtained by solving

$$\frac{1}{n} \sum_{i=1}^{n} s\left(z_i, \widehat{\theta}\right) = 0. \tag{2}$$

Examples for this class of estimators include the MLE, the least squares and IV estimation. In the MLE, $s(z_i, \theta)$ is the single observation score function. For the linear or nonlinear regression model of $y_i = f(X_i; \theta_0) + \varepsilon_i$, we set $s(z_i, \theta) = \frac{\partial f(X_i;\theta)}{\partial \theta}(y_i - f(X_i; \theta))$ and $z_i = (y_i\ X_i')'$ for a known function $f(\cdot)$. In the linear IV model, we have $s(z_i, \theta) = w_i(y_i - X_i'\theta)$ and $z_i = (y_i\ X_i'\ w_i')'$ for some instruments w_i with $\dim(w_i) = \dim(\theta)$. Two-step estimators such as two-stage least squares, feasible generalized least squares (GLS) and Heckman (1979)'s two-step estimator also fit into this framework (see Newey (1984)). Rilstone, Srivastava, and Ullah (1996) provide some special cases of GMM estimators that can be put into the M-estimation but the examples are not restricted to those. Actually the popular two-step GMM estimations and the generalized empirical likelihood estimations (GEL, Newey and Smith (2004)) can also fit into the M-estimation. Partly motivated with this wide applicability, we study the stochastic expansion and the bias correction of the M-estimator.

We obtain the higher order stochastic expansion of the M-estimator using the iterative approach used in Rilstone, Srivastava, and Ullah (1996) up to a certain order. This approach is convenient analytically and straightforward since the estimators are expressed as functions of sums of random variables. Edgeworth expansion can be considered as an alternative whose validity has been derived in Bhattacharya and Ghosh (1978) but the stochastic expansion approach is noted as a much simpler approach. Moreover, the main purpose of this paper is to provide the comparison of several estimators based on the higher-order variance ($O(n^{-1})$ variance). Noting rankings based on the higher-order variances in a third-order stochastic expansion are equivalent to rankings based on the variances of an Edgeworth expansion as shown in Pfanzagl and Wefelmeyer (1978) and Ghosh et. al.(1980) and discussed in Rothenberg (1984), it suffices to use the simple stochastic expansions for our purpose.

Here we borrow Rilstone, Srivastava, and Ullah (1996)'s notation. We denote the matrix of υ-th order partial derivatives of a matrix $A(\theta)$ as $\nabla^{\upsilon} A(\theta)$. Specifically, if $A(\theta)$ is a $k \times 1$ vector function, $\nabla A(\theta)$ is the usual Jacobian whose l-th row contains the partial derivatives of the l-th element of $A(\theta)$. $\nabla^{\upsilon} A(\theta)$ (a $k \times k^{\upsilon}$ matrix) is defined recursively such that the j-th element of the l-th row of $\nabla^{\upsilon} A(\theta)$ is the $1 \times k$ vector $a_{lj}^{\upsilon}(\theta) = \partial a_{lj}^{\upsilon-1}(\theta)/\partial\theta'$, where $a_{lj}^{\upsilon-1}$ is the l-th row and the j-th element of $\nabla^{\upsilon-1} A(\theta)$. We use $\otimes$ to denote a usual Kronecker product. Using this Kronecker product we can express $\nabla^{\upsilon} A(\theta) = \frac{\partial^{\upsilon} A(\theta)}{\underbrace{\partial\theta' \otimes \partial\theta' \otimes \ldots \otimes \partial\theta'}_{\upsilon \text{ Kronecker product of } \partial\theta'}}$. Finally, we use a matrix norm $\|A\| = \sqrt{tr(A'A)}$ for a matrix A.

Before we derive the second order expansion of the M-estimator to obtain the second-order bias analytically, some definitions are introduced. Denote $H_1(\theta) = E\left[\nabla s(z_i, \theta)\right]$, $H_2(\theta) = E\left[\nabla^2 s(z_i, \theta)\right]$, $Q(\theta) = \left(-E\left[\nabla s(z_i, \theta)\right]\right)^{-1}$ and let $H_1 = H_1(\theta_0)$, $H_2 = H_2(\theta_0)$, $Q = Q(\theta_0)$. The following notation is also used later; $\widehat{H}_1(\theta) = \frac{1}{n}\sum_{i=1}^{n} \nabla s(z_i, \theta)$, $H_2(\theta) = \frac{1}{n}\sum_{i=1}^{n} \nabla^2 s(z_i, \theta)$, $\widehat{Q}(\theta) = (-\widehat{H}_1(\theta))^{-1}$, $\widehat{H}_1 = \widehat{H}_1(\theta_0)$, $\widehat{H}_2 = \widehat{H}_2(\theta_0)$, and $\widehat{Q} = \widehat{Q}(\theta_0)$. Also define $J \equiv \frac{1}{\sqrt{n}}\sum_{i=1}^{n} s(z_i, \theta_0)$, $V \equiv \frac{1}{\sqrt{n}}\sum_{i=1}^{n} \left(\nabla s(z_i, \theta_0) - E\left[\nabla s(z_i, \theta_0)\right]\right)$, $W \equiv \frac{1}{\sqrt{n}}\sum_{i=1}^{n} \left(\nabla^2 s(z_i, \theta_0) - E\left[\nabla^2 s(z_i, \theta_0)\right]\right)$.

Lemma 2.1 *Suppose* $\{z_i\}_{i=1}^n$ *is iid,* θ_0 *is in the interior of* Θ *and is the only* $\theta \in \Theta$ *satisfying (1), and the M-estimator* $\widehat{\theta}$ *defined in (2) is consistent. Further suppose that (i)* $s(z,\theta)$ *is* κ*-times continuously differentiable in the neighborhood of* θ_0*, denoted by* $\Theta_0 \subset \Theta$ *for all* $z \in \mathcal{Z}$*,* $\kappa \geq 3$ *with probability one; (iia)* $\nabla^\upsilon s(z,\theta)$ *is integrable for each fixed* $\theta \in \Theta_0$*,* $\upsilon = \{0,1,2,\ldots\kappa\}$*,* $\kappa \geq 3$ *and (iib)* $E\left[\nabla^3 s(z,\theta)\right]$ *is continuous and bounded at* θ_0*; (iii)* $\left\|\frac{1}{n}\sum_{i=1}^n \nabla^\upsilon s\left(z_i,\overline{\theta}\right) - E\left[\nabla^\upsilon s\left(z_i,\theta_0\right)\right]\right\| = o_p(1)$ *for* $\overline{\theta} = \theta_0 + o_p(1)$ *and* $\upsilon = 1,2$*;*
(iv) $\frac{1}{\sqrt{n}}\sum_{i=1}^n \left(\nabla^2 s\left(z_i,\overline{\theta}\right) - H_2(\overline{\theta})\right) - \frac{1}{\sqrt{n}}\sum_{i=1}^n \left(\nabla^2 s\left(z_i,\theta_0\right) - H_2\left(\theta_0\right)\right) = o_p\left(1\right)$ *for* $\overline{\theta} = \theta_0 + o_p(1)$*;*
(v) $Q(\theta_0)$ *exists, i.e.* $E\left[\nabla s(z_i,\theta_0)\right]$ *is nonsingular; (vi)* $J = O_p(1)$*; (vii)* $V = O_p(1)$*; (viii)* $W = O_p(1)$*.*
Then we have $\sqrt{n}\left(\widehat{\theta} - \theta_0\right) = QJ + O_p\left(\frac{1}{\sqrt{n}}\right)$ *and moreover*
$\sqrt{n}\left(\widehat{\theta} - \theta_0\right) = QJ + \frac{1}{\sqrt{n}}Q\left(VQJ + \frac{1}{2}H_2\left(QJ \otimes QJ\right)\right) + O_p(n^{-1})$*.*

This result and the following Lemma 2.2 are available in Rilstone, Srivastava, and Ullah (1996) but we provide these and their proofs for completeness. Some of these results will be used in later discussion. The proofs to this lemma and others are presented in Appendix B. From this result, the higher order bias of $\widehat{\theta}$ is obtained as

$$Bias(\widehat{\theta}) \equiv \frac{1}{n}Q\left(E\left[VQJ\right] + \frac{1}{2}H_2 E\left[(QJ \otimes QJ)\right]\right).$$

Defining $d_i(\theta) = Q(\theta)s(z_i,\theta)$ and $v_i(\theta) = \nabla s\left(z_i,\theta\right) - E\left[\nabla s\left(z_i,\theta\right)\right]$ and letting $d_i = d_i(\theta_0)$ and $v_i = v_i(\theta_0)$, it is not difficult to see that $Q\left(E\left[VQJ\right] + \frac{1}{2}H_2 E\left[(QJ \otimes QJ)\right]\right) = Q\left(E\left[v_i d_i\right] + \frac{1}{2}H_2 E\left[(d_i \otimes d_i)\right]\right)$ as shown in Lemma 2.2 and thus we put $B(\theta) = Q(\theta)\left(E\left[v_i(\theta)d_i(\theta)\right] + \frac{1}{2}H_2(\theta)E\left[d_i(\theta) \otimes d_i(\theta)\right]\right)$.

Lemma 2.2 *Suppose (1) holds and* $\{z_i\}_{i=1}^n$ *are iid.*
Then, $E\left[VQJ\right] + \frac{1}{2}H_2 E\left[QJ \otimes QJ\right] = E\left[v_i d_i\right] + \frac{1}{2}H_2 E\left[d_i \otimes d_i\right]$*, where* $d_i = Qs(z_i,\theta_0)$ *and* $v_i = \nabla s\left(z_i,\theta_0\right) - E\left[\nabla s\left(z_i,\theta_0\right)\right]$*.*

Therefore, we can eliminate the second-order bias of the M-estimator $\widehat{\theta}$ by subtracting a consistent estimator of the bias. Now let $\widehat{\theta}_{bc}$ denote the bias corrected estimator of this sort defined by

$$\widehat{\theta}_{bc} = \widehat{\theta} - \frac{1}{n}\widehat{B}(\overline{\theta}), \tag{3}$$

for a consistent estimator $\overline{\theta}$ of θ_0 where the function $\widehat{B}(\theta)$ is constructed as

$$\widehat{Q}(\theta)\left(\frac{1}{n}\sum_{i=1}^n \widehat{v}_i(\theta)\widehat{d}_i(\theta) + \frac{1}{2}\widehat{H}_2(\theta)\frac{1}{n}\sum_{i=1}^n\left(\widehat{d}_i(\theta) \otimes \widehat{d}_i(\theta)\right)\right) \tag{4}$$

for $\widehat{d}_i(\theta) = \widehat{Q}(\theta)s(z_i,\theta)$ and $\widehat{v}_i\left(\theta\right) = \nabla s\left(z_i,\theta\right)$. In particular, we can put $\overline{\theta} = \widehat{\theta}$. In this sense, $\widehat{\theta}_{bc}$ is a two-step estimator.

To characterize the higher order efficiency based on the higher-order variance ($O(n^{-1})$ variance) of the bias corrections, we need to expand the M-estimator to the third-order. We use some additional definitions: $H_3\left(\theta\right) = E[\nabla^3 s(z,\theta)]$, $\widehat{H}_3\left(\theta\right) = \frac{1}{n}\sum_{i=1}^n \nabla^3 s(z_i,\theta)$, $H_3 = H_3\left(\theta_0\right)$, $W_3 \equiv \frac{1}{\sqrt{n}}\sum_{i=1}^n\left(\nabla^3 s\left(z_i,\theta_0\right) - E\left[\nabla^3 s\left(z_i,\theta_0\right)\right]\right)$.
Also put $a_{-1/2} = QJ$, $a_{-1} = Q\left(Va_{-1/2} + \frac{1}{2}H_2\left(a_{-1/2} \otimes a_{-1/2}\right)\right)$ and
$a_{-3/2} = QVa_{-1} + \frac{1}{2}QW\left(a_{-1/2} \otimes a_{-1/2}\right) + \frac{1}{2}QH_2\left(a_{-1/2} \otimes a_{-1} + a_{-1} \otimes a_{-1/2}\right) + \frac{1}{6}QH_3\left(a_{-1/2} \otimes a_{-1/2} \otimes a_{-1/2}\right)$
for brevity. First consider

Lemma 2.3 *Suppose $\{z_i\}_{i=1}^n$ is iid, θ_0 is in the interior of Θ and is the only $\theta \in \Theta$ satisfying (1), and the M-estimator $\widehat{\theta}$ that solves (2) is consistent. Further suppose that (i) $s(z,\theta)$ is κ-times continuously differentiable in a neighborhood of θ_0, denoted by $\Theta_0 \subset \Theta$ for all $z \in \mathcal{Z}$, $\kappa \geq 4$ with probability one; (iia) $\nabla^{\upsilon} s(z,\theta)$ is integrable for each fixed $\theta \in \Theta_0$, $\upsilon = \{0,1,2,\ldots\kappa\}$, $\kappa \geq 4$ and (iib) $E[\nabla^4 s(z,\theta)]$ is continuous and bounded at θ_0; (iii) $\frac{1}{\sqrt{n}}\sum_{i=1}^n \left(\nabla^3 s\left(z_i,\overline{\theta}\right) - H_3\left(\overline{\theta}\right)\right) - \frac{1}{\sqrt{n}}\sum_{i=1}^n \left(\nabla^3 s\left(z_i,\theta_0\right) - H_3\left(\theta_0\right)\right) = o_p(1)$ for $\overline{\theta} = \theta_0 + o_p(1)$; (iv) Q is nonsingular; (v) $J = O_p(1)$; (vi) $V = O_p(1)$; (vii) $W = O_p(1)$; (viii) $W_3 = O_p(1)$;*
(ix) $\sqrt{n}(\widehat{\theta} - \theta_0) = a_{-1/2} + \frac{1}{\sqrt{n}}a_{-1} + O_p\left(\frac{1}{n}\right)$.
Then we have $\sqrt{n}\left(\widehat{\theta} - \theta_0\right) = a_{-1/2} + \frac{1}{\sqrt{n}}a_{-1} + \frac{1}{n}a_{-3/2} + O_p(n^{-3/2})$.

In the following section, we propose an alternative one-step estimator which eliminates the second-order bias by adjusting the moment equation inspired by Firth (1993).

3 Bias Corrected Moment Equation

Here we consider an alternative higher order bias reduced estimator that solves a bias corrected moment equation. This idea is proposed in Firth (1993) for the ML with a fixed number of parameters and exploited in Hahn and Newey (2003) and Ferández-Val (2004) for the nonlinear panel data models with individual specific effects. We refer this estimator to Firth's estimator.

To be more precise, consider

$$0 = \frac{1}{n}\sum_{i=1}^n s\left(z_i,\theta\right) - \frac{1}{n}c(\theta) \tag{5}$$

for a known function $c(\theta)$ that is given by

$$c(\theta) = Q(\theta)^{-1}B(\theta) = \frac{1}{2}H_2(\theta)E\left[Q(\theta)s(z_i,\theta) \otimes Q\left(\theta\right)s(z_i,\theta)\right] + E\left[\nabla s\left(z_i,\theta\right)Q(\theta)s(z_i,\theta)\right]. \tag{6}$$

In the ML context, Firth (1993) shows that by adjusting the score function (he refers this as a modified score function) with the correction term defined by the product of the Fisher information matrix and the bias term. $c(\theta)$ has the same interpretation in the ML, since $-Q(\theta)^{-1}$ is the Hessian matrix and hence $Q(\theta)^{-1}$ is the Fisher information in the ML. Therefore (6) is a generalization of Firth (1993)'s idea to the M-estimation. In general $c(\theta)$ is unknown and hence to implement this alternative estimator, we need to estimate the function $c\left(\theta\right)$. We use a sample analogue of (6) as

$$\begin{aligned} \widehat{c}(\theta) &= \widehat{Q}(\theta)^{-1}\widehat{B}(\theta) \\ &= \frac{1}{2}\widehat{H}_2(\theta)\left(\frac{1}{n}\sum_{i=1}^n\left[\widehat{Q}(\theta)s(z_i,\theta) \otimes \widehat{Q}(\theta)s(z_i,\theta)\right]\right) + \frac{1}{n}\sum_{i=1}^n\left[\nabla s\left(z_i,\theta\right)\widehat{Q}(\theta)s(z_i,\theta)\right]. \end{aligned} \tag{7}$$

Now we estimate θ_0 by solving

$$0 = \frac{1}{n}\sum_{i=1}^n s\left(z_i,\theta\right) - \frac{1}{n}\widehat{c}(\theta), \tag{8}$$

and claim that this eliminates the second order bias of $\widehat{\theta}$ that solves (2) under following conditions;

Assumption 3.1 *(i)* $\{z_i : i = 1, \ldots, n\}$ *are iid; (ii)* $s(z, \theta)$ *is* κ*-times continuously differentiable in a neighborhood of* θ_0*, denoted by* Θ_0 *for all* $z \in \mathcal{Z}$*,* $\kappa \geq 4$*; (iii)* $E\left[\sup_{\theta\in\Theta_0} \|\nabla^{\upsilon} s(z,\theta)\|^2\right] < \infty$ $\upsilon = \{0, 1, 2, \ldots \kappa\}$*,* $\kappa \geq 4$*; (iv)* Θ *is compact; (v)* θ_0 *is in the interior of* Θ *and is the only* $\theta \in \Theta$ *satisfying (1); (vi)* $E\left[\left\|\nabla^{\overline{\upsilon}} s(z,\theta_0)\right\|^4\right] < \infty$ *for* $\overline{\upsilon} = \{0, 1, 2, \ldots, \overline{\kappa}\}, \overline{\kappa} \geq 3$.

Assumption 3.2 *For* $\theta \in \Theta_0$*,* $E\left[\frac{\partial s(z_i,\theta)}{\partial \theta'}\right]$ *is nonsingular.*

or alternatively instead of Assumption 3.1,

Assumption 3.3 *(i)* $\{z_i : i = 1, \ldots, n\}$ *are iid; (ii)* $\nabla^{\upsilon} s(z, \theta)$ *satisfies the Lipschitz condition in* θ *as*

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

for some function $B_{\upsilon}(\cdot)\colon \mathcal{Z} \to R$ *and* $E\left[B_{\upsilon}(\cdot)^{2t+\delta}\right] < \infty$*,* $\upsilon = \{0, 1, 2, \ldots \kappa\}$*, with positive integer* $t \geq 2$ *and for some* $\delta > 0$ *and* $\kappa \geq 4$ *in a neighborhood of* θ_0*, (iii)* $E\left[\sup_{\theta\in\Theta_0} \|\nabla^{\upsilon} s(z,\theta)\|^{2t+\delta}\right] < \infty$*,* $\upsilon = \{0, 1, 2, \ldots \kappa\}$*,* $\kappa \geq 4$ *with positive integer* $t \geq 2$ *and for some* $\delta > 0$*; (iv)* Θ *is bounded; (v)* θ_0 *is in the interior of* Θ *and is the only* $\theta \in \Theta$ *satisfying (1).*

Under Assumption 3.1-3.2 or Assumption 3.2-3.3, the following three conditions are satisfied (see Lemma A.9 in the appendix).

Condition 1 *(i)* $\widehat{c}(\theta_0) = O_p(1)$*;(ii)* $\widehat{c}(\theta_0) = c(\theta_0) + O_p\left(\frac{1}{\sqrt{n}}\right)$.

Condition 2 $\nabla\widehat{c}(\theta) = O_p(1)$ *around the* $n^{-1/2}$ *neighborhood of* θ_0.

Condition 3 $\nabla^2\widehat{c}(\theta) = O_p(1)$ *around the* $n^{-1/2}$ *neighborhood of* θ_0.

Now we are ready to present one of our main findings.

Proposition 3.1 *Suppose* θ^* *solves*

$$0 = \frac{1}{n}\sum_{i=1}^{n} s\left(z_i, \theta^*\right) - \frac{1}{n}\widehat{c}(\theta^*), \tag{9}$$

where $\widehat{c}(\theta)$ *is given by (7) and that* θ^* *is a consistent estimator of* θ_0*. Further suppose that Condition 1-3 and Condition (i)-(viii) in Lemma 2.1 are satisfied, then we have*

$$\sqrt{n}\left(\theta^* - \theta_0\right) = QJ + \frac{1}{\sqrt{n}}Q\left(VQJ + \frac{1}{2}H_2\left(QJ \otimes QJ\right) - c(\theta_0)\right) + O_p\left(\frac{1}{n}\right),$$

where $c(\theta_0) = \frac{1}{2}H_2 E\left[Qs(z_i,\theta_0) \otimes Qs(z_i,\theta_0)\right] + E\left[\nabla s\left(z_i,\theta_0\right) Qs(z_i,\theta_0)\right]$ *and hence the second-order bias of* θ^* *is* $Bias(\theta^*) \equiv \frac{1}{n}E\left[Q\left(VQJ + \frac{1}{2}H_2\left(QJ \otimes QJ\right) - c(\theta_0)\right)\right] = 0$.

This concludes that we can eliminate the second order bias by adjusting the moment equation as (8) and it is a proper alternative to the analytic bias-correction of (3). Now we derive the higher order expansion of the Firth's estimator up to the third order. For this, we need an additional condition that is satisfied under Assumption 3.1-3.2 or 3.2-3.3 with $\kappa \geq 5$ as shown in Lemma A.11 in the appendix.

Condition 4 *(i)* $\nabla \widehat{c}(\theta_0) = \nabla c(\theta_0) + O_p\left(\frac{1}{\sqrt{n}}\right)$*;(ii)* $\nabla^3 \widehat{c}(\theta) = O_p(1)$ *around the* $n^{-1/2}$ *neighborhood of* θ_0.

Recall that $c(\theta) = Q^{-1}(\theta)B(\theta)$ and $\widehat{c}(\theta) = \widehat{Q}(\theta)^{-1}\widehat{B}(\theta)$ and we obtain

Proposition 3.2 *Suppose* θ^* *solves* $0 = \frac{1}{n}\sum_{i=1}^n s\left(z_i, \theta^*\right) - \frac{1}{n}\widehat{c}(\theta^*)$*, where* $\widehat{c}(\theta)$ *is given in (7) and that* θ^* *is consistent. Further suppose that Condition 1-4 and Condition (i)-(viii) in Lemma 2.3 are satisfied and assume* $\sqrt{n}(\widehat{\theta} - \theta_0) = a_{-1/2} + \frac{1}{\sqrt{n}}\left(a_{-1} - B(\theta_0)\right) + O_p\left(\frac{1}{n}\right)$*. Then, we have*

$$\begin{aligned} &\sqrt{n}\left(\theta^* - \theta_0\right) \\ &= a_{-1/2} + \frac{1}{\sqrt{n}}\left(a_{-1} - B(\theta_0)\right) + \frac{1}{n}\left(a_{-3/2} - \nabla B(\theta_0)a_{-1/2} - \sqrt{n}(\widehat{B}(\theta_0) - B(\theta_0))\right) + O_p(n^{-3/2}) \end{aligned} \tag{10}$$

4 Higher Order Efficiency

Asymptotic bias corrections can provide estimators that have better bias properties in the finite sample. There are several ways of achieving these bias corrections including analytical corrections that we focus on in this paper, jackknife and bootstrap methods. This abundant ways of bias correction methods evoke the issue which method is preferable to others at least on asymptotic efficiency grounds. Hahn, Kuersteiner, and Newey (2004) deals with this issue. For the maximum likelihood (ML) estimation, they show that the method of bias correction does not affect the higher-order efficiency of any estimator that is first-order efficient in a parametric or semiparametric model. The ML estimator is a class of the M-estimator and here we try to extend their intuition to a general M-estimator. In this section we compare the higher order efficiency of several first-order efficient bias-corrected estimators by comparing the higher order variance, which is defined by the $O\left(\frac{1}{n}\right)$ variance in a third-order stochastic expansion of the estimator.

4.1 Third Order Expansion of the Bias-Corrected Estimator

To compare with the estimator of interest θ^*, first we consider a bias-corrected estimator $\widehat{\theta}_{bc}$ defined in (3) as $\widehat{\theta}_{bc} = \widehat{\theta} - \frac{1}{n}\widehat{B}(\widehat{\theta})$ and observe that $\widehat{B}(\widehat{\theta}) = \widehat{Q}(\widehat{\theta})\widehat{c}(\widehat{\theta})$ from (4) and (7). We also consider its infeasible version $\widehat{\theta}_b$ as $\widehat{\theta}_b = \widehat{\theta} - \frac{1}{n}B(\widehat{\theta})$, where the function $B(\widehat{\theta})$ is constructed as $B(\widehat{\theta}) = Q(\widehat{\theta})c(\widehat{\theta})$ provided that both $\widehat{B}(\widehat{\theta})$ and $B(\widehat{\theta})$ are consistent estimators of the higher order bias term $B(\theta_0) = Q(\theta_0)c(\theta_0)$. Note that for $\widetilde{\theta}$ between $\widehat{\theta}$ and θ_0, the mean value theorem gives us

$$c(\widehat{\theta}) - c(\theta_0) = \nabla c(\widetilde{\theta})\left(\widehat{\theta} - \theta_0\right) = O_p(1)O_p\left(1/\sqrt{n}\right) = o_p(1)$$

under Condition 2 and since $\widehat{\theta} - \theta_0 = O_p\left(\frac{1}{\sqrt{n}}\right)$. Also we have

$$\begin{aligned} \left\|\widehat{c}(\widehat{\theta}) - c(\theta_0)\right\| &\leq \left\|\widehat{c}(\widehat{\theta}) - c(\widehat{\theta})\right\| + \left\|c(\widehat{\theta}) - c(\theta_0)\right\| \\ &\leq \sup_{\theta \in \Theta_0}\left\|\widehat{c}(\theta) - c(\theta)\right\| + \left\|c(\widehat{\theta}) - c(\theta_0)\right\| = o_p(1) + o_p(1) = o_p(1) \end{aligned}$$

by Triangle Inequality, Lemma A.7, and the continuity of $c(\theta)$ at θ_0 (applying the Slutsky theorem) and hence both $B(\widehat{\theta})$ and $\widehat{B}(\widehat{\theta})$ are indeed consistent estimators of the higher order bias noting $Q(\widehat{\theta}) = Q(\theta_0) + o_p(1)$ by the continuity of $Q(\theta)$ at θ_0 and $\widehat{Q}(\widehat{\theta}) = Q(\theta_0) + o_p(1)$[2]. Now from the result of Lemma 2.3, it follows that

$$\begin{aligned}\sqrt{n}(\widehat{\theta}_b - \theta_0) &= \sqrt{n}(\widehat{\theta} - \theta_0) - \frac{1}{\sqrt{n}} B(\widehat{\theta}) \\ &= \begin{array}{l} a_{-1/2} + \frac{1}{\sqrt{n}} a_{-1} + \frac{1}{n} a_{-3/2} + O_p(n^{-3/2}) \\ -\frac{1}{\sqrt{n}} B(\theta_0) - \frac{1}{\sqrt{n}} \nabla B(\theta_0)(\widehat{\theta} - \theta_0) - \frac{1}{2\sqrt{n}} \nabla^2 B(\widetilde{\theta})((\widehat{\theta} - \theta_0) \otimes (\widehat{\theta} - \theta_0)) \end{array}\end{aligned}$$

and hence

$$\sqrt{n}(\widehat{\theta}_b - \theta_0) = a_{-1/2} + \frac{1}{\sqrt{n}} (a_{-1} - B(\theta_0)) + \frac{1}{n} \left(a_{-3/2} - \nabla B(\theta_0) a_{-1/2}\right) + O_p(n^{-3/2}), \tag{11}$$

since $\sqrt{n}(\widehat{\theta} - \theta_0) = a_{-1/2} + O_p\left(\frac{1}{\sqrt{n}}\right)$ and $\nabla^2 B(\widetilde{\theta}) = \nabla^2 B(\theta_0) + o_p(1) = O_p(1)$ by the Slutsky theorem, from which we have $\frac{1}{2\sqrt{n}} \nabla^2 B(\widetilde{\theta})((\widehat{\theta} - \theta_0) \otimes (\widehat{\theta} - \theta_0)) = O_p(n^{-3/2})$. Similarly for $\widehat{\theta}_{bc}$, consider

$$\begin{aligned}\sqrt{n}(\widehat{\theta}_{bc} - \theta_0) &= \sqrt{n}(\widehat{\theta} - \theta_0) - \frac{1}{\sqrt{n}} \widehat{B}(\widehat{\theta}) \\ &= \begin{array}{l} a_{-1/2} + \frac{1}{\sqrt{n}} a_{-1} + \frac{1}{n} a_{-3/2} + O_p(n^{-3/2}) \\ -\frac{1}{\sqrt{n}} \widehat{B}(\theta_0) - \frac{1}{\sqrt{n}} \nabla \widehat{B}(\theta_0)(\widehat{\theta} - \theta_0) - \frac{1}{2\sqrt{n}} \nabla^2 \widehat{B}(\widetilde{\theta})((\widehat{\theta} - \theta_0) \otimes (\widehat{\theta} - \theta_0)). \end{array}\end{aligned} \tag{12}$$

From (12) and the following results (that hold under Assumption 3.1-3.2 or 3.2-3.3 as shown in Lemma A.12 in the appendix),

Condition 5 $\widehat{B}(\theta_0) = B(\theta_0) + O_p\left(\frac{1}{\sqrt{n}}\right)$

Condition 6 $\nabla \widehat{B}(\theta_0) = \nabla B(\theta_0) + O_p\left(\frac{1}{\sqrt{n}}\right)$

Condition 7 $\nabla^2 \widehat{B}(\theta) = O_p(1)$ *around the neighborhood of* θ_0.

We obtain

$$\begin{aligned}&\sqrt{n}(\widehat{\theta}_{bc} - \theta_0) \\ = \; & a_{-1/2} + \frac{1}{\sqrt{n}} (a_{-1} - B(\theta_0)) + \frac{1}{n} \left(a_{-3/2} - \nabla B(\theta_0) a_{-1/2} - \sqrt{n}(\widehat{B}(\theta_0) - B(\theta_0))\right) + O_p(n^{-3/2})\end{aligned} \tag{13}$$

noting $\frac{1}{\sqrt{n}} \nabla \widehat{B}(\theta_0)(\widehat{\theta} - \theta_0) = \frac{1}{n} \nabla \widehat{B}(\theta_0) \left(a_{-1/2} + O_p\left(\frac{1}{\sqrt{n}}\right)\right) = \frac{1}{n} \nabla B(\theta_0) a_{-1/2} + O_p(n^{-3/2})$ by Condition 6 and noting $\frac{1}{2\sqrt{n}} \nabla^2 \widehat{B}(\widetilde{\theta})((\widehat{\theta} - \theta_0) \otimes (\widehat{\theta} - \theta_0)) = O_p(n^{-3/2})$ by Condition 7 and $\widehat{\theta} - \theta_0 = O_p\left(\frac{1}{\sqrt{n}}\right)$. Comparing (10) and (13), we conclude that $\sqrt{n}\left(\theta^* - \theta_0\right)$ and $\sqrt{n}\left(\widehat{\theta}_{bc} - \theta_0\right)$ are identical up to $O_p(\frac{1}{n})$ order terms. This means that θ^* and $\widehat{\theta}_{bc}$ have the same higher order variances at least.

[2] For a formal proof, see the result of (75) in the proof of Lemma 2.1 in the appendix.

4.2 Higher Order Variances

For a three term stochastic expansion of an estimator such as

$$\sqrt{n}\left(\check{\theta}-\theta_0\right)=T_{-1/2}+\frac{1}{\sqrt{n}}T_{-1}+\frac{1}{n}T_{-3/2}+O_p(n^{-3/2}),$$

the higher-order variance is given by

$$\Lambda_{\check{\theta}}\equiv\Sigma+\frac{1}{n}\Xi,$$

where Σ =Var$[T_{-1/2}]$ and $\Xi = \left(\text{Var}[T_{-1}]+E\left[\left(\sqrt{n}T_{-1}+T_{-3/2}\right)T_{-1/2}'\right]+E\left[T_{-1/2}\left(\sqrt{n}T_{-1}+T_{-3/2}\right)'\right]\right)$. From (10), (11), and (13), we obtain the higher order variances of three alternative estimators, denoted by $\Lambda_{\widehat{\theta}_b}$, $\Lambda_{\widehat{\theta}_{bc}}$, and Λ_{θ^*}, respectively as[3]

$$\begin{aligned}\Lambda_{\widehat{\theta}_b} &= \left\{\begin{array}{l} E\left[a_{-1/2}a_{-1/2}'\right]+\frac{1}{n}E\left[\left(a_{-1}-B(\theta_0)\right)\left(a_{-1}-B(\theta_0)\right)'\right] \\ +\frac{1}{n}E\left[a_{-1/2}\left(a_{-3/2}-\nabla B(\theta_0)a_{-1/2}\right)'\right]+\frac{1}{n}E\left[\left(a_{-3/2}-\nabla B(\theta_0)a_{-1/2}\right)a_{-1/2}'\right] \\ +\frac{1}{n}E\left[\sqrt{n}a_{-1/2}\left(a_{-1}-B(\theta_0)\right)'\right]+\frac{1}{n}E\left[\sqrt{n}\left(a_{-1}-B(\theta_0)\right)a_{-1/2}'\right]\end{array}\right\} \\ \Lambda_{\widehat{\theta}_{bc}} &= \Lambda_{\widehat{\theta}_b}-\frac{1}{n}E\left[a_{-1/2}\sqrt{n}(\widehat{B}(\theta_0)-B(\theta_0))'\right]-\frac{1}{n}E\left[\sqrt{n}(\widehat{B}(\theta_0)-B(\theta_0))a_{-1/2}'\right] \qquad (14) \\ \Lambda_{\theta^*} &= \Lambda_{\widehat{\theta}_{bc}}.\end{aligned}$$

The result of (14) reveals that the higher order variance of $\widehat{\theta}_{bc}$ has additional terms compared with $\widehat{\theta}_b$ due to the fact that we use the sample analogue of the second order bias, unless $E\left[a_{-1/2}\sqrt{n}(\widehat{B}(\theta_0)-B(\theta_0))'\right]=0$. It is quite remarkable that comparing the third order expansions of (10) and (13), we have concluded that

$$n^{3/2}(\widehat{\theta}_{bc}-\theta^*)=o_p(1). \qquad (15)$$

This is a stronger result, since it implies that these two estimators do not only have the same higher order variance but also agree upon more properties in terms of their stochastic expansions. In the literature (see Hahn and Newey (2003) and Ferández-Val (2004)), it has been argued that removing the bias directly from the moment equations has the attractive features that it does not use pre-estimated parameters that are not bias corrected, though this alternative approach requires more intensive computations since it requires to solve some nonlinear equation. From the results of (10) and (13), this paper concerns that at least for the third order stochastic expansion comparison, there is no benefit of using such bias correction of the moment equations over the simple bias-corrected estimator. Though our result is for the fixed number of parameters, we conjecture this is also true for the panel data models with individual specific parameters.

4.3 Comparison of Alternative Estimators

To have a better understanding for the result of (15). Here we compare several versions of bias-corrected estimators though these are infeasible in most of cases. First, let θ_1^* be the solution of $0=\frac{1}{n}\sum_{i=1}^n s\left(z_i,\theta\right)-\frac{1}{n}c(\theta)$ and we also define

$$\widehat{\theta}_2=\widehat{\theta}-\widehat{Q}(\widehat{\theta})c(\widehat{\theta}) \text{ and } \widehat{\theta}_3=\widehat{\theta}-Q(\widehat{\theta})\widehat{c}(\widehat{\theta}).$$

[3] The analytic forms of these variances are given in the appendix (see Appendix C).

From the previous results of (10) and (13), it is not difficult to see that

$$\sqrt{n}(\theta_1^* - \theta_0) = a_{-1/2} + \frac{1}{\sqrt{n}}(a_{-1} - B(\theta_0)) + \frac{1}{n}\left(a_{-3/2} - \nabla B(\theta_0)a_{-1/2}\right) - \frac{1}{n}QVB(\theta_0) + O_p(n^{-3/2})$$
$$\sqrt{n}(\widehat{\theta}_2 - \theta_0) = a_{-1/2} + \frac{1}{\sqrt{n}}(a_{-1} - B(\theta_0)) + \frac{1}{n}\left(a_{-3/2} - \nabla B(\theta_0)a_{-1/2}\right) - \frac{1}{n}\sqrt{n}(\widehat{Q}(\theta_0) - Q(\theta_0))c(\theta_0) + O_p(n^{-3/2})$$
$$\sqrt{n}(\widehat{\theta}_3 - \theta_0) = a_{-1/2} + \frac{1}{\sqrt{n}}(a_{-1} - B(\theta_0)) + \frac{1}{n}\left(a_{-3/2} - \nabla B(\theta_0)a_{-1/2}\right) - \frac{1}{n}\sqrt{n}Q(\theta_0)\left(\widehat{c}(\theta_0) - c(\theta_0)\right) + O_p(n^{-3/2}).$$

Now note that $\widehat{Q}(\theta_0) = Q(\theta_0) + \frac{1}{\sqrt{n}}Q(\theta_0)VQ(\theta_0) + O_p(\frac{1}{n})$ from (79) and hence

$$\sqrt{n}(\widehat{Q}(\theta_0) - Q(\theta_0))c(\theta_0) = Q(\theta_0)VQ(\theta_0)c(\theta_0) + O_p(1/n) = Q(\theta_0)VB(\theta_0) + O_p(1/n).$$

This implies that $\sqrt{n}(\widehat{\theta}_2 - \theta_0)$ and $\sqrt{n}(\theta_1^* - \theta_0)$ have the same asymptotic expansion up to $O_p\left(\frac{1}{n}\right)$ order. Now from Lemma B.1, we derive

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

which implies $\sqrt{n}(\theta^* - \theta_0) = \sqrt{n}(\widehat{\theta}_3 - \theta_0) + QVB(\theta_0) + O_p(n^{-3/2})$. To sum up, together with previous results we conclude

$$\sqrt{n}(\theta_1^* - \theta_0) = \sqrt{n}(\widehat{\theta}_b - \theta_0) - \frac{1}{n}QVB(\theta_0) + O_p(n^{-3/2})$$
$$\sqrt{n}(\theta_1^* - \theta_0) = \sqrt{n}(\widehat{\theta}_2 - \theta_0) + O_p(n^{-3/2})$$
$$\sqrt{n}(\theta^* - \theta_0) = \sqrt{n}(\widehat{\theta}_3 - \theta_0) + \frac{1}{n}QVB(\theta_0) + O_p(n^{-3/2})$$
$$\sqrt{n}(\theta^* - \theta_0) = \sqrt{n}(\widehat{\theta}_{bc} - \theta_0) + O_p(n^{-3/2}).$$

It illustrates that using $\widehat{Q}(\cdot)$ rather than $Q(\cdot)$ plays a critical role for equating the stochastic expansions (up to the third order) of the bias-corrected estimator and the estimator that solves the bias-corrected moment equation.

More interestingly we consider the iteration of the bias correction. Hahn and Newey (2004) discusses the relationship between the bias corrections of moment equations and the iterated bias correction. The iteration idea is that one can update $\widehat{B}$ several times using the previous estimator of $\widehat{\theta}$. To be more precise, denoting $\widehat{B}(\theta)$ as a function of θ, we can write the one step bias-corrected estimator as $\widehat{\theta}_{bc}^1 = \widehat{\theta} - \widehat{B}(\widehat{\theta})/n$. The k-th iteration will give us $\widehat{\theta}_{bc}^k = \widehat{\theta} - \widehat{B}(\widehat{\theta}_{bc}^{k-1})/n$ ($\widehat{\theta}_{bc}^1 = \widehat{\theta}_{bc}$) for $k = 2, 3, \ldots$. If we would iterate this procedure until achieving the convergence, we will obtain $\widehat{\theta}_{bc}^\infty = \widehat{\theta} - \widehat{B}(\widehat{\theta}_{bc}^\infty)/n$, which imply that $\widehat{\theta}_{bc}^\infty$ solves (note $\widehat{B}(\theta) = \widehat{Q}(\theta)\widehat{c}(\theta)$)

$$0 = \widehat{Q}(\theta)^{-1}(\widehat{\theta} - \theta) - \frac{1}{n}\widehat{c}(\theta) = \frac{1}{n}\sum_{i=1}^n s\left(z_i, \widehat{\theta}\right) + \widehat{Q}(\theta)^{-1}(\widehat{\theta} - \theta) - \frac{1}{n}\widehat{c}(\theta), \tag{16}$$

where the second equality is from the definition of $\widehat{\theta}$ in (2). Noting $\widehat{Q}(\theta)^{-1} = -\frac{1}{n}\sum_{i=1}^n \nabla s(z_i, \theta)$, if $s(z_i, \theta)$ is linear in θ then we find that (16) is the same as (8): the bias corrected moment equation and hence $\widehat{\theta}_{bc}^\infty$ is exactly same with θ^*. Otherwise (16) is an approximation of (8). From this we conclude that the fully iterated bias-corrected estimator $\widehat{\theta}_{bc}^\infty$ can be interpreted as the solution to an approximation of the bias-corrected moment equation (8). Similarly with (12), for $\widetilde{\widetilde{\theta}}$ between $\widehat{\theta}_{bc}^{\infty}$ and θ_0, we can show that

$$
\begin{aligned}
\sqrt{n}(\widehat{\theta}_{bc}^{\infty} - \theta_0) &= \sqrt{n}(\widehat{\theta} - \theta_0) - \tfrac{1}{\sqrt{n}}\widehat{B}(\widehat{\theta}_{bc}^{\infty}) \\
&= \begin{array}{l} a_{-1/2} + \frac{1}{\sqrt{n}}a_{-1} + \frac{1}{n}a_{-3/2} \\ -\frac{1}{\sqrt{n}}\widehat{B}(\theta_0) - \frac{1}{\sqrt{n}}\nabla\widehat{B}(\theta_0)(\widehat{\theta}_{bc}^{\infty} - \theta_0) - \frac{1}{2\sqrt{n}}\nabla^2\widehat{B}(\widetilde{\widetilde{\theta}})((\widehat{\theta}_{bc}^{\infty} - \theta_0)\otimes(\widehat{\theta}_{bc}^{\infty} - \theta_0)) + O_p(n^{-3/2}) \end{array} \\
&= a_{-1/2} + \tfrac{1}{\sqrt{n}}\left(a_{-1} - B(\theta_0)\right) + \tfrac{1}{n}\left(a_{-3/2} - \nabla B(\theta_0)a_{-1/2} - \sqrt{n}(\widehat{B}(\theta_0) - B(\theta_0))\right) + O_p(n^{-3/2})
\end{aligned}
$$

using Condition 5, 6, and 7 and $\sqrt{n}\left(\widehat{\theta}_{bc}^{\infty} - \theta_0\right) = QJ + O_p\left(1/\sqrt{n}\right)$. This result confirms that $\sqrt{n}(\widehat{\theta}_{bc}^{\infty} - \widehat{\theta}_{bc}) = O_p(n^{-3/2})$, which actually holds for all $\widehat{\theta}_{bc}^{k}$ ($k = 2, 3, \ldots$).

Noting this equivalence of the higher order expansions for $\widehat{\theta}_{bc}^{\infty}$ and $\widehat{\theta}_{bc}^{1}$ at least up to the third order term, one would expect that the higher order expansion of θ^{*} will be equivalent to that of $\widehat{\theta}_{bc}^{1}$ at least up to the third order and we confirm this intuition in this paper. However, as observed in some Monte Carlo examples of Hahn and Newey (2004) and Ferández-Val (2004), the iterative bias correction can lower bias for small samples and so can the bias correction of the moment equations. This suggests that the comparison between the one-step bias correction and the method of correcting the moment equation (or the fully iterated bias correction) should be based on the stochastic expansions higher than the third order. As a related estimator, Hahn and Newey (2004) discusses the asymptotic equivalence of the bias-corrected moment equation method to Woutersen's (2002) approach and hence we conjecture that Woutersen's (2002) estimator will not improve over the simple one-step bias correction either at least in the third order stochastic expansion sense.

5 Conclusion

This paper considers an alternative bias correction for the M-estimator, which is achieved by correcting the moment equation in the spirit of Firth (1993). In particular, this paper compares the stochastic expansions of the analytically bias-corrected estimator (which is referred to one-step bias correction) and the alternative estimator and finds that the third-order stochastic expansions of these two estimators are identical. This implies that these two estimators do not only have the same higher order variances but also agree upon more properties in terms of their stochastic expansions.

We conclude that at least in terms of the third order stochastic expansion, we cannot improve on the simple one-step bias-correction by using the bias correction of the moment equations. Though our result is for the fixed number of parameters, we conjecture this is also true for the panel data models with individual specific parameters. The intuition is that the fully iterated bias-corrected estimator can be interpreted as the solution of an approximation to the bias corrected moment equations and the iteration will not improve asymptotic properties in general and neither will the alternative estimator. We have verified this intuition in this paper. Noting the M-estimation framework is quite general, this is a rather strong result.

Appendix

A Technical Lemmas and Proofs

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)* Θ *is compact.*
Then, $E[\|m(z_i, \theta)\|]$ *is continuous for all* $\theta \in \Theta$ *and* $\sup_{\theta\in\Theta} \left\|\frac{1}{n}\sum_{i=1}^{n} m(z_i, \theta) - E[m(z_i, \theta)]\right\| = o_p(1)$.

Proof. This result is implied by Lemma 1 of Tauchen (1985) or can be verified by showing the stochastic equicontinuity of $\left\{\frac{1}{n}\sum_{i=1}^{n}(m(z_i, \theta) - E[m(z_i, \theta)]) : n \geq 1\right\}$ for $\theta \in \Theta$ as in Newey (1991) observing that $E\left[\sup_{\theta\in\Theta} \|m(z_i, \theta)\|\right] < \infty$ is stronger than the Lipschitz condition used in Newey (1991). The continuity of $E[\|m(z_i, \theta)\|]$ 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* Θ *if* $\forall\varepsilon > 0$ $\exists\delta > 0$ *such that*

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

Now define $M_n(\theta) = \frac{1}{n}\sum_{i=1}^{n} m(z_i, \theta) - E[m(z_i, \theta)]$ and $Y_{i\delta} = \sup_{\theta\in\Theta} \sup_{\theta'\in B(\theta,\delta)} \|m(z_i, \theta') - m(z_i, \theta)\|$. Note $E[Y_{i\delta}] \leq 2E\left[\sup_{\theta\in\Theta} \|m(z_i, \theta)\|\right] < \infty$ by Condition (iii). We claim that $E[Y_{i\delta}] \to 0$ 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

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

where the first inequality follows by Triangle inequality, the second holds by Markov inequality, and the last equality holds by $E[Y_{i\delta}] \to 0$ 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 Θ 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* θ *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)* Θ *is bounded.*
Then, $\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(m(z_i, \theta) - E[m(z_i, \theta)])$ *is stochastically equicontinuous and thus*
$\sup_{\theta\in\Theta} \left\|\frac{1}{n}\sum_{i=1}^{n} m(z_i, \theta) - E[m(z_i, \theta)]\right\| = o_p(1)$.

Proof. From condition (ii), we note that $m(\cdot, \cdot)$ belongs to Type II class in Andrews (1994) 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), which is Assumption A for Theorem 1 in Andrews (1994). Condition (iii) implies Assumption B of Theorem 1 in Andrews (1994). Condition (i) is stronger than Assumption C for Theorem 1 in Andrews (1994) and hence stochastic equicontinuity follows. Now noting Condition (iii) is sufficient for pointwise weak convergences of $\frac{1}{n}\sum_{i=1}^{n} m(z_i, \theta)$ to $E[m(z_i, \theta)]$ for all $\theta \in \Theta$ and combining this with the stochastic equicontinuity result, we have the uniform convergence as assuming Θ is bounded. To be more precise, first, note that the stochastic equicontinuity of $\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(m(z_i, \theta) - E[m(z_i, \theta)])$ implies the stochastic equicontinuity of $\frac{1}{n}\sum_{i=1}^{n}(m(z_i, \theta) - E[m(z_i, \theta)])$. Now define $v_n(\theta) = \frac{1}{n}\sum_{i=1}^{n}(m(z_i, \theta) - E[m(z_i, \theta)])$ and let $\varepsilon > 0$ and take a δ 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 a δ exists by the definition of the stochastic equicontinuity. Now note that from the boundedness of Θ, we can construct a finite cover of Θ 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 2.1. The first possible set of primitive conditions is

Assumption A.1 *(i) $\{z_i\}_{i=1}^n$ are iid; (ii) $s(z, \theta)$ is κ-times continuously differentiable in a neighborhood of θ_0, denoted by Θ_0 for all $z \in \mathcal{Z}$, $\kappa \geq 3$ with probability one; (iii) $E\left[\sup_{\theta\in\Theta_0} \|\nabla^\upsilon s(z, \theta)\|\right] < \infty$, $\upsilon = \{0, 1, 2, \ldots \kappa\}$, $\kappa \geq 3$; (iv) Θ is compact; (v) θ_0 is in the interior of Θ and is the only θ satisfying (1).*

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

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) $\nabla^\upsilon s(z, \theta)$ satisfies the Lipschitz condition in θ as*

$$\|\nabla^\upsilon s(z, \theta_1) - \nabla^\upsilon s(z, \theta_2)\| \leq B_\upsilon(z)\, \|\theta_1 - \theta_2\| \quad \forall \theta_1, \theta_2 \in \Theta_0$$

for some function $B_\upsilon(\cdot)\colon \mathcal{Z} \to R$ and $E\left[B_\upsilon(\cdot)^{2+\delta}\right] < \infty$, $\upsilon = \{0, 1, 2, \ldots \kappa\}$ in a neighborhood of θ_0, denoted by Θ_0 for all $z \in \mathcal{Z}$, $\kappa \geq 3$ with probability one; (iii) $E\left[\sup_{\theta\in\Theta_0} \|\nabla^\upsilon s(z, \theta)\|^{2+\delta}\right] < \infty$ for $\exists\ \delta > 0$, $\upsilon = \{0, 1, 2, \ldots \kappa\}$, $\kappa \geq 3$; (iv) Θ is bounded; (v) θ_0 is in the interior of Θ and is the only θ satisfying (1).

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

Proof. Consider

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

We have $\sup_{\theta\in\Theta_0} ||\frac{1}{n}\sum_{i=1}^n \nabla^\upsilon s\left(z_i, \theta\right) - E\left[\nabla^\upsilon s\left(z_i, \theta\right)\right]|| = o_p(1)$ from Lemma A.1 by letting $m(z, \theta) = \nabla^\upsilon s\left(z, \theta\right)$ and noting Assumption A.1 satisfies all the conditions in Lemma A.1 for $\theta \in \Theta_0$. The continuity of $E\left[\nabla^\upsilon s\left(z_i, \theta\right)\right]$ at θ_0 (by the Dominated Convergence theorem with the dominating function $\sup_{\theta\in\Theta_0} \|\nabla^\upsilon s(z_i, \theta)\|$) implies that $\left\|E\left[\nabla^\upsilon s\left(z_i, \overline{\theta}\right)\right] - E\left[\nabla^\upsilon s\left(z_i, \theta_0\right)\right]\right\| = o_p(1)$, since $\overline{\theta} = \theta_0 + o_p(1)$ and hence from this and the result above, it follows that $\left\|\frac{1}{n}\sum_{i=1}^n \nabla^\upsilon s\left(z_i, \overline{\theta}\right) - E\left[\nabla^\upsilon s\left(z_i, \theta_0\right)\right]\right\| = o_p(1)$. ■

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

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 $E\left[\nabla^{\upsilon} s\left(z_i, \theta\right)\right]$ for $\theta \in \Theta_0$ and hence the result follows. ■

Now we show that conditions (i)-(viii) in Lemma 2.1 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

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

in the proof of Lemma 2.1. Alternatively, it can be shown as

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

where $\widetilde{\widetilde{\theta}}$ $(\widetilde{\theta})$ lies between $\widetilde{\theta}$ $(\widehat{\theta})$ and θ_0 noting $\widehat{\theta}-\theta_0 = O_p\left(\frac{1}{\sqrt{n}}\right)$. 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 $\left\|\frac{1}{n}\sum_{i=1}^n \nabla^3 s\left(z_i,\overline{\theta}\right) - E\left[\nabla^3 s\left(z_i,\theta_0\right)\right]\right\| = o_p(1)$ for $\overline{\theta} = \theta_0 + o_p(1)$ under Assumption A.1. Condition (v) is assumed in Assumption A.3. Condition (vi)-(viii) are by CLT provided that $E\left[\left\|\nabla^{\upsilon} s(z,\theta_0)\right\|^2\right] < \infty$, $\upsilon = \{0,1,2\}$ respectively, which are satisfied under Assumption A.2.

Now to establish additional preliminary lemmas, we need a stronger set of conditions as Assumption 3.1-3.2 or 3.3-3.2. Note that Assumption 3.1-3.2 implies Assumption A.1-A.3 and Assumption A.4 is weaker than Assumption 3.3. First, under Assumption 3.1 or 3.3, we have the uniform weak convergences (U-WCON) for the normalized sums of functions in $\nabla^{\upsilon} s(z,\theta), \upsilon = \{0,1,2,\ldots\kappa\}$ up to the second order as in a neighborhood of θ_0, denoted by Θ_0 and hence it is not difficult to show that

Lemma A.5 *Under Assumption 3.1 or 3.3, we have*

$$\sup_{\theta\in\Theta}\left|\frac{1}{n}\sum_{i=1}^n \left\|\nabla^{\upsilon_1} s\left(z_i,\theta\right)\right\|\left\|\nabla^{\upsilon_2} s\left(z_i,\theta\right)\right\| - E\left[\left\|\nabla^{\upsilon_1} s\left(z_i,\theta\right)\right\|\left\|\nabla^{\upsilon_2} s\left(z_i,\theta\right)\right\|\right]\right| = o_p(1), \tag{17}$$

for $\upsilon_1, \upsilon_2 \in \{0,1,2,\ldots,\kappa\}, \kappa \geq 4$.

Proof. Provided Assumption 3.1 holds, (17) is obtained by applying the Uniform Convergence theorem of Lemma A.1 by letting $m(z,\theta) = \left\|\nabla^{\upsilon_1} s\left(z_i,\theta\right)\right\|\left\|\nabla^{\upsilon_2} s\left(z_i,\theta\right)\right\|$. Noting

$$\begin{aligned}&E\left[\sup_{\theta\in\Theta_0}\left\|\nabla^{\upsilon_1} s\left(z_i,\theta\right)\right\|\left\|\nabla^{\upsilon_2} s\left(z_i,\theta\right)\right\|\right] \leq E\left[\sup_{\theta\in\Theta_0}\tfrac{\left\|\nabla^{\upsilon_1} s(z_i,\theta)\right\|^2 + \left\|\nabla^{\upsilon_2} s(z_i,\theta)\right\|^2}{2}\right] \\ \leq &E\left[\sup_{\theta\in\Theta_0}\left\|\nabla^{\upsilon_1} s\left(z_i,\theta\right)\right\|^2\right]/2 + E\left[\sup_{\theta\in\Theta_0}\left\|\nabla^{\upsilon_2} s\left(z_i,\theta\right)\right\|^2\right]/2 < \infty,\end{aligned} \tag{18}$$

which is satisfied by Assumption 3.1 (iii), all the conditions for Lemma A.1 are trivially satisfied.

Alternatively under Assumption 3.3, we obtain (17) directly from Theorem 1-3 in Andrews (1994), 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 3.3, the uniform convergence result of (17) is obtained upon verifying the Lipschitz condition for $\forall\theta_1,\theta_2 \in \Theta$ as

$$\begin{aligned}&\left|\left\|\nabla^{\upsilon_1} s\left(z,\theta_1\right)\right\|\left\|\nabla^{\upsilon_2} s(z,\theta_1)\right\| - \left\|\nabla^{\upsilon_1} s\left(z,\theta_2\right)\right\|\left\|\nabla^{\upsilon_2} s(z,\theta_2)\right\|\right| \\ \leq &\left|\left\|\nabla^{\upsilon_1} s\left(z,\theta_1\right)\right\|\left\|\nabla^{\upsilon_2} s(z,\theta_1)\right\| - \left\|\nabla^{\upsilon_1} s\left(z,\theta_1\right)\right\|\left\|\nabla^{\upsilon_2} s(z,\theta_2)\right\|\right| \\ &+ \left|\left\|\nabla^{\upsilon_1} s\left(z,\theta_1\right)\right\|\left\|\nabla^{\upsilon_2} s(z,\theta_2)\right\| - \left\|\nabla^{\upsilon_1} s\left(z,\theta_2\right)\right\|\left\|\nabla^{\upsilon_2} s(z,\theta_2)\right\|\right| \\ = &\left\|\nabla^{\upsilon_1} s\left(z,\theta_1\right)\right\|\left|\left\|\nabla^{\upsilon_2} s(z,\theta_1)\right\| - \left\|\nabla^{\upsilon_2} s(z,\theta_2)\right\|\right| + \left\|\nabla^{\upsilon_2} s\left(z,\theta_2\right)\right\|\left|\left\|\nabla^{\upsilon_1} s(z,\theta_1)\right\| - \left\|\nabla^{\upsilon_1} s(z,\theta_2)\right\|\right| \\ \leq &\left\|\nabla^{\upsilon_1} s\left(z,\theta_1\right)\right\| B_{\upsilon_2}(z)\left\|\theta_1-\theta_2\right\| + \left\|\nabla^{\upsilon_2} s\left(z,\theta_2\right)\right\| B_{\upsilon_1}(z)\left\|\theta_1-\theta_2\right\| \\ \leq &\sup_{\theta\in\Theta}\left\|\nabla^{\upsilon_1} s\left(z,\theta\right)\right\| B_{\upsilon_2}(z)\left\|\theta_1-\theta_2\right\| + \sup_{\theta\in\Theta}\left\|\nabla^{\upsilon_2} s\left(z,\theta\right)\right\| B_{\upsilon_1}(z)\left\|\theta_1-\theta_2\right\| \\ = &\left(\sup_{\theta\in\Theta}\left\|\nabla^{\upsilon_1} s\left(z,\theta\right)\right\| B_{\upsilon_2}(z) + \sup_{\theta\in\Theta}\left\|\nabla^{\upsilon_2} s\left(z,\theta\right)\right\| B_{\upsilon_1}(z)\right)\left\|\theta_1-\theta_2\right\| \equiv M(z)\left\|\theta_1-\theta_2\right\|,\end{aligned}$$

where the first inequality is by Triangle Inequality and the second inequality is obtained by the Lipschitz conditions for $\nabla^{\upsilon_1} s(z,\theta)$ and $\nabla^{\upsilon_2} s(z,\theta)$, since for $\upsilon = \upsilon_1, \upsilon_2$, $|\|\nabla^{\upsilon} s(z,\theta_1)\| - \|\nabla^{\upsilon} s(z,\theta_2)\|| \leq \|\nabla^{\upsilon} s(z,\theta_1) - \nabla^{\upsilon} s(z,\theta_2)\|$ by Triangle Inequality. Now we need to verify that $E\left[M(z)^{2+\delta}\right] < \infty$, which is true, since

$$E\left[\sup_{\theta\in\Theta} \|\nabla^{\upsilon_a} s(z,\theta)\|^{2+\delta} B_{\upsilon_b}(z)^{2+\delta}\right] \leq E\left[\left(\sup_{\theta\in\Theta} \|\nabla^{\upsilon_a} s(z,\theta)\|^{4+2\delta} + B_{\upsilon_b}(z)^{4+2\delta}\right)/2\right] < \infty \tag{19}$$

for $(\upsilon_a, \upsilon_b) \in \{(\upsilon_1,\upsilon_2),(\upsilon_2,\upsilon_1)\}$ under $E\left[\sup_{\theta\in\Theta} \|\nabla^{\upsilon_1} s(z,\theta)\|^{4+\delta'}\right] < \infty$, $E\left[\sup_{\theta\in\Theta} \|\nabla^{\upsilon_2} s(z,\theta)\|^{4+\delta'}\right] < \infty$, $E\left[B_{\upsilon_1}(z)^{4+\delta'}\right] < \infty$, and $E\left[B_{\upsilon_2}(z)^{4+\delta'}\right] < \infty$ with $\delta' = 2\delta$. ■

Lemma A.6 *(Consistency of θ^*) Suppose θ_0 is the unique solution of (1) and θ^* solves (8) 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 θ^* is a consistent estimator of θ_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 θ_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 3.1-3.2 or Assumption 3.3-3.2 hold for the whole parameter space Θ instead of Θ_0 similarly with Lemma A.7. ■

Lemma A.7 *Under Assumption 3.1-3.2 or 3.3-3.2, (a) we have*

$$\widehat{c}(\theta) = c(\theta) + o_p(1) \tag{20}$$

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 3.1 (ii)-(iii) and Assumption 3.2. This is evident, 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. It is worthwhile to remark that $\widehat{c}(\theta) = c(\theta) + o_p(1)$ is not necessary to show Condition 1 and hence not necessary in proving Proposition 3.1. However, nonetheless we present this result, since Assumption 3.1-3.2 or 3.3-3.2 that are sufficient for Proposition 3.1 imply (20) and it is useful to show $\widehat{c}(\theta_0) = c(\theta_0) + O_p\left(\frac{1}{\sqrt{n}}\right)$. 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

$$\begin{aligned}
&\widehat{c}(\theta) - c(\theta)\\
=\;&\left(\frac{1}{2}\widehat{H}_2(\theta)\left(\frac{1}{n}\textstyle\sum_{i=1}^n\left[\widehat{Q}(\theta)s(z_i,\theta)\otimes\widehat{Q}(\theta)s(z_i,\theta)\right]\right) - \frac{1}{2}H_2(\theta)\left(E\left[Q(\theta)s(z_i,\theta)\otimes Q(\theta)s(z_i,\theta)\right]\right)\right) && (21)\\
&+\left(\frac{1}{n}\textstyle\sum_{i=1}^n\left[\nabla s(z_i,\theta)\widehat{Q}(\theta)s(z_i,\theta)\right] - E\left[\nabla s(z_i,\theta)Q(\theta)s(z_i,\theta)\right]\right) && (22)
\end{aligned}$$

Now rewrite (22) as

$$\begin{aligned}
&\frac{1}{n}\textstyle\sum_{i=1}^n\left[\nabla s(z_i,\theta)\widehat{Q}(\theta)s(z_i,\theta)\right] - E\left[\nabla s(z_i,\theta)Q(\theta)s(z_i,\theta)\right]\\
=\;&\frac{1}{n}\textstyle\sum_{i=1}^n\left[\nabla s(z_i,\theta)\widehat{Q}(\theta)s(z_i,\theta) - \nabla s(z_i,\theta)Q(\theta)s(z_i,\theta)\right] && (23)\\
&+\frac{1}{n}\textstyle\sum_{i=1}^n\left[\nabla s(z_i,\theta)Q(\theta)s(z_i,\theta) - E\left[\nabla s(z_i,\theta)Q(\theta)s(z_i,\theta)\right]\right]. && (24)
\end{aligned}$$

Then we have for (23),

$$\left\| \frac{1}{n} \sum_{i=1}^{n} \left[\nabla s(z_i,\theta) \left(\widehat{Q}(\theta) - Q(\theta) \right) s(z_i,\theta) \right] \right\| \le \frac{1}{n} \sum_{i=1}^{n} \| \nabla s(z_i,\theta) \| \| s(z_i,\theta) \| \left\| \widehat{Q}(\theta) - Q(\theta) \right\| \tag{25}$$

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 3.2, with probability approaching to one, $\widehat{H}_1(\theta)$ is nonsingular for $\theta \in \Theta_0$. Now note

$$\begin{aligned} & \left\| \widehat{Q}(\theta) - Q(\theta) \right\| = \left\| Q(\theta) \left(\widehat{Q}(\theta)^{-1} - Q(\theta)^{-1} \right) \widehat{Q}(\theta) \right\| \\ & \le \| Q(\theta) \| \left\| \widehat{Q}(\theta) \right\| \left\| \widehat{Q}(\theta)^{-1} - Q(\theta)^{-1} \right\| \le C O_p(1) o_p(1) = o_p(1) \end{aligned} \tag{26}$$

by the uniform convergence of $\widehat{Q}(\theta)^{-1}$ to $Q(\theta)^{-1}$ and Assumption 3.2 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 (18) and Lemma A.5 with $v_1 = 1$ and $v_2 = 0$. Together with (26), this implies (23) is $o_p(1)$. Now note we have

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

from (18) and $\sup_{\theta \in \Theta_0} \| Q(\theta) \| < \infty$ or Lipschitz condition as

$$\begin{aligned} & \| \nabla s(z,\theta_1) Q(\theta_1) s(z,\theta_1) - \nabla s(z,\theta_2) Q(\theta_2) s(z,\theta_2) \| \\ & \le \sup_{\theta \in \Theta_0} \| Q(\theta) s(z,\theta) \| \| \nabla s(z,\theta_1) - \nabla s(z,\theta_2) \| \\ & \quad + \sup_{\theta \in \Theta_0} \| Q(\theta) \nabla s(z,\theta) \| \| s(z,\theta_1) - s(z,\theta_2) \| + \sup_{\theta \in \Theta_0} \| s(z,\theta) \nabla s(z,\theta) \| \| Q(\theta_1) - Q(\theta_2) \| \\ & \le \sup_{\theta \in \Theta_0} \| Q(\theta) \| \sup_{\theta \in \Theta_0} \| s(z,\theta) \| B_1(z) \| \theta_1 - \theta_2 \| + \sup_{\theta \in \Theta_0} \| Q(\theta) \| \sup_{\theta \in \Theta_0} \| \nabla s(z,\theta) \| B_0(z) \| \theta_1 - \theta_2 \| \\ & \quad + \sup_{\theta \in \Theta_0} \| s(z,\theta) \| \sup_{\theta \in \Theta_0} \| \nabla s(z,\theta) \| \left(\sup_{\theta \in \Theta_0} \| Q(\theta) \| \right)^2 \sup_{\theta \in \Theta_0} \| \nabla H_1(\theta) \| \| \theta_1 - \theta_2 \| \\ & \equiv \overline{M}(z) \| \theta_1 - \theta_2 \|, \end{aligned} \tag{27}$$

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 (19) 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 (24) and have

$$\sup_{\theta \in \Theta_0} \left\| \frac{1}{n} \sum_{i=1}^{n} \left[\nabla s(z_i,\theta) Q(\theta) s(z_i,\theta) - E \left[\nabla s(z_i,\theta) Q(\theta) s(z_i,\theta) \right] \right] \right\| = o_p(1). \tag{28}$$

From (23)=$o_p(1)$ and (28), we conclude (22) is $o_p(1)$ uniformly over $\theta \in \Theta_0$. Now consider

$$\sup_{\theta \in \Theta_0} \left\| \widehat{H}_2(\theta) - H_2(\theta) \right\| = o_p(1) \tag{29}$$

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

$$\left\| \frac{1}{n} \sum_{i=1}^{n} s(z_i,\theta) s(z_i,\theta)' - E \left[s(z_i,\theta) s(z_i,\theta)' \right] \right\| = o_p(1) \tag{30}$$

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

$$\begin{aligned} & \left\|s(z,\theta_1)s(z,\theta_1)' - s(z,\theta_2)\, s(z,\theta_2)'\right\| \\ \leq\ & \left\|s(z,\theta_1)s(z,\theta_1)' - s(z,\theta_1)\, s(z,\theta_2)'\right\| + \left\|s(z,\theta_1)s(z,\theta_2)' - s(z,\theta_2)\, s(z,\theta_2)'\right\| \\ \leq\ & 2\sup_{\theta\in\Theta} \|s(z,\theta)\| \, \|s(z,\theta_1) - s(z,\theta_2)\| \leq 2\sup_{\theta\in\Theta} \|s(z,\theta)\| \, B_0(z) \, \|\theta_1 - \theta_2\| \end{aligned} \tag{31}$$

by Triangle Inequality and noting $E\left[\sup_{\theta\in\Theta} \|s(z,\theta)\|^{2+\delta} B_0(z)^{2+\delta}\right] < \infty$ 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 (26), (29), and (30), it follows that

$$\begin{aligned} & \widehat{H}_2(\theta)\left(\tfrac{1}{n}\textstyle\sum_{i=1}^n \left[\widehat{Q}(\theta)s(z_i,\theta) \otimes \widehat{Q}(\theta)s(z_i,\theta)\right]\right) \\ &= \widehat{H}_2(\theta)\left(vec\left(\widehat{Q}(\theta)\left(\tfrac{1}{n}\textstyle\sum_{i=1}^n s(z_i,\theta)\, s(z_i,\theta)'\right)\widehat{Q}(\theta)'\right)\right) \\ &= (H_2(\theta) + o_p(1))\left(vec\left((Q(\theta) + o_p(1))\left(E\left[s(z_i,\theta)\, s(z_i,\theta)'\right] + o_p(1)\right)(Q(\theta) + o_p(1))'\right)\right) \\ &= H_2(\theta)\left(vec\left(Q(\theta)\left(E\left[s(z_i,\theta)\, s(z_i,\theta)'\right]\right)Q(\theta)'\right)\right) + o_p(1) \\ &= H_2(\theta)\left(E\left[Q(\theta)s(z_i,\theta) \otimes Q(\theta)s(z_i,\theta)\right]\right) + o_p(1), \end{aligned} \tag{32}$$

where the first and the last equality come from $vec(gg') = g \otimes g$ for a column vector g and hence we bound (21) 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 (ii)

Note

$$\sqrt{n}\left(\widehat{Q}(\theta_0) - Q(\theta_0)\right) = O_p(1) \tag{33}$$

by the Slutsky theorem and that

$$\sqrt{n}\left(\widehat{H}_2(\theta_0) - H_2(\theta_0)\right) = \frac{1}{\sqrt{n}}\textstyle\sum_{i=1}^n \left(\nabla^2 s(z_i,\theta_0) - E\left[\nabla^2 s(z_i,\theta_0)\right]\right) = O_p(1) \tag{34}$$

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

$$\frac{1}{\sqrt{n}}\textstyle\sum_{i=1}^n s(z_i,\theta_0)s(z_i,\theta_0)' - E\left[s(z_i,\theta_0)s(z_i,\theta_0)'\right] = O_p(1) \tag{35}$$

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

$$\frac{1}{\sqrt{n}}\textstyle\sum_{i=1}^n \|\nabla s(z_i,\theta_0)\| \, \|s(z_i,\theta_0)\| = E\left[\|\nabla s(z_i,\theta_0)\| \, \|s(z_i,\theta_0)\|\right] + O_p(1) \tag{36}$$

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

$$\frac{1}{\sqrt{n}}\textstyle\sum_{i=1}^n \left[\nabla s(z_i,\theta_0)\, Q(\theta_0)s(z_i,\theta_0) - E\left[\nabla s(z_i,\theta_0)\, Q(\theta_0)s(z_i,\theta_0)\right]\right] = O_p(1) \tag{37}$$

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

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

by Cauchy-Schwarz Inequality. Applying the results of (33), (34), and (35) to (32), we can show that (21)=$O_p(1/\sqrt{n})$ for $\theta = \theta_0$. Similarly, plugging the results of (33), (36), and (37) into (24) and (25), we obtain (22)=$O_p(1/\sqrt{n})$ for $\theta = \theta_0$ and hence we conclude that $\widehat{c}(\theta_0) = c(\theta_0) + O_p(1/\sqrt{n})$. ■

To characterize $\nabla\widehat{c}(\theta)$, 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 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

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

■

Remark 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

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

■

Remark 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

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

■

Remark 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

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

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

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

■

Remark 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 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 $\nabla\left(A^{-1}\right) = -A^{-1}(A')^{-1}\nabla(A')$. ■

Lemma A.8 *Under Assumption 3.1-3.2 or 3.3-3.2, we have (a)* $\left\|\nabla^{\upsilon}\left(\widehat{Q}(\theta)'\right)\right\| = \left\|\nabla^{\upsilon}\widehat{Q}(\theta)\right\| = O_p(1)$ *and (b)* $\nabla^{\upsilon-1}\left(\widehat{Q}(\theta_0)'\right) - \nabla^{\upsilon-1}\left(Q(\theta_0)'\right) = O_p\left(1/\sqrt{n}\right)$ *for* $\theta\in\Theta_0$ *and* $\upsilon = \{1,2,3\}$

Proof. For $\theta\in\Theta_0$, note Remark 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)

$$\nabla\left(\widehat{Q}(\theta)'\right) = -\nabla\left(\left(\widehat{H}_1(\theta)'\right)^{-1}\right) = \left(\widehat{H}_1(\theta)'\right)^{-1}\widehat{H}_1(\theta)^{-1}\nabla\widehat{H}_1(\theta) = \widehat{Q}(\theta)'\widehat{Q}(\theta)\widehat{H}_2(\theta) \tag{38}$$

and hence $||\nabla(\widehat{Q}(\theta)')|| \le ||\widehat{Q}(\theta)||^2||\widehat{H}_2(\theta)|| = O_p(1)$ by (26) and (29). Now consider

$$\begin{aligned} &\left\|\nabla^2\left(\widehat{Q}(\theta)'\right)\right\| = \left\|\nabla\left(\widehat{Q}(\theta)'\widehat{Q}(\theta)\widehat{H}_2(\theta)\right)\right\| \le \left\|\nabla\left(\widehat{Q}(\theta)'\widehat{Q}(\theta)\right)\right\|\left\|\widehat{H}_2(\theta)\right\| + \left\|\widehat{Q}(\theta)'\widehat{Q}(\theta)\right\|\left\|\nabla\widehat{H}_2(\theta)\right\| \\ &\le 2\left\|\widehat{Q}(\theta)'\right\|\left\|\nabla\widehat{Q}(\theta)\right\|\left\|\widehat{H}_2(\theta)\right\| + \left\|\widehat{Q}(\theta)'\widehat{Q}(\theta)\right\|\left\|\nabla\widehat{H}_2(\theta)\right\| \\ &\le 2\left\|\widehat{Q}(\theta)\right\|\left\|\nabla\widehat{Q}(\theta)\right\|\left\|\widehat{H}_2(\theta)\right\| + \left\|\widehat{Q}(\theta)\right\|^2\left\|\nabla\widehat{H}_2(\theta)\right\| = O_p(1), \end{aligned} \tag{39}$$

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

$$\left\|\nabla\widehat{H}_2(\theta)\right\| = \left\|\frac{1}{n}\sum_{i=1}^n \nabla^3 s(z_i,\theta)\right\| = \left\|E\left[\nabla^3 s(z_i,\theta)\right]\right\| + o_p(1) = O_p(1) \tag{40}$$

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

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

from (26), (38), (39), 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 (26), (38), and (39). Now to show the second result, first note that we can rewrite

$$\widehat{Q}(\theta_0) = \left(-H_1 - V/\sqrt{n}\right)^{-1} = Q + O_p\left(1/\sqrt{n}\right) \tag{41}$$

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

$$\widehat{H}_2(\theta_0) = H_2 + W/\sqrt{n} = H_2 + O_p\left(1/\sqrt{n}\right), \tag{42}$$

since $W = O_p(1)$ by CLT. From (38), consider

$$\begin{aligned}\nabla\left(\widehat{Q}(\theta_0)'\right) &= \widehat{Q}(\theta_0)'\widehat{Q}(\theta_0)\widehat{H}_2(\theta_0) \\ &= \left(Q + O_p\left(1/\sqrt{n}\right)\right)'\left(Q + O_p\left(1/\sqrt{n}\right)\right)\left(H_2 + O_p\left(1/\sqrt{n}\right)\right) \\ &= Q'QH_2 + O_p\left(1/\sqrt{n}\right) = \nabla\left(Q(\theta_0)'\right) + O_p\left(1/\sqrt{n}\right)\end{aligned} \tag{43}$$

using (41) and (42). Similarly from (39), 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 (41) and (42)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 3.1 -3.2 or 3.3 -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 $\theta \in \Theta_0$

$$\begin{aligned}&\nabla\widehat{c}(\theta) \\ &= \nabla\left(\tfrac{1}{2}\widehat{H}_2(\theta)\left(\tfrac{1}{n}\sum_{i=1}^n\left[\widehat{Q}(\theta)s(z_i,\theta)\otimes\widehat{Q}(\theta)s(z_i,\theta)\right]\right) + \tfrac{1}{n}\sum_{i=1}^n\left[\nabla s(z_i,\theta)\widehat{Q}(\theta)s(z_i,\theta)\right]\right) \\ &= \nabla\left(\tfrac{1}{2}\widehat{H}_2(\theta)\left(\tfrac{1}{n}\sum_{i=1}^n\left[\widehat{Q}(\theta)s(z_i,\theta)\otimes\widehat{Q}(\theta)s(z_i,\theta)\right]\right)\right) + \nabla\left(\tfrac{1}{n}\sum_{i=1}^n\left[\nabla s(z_i,\theta)\widehat{Q}(\theta)s(z_i,\theta)\right]\right) \\ &= \tfrac{1}{2}vec^*\left(\left(\tfrac{1}{n}\sum_{i=1}^n\left[\widehat{Q}(\theta)s(z_i,\theta)\otimes\widehat{Q}(\theta)s(z_i,\theta)\right]\right)'\left(\nabla\left(\widehat{H}_2(\theta)'\right)\right)\right) \\ &\quad + \tfrac{1}{2}\widehat{H}_2(\theta)\nabla\left(\tfrac{1}{n}\sum_{i=1}^n\left[\widehat{Q}(\theta)s(z_i,\theta)\otimes\widehat{Q}(\theta)s(z_i,\theta)\right]\right) + \nabla\left(\tfrac{1}{n}\sum_{i=1}^n\left[\nabla s(z_i,\theta)\widehat{Q}(\theta)s(z_i,\theta)\right]\right),\end{aligned} \tag{44}$$

using Remark 2. For the first RHS term of the last equality in (44), note

$$\begin{aligned}&\left\|\nabla\left(\widehat{H}_2(\theta)'\right)\right\| = \left\|\tfrac{1}{n}\sum_{i=1}^n\nabla\left(\left(\nabla^2 s(z_i,\theta)\right)'\right)\right\| \\ &\leq \tfrac{1}{n}\sum_{i=1}^n\left\|\nabla\left(\left(\nabla^2 s(z_i,\theta)\right)'\right)\right\| = \tfrac{1}{n}\sum_{i=1}^n\left\|\nabla^3 s(z_i,\theta)\right\| = E\left[\left\|\nabla^3 s(z_i,\theta)\right\|\right] + o_p(1) = O_p(1)\end{aligned} \tag{45}$$

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

$$\left\|\frac{1}{n}\sum_{i=1}^n\left[\widehat{Q}(\theta)s(z_i,\theta)(\theta)\otimes\widehat{Q}(\theta)s(z_i,\theta)(\theta)\right] - Q(\theta)E\left[s(z_i,\theta)s(z_i,\theta)'\right]Q(\theta)'\right\| = o_p(1) \tag{46}$$

uniformly over $\theta \in \Theta_0$ in (32) and from this result with (45), we bound the first RHS term of the last equality in (44) as

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

Now consider

$$\nabla\left(\frac{1}{n}\sum_{i=1}^n\left[\widehat{Q}(\theta)s(z_i,\theta)\otimes\widehat{Q}(\theta)s(z_i,\theta)\right]\right)=\frac{1}{n}\sum_{i=1}^n\nabla\left[\widehat{Q}(\theta)s(z_i,\theta)\otimes\widehat{Q}(\theta)s(z_i,\theta)\right] \tag{47}$$

$$= \frac{1}{n}\sum_{i=1}^n\left(\nabla\left(\widehat{Q}(\theta)s(z_i,\theta)\right)\otimes^*\widehat{Q}(\theta)s(z_i,\theta)\right) \tag{48}$$

$$+\frac{1}{n}\sum_{i=1}^n\left(\widehat{Q}(\theta)s(z_i,\theta)\otimes\nabla\left(\widehat{Q}(\theta)s(z_i,\theta)\right)\right) \tag{49}$$

from Remark 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 2, we rewrite (48) as

$$\frac{1}{n}\sum_{i=1}^n\left(\widehat{Q}(\theta)\nabla s(z_i,\theta)+vec^*\left(s(z_i,\theta)'\nabla\left(\widehat{Q}(\theta)'\right)\right)\otimes^*\widehat{Q}(\theta)s(z_i,\theta)\right) \tag{50}$$

$$= \frac{1}{n}\sum_{i=1}^n\left(\widehat{Q}(\theta)\nabla s(z_i,\theta)\otimes^*\widehat{Q}(\theta)s(z_i,\theta)\right) \tag{51}$$

$$+\frac{1}{n}\sum_{i=1}^n\left(vec^*\left(s(z_i,\theta)'\nabla\left(\widehat{Q}(\theta)'\right)\right)\otimes^*\widehat{Q}(\theta)s(z_i,\theta)\right). \tag{52}$$

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 (51)

$$\begin{aligned}&\left\|\tfrac{1}{n}\sum_{i=1}^n\left(\widehat{Q}(\theta)\nabla s(z_i,\theta)\otimes^*\widehat{Q}(\theta)s(z_i,\theta)\right)\right\|\leq\tfrac{1}{n}\sum_{i=1}^n\left\|\left(\widehat{Q}(\theta)s(z_i,\theta)\otimes^*\widehat{Q}(\theta)s(z_i,\theta)\right)\right\|\\&=\tfrac{1}{n}\sum_{i=1}^n\left\|\widehat{Q}(\theta)\nabla s(z_i,\theta)\right\|\left\|\widehat{Q}(\theta)s(z_i,\theta)\right\|\leq\tfrac{1}{n}\sum_{i=1}^n\|\nabla s(z_i,\theta)\|\,\|s(z_i,\theta)\|\left\|\widehat{Q}(\theta)\right\|^2=O_p(1)\end{aligned} \tag{53}$$

uniformly over $\theta\in\Theta_0$ by Lemma A.5 for $(\upsilon_1,\upsilon_2)=(1,0)$ under $E\left[\sup_{\theta\in\Theta_0}\|\nabla^\upsilon s(z_i,\theta)\|^2\right]<\infty,\upsilon=1,0$ and by (26). This gives

$$\begin{aligned}&\left\|\tfrac{1}{n}\sum_{i=1}^n\left(vec^*\left(s(z_i,\theta)'\nabla\left(\widehat{Q}(\theta)'\right)\right)\otimes^*\widehat{Q}(\theta)s(z_i,\theta)\right)\right\|\\&\leq\tfrac{1}{n}\sum_{i=1}^n\left\|vec^*\left(s(z_i,\theta)'\nabla\left(\widehat{Q}(\theta)'\right)\right)\right\|\left\|\widehat{Q}(\theta)s(z_i,\theta)\right\|\\&\leq\tfrac{1}{n}\sum_{i=1}^n\left\|s(z_i,\theta)'\nabla\left(\widehat{Q}(\theta)'\right)\right\|\left\|\widehat{Q}(\theta)s(z_i,\theta)\right\|\leq\tfrac{1}{n}\sum_{i=1}^n\|s(z_i,\theta)\|^2\left\|\widehat{Q}(\theta)\right\|\left\|\nabla\left(\widehat{Q}(\theta)'\right)\right\|=O_p(1)\end{aligned} \tag{54}$$

by the uniform convergence of $\frac{1}{n}\sum_{i=1}^n\|s(z_i,\theta)\|^2$ to $E\left[\|s(z,\theta)\|^2\right]<\infty$, (26), and Lemma A.8 noting $\|vec^*(\cdot)\|=\|\cdot\|$ and hence we show that (48) is $O_p(1)$ uniformly over $\theta\in\Theta_0$ from (53) and (54). Similarly we can show that (49) is $O_p(1)$ around the neighborhood of θ_0 and hence we have

$$\nabla\left(\frac{1}{n}\sum_{i=1}^n\left[\widehat{Q}(\theta)s(z_i,\theta)\otimes\widehat{Q}(\theta)s(z_i,\theta)\right]\right)=O_p(1). \tag{55}$$

Together with (29), this shows the second RHS term of the last equality in (44) is $O_p(1)$. Now consider for the third RHS term of the last equality in (44),

$$\begin{aligned}&\nabla\left(\tfrac{1}{n}\sum_{i=1}^n\left[\nabla s(z_i,\theta)\widehat{Q}(\theta)s(z_i,\theta)\right]\right)\\&=\tfrac{1}{n}\sum_{i=1}^n\nabla s(z_i,\theta)\nabla\left(\widehat{Q}(\theta)s(z_i,\theta)\right)+\tfrac{1}{n}\sum_{i=1}^n vec^*\left(\left(\widehat{Q}(\theta)s(z_i,\theta)\right)'\nabla\left((\nabla s(z_i,\theta))'\right)\right)\\&=\tfrac{1}{n}\sum_{i=1}^n\nabla s(z_i,\theta)\left(\widehat{Q}(\theta)\nabla s(z_i,\theta)+vec^*\left(s(z_i,\theta)'\nabla\left(\widehat{Q}(\theta)'\right)\right)\right)\\&\quad+\tfrac{1}{n}\sum_{i=1}^n vec^*\left(s(z_i,\theta)'\widehat{Q}(\theta)'\nabla\left((\nabla s(z_i,\theta))'\right)\right).\end{aligned} \tag{56}$$

This implies

$$\begin{aligned}
&\left\|\nabla\left(\frac{1}{n}\sum_{i=1}^{n}\left[\nabla s(z_i,\theta)\widehat{Q}(\theta)s(z_i,\theta)\right]\right)\right\| \le \frac{1}{n}\sum_{i=1}^{n}\|\nabla s(z_i,\theta)\|^2\left\|\widehat{Q}(\theta)\right\| \\
&\quad+\frac{1}{n}\sum_{i=1}^{n}\|\nabla s(z_i,\theta)\|\left\|vec^*\left(s(z_i,\theta)'\nabla\left(\widehat{Q}(\theta)'\right)\right)\right\|+\frac{1}{n}\sum_{i=1}^{n}\left\|vec^*\left(s(z_i,\theta)'\widehat{Q}(\theta)'\nabla\left((\nabla s(z_i,\theta))'\right)\right)\right\| \\
&=\frac{1}{n}\sum_{i=1}^{n}\|\nabla s(z_i,\theta)\|^2\left\|\widehat{Q}(\theta)\right\|+\frac{1}{n}\sum_{i=1}^{n}\|\nabla s(z_i,\theta)\|\left\|s(z_i,\theta)'\nabla\left(\widehat{Q}(\theta)'\right)\right\| \\
&\quad+\frac{1}{n}\sum_{i=1}^{n}\left\|s(z_i,\theta)'\widehat{Q}(\theta)'\nabla\left((\nabla s(z_i,\theta))'\right)\right\| \\
&\le\frac{1}{n}\sum_{i=1}^{n}\|\nabla s(z_i,\theta)\|^2\left\|\widehat{Q}(\theta)\right\|+\frac{1}{n}\sum_{i=1}^{n}\|\nabla s(z_i,\theta)\|\|s(z_i,\theta)\|\left\|\nabla\left(\widehat{Q}(\theta)'\right)\right\| \\
&\quad+\frac{1}{n}\sum_{i=1}^{n}\|s(z_i,\theta)\|\left\|\nabla^2 s(z_i,\theta)\right\|\left\|\widehat{Q}(\theta)\right\|.
\end{aligned}\tag{57}$$

We have the first RHS term in the last inequality of (57) equals to $O_p(1)$ uniformly over $\theta\in\Theta_0$ by (26) 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 (31) 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 (57) equals to $O_p(1)$ uniformly over $\theta\in\Theta_0$ from (26) and Lemma A.5 with $(v_1,v_2)=(0,2)$ and thus we bound the third RHS term of the last equality in (44) 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 $\nabla\widehat{c}(\theta)$ and $\nabla c(\theta)$, respectively, as

$$\nabla\widehat{c}(\theta)=\left\{\begin{array}{l}
\frac{1}{2}vec^*\left(\left(\frac{1}{n}\sum_{i=1}^{n}\left[\widehat{Q}(\theta)s(z_i,\theta)\otimes\widehat{Q}(\theta)s(z_i,\theta)\right]\right)'\left(\nabla\left(\widehat{H}_2(\theta)'\right)\right)\right) \\
+\frac{1}{2}\widehat{H}_2(\theta)\left(\begin{array}{c}\frac{1}{n}\sum_{i=1}^{n}\left(\widehat{Q}(\theta)\nabla s(z_i,\theta)\otimes^*\widehat{Q}(\theta)s(z_i,\theta)\right) \\ +\frac{1}{n}\sum_{i=1}^{n}\left(vec^*\left(s(z_i,\theta)'\nabla\left(\widehat{Q}(\theta)'\right)\right)\otimes^*\widehat{Q}(\theta)s(z_i,\theta)\right)\end{array}\right) \\
+\frac{1}{2}\widehat{H}_2(\theta)\left(\begin{array}{c}\frac{1}{n}\sum_{i=1}^{n}\left(\widehat{Q}(\theta)s(z_i,\theta)\otimes\widehat{Q}(\theta)\nabla s(z_i,\theta)\right) \\ +\frac{1}{n}\sum_{i=1}^{n}\left(\widehat{Q}(\theta)s(z_i,\theta)\right)\otimes vec^*\left(s(z_i,\theta)'\nabla\left(\widehat{Q}(\theta)'\right)\right)\end{array}\right) \\
+\frac{1}{n}\sum_{i=1}^{n}\nabla s(z_i,\theta)\left(\widehat{Q}(\theta)\nabla s(z_i,\theta)+vec^*\left(s(z_i,\theta)'\nabla\left(\widehat{Q}(\theta)'\right)\right)\right) \\
+\frac{1}{n}\sum_{i=1}^{n}vec^*\left(s(z_i,\theta)'\widehat{Q}(\theta)'\nabla\left((\nabla s(z_i,\theta))'\right)\right)
\end{array}\right\}\tag{58}$$

$$\begin{aligned}
&\nabla c(\theta)= \\
&\frac{1}{2}vec^*\left(\left(E\left[Q(\theta)s(z_i,\theta)\otimes Q(\theta)s(z_i,\theta)\right]\right)'\left(\nabla\left(H_2(\theta)'\right)\right)\right) \\
&+\frac{1}{2}H_2(\theta)\left(E\left[Q(\theta)\nabla s(z_i,\theta)\otimes^* Q(\theta)s(z_i,\theta)\right]+E\left[vec^*\left(s(z_i,\theta)'\nabla\left(Q(\theta)'\right)\right)\otimes^* Q(\theta)s(z_i,\theta)\right]\right) \\
&+\frac{1}{2}H_2(\theta)\left(E\left[Q(\theta)s(z_i,\theta)\otimes Q(\theta)\nabla s(z_i,\theta)\right]+E\left[\left(Q(\theta)s(z_i,\theta)\right)\otimes vec^*\left(s(z_i,\theta)'\nabla\left(Q(\theta)'\right)\right)\right]\right) \\
&+E\left[\nabla s(z_i,\theta)\left(Q(\theta)\nabla s(z_i,\theta)+vec^*\left(s(z_i,\theta)'\nabla\left(Q(\theta)'\right)\right)\right)\right]+E\left[vec^*\left(s(z_i,\theta)'Q(\theta)'\nabla\left((\nabla s(z_i,\theta))'\right)\right)\right].
\end{aligned}\tag{59}$$

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 (44), consider

$$\nabla^2\widehat{c}(\theta)=\left\{\begin{array}{l}
\nabla\left(\frac{1}{2}vec^*\left(\left(\frac{1}{n}\sum_{i=1}^{n}\left[\widehat{Q}(\theta)s(z_i,\theta)\otimes\widehat{Q}(\theta)s(z_i,\theta)\right]\right)'\left(\nabla\left(\widehat{H}_2(\theta)'\right)\right)\right)\right) \\
+\nabla\left(\frac{1}{2}\widehat{H}_2(\theta)\nabla\left(\frac{1}{n}\sum_{i=1}^{n}\left[\widehat{Q}(\theta)s(z_i,\theta)\otimes\widehat{Q}(\theta)s(z_i,\theta)\right]\right)\right) \\
+\nabla^2\left(\frac{1}{n}\sum_{i=1}^{n}\left[\nabla s(z_i,\theta)\widehat{Q}(\theta)s(z_i,\theta)\right]\right).
\end{array}\right\}\tag{60}$$

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 (60), we have

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

uniformly over $\theta \in \Theta_0$ from (40), (46), (55), 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

$$\begin{aligned}
&\left\|\nabla\left(\tfrac{1}{2}\widehat{H}_2(\theta)\nabla\left(\tfrac{1}{n}\sum_{i=1}^n\left[\widehat{Q}(\theta)s(z_i,\theta)\otimes\widehat{Q}(\theta)s(z_i,\theta)\right]\right)\right)\right\| \\
&\le \tfrac{1}{2}\left\|\nabla\widehat{H}_2(\theta)\right\|\left\|\nabla\left(\tfrac{1}{n}\sum_{i=1}^n\left[\widehat{Q}(\theta)s(z_i,\theta)\otimes\widehat{Q}(\theta)s(z_i,\theta)\right]\right)\right\| + \tfrac{1}{2}\left\|\widehat{H}_2(\theta)\right\|\left\|\nabla^2\left(\tfrac{1}{n}\sum_{i=1}^n\left[\widehat{Q}(\theta)s(z_i,\theta)\otimes\widehat{Q}(\theta)s(z_i,\theta)\right]\right)\right\|
\end{aligned} \tag{61}$$

Note, for the first RHS term in (61)

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

by (40) and (55). From (47) and (50), for the second RHS term in (61), we have

$$\begin{aligned}
&\left\|\nabla^2\left(\tfrac{1}{n}\sum_{i=1}^n\left[\widehat{Q}(\theta)s(z_i,\theta)\otimes\widehat{Q}(\theta)s(z_i,\theta)\right]\right)\right\| \\
&\le \left\|\nabla\left(\tfrac{1}{n}\sum_{i=1}^n\left(\widehat{Q}(\theta)\nabla s(z_i,\theta)\otimes^*\widehat{Q}(\theta)s(z_i,\theta)\right)\right)\right\| \\
&\quad + \left\|\nabla\left(\tfrac{1}{n}\sum_{i=1}^n\left(vec^*\left(s(z_i,\theta)'\nabla\left(\widehat{Q}(\theta)'\right)\right)\otimes^*\widehat{Q}(\theta)s(z_i,\theta)\right)\right)\right\| \\
&\quad + \left\|\nabla\left(\tfrac{1}{n}\sum_{i=1}^n\left(\widehat{Q}(\theta)s(z_i,\theta)\otimes\nabla\left(\widehat{Q}(\theta)s(z_i,\theta)\right)\right)\right)\right\|.
\end{aligned} \tag{62}$$

First we bound the first RHS term of (62) uniformly over $\theta \in \Theta_0$ as

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

where the second inequality is from Remark 4, the first equality is from Remark 1 and Remark 2, the third inequality is from Remark 3. The second last equality comes from Lemma A.5, Lemma A.8 and (26). For the second RHS term of (62), from Remark 2-4, it follows that

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

uniformly over $\theta \in \Theta_0$. The last equality comes from Lemma A.5, Lemma A.8, and (26). Similarly we can also bound the last RHS term of (62) uniformly over $\theta \in \Theta_0$ as

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

using Remark 2 and Remark 4. The second last equality is obtained from Lemma A.5, Lemma A.8, (26), 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 (60) to be $O_p(1)$. Finally we rewrite the third RHS term of (60) as

$$
\begin{aligned}
&\nabla^2\left(\frac{1}{n}\sum_{i=1}^{n}\left[\nabla s(z_i,\theta)\,\widehat{Q}(\theta)s(z_i,\theta)\right]\right) \\
&= \nabla\begin{pmatrix}\frac{1}{n}\sum_{i=1}^{n}\nabla s(z_i,\theta)\left(\widehat{Q}(\theta)\nabla s(z_i,\theta)+vec^{*}\left(s(z_i,\theta)'\nabla\left(\widehat{Q}(\theta)'\right)\right)\right) \\ +\frac{1}{n}\sum_{i=1}^{n}vec^{*}\left(s(z_i,\theta)'\widehat{Q}(\theta)'\nabla\left((\nabla s(z_i,\theta))'\right)\right)\end{pmatrix} \\
&= \nabla\left(\frac{1}{n}\sum_{i=1}^{n}\nabla s(z_i,\theta)\left(\widehat{Q}(\theta)\nabla s(z_i,\theta)+vec^{*}\left(s(z_i,\theta)'\nabla\left(\widehat{Q}(\theta)'\right)\right)\right)\right) \qquad (63) \\
&\quad +\nabla\left(\frac{1}{n}\sum_{i=1}^{n}vec^{*}\left(s(z_i,\theta)'\widehat{Q}(\theta)'\nabla\left((\nabla s(z_i,\theta))'\right)\right)\right) \qquad (64)
\end{aligned}
$$

from (56). For (63), note

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

by Remark 1 and Remark 3 and hence

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

from (26), (18), and Lemma A.8 and since i) $\frac{1}{n}\sum_{i=1}^{n}\|\nabla^{\upsilon}s(z_i,\theta)\|\left\|\nabla^2 s(z_i,\theta)\right\|=O_p(1)$ for $\upsilon\in\{0,1\}$ by Lemma A.5 and since ii)$\frac{1}{n}\sum_{i=1}^{n}\|\nabla s(z_i,\theta)\|^2=O_p(1)$ by the uniform convergence. Now consider for (64)

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

by (26), Lemma A.8, and Lemma A.5 and hence we have the third RHS term of (60) equals to $O_p(1)$ uniformly over $\theta \in \Theta_0$. This completes the proof. ■

A.2 Additional Preliminary Lemmas for the Third Order Expansion

First, note that Lemma A.3 and Lemma A.4 trivially hold under Assumption 3.1 and Assumption 3.3, respectively considering that Assumption 3.1 and Assumption 3.3 are stronger than Assumption A.1 and Assumption A.4, respectively. We establish conditions (i)-(ix) in Lemma 2.3 are satisfied under Assumption 3.1-3.2 or Assumption 3.3 and 3.2. Again Condition (i) and (iia) are directly assumed. Condition (iib) is by the dominated convergence theorem with the dominating function given by $\sup_{\theta\in\Theta_0}\left\|\nabla^4 s(z,\theta)\right\|$ under Condition (i), (iia), and $E[\sup_{\theta\in\Theta_0}\left\|\nabla^4 s(z,\theta)\right\|]<\infty$. Condition (iii) holds by the stochastic equicontinuity of $\frac{1}{\sqrt{n}}\sum_{i=1}^n\left(\nabla^3 s\left(z_i,\theta\right)-E[\nabla^3 s\left(z_i,\theta\right)]\right)$ for $\theta\in\Theta_0$ as discussed in Lemma A.2 with $m(z,\theta)=\nabla^3 s\left(z,\theta\right)$ under Assumption 3.3. Instead, under Assumption 3.1, 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 2.1 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 3.2. 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 3.1(iii) or 3.3(iii). Condition (ix) is the result of Lemma 2.1. We also need to verify following lemmas.

Lemma A.10 *Under Assumption 3.1-3.2 or 3.3-3.2, Condition 4 (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 (58) and (59), it follows that

$$
\begin{aligned}
&\|\nabla\widehat{c}(\theta_0)-\nabla c(\theta_0)\| \\
\le &\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\| &(65)\\
&+\begin{array}{l}\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} &(66)\\
&+\left\|\begin{array}{l}\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\| &(67)\\
&+\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\left(z_i,\theta_0\right)\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\| &(68)\\
&+\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\left(z_i,\theta_0\right)\right)'\right)\right)\\-E\left[vec^*\left(s(z_i,\theta_0)'Q(\theta_0)'\nabla\left(\left(\nabla s\left(z_i,\theta_0\right)\right)'\right)\right)\right]\end{array}\right\|. &(69)
\end{aligned}
$$

We show (65), (66), (67), (68), and (69) 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 $E\left[\left\|\nabla^3 s(z_i,\theta_0)\right\|^2\right]<\infty$. Recalling (41), this implies

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

and hence (65) is $O_p(1/\sqrt{n})$. 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

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

applying the CLT from

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

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

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

by the CLT under

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

recalling that $\|\nabla(Q(\theta_0)')\| = \|Q(\theta_0)\|^2 \|H_2(\theta_0)\| < \infty$. Applying these two results together with (41), (42), and Lemma A.8 (b), we have (66) equals to $O_p(1/\sqrt{n})$ by the Triangle inequality. Similarly we have (67)=$O_p(1/\sqrt{n})$.

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

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

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

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

by applying the CLT provided that $\|\nabla(Q(\theta_0)')\|^{2t_1} E\left[\|\nabla s(z_i,\theta_0)\|^2 \|s(z_i,\theta_0)\|^2\right] < \infty$ that holds under $\|\nabla(Q(\theta_0)')\| < \infty$, $E\left[\|\nabla s(z_i,\theta_0)\|^4\right] < \infty$, and $E\left[\|s(z_i,\theta_0)\|^4\right] < \infty$. Applying these two results together with (41) and Lemma A.8 (b), we have (68)=$O_p(1/\sqrt{n})$. Finally, for $j_1, j_2 \in \{0,1\}$, note

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

by the CLT since $\|Q(\theta_0)\|^{2j_1} E\left[\left\|\nabla^2 s(z_i,\theta_0)\right\|^2 \|s(z_i,\theta_0)\|^2\right] < \infty$ holds by $\|Q(\theta_0)\| < \infty$, $E\left[\left\|\nabla^2 s(z_i,\theta_0)\right\|^4\right] < \infty$, and $E\left[\|s(z_i,\theta_0)\|^4\right] < \infty$. It implies (69)=$O_p(1/\sqrt{n})$ together with (41). This completes the proof. ■

Lemma A.11 *Under Assumption 3.1-3.2 or 3.3-3.2 with* $\kappa \geq 5$, *Condition 4 (ii):*$\nabla^3 \widehat{c}(\theta) = O_p(1)$ *around the neighborhood of* θ_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 1-4 when $\dim(\theta) = 1$ as an illustrational purpose. With $\dim(\theta) = 1$, we can rewrite the correction term (6) as

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

Now define $c(\theta) \equiv \tau(E[m(z_i, \theta)])$
where $\tau(t_1, t_2, t_3, t_4) \equiv \frac{1}{2t_1^2} t_2 t_3 - \frac{1}{t_1} t_4$, $t_1 = E[\nabla s(z_i, \theta)]$, $t_2 = E\left[\nabla^2 s(z_i, \theta)\right]$, $t_3 = E\left[s(z_i, \theta)^2\right]$, $t_4 = E[\nabla s(z_i, \theta) s(z_i, \theta)]$, 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 θ.

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 $E\left[\sup_{\theta \in \Theta_0} \|\nabla s(z_i, \theta)\|^2\right] < \infty$, $E\left[\sup_{\theta \in \Theta_0} \left\|\nabla^2 s(z_i, \theta)\right\|\right] < \infty$, and $E\left[\sup_{\theta \in \Theta_0} \|s(z_i, \theta)\|^2\right] < \infty$, if we assume Assumption 3.1. 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 $E[\sup_{\theta \in \Theta_0} \left\|\nabla^{\overline{\upsilon}} s(z_i, \theta)\right\|^2] < \infty$ for $\overline{\upsilon} = \{0, 1, 2, 3, 4\}$ and $E[\sup_{\theta \in \Theta_0} \left\|\nabla^5 s(z_i, \theta)\right\|]$
$< \infty$, if we assume Assumption 3.1. 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, 4 (ii), respectively. Moreover, it is also clear that $\widehat{m}(\theta_0) = \overline{m}(\theta_0) + O_p(1/\sqrt{n})$ 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 $\widehat{m}_\theta(\theta_0) = \overline{m}_\theta(\theta_0) + O_p(1/\sqrt{n})$ by the CLT under $E[\left\|\nabla^{\overline{\overline{\upsilon}}} s(z_i, \theta_0)\right\|^4] < \infty$ for $\overline{\overline{\upsilon}} = \{0, 1, 2\}$ and $E[\left\|\nabla^3 s(z_i, \theta_0)\right\|^2] < \infty$. Also note that $\widehat{\tau}_m(\theta_0) = \overline{\tau}_m(\theta_0) + O_p(1/\sqrt{n})$ by the Slutsky theorem and $\widehat{m}(\theta_0) = \overline{m}(\theta_0) + O_p(1/\sqrt{n})$. These imply that $\widehat{c}(\theta_0) = c(\theta_0) + O_p(1/\sqrt{n})$ and $\nabla \widehat{c}(\theta_0) = \nabla c(\theta_0) + O_p(1/\sqrt{n})$, which are Condition 1 (ii) and 4 (i), respectively.
■

Lemma A.12 *Under Assumption 3.1-3.2 or 3.3-3.2, Condition 5, 6, 7 are satisfied.*

Proof. Condition 5: Note

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

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

Condition 6: From Remark 2, Condition 1 (ii), Condition 4 (i), $\widehat{Q}(\theta_0) = Q(\theta_0) + O_p(1/\sqrt{n})$, and $\nabla\left(\widehat{Q}(\theta_0)'\right) = \nabla\left(Q(\theta_0)'\right) + O_p(1/\sqrt{n})$ by Lemma A.8, we have

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

Condition 7: From Remark 1-5, 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 2, Lemma A.8, $\left\|\widehat{Q}(\widetilde{\theta})\right\| = O_p(1)$, and Condition 2 and 3. ■

B Proofs of Main Lemmas and Propositions

B.1 Lemma 2.1

Proof. Lemma 2.1 (i)

Consider a first order Taylor expansion of $0 = \frac{1}{n}\sum_{i=1}^n s(z_i,\theta_0) + \left(\frac{1}{n}\sum_{i=1}^n \nabla s\left(z_i,\widetilde{\theta}\right)\right)\left(\widehat{\theta}-\theta_0\right)$, where $\widetilde{\theta}$ lies between θ_0 and $\widehat{\theta}$ and hence

$$\sqrt{n}\left(\widehat{\theta}-\theta_0\right) = \left(-\frac{1}{n}\sum_{i=1}^n \nabla s\left(z_i,\widetilde{\theta}\right)\right)^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^n s(z_i,\theta_0), \tag{70}$$

assuming $\frac{1}{n}\sum_{i=1}^n \nabla s\left(z_i,\widetilde{\theta}\right)$ is nonsingular[4]. Note that $\left\|\frac{1}{n}\sum_{i=1}^n \nabla s\left(z_i,\widetilde{\theta}\right) - E\left[\nabla s(z_i,\theta_0)\right]\right\| = o_p(1)$ by Condition (iii) for $v=1$ and $\widetilde{\theta} = \theta_0 + o_p(1)$. Therefore, by the Slutsky theorem, we have

$$\sqrt{n}\left(\widehat{\theta}-\theta_0\right) = \left(-E\left[\nabla s(z_i,\theta_0)\right]\right)^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^n s(z_i,\theta_0) + o_p(1) = QJ + o_p(1), \tag{71}$$

since we assume $J = O_p(1)$ and $E\left[\nabla s(z_i,\theta_0)\right]$ is nonsingular. (71) implies $\widehat{\theta}-\theta_0 = O_p\left(\frac{1}{\sqrt{n}}\right)$. Define $\widehat{H}_1(\theta) \equiv \frac{1}{n}\sum_{i=1}^n \nabla s(z_i,\theta)$ and $H_1(\theta_0)(= -Q^{-1}) \equiv E[\nabla s(z_i,\theta_0)]$. In what follows, we treat $\widehat{H}_1(\theta)$ as nonsingular for θ around the neighborhood of θ_0. This is innocuous, since with probability approaching to one, $\widehat{H}_1(\widetilde{\theta})$ is nonsingular for $\widetilde{\theta} = \theta_0 + o_p(1)$ as long as $H_1(\theta_0)$ is nonsingular. Now note that we can rewrite (71) as $\sqrt{n}\left(\widehat{\theta}-\theta_0\right) = -H_1(\theta_0)^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^n s(z_i,\theta_0) - \left(\widehat{H}_1(\widetilde{\theta})^{-1} - H_1(\theta_0)^{-1}\right)\frac{1}{\sqrt{n}}\sum_{i=1}^n s(z_i,\theta_0)$ from (70). Consider

$$\begin{aligned}
&\left\|\widehat{H}_1(\widetilde{\theta})^{-1} - H_1(\theta_0)^{-1}\right\| \leq \left\|\widehat{H}_1(\widetilde{\theta})^{-1} - \widehat{H}_1(\theta_0)^{-1}\right\| + \left\|\widehat{H}_1(\theta_0)^{-1} - H_1(\theta_0)^{-1}\right\| \\
= &\left\|\widehat{H}_1(\theta_0)^{-1}\left(\widehat{H}_1(\widetilde{\theta}) - \widehat{H}_1(\theta_0)\right)\widehat{H}_1(\widetilde{\theta})^{-1}\right\| + \left\|H_1(\theta_0)^{-1}\left(\widehat{H}_1(\theta_0) - H_1(\theta_0)\right)\widehat{H}_1(\theta_0)^{-1}\right\| \\
\leq &\left\|\widehat{H}_1(\theta_0)^{-1}\right\|\left\|\widehat{H}_1(\widetilde{\theta}) - \widehat{H}_1(\theta_0)\right\|\left\|\widehat{H}_1(\widetilde{\theta})^{-1}\right\| + \left\|H_1(\theta_0)^{-1}\right\|\left\|\widehat{H}_1(\theta_0) - H_1(\theta_0)\right\|\left\|\widehat{H}_1(\theta_0)^{-1}\right\|,
\end{aligned} \tag{72}$$

by Triangle Inequality and Cauchy-Schwarz Inequality. Now applying the mean value theorem, we have

$$\begin{aligned}
&\left\|\left(\widehat{H}_{1m,n}(\widetilde{\theta}) - \widehat{H}_{1m,n}(\theta_0)\right)\right\| = \left\|\frac{1}{n}\sum_{i=1}^n \nabla\frac{\partial s_m\left(z_i,\overline{\theta}\right)}{\partial\theta_n}(\widetilde{\theta}-\theta_0)\right\| \\
\leq &\left\|\frac{1}{n}\sum_{i=1}^n \nabla\frac{\partial s_m\left(z_i,\overline{\theta}\right)}{\partial\theta_n}\right\|\left\|(\widetilde{\theta}-\theta_0)\right\| = \left\|E\left[\nabla\frac{\partial s_m(z_i,\theta_0)}{\partial\theta_n}\right] + o_p(1)\right\|\left\|(\widetilde{\theta}-\theta_0)\right\| \leq C\left\|(\widetilde{\theta}-\theta_0)\right\|
\end{aligned}$$

where $\overline{\theta}$ lies between $\widetilde{\theta}$ and θ_0, $\widehat{H}_{1m,n}(\cdot)$ denotes the m-th row and the n-th column element of $\widehat{H}_1(\cdot)$, $s_m(z_i,\cdot)$ is the m-th element of $s(z_i,\cdot)$, and θ_n is the n-th element of θ. The second equality comes from Condition

[4] Instead of the usual inverse operator, we can use the Moore-Penrose generalized inverse allowing singularity. Since the probability of the event that $\frac{1}{n}\sum_{i=1}^n \nabla s\left(z_i,\widetilde{\theta}\right)$ is nonsingular approaches to one as the sample size goes to infinity by Condition (iii) and (v), we can simply assume nonsingularity. Some technical proof allowing the singularity can be provided as Newey and McFadden (1994) (see p.2152).

(iii):$\left\| \frac{1}{n} \sum_{i=1}^{n} \nabla^2 s\left(z_i, \overline{\theta}\right) - E\left[\nabla^2 s\left(z_i, \theta_0\right)\right] \right\| = o_p(1)$ for $\overline{\theta} = \theta_0 + o_p(1)$. The last inequality is by $E\left[\left\|\nabla^2 s(z, \theta_0)\right\|\right] < \infty$ from Condition (ii). This implies

$$\left\| \left(\widehat{H}_{1m,n}(\widetilde{\theta}) - \widehat{H}_{1m,n}(\theta_0) \right) \right\| = O_p\left(1/\sqrt{n}\right), \tag{73}$$

since $\widetilde{\theta} - \theta_0 = O_p(1/\sqrt{n})$. Also from Condition (vii), we have

$$\sqrt{n}\left(\widehat{H}_1(\theta_0) - H_1(\theta_0) \right) = V = O_p(1). \tag{74}$$

By Condition (v) and $\widetilde{\theta} = \theta_0 + o_p(1)$, we have $\left\| \widehat{H}_1(\theta_0)^{-1} \right\| = O_p(1)$, $\left\| \widehat{H}_1(\widetilde{\theta})^{-1} \right\| = O_p(1)$, and $\left\| H_1(\theta_0)^{-1} \right\| < \infty$. Applying these results together with (73) and (74) to (72), it follows

$$\left\| \widehat{H}_1(\widetilde{\theta})^{-1} - H_1(\theta_0)^{-1} \right\| = O_p\left(1/\sqrt{n}\right) \tag{75}$$

and hence

$$\begin{aligned} \sqrt{n}\left(\widehat{\theta} - \theta_0 \right) &= -H_1(\theta_0)^{-1} \frac{1}{\sqrt{n}} \sum_{i=1}^{n} s\left(z_i, \theta_0\right) \\ &\quad - \left(\widehat{H}_1(\widetilde{\theta})^{-1} - H_1(\theta_0)^{-1} \right) \frac{1}{\sqrt{n}} \sum_{i=1}^{n} s\left(z_i, \theta_0\right) = QJ + O_p\left(1/\sqrt{n}\right), \end{aligned} \tag{76}$$

noting $J = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} s\left(z_i, \theta_0\right) = O_p(1)$. This completes the proof of Lemma 2.1 (i). ■

Proof. Lemma 2.1 (ii)

Consider a second order Taylor expansion of (2) around the true value of $\theta = \theta_0$ as

$$\begin{aligned} 0 &= \tfrac{1}{n} \textstyle\sum_{i=1}^{n} s\left(z_i, \theta_0\right) + \left(\tfrac{1}{n} \textstyle\sum_{i=1}^{n} \nabla s\left(z_i, \theta_0\right) \right) \left(\widehat{\theta} - \theta_0 \right) \\ &\quad + \tfrac{1}{2} \left(\tfrac{1}{n} \textstyle\sum_{i=1}^{n} \nabla^2 s\left(z_i, \widetilde{\theta}\right) \right) \left(\left(\widehat{\theta} - \theta_0 \right) \otimes \left(\widehat{\theta} - \theta_0 \right) \right) \end{aligned},$$

where $\widetilde{\theta}$ lies between θ_0 and $\widehat{\theta}$. Now using the stochastic equicontinuity Condition (iv);

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left(\nabla^2 s\left(z_i, \widetilde{\theta}\right) - H_2(\widetilde{\theta}) \right) - \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left(\nabla^2 s\left(z_i, \theta_0\right) - H_2(\theta_0) \right) = o_p(1) \text{ for } \widetilde{\theta} = \theta_0 + o_p(1)$$

we have

$$\begin{aligned} &\left(\tfrac{1}{n} \textstyle\sum_{i=1}^{n} \nabla^2 s\left(z_i, \widetilde{\theta}\right) - \tfrac{1}{n} \textstyle\sum_{i=1}^{n} \nabla^2 s\left(z_i, \theta_0\right) \right) \left(\left(\widehat{\theta} - \theta_0 \right) \otimes \left(\widehat{\theta} - \theta_0 \right) \right) \\ &= \left(H_2(\widetilde{\theta}) - H_2\left(\theta_0\right) + o_p\left(\tfrac{1}{\sqrt{n}}\right) \right) \left(\left(\widehat{\theta} - \theta_0 \right) \otimes \left(\widehat{\theta} - \theta_0 \right) \right) \\ &= \left(\nabla \left(E\left[\nabla^2 s\left(z_i, \theta\right)\right] \right)\big|_{\theta = \widetilde{\widetilde{\theta}}} (\widetilde{\theta} - \theta_0) + o_p\left(\tfrac{1}{\sqrt{n}}\right) \right) \left(\left(\widehat{\theta} - \theta_0 \right) \otimes \left(\widehat{\theta} - \theta_0 \right) \right) \\ &= \left(E\left[\nabla^3 s\left(z_i, \theta_0\right)\right] (\widetilde{\theta} - \theta_0) + o_p(1) O_p\left(\tfrac{1}{\sqrt{n}}\right) + o_p\left(\tfrac{1}{\sqrt{n}}\right) \right) \left(\left(\widehat{\theta} - \theta_0 \right) \otimes \left(\widehat{\theta} - \theta_0 \right) \right) = O_p(n^{-3/2}), \end{aligned}$$

where the third equality is from standard results on differentiating inside the integral and the Slutsky theorem. We obtain the second equality by applying the mean value theorem where $\widetilde{\widetilde{\theta}}$ lies between $\widetilde{\theta}$ and θ_0. The second last equality is from the continuity of $E\left[\nabla^3 s\left(z_i, \theta\right)\right]$ at θ_0 and since $\widetilde{\widetilde{\theta}} = \theta_0 + o_p(1)$. We, thus, obtain

$$\begin{aligned} 0 &= \frac{1}{n} \sum_{i=1}^{n} s\left(z_i, \theta_0\right) + \left(\frac{1}{n} \sum_{i=1}^{n} \nabla s\left(z_i, \theta_0\right) \right) \left(\widehat{\theta} - \theta_0 \right) \\ &\quad + \frac{1}{2} \left(\frac{1}{n} \sum_{i=1}^{n} \nabla^2 s\left(z_i, \theta_0\right) \right) \left(\left(\widehat{\theta} - \theta_0 \right) \otimes \left(\widehat{\theta} - \theta_0 \right) \right) + O_p(n^{-3/2}). \end{aligned} \tag{77}$$

Write $\frac{1}{n}\sum_{i=1}^{n}\nabla s(z_i,\theta_0) = -Q^{-1} + \frac{1}{\sqrt{n}}V$ and $\frac{1}{n}\sum_{i=1}^{n}\nabla^2 s(z_i,\theta_0) = H_2 + \frac{1}{\sqrt{n}}W$ for $Q \equiv -(E[\nabla s(z_i,\theta_0)])^{-1}$, $H_2 \equiv E\left[\nabla^2 s(z_i,\theta_0)\right]$, and

$$\begin{array}{l} V \equiv \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(\nabla s(z_i,\theta_0) - E\left[\nabla s(z_i,\theta_0)\right]\right) = O_p(1) \\ W \equiv \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(\nabla^2 s(z_i,\theta_0) - E\left[\nabla^2 s(z_i,\theta_0)\right]\right) = O_p(1) \end{array}$$

from Condition (vii) and (viii). We then obtain

$$\begin{array}{l} 0 = \frac{1}{\sqrt{n}}J + \left(-Q^{-1} + \frac{1}{\sqrt{n}}V\right)\left(\widehat{\theta} - \theta_0\right) \\ \quad + \frac{1}{2}\left(H_2 + \frac{1}{\sqrt{n}}W\right)\left(\left(\widehat{\theta} - \theta_0\right) \otimes \left(\widehat{\theta} - \theta_0\right)\right) + O_p(n^{-3/2}). \end{array} \tag{78}$$

Now note that we can expand

$$\begin{aligned} \left(-Q^{-1} + \frac{1}{\sqrt{n}}V\right)^{-1} &= \left(I - Q\frac{1}{\sqrt{n}}V\right)^{-1}(-Q) \qquad (79) \\ &= -Q + \left\{ \begin{array}{c} O_p(\frac{1}{\sqrt{n}}) \\ -\frac{1}{\sqrt{n}}QVQ + O_p(n^{-1}) \\ -\frac{1}{\sqrt{n}}QVQ - \frac{1}{n}QVQVQ + O_p(n^{-3/2}) \end{array} \right. \end{aligned}$$

depending on the orders we need. Using this result, from (78) we have

$$\begin{aligned} \widehat{\theta} - \theta_0 &= -\left(-Q^{-1} + \tfrac{1}{\sqrt{n}}V\right)^{-1}\tfrac{1}{\sqrt{n}}J \\ &\quad - \tfrac{1}{2}\left(-Q^{-1} + \tfrac{1}{\sqrt{n}}V\right)^{-1}\left(H_2 + \tfrac{1}{\sqrt{n}}W\right)\left(\left(\widehat{\theta} - \theta_0\right) \otimes \left(\widehat{\theta} - \theta_0\right)\right) + O_p(n^{-3/2}) \\ &= -\left(-Q - \tfrac{1}{\sqrt{n}}QVQ + O_p(n^{-1})\right)\tfrac{1}{\sqrt{n}}J \\ &\quad - \tfrac{1}{2}\left(-Q + O_p(\tfrac{1}{\sqrt{n}})\right)\left(H_2 + \tfrac{1}{\sqrt{n}}W\right)\left(\left(\widehat{\theta} - \theta_0\right) \otimes \left(\widehat{\theta} - \theta_0\right)\right) + O_p(n^{-3/2}) \\ &= \left(\tfrac{1}{\sqrt{n}}QJ + \tfrac{1}{n}QVQJ\right) + \tfrac{1}{2}QH_2\left(\left(\widehat{\theta} - \theta_0\right) \otimes \left(\widehat{\theta} - \theta_0\right)\right) + O_p(n^{-3/2}). \end{aligned} \tag{80}$$

Now plugging $\sqrt{n}\left(\widehat{\theta} - \theta_0\right) = QJ + O_p\left(1/\sqrt{n}\right)$ in (80), we obtain

$$\begin{aligned} \widehat{\theta} - \theta_0 &= \left(\tfrac{1}{\sqrt{n}}QJ + \tfrac{1}{n}QVQJ\right) + \tfrac{1}{n}\tfrac{1}{2}QH_2\left(\sqrt{n}\left(\widehat{\theta} - \theta_0\right) \otimes \sqrt{n}\left(\widehat{\theta} - \theta_0\right)\right) + O_p(n^{-3/2}) \\ &= \left(\tfrac{1}{\sqrt{n}}QJ + \tfrac{1}{n}QVQJ\right) + \tfrac{1}{n}\tfrac{1}{2}QH_2\left(\left(QJ + O_p\left(1/\sqrt{n}\right)\right) \otimes \left(QJ + O_p\left(1/\sqrt{n}\right)\right)\right) + O_p(n^{-3/2}) \\ &= \left(\tfrac{1}{\sqrt{n}}QJ + \tfrac{1}{n}QVQJ\right) + \tfrac{1}{n}\tfrac{1}{2}QH_2\left(QJ \otimes QJ\right) + O_p(n^{-3/2}). \end{aligned} \tag{81}$$

By rearranging (81), we have $\sqrt{n}\left(\widehat{\theta} - \theta_0\right) = QJ + \frac{1}{\sqrt{n}}Q\left(VQJ + \frac{1}{2}H_2\left(QJ \otimes QJ\right)\right) + O_p(n^{-1})$. ■

B.2 Lemma 2.2

Proof. Consider

$$\begin{array}{l} E[VQJ] = E\left[\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(\nabla s(z_i,\theta_0) - E\left[\nabla s(z_i,\theta_0)\right]\right)\right)Q\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}s(z_i,\theta_0)\right)\right] \\ = E\left[\frac{1}{n}\sum_{i=1}^{n}\left(\nabla s(z_i,\theta_0) - E\left[\nabla s(z_i,\theta_0)\right]\right)Qs(z_i,\theta_0)\right] + \frac{1}{n}\sum_{i\neq j}^{n}E\left[\left(\nabla s(z_i,\theta_0) - E\left[\nabla s(z_i,\theta_0)\right]\right)Qs(z_j,\theta_0)\right] \end{array}$$

Now noting for $i \neq j$, $\nabla s(z_i,\theta_0)$ and $s(z_j,\theta_0)$ are independent by iid assumption, we have

$$\begin{aligned} &\frac{1}{n}\sum_{i\neq j}^{n}E\left[\left(\nabla s(z_i,\theta_0) - E\left[\nabla s(z_i,\theta_0)\right]\right)Qs(z_j,\theta_0)\right] \\ = \;&\frac{1}{n}\sum_{i\neq j}^{n}E\left[\nabla s(z_i,\theta_0) - E\left[\nabla s(z_i,\theta_0)\right]\right]QE\left[s(z_j,\theta_0)\right] = 0 \end{aligned}$$

and hence $E[VQJ] = E[v_i d_i]$. Similarly

$$\begin{aligned}
&E[QJ \otimes QJ] = E\left[Q\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n} s(z_i,\theta_0)\right) \otimes Q\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n} s(z_i,\theta_0)\right)\right] \\
&= E\left[\frac{1}{n}\sum_{i=1}^{n} Qs(z_i,\theta_0) \otimes Qs(z_i,\theta_0)\right] + \frac{1}{n}\sum_{i\neq j}^{n} E[Qs(z_i,\theta_0) \otimes Qs(z_j,\theta_0)] \\
&= E[Qs(z_i,\theta_0) \otimes Qs(z_i,\theta_0)] = E[(d_i \otimes d_i)],
\end{aligned} \tag{82}$$

since $E[Qs(z_i,\theta_0) \otimes Qs(z_j,\theta_0)] = QE[s(z_i,\theta_0)] \otimes QE[s(z_j,\theta_0)] = 0$ for $i \neq j$ by (1). This completes the proof. ■

B.3 Proposition 3.1

Proof. By the first order Taylor series approximation of (9), we have

$$0 = \frac{1}{n}\sum_{i=1}^{n} s(z_i,\theta_0) + \left(\frac{1}{n}\sum_{i=1}^{n} \nabla s\left(z_i,\widetilde{\theta}\right)\right)\left(\theta^* - \theta_0\right) - \frac{1}{n}\widehat{c}(\theta_0) - \frac{1}{n}\nabla\widehat{c}\left(\widetilde{\theta}\right)\left(\theta^* - \theta_0\right)$$

for $\widetilde{\theta}$ between θ^* and θ_0 and hence

$$\begin{aligned}
&\sqrt{n}\left(\theta^* - \theta_0\right) \\
&= -\left(\frac{1}{n}\sum_{i=1}^{n} \nabla s\left(z_i,\widetilde{\theta}\right) - \frac{1}{n}\nabla\widehat{c}(\widetilde{\theta})\right)^{-1}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n} s(z_i,\theta_0) - \frac{1}{\sqrt{n}}\widehat{c}(\theta_0)\right) \\
&= -\left(E[\nabla s(z_i,\theta_0)] + o_p(1) + O_p\left(\tfrac{1}{n}\right)\right)^{-1}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n} s(z_i,\theta_0) + O_p\left(\tfrac{1}{\sqrt{n}}\right)\right) = QJ + o_p(1),
\end{aligned} \tag{83}$$

by Condition 1(i), 2 and $\widetilde{\theta} = \theta_0 + o_p(1)$ provided that $\left\|\frac{1}{n}\sum_{i=1}^{n} \nabla s\left(z_i,\overline{\theta}\right) - E[\nabla s(z_i,\theta_0)]\right\| = o_p(1)$ for $\overline{\theta} = \theta_0 + o_p(1)$. This confirms that the estimator has the same first order asymptotic distribution as $\sqrt{n}(\widehat{\theta} - \theta_0)$ in (71). Recalling $\widehat{H}_1(\theta) \equiv \frac{1}{n}\sum_{i=1}^{n} \nabla s(z_i,\theta)$ and $H_1(\theta_0)(= -Q^{-1}) \equiv E[\nabla s(z_i,\theta_0)]$, we can rewrite (83) as

$$\begin{aligned}
&\sqrt{n}\left(\theta^* - \theta_0\right) \\
&= -\left(H_1(\theta_0) - \tfrac{1}{n}\nabla\widehat{c}(\theta_0)\right)^{-1}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n} s(z_i,\theta_0) - \frac{1}{\sqrt{n}}\widehat{c}(\theta_0)\right) \\
&\quad - \left(\left(\widehat{H}_1(\widetilde{\theta}) - \tfrac{1}{n}\nabla\widehat{c}(\widetilde{\theta})\right)^{-1} - \left(H_1(\theta_0) - \tfrac{1}{n}\nabla\widehat{c}(\theta_0)\right)^{-1}\right)\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n} s(z_i,\theta_0) - \frac{1}{\sqrt{n}}\widehat{c}(\theta_0)\right) \\
&= -\left(H_1(\theta_0) + O_p(1/n)\right)^{-1}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n} s(z_i,\theta_0) + O_p(1/\sqrt{n})\right) \\
&\quad - \left(\left(\widehat{H}_1(\widetilde{\theta}) + O_p(1/n)\right)^{-1} - \left(H_1(\theta_0) + O_p\left(\tfrac{1}{n}\right)\right)^{-1}\right)\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n} s(z_i,\theta_0) + O_p(1/\sqrt{n})\right) \\
&= -\left(H_1(\theta_0)^{-1} + O_p(1/n)\right)\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n} s(z_i,\theta_0) + O_p(1/\sqrt{n})\right) \\
&\quad - \left(\widehat{H}_1(\widetilde{\theta})^{-1} - H_1(\theta_0)^{-1} + O_p(1/n)\right)\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n} s(z_i,\theta_0) + O_p(1/\sqrt{n})\right) \\
&= -H_1(\theta_0)^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^{n} s(z_i,\theta_0) - \left(\widehat{H}_1(\widetilde{\theta})^{-1} - H_1(\theta_0)^{-1}\right)\frac{1}{\sqrt{n}}\sum_{i=1}^{n} s(z_i,\theta_0) + O_p(1/n) \\
&= -H_1(\theta_0)^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^{n} s(z_i,\theta_0) + O_p(1/n),
\end{aligned}$$

where the second inequality is by Condition 2 and the last equality is obtained by $\left(\widehat{H}_1(\widetilde{\theta})^{-1} - H_1(\theta_0)^{-1}\right) = O_p(1/\sqrt{n})$ from (75) and $\frac{1}{\sqrt{n}}\sum_{i=1}^{n} s(z_i,\theta_0) = O_p(1)$ and hence we have

$$\sqrt{n}\left(\theta^* - \theta_0\right) = QJ + O_p\left(1/\sqrt{n}\right). \tag{84}$$

This implies that θ^* and $\widehat{\theta}$ have the same first order asymptotics. In order to analyze the higher order asymptotic distribution, we make a second order Taylor series expansion:

$$0 = \left\{\begin{array}{l}
\frac{1}{n}\sum_{i=1}^{n} s(z_i,\theta_0) + \left(\frac{1}{n}\sum_{i=1}^{n} \nabla s(z_i,\theta_0)\right)\left(\theta^* - \theta_0\right) \\
+\frac{1}{2}\left(\frac{1}{n}\sum_{i=1}^{n} \nabla^2 s\left(z_i,\widetilde{\theta}\right)\right)\left(\left(\theta^* - \theta_0\right) \otimes \left(\theta^* - \theta_0\right)\right) \\
-\frac{1}{n}\widehat{c}(\theta_0) - \frac{1}{n}\nabla\widehat{c}(\theta_0)\left(\theta^* - \theta_0\right) - \frac{1}{2n}\nabla^2\widehat{c}\left(\widetilde{\theta}\right)\left(\left(\theta^* - \theta_0\right) \otimes \left(\theta^* - \theta_0\right)\right)
\end{array}\right\}. \tag{85}$$

As we rewrite (77) into the form of (78) in Lemma 2.1's proof, we rewrite (85) as

$$\begin{aligned} 0 &= \left\{ \begin{array}{l} \frac{1}{\sqrt{n}}J + \left(-Q^{-1} + \frac{1}{\sqrt{n}}V\right)\left(\theta^{*} - \theta_0\right) \\ +\frac{1}{2}\left(H_2 + \frac{1}{\sqrt{n}}W\right)\left(\left(\theta^{*} - \theta_0\right) \otimes \left(\theta^{*} - \theta_0\right)\right) - \frac{1}{n}\widehat{c}(\theta_0) - \frac{1}{n}\nabla\widehat{c}(\theta_0)\left(\theta^{*} - \theta_0\right) \\ -\frac{1}{2n}\nabla^2\widehat{c}\left(\widetilde{\theta}\right)\left(\left(\theta^{*} - \theta_0\right) \otimes \left(\theta^{*} - \theta_0\right)\right) + O_p(n^{-3/2}) \end{array} \right\} \\ &= \left\{ \begin{array}{l} \frac{1}{\sqrt{n}}J + \left(-Q^{-1} + \frac{1}{\sqrt{n}}V\right)\left(\theta^{*} - \theta_0\right) \\ +\frac{1}{2}\left(H_2 + \frac{1}{\sqrt{n}}W\right)\left(\left(\theta^{*} - \theta_0\right) \otimes \left(\theta^{*} - \theta_0\right)\right) - \frac{1}{n}\widehat{c}(\theta_0) + O_p(n^{-3/2}) \end{array} \right\} \end{aligned} \tag{86}$$

since (a) $\frac{1}{n}\nabla\widehat{c}(\theta_0)\left(\theta^{*} - \theta_0\right) = O_p(n^{-3/2})$ by Condition 2 and $\theta^{*} = \theta_0 + O_p(\frac{1}{\sqrt{n}})$ from (83) noting $J = O_p(1)$ and since (b)

$$\begin{aligned} &\left\|\tfrac{1}{2n}\nabla^2\widehat{c}\left(\widetilde{\theta}\right)\left(\left(\theta^{*} - \theta_0\right) \otimes \left(\theta^{*} - \theta_0\right)\right)\right\| \\ &\leq \tfrac{1}{2n}\left\|\nabla^2\widehat{c}\left(\widetilde{\theta}\right)\right\|\left\|\theta^{*} - \theta_0\right\|^2 = O(n^{-1})O_p(1)O_p(n^{-1}) = O_p(n^{-2}) \end{aligned}$$

by Condition 3 and $\theta^{*} = \theta_0 + O_p\left(\frac{1}{\sqrt{n}}\right)$.

From (86), by observing that θ^{*} and $\widehat{\theta}$ have the same first order asymptotics, we obtain

$$\begin{aligned} \sqrt{n}\left(\theta^{*} - \theta_0\right) &= QJ + \frac{1}{\sqrt{n}}Q\left(VQJ + \frac{1}{2}H_2\left(QJ \otimes QJ\right) - \widehat{c}(\theta_0)\right) + O_p\left(\frac{1}{n}\right) \\ &= QJ + \frac{1}{\sqrt{n}}Q\left(VQJ + \frac{1}{2}H_2\left(QJ \otimes QJ\right) - c(\theta_0)\right) + O_p\left(\frac{1}{n}\right), \end{aligned}$$

as in Lemma 2.1. The second equality comes from Condition 1 (ii) ($\widehat{c}(\theta_0) = c(\theta_0) + O_p\left(1/\sqrt{n}\right)$) and thus the second-order bias $Bias(\theta^{*}) \equiv \frac{1}{n}E\left[Q\left(VQJ + \frac{1}{2}H_2\left(QJ \otimes QJ\right) - c(\theta_0)\right)\right] = 0$ since (noting $Q \equiv Q(\theta_0)$ and $H_2 \equiv H_2(\theta_0)$)

$$\begin{aligned} &E\left[VQJ + \tfrac{1}{2}H_2\left(QJ \otimes QJ\right)\right] \\ &= E\left[\nabla s\left(z_i, \theta_0\right)Q(\theta_0)s(z_i, \theta_0)\right] + \tfrac{1}{2}H_2E\left[Q(\theta_0)s(z_i, \theta_0) \otimes Q\left(\theta_0\right)s(z_i, \theta_0)\right] = c\left(\theta_0\right) \end{aligned}$$

by definition of $c(\theta)$ in (6) and Lemma 2.2. ∎

B.4 Lemma 2.3

Proof. Consider a higher order Taylor expansion of (2) around the true value of $\theta = \theta_0$ up to the third order as

$$\begin{aligned} 0 = &\tfrac{1}{n}\textstyle\sum_{i=1}^{n} s\left(z_i, \theta_0\right) + \left(\frac{1}{n}\sum_{i=1}^{n}\nabla s\left(z_i, \theta_0\right)\right)\left(\widehat{\theta} - \theta_0\right) + \frac{1}{2}\left(\frac{1}{n}\sum_{i=1}^{n}\nabla^2 s\left(z_i, \theta_0\right)\right)\left(\left(\widehat{\theta} - \theta_0\right) \otimes \left(\widehat{\theta} - \theta_0\right)\right) \\ &+ \tfrac{1}{6}\left(\tfrac{1}{n}\textstyle\sum_{i=1}^{n}\nabla^3 s\left(z_i, \widetilde{\theta}\right)\right)\left(\left(\widehat{\theta} - \theta_0\right) \otimes \left(\widehat{\theta} - \theta_0\right) \otimes \left(\widehat{\theta} - \theta_0\right)\right), \end{aligned}$$

where $\widetilde{\theta}$ lies between θ_0 and $\widehat{\theta}$. Now by the stochastic equicontinuity Condition (iii) and $\widetilde{\theta} = \theta_0 + o_p(1)$, we have

$$\frac{1}{\sqrt{n}}\textstyle\sum_{i=1}^{n}\left(\nabla^3 s\left(z_i, \widetilde{\theta}\right) - H_3\left(\widetilde{\theta}\right)\right) - \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(\nabla^3 s\left(z_i, \theta_0\right) - H_3(\theta_0)\right) = o_p\left(1\right)$$

and hence

$$\begin{aligned} &\left(\tfrac{1}{n}\textstyle\sum_{i=1}^{n}\nabla^3 s\left(z_i, \widetilde{\theta}\right) - \frac{1}{n}\sum_{i=1}^{n}\nabla^3 s\left(z_i, \theta_0\right)\right)\left(\left(\widehat{\theta} - \theta_0\right) \otimes \left(\widehat{\theta} - \theta_0\right) \otimes \left(\widehat{\theta} - \theta_0\right)\right) \\ &= \left(H_3\left(\widetilde{\theta}\right) - H_3\left(\theta_0\right) + o_p\left(\tfrac{1}{\sqrt{n}}\right)\right)\left(\left(\widehat{\theta} - \theta_0\right) \otimes \left(\widehat{\theta} - \theta_0\right) \otimes \left(\widehat{\theta} - \theta_0\right)\right) \\ &= \left(\nabla\left(E\left[\nabla^3 s\left(z_i, \theta\right)\right]\right)\big|_{\theta = \widetilde{\widetilde{\theta}}}\left(\widetilde{\theta} - \theta_0\right) + o_p\left(\tfrac{1}{\sqrt{n}}\right)\right)\left(\left(\widehat{\theta} - \theta_0\right) \otimes \left(\widehat{\theta} - \theta_0\right) \otimes \left(\widehat{\theta} - \theta_0\right)\right) \\ &= \left(E\left[\nabla^4 s\left(z_i, \theta_0\right)\right]\left(\widetilde{\theta} - \theta_0\right) + o_p\left(\tfrac{1}{\sqrt{n}}\right)\right)\left(\left(\widehat{\theta} - \theta_0\right) \otimes \left(\widehat{\theta} - \theta_0\right) \otimes \left(\widehat{\theta} - \theta_0\right)\right) = O_p(n^{-2}), \end{aligned}$$

applying the mean value theorem where $\widetilde{\widetilde{\theta}}$ lies between $\widetilde{\theta}$ and θ_0 and from standard results on differentiating inside the integral. The second last equality is from the continuity of $E\left[\nabla^4 s\left(z_i,\theta_0\right)\right]$ at θ_0 and since $\widetilde{\widetilde{\theta}}=\theta_0+o_p\left(1\right)$. We, thus, obtain

$$0=\left\{\begin{array}{l}\frac{1}{n}\sum_{i=1}^{n}s\left(z_i,\theta_0\right)+\left(\frac{1}{n}\sum_{i=1}^{n}\nabla s\left(z_i,\theta_0\right)\right)\left(\widehat{\theta}-\theta_0\right)\\+\frac{1}{2}\left(\frac{1}{n}\sum_{i=1}^{n}\nabla^2 s\left(z_i,\theta_0\right)\right)\left(\left(\widehat{\theta}-\theta_0\right)\otimes\left(\widehat{\theta}-\theta_0\right)\right)\\+\frac{1}{6}\left(\frac{1}{n}\sum_{i=1}^{n}\nabla^3 s\left(z_i,\theta_0\right)\right)\left(\left(\widehat{\theta}-\theta_0\right)\otimes\left(\widehat{\theta}-\theta_0\right)\otimes\left(\widehat{\theta}-\theta_0\right)\right)+O_p(n^{-2})\end{array}\right\}. \tag{87}$$

Now note

$$\begin{array}{l}\frac{1}{n}\sum_{i=1}^{n}\nabla s\left(z_i,\theta_0\right)=-Q^{-1}+\frac{1}{\sqrt{n}}V,\ \frac{1}{n}\sum_{i=1}^{n}\nabla^2 s\left(z_i,\theta_0\right)=H_2+\frac{1}{\sqrt{n}}W\\\frac{1}{n}\sum_{i=1}^{n}\nabla^3 s\left(z_i,\theta_0\right)=H_3+\frac{1}{\sqrt{n}}W_3\text{ with }W_3\equiv\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(\nabla^3 s\left(z_i,\theta_0\right)-E\left[\nabla^3 s\left(z_i,\theta_0\right)\right]\right)=O_p(1).\end{array}$$

We then rewrite (87) as

$$0=\left\{\begin{array}{l}\frac{1}{\sqrt{n}}J+\left(-Q^{-1}+\frac{1}{\sqrt{n}}V\right)\left(\widehat{\theta}-\theta_0\right)+\frac{1}{2}\left(H_2+\frac{1}{\sqrt{n}}W\right)\left(\left(\widehat{\theta}-\theta_0\right)\otimes\left(\widehat{\theta}-\theta_0\right)\right)\\+\frac{1}{6}\left(H_3+\frac{1}{\sqrt{n}}W_3\right)\left(\left(\widehat{\theta}-\theta_0\right)\otimes\left(\widehat{\theta}-\theta_0\right)\otimes\left(\widehat{\theta}-\theta_0\right)\right)+O_p(n^{-2})\end{array}\right\}. \tag{88}$$

Plugging (79) into (88) (depending the orders we need) and inspecting the orders, we have

$$\widehat{\theta}-\theta_0 \tag{89}$$

$$\begin{aligned}&=\left\{\begin{array}{l}-\left(-Q^{-1}+\frac{1}{\sqrt{n}}V\right)^{-1}\frac{1}{\sqrt{n}}J\\-\frac{1}{2}\left(-Q^{-1}+\frac{1}{\sqrt{n}}V\right)^{-1}\left(H_2+\frac{1}{\sqrt{n}}W\right)\left(\left(\widehat{\theta}-\theta_0\right)\otimes\left(\widehat{\theta}-\theta_0\right)\right)\\-\frac{1}{6}\left(-Q^{-1}+\frac{1}{\sqrt{n}}V\right)^{-1}\left(H_3+\frac{1}{\sqrt{n}}W_3\right)\left(\left(\widehat{\theta}-\theta_0\right)\otimes\left(\widehat{\theta}-\theta_0\right)\otimes\left(\widehat{\theta}-\theta_0\right)\right)+O_p(n^{-2})\end{array}\right\}\\&=\left\{\begin{array}{l}-\left(-Q-\frac{1}{\sqrt{n}}QVQ-\frac{1}{n}QVQVQ+O_p(n^{-3/2})\right)\frac{1}{\sqrt{n}}J\\-\frac{1}{2}\left(-Q-\frac{1}{\sqrt{n}}QVQ+O_p(n^{-1})\right)\left(H_2+\frac{1}{\sqrt{n}}W\right)\left(\left(\widehat{\theta}-\theta_0\right)\otimes\left(\widehat{\theta}-\theta_0\right)\right)\\-\frac{1}{6}\left(-Q+O_p\left(\frac{1}{\sqrt{n}}\right)\right)\left(H_3+\frac{1}{\sqrt{n}}W_3\right)\left(\left(\widehat{\theta}-\theta_0\right)\otimes\left(\widehat{\theta}-\theta_0\right)\otimes\left(\widehat{\theta}-\theta_0\right)\right)+O_p(n^{-2})\end{array}\right\}\\&=\left\{\begin{array}{l}\left(\frac{1}{\sqrt{n}}QJ+\frac{1}{n}QVQJ+\frac{1}{n^{3/2}}QVQVQJ\right)\\+\frac{1}{2}\left(Q\left(H_2+\frac{1}{\sqrt{n}}W\right)+\frac{1}{\sqrt{n}}QVQH_2\right)\left(\left(\widehat{\theta}-\theta_0\right)\otimes\left(\widehat{\theta}-\theta_0\right)\right)\\+\frac{1}{6}QH_3\left(\left(\widehat{\theta}-\theta_0\right)\otimes\left(\widehat{\theta}-\theta_0\right)\otimes\left(\widehat{\theta}-\theta_0\right)\right)+O_p(n^{-2})\end{array}\right\}.\end{aligned} \tag{90}$$

Now plugging $\sqrt{n}(\widehat{\theta}-\theta_0)=a_{-1/2}+O_p\left(1/\sqrt{n}\right)$ or $\sqrt{n}(\widehat{\theta}-\theta_0)=a_{-1/2}+\frac{1}{\sqrt{n}}a_{-1}+O_p\left(1/n\right)$ in (90) depending on the orders required, we obtain

$$\begin{aligned}&\widehat{\theta}-\theta_0\\&=\left\{\begin{array}{l}\left(\frac{1}{\sqrt{n}}QJ+\frac{1}{n}QVQJ+\frac{1}{n^{3/2}}QVQVQJ\right)\\+\frac{1}{2}\frac{1}{n}\left(Q\left(H_2+\frac{W}{\sqrt{n}}\right)+\frac{QVQH_2}{\sqrt{n}}\right)\left(\left(a_{-1/2}+\frac{a_{-1}}{\sqrt{n}}+O_p\left(\frac{1}{n}\right)\right)\otimes\left(a_{-1/2}+\frac{a_{-1}}{\sqrt{n}}+O_p\left(\frac{1}{n}\right)\right)\right)\\+\frac{1}{6}\frac{1}{n^{3/2}}QH_3\left(\left(a_{-1/2}+O_p\left(\frac{1}{\sqrt{n}}\right)\right)\otimes\left(a_{-1/2}+O_p\left(\frac{1}{\sqrt{n}}\right)\right)\otimes\left(a_{-1/2}+O_p\left(\frac{1}{\sqrt{n}}\right)\right)\right)+O_p(n^{-2})\end{array}\right\}\\&=\left\{\begin{array}{l}\left(\frac{1}{\sqrt{n}}QJ+\frac{1}{n}QVQJ+\frac{1}{n^{3/2}}QVQVQJ\right)\\+\frac{1}{2}\frac{1}{n}\left(Q\left(H_2+\frac{1}{\sqrt{n}}W\right)+\frac{1}{\sqrt{n}}QVQH_2\right)\left(\left(a_{-1/2}+\frac{1}{\sqrt{n}}a_{-1}\right)\otimes\left(a_{-1/2}+\frac{1}{\sqrt{n}}a_{-1}\right)\right)\\+\frac{1}{6}\frac{1}{n^{3/2}}QH_3\left(a_{-1/2}\otimes a_{-1/2}\otimes a_{-1/2}\right)+O_p(n^{-2})\end{array}\right\} \qquad (91)\\&=\left\{\begin{array}{l}\left(\frac{1}{\sqrt{n}}QJ+\frac{1}{n}QVQJ+\frac{1}{n^{3/2}}QVQVQJ\right)+\frac{1}{2}\frac{1}{n}QH_2\left(a_{-1/2}\otimes a_{-1/2}\right)\\+\frac{1}{2}\frac{1}{n^{3/2}}\left(QH_2\left(\left(a_{-1/2}\otimes a_{-1}\right)+\left(a_{-1}\otimes a_{-1/2}\right)\right)+\left(QW+QVQH_2\right)\left(a_{-1/2}\otimes a_{-1/2}\right)\right)\\+\frac{1}{6}\frac{1}{n^{3/2}}QH_3\left(a_{-1/2}\otimes a_{-1/2}\otimes a_{-1/2}\right)+O_p(n^{-2})\end{array}\right\}.\end{aligned}$$

Finally, rearranging (91) according to the orders, we have

$$
\begin{aligned}
&\widehat{\theta}-\theta_0\\
&= \left\{ \begin{array}{l} \frac{1}{\sqrt{n}}QJ + \frac{1}{n}Q\left(VQJ + \frac{1}{2}H_2\left(a_{-1/2}\otimes a_{-1/2}\right)\right) \\ +\frac{1}{n^{3/2}}\left(QVQ\left(VQJ + \frac{1}{2}H_2\left(a_{-1/2}\otimes a_{-1/2}\right)\right) + \frac{1}{2}QW\left(a_{-1/2}\otimes a_{-1/2}\right)\right) \\ +\frac{1}{n^{3/2}}\left(\frac{1}{2}QH_2\left(\left(a_{-1/2}\otimes a_{-1}\right) + \left(a_{-1}\otimes a_{-1/2}\right)\right) + \frac{1}{6}QH_3\left(a_{-1/2}\otimes a_{-1/2}\otimes a_{-1/2}\right)\right) + O_p(n^{-2}) \end{array} \right\}\\
&= \left\{ \begin{array}{l} \frac{1}{\sqrt{n}}a_{-1/2} + \frac{1}{n}a_{-1} \\ +\frac{1}{n^{3/2}}\left(QVa_{-1} + \frac{1}{2}QW\left(a_{-1/2}\otimes a_{-1/2}\right) + \frac{1}{2}QH_2\left(\left(a_{-1/2}\otimes a_{-1}\right) + \left(a_{-1}\otimes a_{-1/2}\right)\right)\right) \\ +\frac{1}{n^{3/2}}\left(\frac{1}{6}QH_3\left(a_{-1/2}\otimes a_{-1/2}\otimes a_{-1/2}\right)\right) + O_p(n^{-2}) \end{array} \right\}.
\end{aligned}
$$

■

B.5 Proposition 3.2

Proof. Now consider a third order Taylor series expansion of $0 = \frac{1}{n}\sum_{i=1}^{n} s\left(z_i, \theta^*\right) - \frac{1}{n}\widehat{c}(\theta^*)$:

$$
\begin{aligned}
0 = &\frac{1}{n}\sum_{i=1}^{n} s\left(z_i,\theta_0\right) + \left(\frac{1}{n}\sum_{i=1}^{n}\nabla s\left(z_i,\theta_0\right)\right)\left(\theta^* - \theta_0\right)\\
&+\frac{1}{2}\left(\frac{1}{n}\sum_{i=1}^{n}\nabla^2 s\left(z_i,\theta_0\right)\right)\left(\left(\theta^*-\theta_0\right)\otimes\left(\theta^*-\theta_0\right)\right)\\
&+\frac{1}{6}\left(\frac{1}{n}\sum_{i=1}^{n}\nabla^3 s\left(z_i,\widetilde{\theta}\right)\right)\left(\left(\theta^*-\theta_0\right)\otimes\left(\theta^*-\theta_0\right)\otimes\left(\theta^*-\theta_0\right)\right)\\
&-\frac{1}{n}\widehat{c}(\theta_0) - \frac{1}{n}\nabla\widehat{c}\left(\theta_0\right)\left(\theta^*-\theta_0\right) - \frac{1}{2}\frac{1}{n}\nabla^2\widehat{c}\left(\theta_0\right)\left(\left(\theta^*-\theta_0\right)\otimes\left(\theta^*-\theta_0\right)\right)\\
&-\frac{1}{6}\frac{1}{n}\nabla^3\widehat{c}\left(\widetilde{\theta}\right)\left(\left(\theta^*-\theta_0\right)\otimes\left(\theta^*-\theta_0\right)\otimes\left(\theta^*-\theta_0\right)\right)
\end{aligned}
$$

From this, similarly with (87) to (88), we obtain

$$
\begin{aligned}
0 = &\frac{1}{\sqrt{n}}J + \left(-Q^{-1} + \frac{1}{\sqrt{n}}V\right)\left(\theta^*-\theta_0\right) + \frac{1}{2}\left(H_2 + \frac{1}{\sqrt{n}}W\right)\left(\left(\theta^*-\theta_0\right)\otimes\left(\theta^*-\theta_0\right)\right)\\
&-\frac{1}{6}\left(H_3 + \frac{1}{\sqrt{n}}W_3\right)\left(\left(\theta^*-\theta_0\right)\otimes\left(\theta^*-\theta_0\right)\otimes\left(\theta^*-\theta_0\right)\right) - \frac{1}{n}\widehat{c}(\theta_0) - \frac{1}{n}\nabla\widehat{c}\left(\theta_0\right)\left(\theta^*-\theta_0\right) + O_p(n^{-2}),
\end{aligned}
$$

since $\frac{1}{2}\frac{1}{n}\nabla^2\widehat{c}\left(\theta_0\right)\left(\left(\theta^*-\theta_0\right)\otimes\left(\theta^*-\theta_0\right)\right) = O_p(n^{-2})$ by Condition 3 and $\theta^* = \theta_0 + O_p(\frac{1}{\sqrt{n}})$ and since

$$
\begin{aligned}
&\left\|\frac{1}{2}\frac{1}{n}\nabla^3\widehat{c}\left(\widetilde{\theta}\right)\left(\left(\theta^*-\theta_0\right)\otimes\left(\theta^*-\theta_0\right)\otimes\left(\theta^*-\theta_0\right)\right)\right\|\\
&\leq \frac{1}{2}\frac{1}{n}\left\|\nabla^3\widehat{c}\left(\widetilde{\theta}\right)\right\|\left\|\theta^*-\theta_0\right\|^3 = O(n^{-1})O_p(1)O_p(n^{-3/2}) = O_p(n^{-5/2})
\end{aligned}
$$

by Condition 4 (ii) and $\theta^* = \theta_0 + O_p(\frac{1}{\sqrt{n}})$. Similarly with (89), we obtain

$$
\begin{aligned}
\theta^* - \theta_0 &= \left\{ \begin{array}{l} -\left(-Q - \frac{1}{\sqrt{n}}QVQ - \frac{1}{n}QVQVQ + O_p(n^{-3/2})\right)\frac{1}{\sqrt{n}}J \\ -\frac{1}{2}\left(-Q - \frac{1}{\sqrt{n}}QVQ + O_p(n^{-1})\right)\left(H_2 + \frac{1}{\sqrt{n}}W\right)\left(\left(\theta^*-\theta_0\right)\otimes\left(\theta^*-\theta_0\right)\right) \\ -\frac{1}{6}\left(-Q + O_p\left(\frac{1}{\sqrt{n}}\right)\right)\left(H_3 + \frac{1}{\sqrt{n}}W_3\right)\left(\left(\theta^*-\theta_0\right)\otimes\left(\theta^*-\theta_0\right)\otimes\left(\theta^*-\theta_0\right)\right) \\ +\frac{1}{n}\left(-Q - \frac{1}{\sqrt{n}}QVQ + O_p(n^{-1})\right)\widehat{c}(\theta_0) + \frac{1}{n}\left(-Q + O_p\left(\frac{1}{\sqrt{n}}\right)\right)\nabla\widehat{c}\left(\theta_0\right)\left(\theta^*-\theta_0\right) + O_p(n^{-2}) \end{array} \right\}\\
&= \left\{ \begin{array}{l} \left(\frac{1}{\sqrt{n}}QJ + \frac{1}{n}QVQJ + \frac{1}{n^{3/2}}QVQVQJ\right) \\ +\frac{1}{2}\left(Q\left(H_2 + \frac{1}{\sqrt{n}}W\right) + \frac{QVQH_2}{\sqrt{n}}\right)\left(\left(\theta^*-\theta_0\right)\otimes\left(\theta^*-\theta_0\right)\right) \\ +\frac{1}{6}QH_3\left(\left(\theta^*-\theta_0\right)\otimes\left(\theta^*-\theta_0\right)\otimes\left(\theta^*-\theta_0\right)\right) \\ -\frac{1}{n}\left(Q + \frac{1}{\sqrt{n}}QVQ\right)\widehat{c}(\theta_0) - \frac{1}{n}Q\nabla\widehat{c}\left(\theta_0\right)\left(\theta^*-\theta_0\right) + O_p(n^{-2}). \end{array} \right\} \qquad (92)
\end{aligned}
$$

Now replacing $\sqrt{n}(\theta^* - \theta_0) = a_{-1/2} + O_p\left(\frac{1}{\sqrt{n}}\right)$ or $\sqrt{n}(\theta^* - \theta_0) = a_{-1/2} + \frac{1}{\sqrt{n}}\left(a_{-1} - Qc(\theta_0)\right) + O_p\left(\frac{1}{n}\right)$ in (92) depending on the orders required, we obtain

$$
\begin{aligned}
&\theta^* - \theta_0 \\
= &\left\{ \begin{array}{l}
\left(\frac{1}{\sqrt{n}}QJ + \frac{1}{n}QVQJ + \frac{1}{n^{3/2}}QVQVQJ\right) \\
+\frac{1}{2n}\left(Q\left(H_2 + \frac{1}{\sqrt{n}}W\right) + \frac{QVQH_2}{\sqrt{n}}\right)\left(\begin{array}{c} \left(a_{-1/2} + \frac{a_{-1}-Qc(\theta_0)}{\sqrt{n}} + O_p\left(n^{-1}\right)\right) \\ \otimes\left(a_{-1/2} + \frac{a_{-1}-Qc(\theta_0)}{\sqrt{n}} + O_p\left(n^{-1}\right)\right) \end{array}\right) \\
+\frac{1}{6n^{3/2}}QH_3\left(\left(a_{-1/2} + O_p\left(\frac{1}{\sqrt{n}}\right)\right) \otimes \left(a_{-1/2} + O_p\left(\frac{1}{\sqrt{n}}\right)\right) \otimes \left(a_{-1/2} + O_p\left(\frac{1}{\sqrt{n}}\right)\right)\right) \\
-\frac{1}{n}\left(Q + \frac{1}{\sqrt{n}}QVQ\right)\widehat{c}(\theta_0) - \frac{1}{n^{3/2}}Q\nabla\widehat{c}(\theta_0)\left(a_{-1/2} + O_p\left(\frac{1}{\sqrt{n}}\right)\right) + O_p(n^{-2})
\end{array} \right\} \\
= &\left\{ \begin{array}{l}
\left(\frac{1}{\sqrt{n}}QJ + \frac{1}{n}QVQJ + \frac{1}{n^{3/2}}QVQVQJ\right) \\
+\frac{1}{2n}\left(Q\left(H_2 + \frac{W}{\sqrt{n}}\right) + \frac{QVQH_2}{\sqrt{n}}\right)\left(\left(a_{-1/2} + \frac{a_{-1}-Qc(\theta_0)}{\sqrt{n}}\right) \otimes \left(a_{-1/2} + \frac{a_{-1}-Qc(\theta_0)}{\sqrt{n}}\right)\right) \\
+\frac{1}{6n^{3/2}}QH_3\left(a_{-1/2} \otimes a_{-1/2} \otimes a_{-1/2}\right) - \frac{1}{n}\left(Q + \frac{1}{\sqrt{n}}QVQ\right)\widehat{c}(\theta_0) - \frac{1}{n^{3/2}}Q\nabla c(\theta_0)a_{-1/2} \\
+O_p(n^{-2}),
\end{array} \right\}
\end{aligned}
$$

where we replaced $\nabla\widehat{c}(\theta_0)$ with $\nabla c(\theta_0) + O_p\left(\frac{1}{\sqrt{n}}\right)$ from Condition 4 (i). Rearranging terms according to the orders, we have

$$
\begin{aligned}
&\theta^* - \theta_0 \qquad (93) \\
= &\left\{ \begin{array}{l}
\frac{1}{\sqrt{n}}QJ + \frac{1}{n}\left(QVQJ + \frac{1}{2}QH_2(a_{-1/2} \otimes a_{-1/2}) - Q\widehat{c}(\theta_0)\right) \\
+\frac{1}{n^{3/2}}\left(\begin{array}{l} QVQVQJ + \frac{1}{2}QH_2\left(\left(a_{-1/2} \otimes (a_{-1} - Qc(\theta_0))\right) + \left((a_{-1} - Qc(\theta_0)) \otimes a_{-1/2}\right)\right) \\ +\frac{1}{2}(QW + QVQH_2)\left(a_{-1/2} \otimes a_{-1/2}\right) \\ +\frac{1}{6}QH_3\left(a_{-1/2} \otimes a_{-1/2} \otimes a_{-1/2}\right) - QVQ\widehat{c}(\theta_0) - Q\nabla c(\theta_0)a_{-1/2} \end{array}\right) \\
+O_p(n^{-2})
\end{array} \right\} \\
= &\left\{ \begin{array}{l}
\frac{1}{\sqrt{n}}QJ + \frac{1}{n}\left(QVQJ + \frac{1}{2}QH_2(a_{-1/2} \otimes a_{-1/2}) - Qc(\theta_0)\right) \\
+\frac{1}{n^{3/2}}\left(\begin{array}{l} QVa_{-1} + \frac{1}{2}QH_2\left(\left(a_{-1/2} \otimes (a_{-1} - Qc(\theta_0))\right) + \left((a_{-1} - Qc(\theta_0)) \otimes a_{-1/2}\right)\right) \\ +\frac{1}{2}QW\left(a_{-1/2} \otimes a_{-1/2}\right) + \frac{1}{6}QH_3\left(a_{-1/2} \otimes a_{-1/2} \otimes a_{-1/2}\right) \\ -QVQ\left(c(\theta_0) + O_p\left(\frac{1}{\sqrt{n}}\right)\right) - Q\nabla c(\theta_0)a_{-1/2} - \sqrt{n}Q(\widehat{c}(\theta_0) - c(\theta_0)) \end{array}\right) + O_p(n^{-2})
\end{array} \right\} \\
= &\left\{ \begin{array}{l}
\frac{1}{\sqrt{n}}QJ + \frac{1}{n}\left(a_{-1} - Qc(\theta_0)\right) \\
+\frac{1}{n^{3/2}}\left(\begin{array}{l} a_{-3/2} - \frac{1}{2}QH_2\left(\left(a_{-1/2} \otimes Qc(\theta_0)\right) + \left(Qc(\theta_0) \otimes a_{-1/2}\right)\right) \\ -QVQc(\theta_0) - Q\nabla c(\theta_0)a_{-1/2} - \sqrt{n}Q(\widehat{c}(\theta_0) - c(\theta_0)) \end{array}\right) + O_p(n^{-2}),
\end{array} \right\}
\end{aligned}
$$

noting $\widehat{c}(\theta_0) = c(\theta_0) + O_p\left(\frac{1}{\sqrt{n}}\right)$.

Now we rewrite the higher order expansion of θ^* in terms of $B(\theta)$ recalling that $Q(\theta)^{-1}B(\theta) = c(\theta)$ and hence

$$
\nabla c(\theta) = Q(\theta)^{-1}\nabla B(\theta) - vec^*\left(B(\theta)'\nabla\left(H_1(\theta)'\right)\right) \qquad (94)
$$

from Remark 2. From (93), note

$$
\begin{aligned}
&\sqrt{n}\left(\theta^* - \theta_0\right) \qquad (95) \\
= &\begin{array}{l}
a_{-1/2} + \frac{1}{\sqrt{n}}\left(a_{-1} - Qc(\theta_0)\right) \\
+\frac{1}{n}\left(a_{-3/2} - \frac{1}{2}QH_2\left(\left(a_{-1/2} \otimes Qc(\theta_0)\right) + \left(Qc(\theta_0) \otimes a_{-1/2}\right)\right) - QVQc(\theta_0) - Q\nabla c(\theta_0)a_{-1/2}\right) \\
-\frac{1}{n}Q\sqrt{n}(\widehat{c}(\theta_0) - c(\theta_0)) + O_p(n^{-3/2})
\end{array}
\end{aligned}
$$

from (10) and also note that

$$\begin{aligned}
&\tfrac{1}{2}QH_2\left(\left(a_{-1/2}\otimes Qc(\theta_0)\right)+\left(Qc(\theta_0)\otimes a_{-1/2}\right)\right)+QVQc\left(\theta_0\right)+Q\nabla c\left(\theta_0\right)a_{-1/2}\\
&=\tfrac{1}{2}QH_2\left(\left(a_{-1/2}\otimes B(\theta_0)\right)+\left(B(\theta_0)\otimes a_{-1/2}\right)\right)+QVB\left(\theta_0\right)\\
&+Q\left(Q(\theta_0)^{-1}\nabla B(\theta_0)-vec^{*}\left(B(\theta_0)'\nabla\left(H_1\left(\theta_0\right)'\right)\right)\right)a_{-1/2}\\
&=\tfrac{1}{2}QH_2\left(\left(a_{-1/2}\otimes Qc(\theta_0)\right)+\left(Qc(\theta_0)\otimes a_{-1/2}\right)\right)-Qvec^{*}\left(B(\theta_0)'\nabla\left(H_1\left(\theta_0\right)'\right)\right)a_{-1/2}\\
&+\nabla B(\theta_0)a_{-1/2}+QVB\left(\theta_0\right)
\end{aligned} \tag{96}$$

from (94) and $B(\theta)=Q(\theta)c(\theta)$. We claim that

$$\frac{1}{2}H_2\left(\left(a_{-1/2}\otimes B(\theta_0)\right)+\left(B(\theta_0)\otimes a_{-1/2}\right)\right)-vec^{*}\left(B(\theta_0)'\nabla\left(H_1\left(\theta_0\right)'\right)\right)a_{-1/2}=0, \tag{97}$$

which simplifies (96) to $\nabla B(\theta_0)a_{-1/2}+QVB\left(\theta_0\right)$. This is obvious when $\dim(\theta_0)=1$, since

$$\frac{1}{2}H_2\left(\left(a_{-1/2}\otimes B(\theta_0)\right)+\left(B(\theta_0)\otimes a_{-1/2}\right)\right)=H_2B(\theta_0)a_{-1/2}$$

and $vec^{*}\left(B(\theta_0)'\nabla\left(H_1\left(\theta_0\right)'\right)\right)a_{-1/2}=B(\theta_0)H_2a_{-1/2}$ noting $\nabla\left(H_1\left(\theta_0\right)'\right)=H_2$ for the scalar case. To verify this for a general case with $\dim(\theta_0)=k$, we note $vec(AB)=(I\otimes A)vec(B)=(B'\otimes I)vec(A)$ and hence

$$\begin{aligned}
&\tfrac{1}{2}H_2\left(\left(a_{-1/2}\otimes B(\theta_0)\right)+\left(B(\theta_0)\otimes a_{-1/2}\right)\right)=\tfrac{1}{2}H_2\left(vec\left(B(\theta_0)a'_{-1/2}\right)+vec\left(a_{-1/2}B(\theta_0)'\right)\right)\\
&=\tfrac{1}{2}H_2\left(\left(I\otimes B(\theta_0)\right)a_{-1/2}+\left(B(\theta_0)\otimes I\right)a_{-1/2}\right)=\tfrac{1}{2}H_2\left(I\otimes B(\theta_0)+B(\theta_0)\otimes I\right)a_{-1/2}.
\end{aligned}$$

Thus, (97) follows upon showing (see Appendix B.6)

$$\frac{1}{2}H_2\left(I\otimes B(\theta_0)+B(\theta_0)\otimes I\right)=vec^{*}\left(B(\theta_0)'\nabla\left(H_1\left(\theta_0\right)'\right)\right). \tag{98}$$

Therefore, we can rewrite (95) as

$$\begin{aligned}
&\sqrt{n}\left(\theta^{*}-\theta_0\right)\\
&=a_{-1/2}+\tfrac{1}{\sqrt{n}}\left(a_{-1}-B(\theta_0)\right)+\tfrac{1}{n}\left(a_{-3/2}-\nabla B(\theta_0)a_{-1/2}-\sqrt{n}(\widehat{B}(\theta_0)-B(\theta_0))\right)\\
&\quad+\tfrac{1}{n}\left(\sqrt{n}(\widehat{Q}(\theta_0)-Q(\theta_0))\widehat{c}(\theta_0)-QVB\left(\theta_0\right)\right)+O_p(n^{-3/2}).
\end{aligned}$$

From Lemma B.1 below, we have $\sqrt{n}(\widehat{Q}(\theta_0)-Q(\theta_0))\widehat{c}(\theta_0)-QVB\left(\theta_0\right)=O_p\left(\frac{1}{\sqrt{n}}\right)$ and hence

$$\begin{aligned}
&\sqrt{n}\left(\theta^{*}-\theta_0\right)\\
&=\quad a_{-1/2}+\frac{1}{\sqrt{n}}\left(a_{-1}-B(\theta_0)\right)+\frac{1}{n}\left(a_{-3/2}-\nabla B(\theta_0)a_{-1/2}-\sqrt{n}(\widehat{B}(\theta_0)-B(\theta_0))\right)+O_p(n^{-3/2}).
\end{aligned} \tag{99}$$

This completes the proof. ■

Lemma B.1 *Suppose Condition 5 holds and* $V\equiv\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(\nabla s\left(z_i,\theta_0\right)-E\left[\nabla s\left(z_i,\theta_0\right)\right]\right)=O_p(1)$, *then we have* $\sqrt{n}(\widehat{Q}(\theta_0)-Q(\theta_0))\widehat{c}(\theta_0)-QVB\left(\theta_0\right)=O_p\left(1/\sqrt{n}\right)$.

Proof. From $\widehat{c}(\theta_0)=\widehat{Q}(\theta_0)^{-1}\widehat{B}\left(\theta_0\right)$, it follows that

$$\begin{aligned}
&\sqrt{n}(\widehat{Q}(\theta_0)-Q(\theta_0))\widehat{c}(\theta_0)=\sqrt{n}(\widehat{Q}(\theta_0)-Q(\theta_0))\widehat{Q}(\theta_0)^{-1}\widehat{B}\left(\theta_0\right)\\
&=Q(\theta_0)\sqrt{n}Q(\theta_0)^{-1}(\widehat{Q}(\theta_0)-Q(\theta_0))\widehat{Q}(\theta_0)^{-1}\widehat{B}\left(\theta_0\right)=Q(\theta_0)\sqrt{n}\left(Q(\theta_0)^{-1}-\widehat{Q}(\theta_0)^{-1}\right)\widehat{B}\left(\theta_0\right)\\
&=Q(\theta_0)\sqrt{n}\left(\widehat{H}_1(\theta_0)-H_1(\theta_0)\right)\widehat{B}\left(\theta_0\right)=QV\widehat{B}\left(\theta_0\right).
\end{aligned}$$

The last result is obtained noting $V\equiv\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(\nabla s\left(z_i,\theta_0\right)-E\left[\nabla s\left(z_i,\theta_0\right)\right]\right)$ and $\widehat{H}_1(\theta_0)=\frac{1}{n}\sum_{i=1}^{n}\nabla s\left(z_i,\theta_0\right)$ and hence we have

$$\sqrt{n}(\widehat{Q}(\theta_0)-Q(\theta_0))\widehat{c}(\theta_0)-QVB\left(\theta_0\right)=QV\left(\widehat{B}\left(\theta_0\right)-B\left(\theta_0\right)\right)=O_p\left(1/\sqrt{n}\right),$$

since $V=O_p(1)$ and $\widehat{B}\left(\theta_0\right)-B\left(\theta_0\right)=O_p\left(1/\sqrt{n}\right)$ by Condition 5. ■

B.6 Derivation of (98)

Let $H_2 = \begin{pmatrix} \nabla H_1^{11} & \nabla H_1^{12} & \cdots & \nabla H_1^{1k} \\ \vdots & \vdots & \vdots & \vdots \\ \nabla H_1^{k1} & \nabla H_1^{k2} & \cdots & \nabla H_1^{kk} \end{pmatrix}$ and $B(\theta_0) \equiv b = \begin{pmatrix} b_1 \\ \vdots \\ b_k \end{pmatrix}$, where H_1^{gh} is the g-th row and the h-th column element of H_1. Note that

$$\begin{aligned}
&\tfrac{1}{2}H_2 \left(I \otimes B(\theta_0) + B(\theta_0) \otimes I\right) = \tfrac{1}{2}\begin{pmatrix} \nabla H_1^{11} & \nabla H_1^{12} & \cdots & \nabla H_1^{1k} \\ \vdots & \vdots & \vdots & \vdots \\ \nabla H_1^{k1} & \nabla H_1^{k2} & \cdots & \nabla H_1^{kk} \end{pmatrix} \left(I \otimes B(\theta_0) + B(\theta_0) \otimes I\right) \\
&= \tfrac{1}{2}\begin{pmatrix} \nabla H_1^{11}b & \nabla H_1^{12}b & \cdots & \nabla H_1^{1k}b \\ \vdots & \vdots & \vdots & \vdots \\ \nabla H_1^{k1}b & \nabla H_k^{k2}b & \cdots & \nabla H_k^{kk}b \end{pmatrix} + \tfrac{1}{2}\begin{pmatrix} \nabla H_1^{11}b & \nabla H_1^{12}b & \cdots & \nabla H_1^{1k}b \\ \vdots & \vdots & \vdots & \vdots \\ \nabla H_1^{k1}b & \nabla H_1^{k2}b & \cdots & \nabla H_k^{kk}b \end{pmatrix} \\
&= \begin{pmatrix} \nabla H_1^{11}b & \nabla H_1^{12}b & \cdots & \nabla H_1^{1k}b \\ \vdots & \vdots & \vdots & \vdots \\ \nabla H_1^{k1}b & \nabla H_1^{k2}b & \cdots & \nabla H_k^{kk}b \end{pmatrix},
\end{aligned}$$

where the second term in the second equality is obtained from $\frac{\partial^2 s(z,\theta)}{\partial\theta_h \partial\theta_p} = \frac{\partial^2 s(z,\theta)}{\partial\theta_p \partial\theta_h}$ which implies $\frac{\partial H_1^{gp}}{\partial\theta_h} = \frac{\partial H_1^{gh}}{\partial\theta_p}$ for $g, h, p \in \{1, 2, \ldots, k\}$.

Similarly we obtain

$$\begin{aligned}
&vec^{*}\left(B(\theta_0)'\nabla\left(H_1\left(\theta_0\right)'\right)\right) = vec^{*}\left(b'\begin{pmatrix} \nabla H_1^{11} & \nabla H_1^{21} & \cdots & \nabla H_1^{k1} \\ \vdots & \vdots & \vdots & \vdots \\ \nabla H_1^{1k} & \nabla H_1^{2k} & \cdots & \nabla H_k^{kk} \end{pmatrix}\right) \\
&= \begin{pmatrix} b'\begin{pmatrix} \nabla H_1^{11} \\ \vdots \\ \nabla H_1^{1k} \end{pmatrix} \\ \vdots \\ b'\begin{pmatrix} \nabla H_1^{k1} \\ \vdots \\ \nabla H_1^{kk} \end{pmatrix} \end{pmatrix} = \begin{pmatrix} \nabla H_1^{11}b & \nabla H_1^{12}b & \cdots & \nabla H_1^{1k}b \\ \vdots & \vdots & \vdots & \vdots \\ \nabla H_1^{k1}b & \nabla H_1^{k2}b & \cdots & \nabla H_k^{kk}b \end{pmatrix}
\end{aligned}$$

again noting $\frac{\partial H_1^{gp}}{\partial\theta_h} = \frac{\partial H_1^{gh}}{\partial\theta_p}$ for $g, h, p \in \{1, 2, \ldots, k\}$ and hence we have established (98).

B.7 Derivation of the Main Result for the Scalar Case

We derive our main result for the scalar case with $\dim(\theta_0) = 1$ where $\widehat{c}(\theta)$ and $c(\theta)$ are simplified as

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

Note Condition 1-4 are easily verified by the proof of Lemma A.11. We derive the third order stochastic expansion iteratively. First, consider the first order Taylor series approximation of (9),

$$0 = \frac{1}{n}\textstyle\sum_{i=1}^n s\left(z_i,\theta_0\right) + \left(\frac{1}{n}\textstyle\sum_{i=1}^n \nabla s\left(z_i,\widetilde{\theta}\right)\right)\left(\theta^{*} - \theta_0\right) - \frac{1}{n}\widehat{c}(\theta_0) - \frac{1}{n}\nabla\widehat{c}\left(\widetilde{\theta}\right)\left(\theta^{*} - \theta_0\right)$$

for $\widetilde{\theta}$ between θ^{*} and θ_0 and hence

$$\begin{aligned}
&\sqrt{n}\left(\theta^{*} - \theta_0\right) \\
&= -\left(\tfrac{1}{n}\textstyle\sum_{i=1}^n \nabla s\left(z_i,\widetilde{\theta}\right) - \tfrac{1}{n}\nabla\widehat{c}(\widetilde{\theta})\right)^{-1}\left(\tfrac{1}{\sqrt{n}}\textstyle\sum_{i=1}^n s\left(z_i,\theta_0\right) - \tfrac{1}{\sqrt{n}}\widehat{c}(\theta_0)\right) \qquad (100) \\
&= -\left(E\left[\nabla s\left(z_i,\theta_0\right)\right] + o_p(1) + O_p\left(1/n\right)\right)^{-1}\left(\tfrac{1}{\sqrt{n}}\textstyle\sum_{i=1}^n s\left(z_i,\theta_0\right) + O_p\left(1/\sqrt{n}\right)\right) = QJ + o_p(1),
\end{aligned}$$

by Condition 1(i), 2 and $\widetilde{\theta} = \theta_0 + o_p(1)$ noting $\left\| \frac{1}{n} \sum_{i=1}^n \nabla s\left(z_i, \overline{\theta}\right) - E\left[\nabla s\left(z_i, \theta_0\right)\right] \right\| = o_p(1)$ for $\overline{\theta} = \theta_0 + o_p(1)$ by Lemma A.3 or A.4. Now rewrite (100) as

$$\begin{aligned}
&\sqrt{n}\left(\theta^* - \theta_0\right) \\
&= -\left(H_1(\theta_0) - \tfrac{\nabla\widehat{c}(\theta_0)}{n}\right)^{-1}\left(J - \tfrac{1}{\sqrt{n}}\widehat{c}(\theta_0)\right) - \left(\left(\widehat{H}_1(\widetilde{\theta}) - \tfrac{\nabla\widehat{c}(\widetilde{\theta})}{n}\right)^{-1} - \left(H_1(\theta_0) - \tfrac{\nabla\widehat{c}(\theta_0)}{n}\right)^{-1}\right)\left(J - \tfrac{1}{\sqrt{n}}\widehat{c}(\theta_0)\right) \\
&= -\left(H_1(\theta_0)^{-1} + O_p\left(\tfrac{1}{n}\right)\right)\left(J + O_p\left(\tfrac{1}{\sqrt{n}}\right)\right) - \left(\widehat{H}_1(\widetilde{\theta})^{-1} - H_1(\theta_0)^{-1} + O_p\left(\tfrac{1}{n}\right)\right)\left(J + O_p\left(\tfrac{1}{\sqrt{n}}\right)\right) \\
&= -H_1(\theta_0)^{-1}J - \left(\widehat{H}_1(\widetilde{\theta})^{-1} - H_1(\theta_0)^{-1}\right)J + O_p\left(\tfrac{1}{n}\right) = -H_1(\theta_0)^{-1}\tfrac{1}{\sqrt{n}}\textstyle\sum_{i=1}^n s\left(z_i, \theta_0\right) + O_p\left(\tfrac{1}{n}\right),
\end{aligned}$$

where the second inequality is by Condition 2 and the last equality is obtained by $(\widehat{H}_1(\widetilde{\theta})^{-1} - H_1(\theta_0)^{-1}) = O_p\left(1/\sqrt{n}\right)$ from (75) and $J = O_p(1)$ and hence we have

$$\sqrt{n}\left(\theta^* - \theta_0\right) = QJ + O_p\left(1/\sqrt{n}\right). \tag{101}$$

Now consider second order Taylor series expansion:

$$0 = \begin{array}{l} \frac{1}{n}\sum_{i=1}^n s\left(z_i, \theta_0\right) + \left(\frac{1}{n}\sum_{i=1}^n \nabla s\left(z_i, \theta_0\right)\right)\left(\theta^* - \theta_0\right) + \frac{1}{2}\left(\frac{1}{n}\sum_{i=1}^n \nabla^2 s\left(z_i, \widetilde{\theta}\right)\right)\left(\theta^* - \theta_0\right)^2 \\ -\frac{1}{n}\widehat{c}(\theta_0) - \frac{1}{n}\nabla\widehat{c}\left(\theta_0\right)\left(\theta^* - \theta_0\right) - \frac{1}{2n}\nabla^2\widehat{c}\left(\widetilde{\theta}\right)\left(\theta^* - \theta_0\right)^2. \end{array} \tag{102}$$

As we rewrite (77) into the form of (78) in Lemma 2.1's proof, we rewrite (102) as

$$\begin{aligned}
0 \;&=\; \begin{array}{l} \frac{1}{\sqrt{n}}J + \left(-Q^{-1} + \frac{1}{\sqrt{n}}V\right)\left(\theta^* - \theta_0\right) + \frac{1}{2}\left(H_2 + \frac{1}{\sqrt{n}}W\right)\left(\theta^* - \theta_0\right)^2 \\ -\frac{1}{n}\widehat{c}(\theta_0) - \frac{1}{n}\nabla\widehat{c}\left(\theta_0\right)\left(\theta^* - \theta_0\right) - \frac{1}{2n}\nabla^2\widehat{c}\left(\widetilde{\theta}\right)\left(\theta^* - \theta_0\right)^2 + O_p(n^{-3/2}) \end{array} && (103) \\
&=\; \begin{array}{l} \frac{1}{\sqrt{n}}J + \left(-Q^{-1} + \frac{1}{\sqrt{n}}V\right)\left(\theta^* - \theta_0\right) \\ +\frac{1}{2}\left(H_2 + \frac{1}{\sqrt{n}}W\right)\left(\theta^* - \theta_0\right)^2 - \frac{1}{n}\widehat{c}(\theta_0) + O_p(n^{-3/2}), \end{array} && (104)
\end{aligned}$$

since $\frac{1}{n}\nabla\widehat{c}\left(\theta_0\right)\left(\theta^* - \theta_0\right) = O_p(n^{-3/2})$ by Condition 2 and $\theta^* = \theta_0 + O_p(\frac{1}{\sqrt{n}})$ from (101) noting $J = O_p(1)$ and since $\frac{1}{2n}\nabla^2\widehat{c}\left(\widetilde{\theta}\right)\left(\theta^* - \theta_0\right)^2 = O(n^{-1})O_p(1)O_p(n^{-1}) = O_p(n^{-2})$ by Condition 3 and $\theta^* = \theta_0 + O_p\left(1/\sqrt{n}\right)$. Multiplying expansions (as (79)) of $(-Q^{-1} + V/\sqrt{n})^{-1}$ to both sides of (104) depending on orders and replacing $\sqrt{n}\left(\theta^* - \theta_0\right) = QJ + O_p\left(1/n\right)$, we obtain

$$\begin{aligned}
\sqrt{n}\left(\theta^* - \theta_0\right) \;&=\; QJ + \frac{1}{\sqrt{n}}\left(Q^2VJ + \frac{1}{2}Q^3H_2J^2 - Q\widehat{c}(\theta_0)\right) + O_p\left(\frac{1}{n}\right) \\
&=\; QJ + \frac{1}{\sqrt{n}}\left(Q^2VJ + \frac{1}{2}H_2Q^3J^2 - Qc(\theta_0)\right) + O_p\left(\frac{1}{n}\right),
\end{aligned}$$

as in Lemma 2.1. The second equality comes from $\widehat{c}(\theta_0) = c(\theta_0) + O_p\left(1/\sqrt{n}\right)$.

Now consider a third order Taylor series expansion of $0 = \frac{1}{n}\sum_{i=1}^n s\left(z_i, \theta^*\right) - \frac{1}{n}\widehat{c}(\theta^*)$:

$$\begin{aligned}
0 &= \tfrac{1}{n}\textstyle\sum_{i=1}^n s\left(z_i, \theta_0\right) + \left(\frac{1}{n}\sum_{i=1}^n \nabla s\left(z_i, \theta_0\right)\right)\left(\theta^* - \theta_0\right) + \frac{1}{2}\left(\frac{1}{n}\sum_{i=1}^n \nabla^2 s\left(z_i, \theta_0\right)\right)\left(\theta^* - \theta_0\right)^2 \\
&\quad + \tfrac{1}{6}\left(\tfrac{1}{n}\textstyle\sum_{i=1}^n \nabla^3 s\left(z_i, \widetilde{\theta}\right)\right)\left(\theta^* - \theta_0\right)^3 - \frac{1}{n}\widehat{c}(\theta_0) - \frac{1}{n}\nabla\widehat{c}\left(\theta_0\right)\left(\theta^* - \theta_0\right) - \frac{1}{2}\frac{1}{n}\nabla^2\widehat{c}\left(\theta_0\right)\left(\theta^* - \theta_0\right)^2 \\
&\quad - \tfrac{1}{6}\tfrac{1}{n}\nabla^3\widehat{c}\left(\widetilde{\theta}\right)\left(\theta^* - \theta_0\right)^3
\end{aligned}$$

From this, similarly with (102) to (103), we obtain

$$\begin{aligned}
0 &= \tfrac{1}{\sqrt{n}}J + \left(-Q^{-1} + \tfrac{1}{\sqrt{n}}V\right)\left(\theta^* - \theta_0\right) + \tfrac{1}{2}\left(H_2 + \tfrac{1}{\sqrt{n}}W\right)\left(\theta^* - \theta_0\right)^2 \\
&\quad - \tfrac{1}{6}\left(H_3 + \tfrac{1}{\sqrt{n}}W_3\right)\left(\theta^* - \theta_0\right)^3 - \tfrac{1}{n}\widehat{c}(\theta_0) - \tfrac{1}{n}\nabla\widehat{c}\left(\theta_0\right)\left(\theta^* - \theta_0\right) + O_p(n^{-2}),
\end{aligned} \tag{105}$$

since $\frac{1}{2}\frac{1}{n}\nabla^2\widehat{c}(\theta_0)\left(\theta^* - \theta_0\right)^2 = O_p(n^{-2})$ by Condition 3 and $\theta^* = \theta_0 + O_p(1/\sqrt{n})$ and since $\frac{1}{2}\frac{1}{n}\nabla^3\widehat{c}\left(\widetilde{\theta}\right)(\theta^* - \theta_0^3)$ $= O_p(n^{-5/2})$ by Condition 4 (ii) and $\theta^* = \theta_0 + O_p(1/\sqrt{n})$. Multiplying expansions (as (79)) of $\left(-Q^{-1} + \frac{1}{\sqrt{n}}V\right)^{-1}$ to both sides of (105) depending on orders, we obtain

$$\begin{aligned}
&\theta^* - \theta_0 \\
&= \left\{\begin{array}{l}
-\left(-Q - \frac{1}{\sqrt{n}}Q^2V - \frac{1}{n}Q^3V^2 + O_p(n^{-3/2})\right)\frac{1}{\sqrt{n}}J \\
-\frac{1}{2}\left(-Q - \frac{1}{\sqrt{n}}Q^2V + O_p(n^{-1})\right)\left(H_2 + \frac{1}{\sqrt{n}}W\right)\left(\theta^* - \theta_0\right)^2 \\
-\frac{1}{6}\left(-Q + O_p\left(\frac{1}{\sqrt{n}}\right)\right)\left(H_3 + \frac{1}{\sqrt{n}}W_3\right)\left(\theta^* - \theta_0\right)^3 \\
+\frac{1}{n}\left(-Q - \frac{1}{\sqrt{n}}Q^2V + O_p(n^{-1})\right)\widehat{c}(\theta_0) + \frac{1}{n}\left(-Q + O_p\left(\frac{1}{\sqrt{n}}\right)\right)\nabla\widehat{c}(\theta_0)(\theta^* - \theta_0) + O_p(n^{-2})
\end{array}\right\} \\
&= \left\{\begin{array}{l}
\left(\frac{1}{\sqrt{n}}QJ + \frac{1}{n}Q^2VJ + \frac{1}{n^{3/2}}Q^3V^2J\right) + \frac{1}{2}\left(Q\left(H_2 + \frac{1}{\sqrt{n}}W\right) + \frac{Q^2H_2V}{\sqrt{n}}\right)\left(\theta^* - \theta_0\right)^2 \\
+\frac{1}{6}QH_3\left(\theta^* - \theta_0\right)^3 - \frac{1}{n}\left(Q + \frac{1}{\sqrt{n}}Q^2V\right)\widehat{c}(\theta_0) - \frac{1}{n}Q\nabla\widehat{c}(\theta_0)(\theta^* - \theta_0) + O_p(n^{-2})
\end{array}\right\}.
\end{aligned}$$

Now replacing $\sqrt{n}(\theta^* - \theta_0) = a_{-1/2} + O_p(1/\sqrt{n})$ or $\sqrt{n}(\theta^* - \theta_0) = a_{-1/2} + \frac{1}{\sqrt{n}}(a_{-1} - Qc(\theta_0)) + O_p(1/n)$ depending on the orders required, we obtain

$$\begin{aligned}
&\theta^* - \theta_0 \\
&= \left\{\begin{array}{l}
\left(\frac{1}{\sqrt{n}}QJ + \frac{1}{n}Q^2VJ + \frac{1}{n^{3/2}}Q^3V^2J\right) \\
+\frac{1}{2n}\left(Q\left(H_2 + \frac{1}{\sqrt{n}}W\right) + \frac{Q^2H_2V}{\sqrt{n}}\right)\left(a_{-1/2} + \frac{1}{\sqrt{n}}(a_{-1} - Qc(\theta_0)) + O_p\left(\frac{1}{n}\right)\right)^2 \\
+\frac{1}{6n^{3/2}}QH_3\left(a_{-1/2} + O_p\left(\frac{1}{\sqrt{n}}\right)\right)^3 - \frac{1}{n}\left(Q + \frac{1}{\sqrt{n}}Q^2V\right)\widehat{c}(\theta_0) \\
-\frac{1}{n^{3/2}}Q\nabla\widehat{c}(\theta_0)\left(a_{-1/2} + O_p\left(\frac{1}{\sqrt{n}}\right)\right) + O_p(n^{-2})
\end{array}\right\} \\
&= \left\{\begin{array}{l}
\left(\frac{1}{\sqrt{n}}QJ + \frac{1}{n}Q^2VJ + \frac{1}{n^{3/2}}Q^3V^2J\right) + \frac{1}{2n}\left(Q\left(H_2 + \frac{W}{\sqrt{n}}\right) + \frac{Q^2H_2V}{\sqrt{n}}\right)\left(a_{-1/2} + \frac{a_{-1} - Qc(\theta_0)}{\sqrt{n}}\right)^2 \\
\quad + \frac{1}{6n^{3/2}}QH_3a_{-1/2}^3 - \frac{1}{n}\left(Q + \frac{1}{\sqrt{n}}Q^2V\right)\widehat{c}(\theta_0) - \frac{1}{n^{3/2}}Q\nabla c(\theta_0)a_{-1/2} + O_p(n^{-2})
\end{array}\right\},
\end{aligned}$$

where we replaced $\nabla\widehat{c}(\theta_0)$ with $\nabla c(\theta_0) + O_p\left(\frac{1}{\sqrt{n}}\right)$ from Condition 4 (i). Rearranging terms according to the orders, we have

$$\begin{aligned}
&\sqrt{n}\left(\theta^* - \theta_0\right) \\
&= a_{-1/2} + \tfrac{1}{\sqrt{n}}(a_{-1} - Qc(\theta_0)) + \tfrac{1}{n}\left(a_{-3/2} - Q^2H_2c(\theta_0)a_{-1/2} - Q^2Vc(\theta_0) - Q\nabla c(\theta_0)a_{-1/2}\right) \\
&\quad - \tfrac{1}{n}Q\sqrt{n}(\widehat{c}(\theta_0) - c(\theta_0)) + O_p(n^{-3/2}).
\end{aligned} \tag{106}$$

Now note $\widehat{c}(\theta_0) = \frac{\widehat{B}(\theta_0)}{\widehat{Q}(\theta_0)}$, $c(\theta_0) = \frac{B(\theta_0)}{Q(\theta_0)}$, $\nabla c(\theta_0) = \frac{\nabla B(\theta_0)}{Q} - \frac{B(\theta_0)\nabla Q}{Q^2}$, $\nabla Q = -\nabla(\frac{1}{H_1}) = H_1^{-2}\nabla H_1 = Q^2H_2$ by definition of H_1 and H_2 and hence $\nabla c(\theta_0) = \frac{\nabla B(\theta_0)}{Q} - B(\theta_0)H_2$. Pugging these results into (106), we have

$$\begin{aligned}
\sqrt{n}\left(\theta^* - \theta_0\right) &= a_{-1/2} + \tfrac{1}{\sqrt{n}}(a_{-1} - B(\theta_0)) \\
&\quad + \tfrac{1}{n}\left(a_{-3/2} - QVB(\theta_0) - \nabla B(\theta_0)a_{-1/2}\right) - \tfrac{1}{n}Q\sqrt{n}(\widehat{c}(\theta_0) - c(\theta_0)) + O_p(n^{-3/2})
\end{aligned} \tag{107}$$

noting $-Q^2H_2c(\theta_0) - Q\nabla c(\theta_0) = -QH_2B(\theta_0) - Q\left(\frac{\nabla B(\theta_0)}{Q} - B(\theta_0)H_2\right) = -\nabla B(\theta_0)$. Finally consider that

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

and that

$$
\begin{aligned}
&\sqrt{n}\left(\widehat{Q}(\theta_0)-Q(\theta_0)\right)\widehat{c}(\theta_0)\\
&=-\sqrt{n}Q(\theta_0)\left(\tfrac{1}{\widehat{Q}(\theta_0)}-\tfrac{1}{Q(\theta_0)}\right)\widehat{Q}(\theta_0)\widehat{c}(\theta_0)=\sqrt{n}Q(\theta_0)\left(\widehat{H}_1(\theta_0)-H_1(\theta_0)\right)\widehat{Q}(\theta_0)\widehat{c}(\theta_0)\\
&=Q(\theta_0)\left(\tfrac{1}{\sqrt{n}}\textstyle\sum_{i=1}^n\left(\nabla s\left(z_i,\theta_0\right)-E\left[\nabla s\left(z_i,\theta_0\right)\right]\right)\right)\widehat{Q}(\theta_0)\widehat{c}(\theta_0)=Q(\theta_0)V\widehat{B}\left(\theta_0\right)=QVB\left(\theta_0\right)+O_p\left(\tfrac{1}{\sqrt{n}}\right)
\end{aligned}
$$

by definition of V and Condition 5:$\widehat{B}\left(\theta_0\right)=B\left(\theta_0\right)+O_p\left(1/\sqrt{n}\right)$. Applying these results to (107), we obtain

$$
\begin{aligned}
\sqrt{n}\left(\theta^*-\theta_0\right)&=a_{-1/2}+\tfrac{1}{\sqrt{n}}\left(a_{-1}-B(\theta_0)\right)\\
&\quad+\tfrac{1}{n}\left(a_{-3/2}-\nabla B(\theta_0)a_{-1/2}\right)-\tfrac{1}{n}\sqrt{n}(\widehat{B}(\theta_0)-B(\theta_0))+O_p(n^{-3/2})
\end{aligned}
$$

and hence $\sqrt{n}\left(\theta^*-\theta_0\right)=\sqrt{n}(\widehat{\theta}_{bc}-\theta_0)+O_p(n^{-3/2})$.

C Higher Order Variances

Here we derive the analytic forms of the higher order variances for several alternative estimators. Note that $E\left[\left(a_{-1}-B(\theta_0)\right)\left(a_{-1}-B(\theta_0)\right)'\right]=E[a_{-1}a_{-1}']-B(\theta_0)B(\theta_0)'$, $E\left[\sqrt{n}a_{-1/2}\left(a_{-1}-B(\theta_0)\right)'\right]=E\left[\sqrt{n}a_{-1/2}a_{-1}'\right]$, $E\left[a_{-1/2}\left(a_{-3/2}-\nabla B(\theta_0)a_{-1/2}\right)'\right]=E\left[a_{-1/2}a_{-3/2}'\right]-E\left[a_{-1/2}a_{-1/2}'\right]\left(\nabla B(\theta_0)\right)'$ from $E[a_{-1}]=B(\theta_0)$ and $E[a_{-1/2}]=0$ and hence

$$
\Lambda_{\widehat{\theta}_b}=\left\{\begin{array}{l}
E\left[a_{-1/2}a_{-1/2}'\right]+\frac{1}{n}\left(E\left[\sqrt{n}a_{-1}a_{-1/2}'\right]+E\left[\sqrt{n}a_{-1/2}a_{-1}'\right]\right)\\
+\frac{1}{n}\left(E[a_{-1}a_{-1}']+E\left[a_{-3/2}a_{-1/2}'\right]+E\left[a_{-1/2}a_{-3/2}'\right]\right)\\
-B(\theta_0)B(\theta_0)'-E\left[a_{-1/2}a_{-1/2}'\right]\left(\nabla B(\theta_0)\right)'-\nabla B(\theta_0)E\left[a_{-1/2}a_{-1/2}'\right]
\end{array}\right\}.
$$

Rilstone, Srivastava, and Ullah (1996) derive the second-order mean squared error (MSE) of the M-estimator that solves the moment condition (2). Proposition 3.4 in Rilstone, Srivastava, and Ullah (1996) implies that

$$
\Lambda_{\widehat{\theta}_b}=\gamma_1+\frac{1}{n}\left(\gamma_2+\gamma_2'\right)+\frac{1}{n}\left(\gamma_3+\gamma_4+\gamma_4'\right)-\frac{1}{n}\left(B(\theta_0)B(\theta_0)'+\gamma_1\left(\nabla B(\theta_0)\right)'+\nabla B(\theta_0)\gamma_1'\right)+O\left(n^{-2}\right)
$$

where (denoting the expectation of a function $A(\theta)$ as $\overline{A(\theta)}=E[A(\theta)]$ for notational convenience)

$$
\begin{aligned}
\gamma_1&=\overline{d_1d_1'},\ \gamma_2=Q\left\{\overline{v_1d_1d_1'}+\frac{1}{2}H_2\overline{\left(d_1\otimes d_1\right)d_1'}\right\}\\
\gamma_3&=\left\{\begin{array}{l}
Q\left\{\overline{v_1d_1d_2'V_2'}+\overline{v_1d_2d_1'v_2'}+\overline{v_1d_2d_2'v_1'}\right\}Q'\\
+QH_2\left\{\overline{(d_1\otimes d_1)(d_2'\otimes d_2')}+\overline{\left(d_1\otimes d_2\right)\left(d_1'\otimes d_2'\right)}+\overline{\left(d_1\otimes d_2\right)\left(d_2'\otimes d_1'\right)}\right\}H_2'Q'\\
-Q\left\{\overline{v_1d_1\left(d_2'\otimes d_2'\right)}+\overline{v_1d_2\left(d_1'\otimes d_2'\right)}+\overline{v_1d_2\left(d_2'\otimes d_1'\right)}\right\}H_2'Q'\\
-QH_2\left\{\overline{d_1\otimes d_1d_2'v_2'}+\overline{\left(d_1\otimes d_2\right)d_1'v_2'}+\overline{\left(d_1\otimes d_2\right)d_2'v_1'}\right\}Q'
\end{array}\right.\\
\gamma_4&=\left\{\begin{array}{l}
Q\left\{\overline{v_1Qv_1d_2d_2'}+\overline{v_1Qv_2d_1d_2'}+\overline{v_1Qv_2d_2d_1'}\right\}\\
+\frac{1}{2}Q\left\{\overline{v_1QH_2\left(d_1\otimes d_2\right)d_2'}+\overline{v_1QH_2\left(d_2\otimes d_1\right)d_2'}+\overline{v_1QH_2\left(d_2\otimes d_2\right)d_1'}\right\}\\
+\frac{1}{2}Q\left\{\overline{w_1\left(d_1\otimes d_2\right)d_2'}+\overline{w_1\left(d_2\otimes d_1\right)d_2'}+\overline{w_1\left(d_2\otimes d_2\right)d_1'}\right\}\\
+\frac{1}{2}QH_2\left\{\overline{\left(d_1\otimes Qv_1d_2\right)d_2'}+\overline{\left(d_1\otimes Qv_2d_1\right)d_2'}+\overline{\left(d_1\otimes Qv_2d_2\right)d_1'}\right\}\\
+\frac{1}{4}QH_2\left\{\overline{d_1\otimes QH_2\left(d_1\otimes d_2\right)d_2'}+\overline{d_1\otimes QH_2\left(d_2\otimes d_1\right)d_2'}+\overline{d_1\otimes QH_2\left(d_2\otimes d_2\right)d_1'}\right\}\\
+\frac{1}{2}QH_2\left\{\overline{\left(QV_1d_1\otimes d_2\right)d_2'}+\overline{\left(QV_1d_2\otimes d_1\right)d_2'}+\overline{\left(QV_1d_2\otimes d_2\right)d_1'}\right\}\\
+\frac{1}{4}QH_2\left\{\overline{\left(QH_2\left(d_1\otimes d_1\right)\otimes d_2\right)d_2'}+\overline{\left(QH_2\left(d_1\otimes d_2\right)\otimes d_1\right)d_2'}+\overline{\left(QH_2\left(d_1\otimes d_2\right)\otimes d_2\right)d_1'}\right\}\\
+\frac{1}{6}QH_3\left\{\overline{\left[d_1\otimes d_1\otimes d_2\right]d_2'}+\overline{\left[d_1\otimes d_2\otimes d_1\right]d_2'}+\overline{\left[d_1\otimes d_2\otimes d_2\right]d_1'}\right\}
\end{array}\right\}.
\end{aligned}
$$

for $d_i = Qs(z_i,\theta_0)$, $v_i = \nabla s(z_i,\theta_0) - E\left[\nabla s(z_i,\theta_0)\right]$ and $w_i = \nabla^2 s(z_i,\theta_0) - E\left[\nabla^2 s(z_i,\theta_0)\right]$. We also note $B(\theta_0) = \left(Q\overline{v_1 d_1} + \frac{1}{2}H_2\overline{d_1 \otimes d_1}\right)$ from Lemma 2.2. Finally we derive $\nabla B(\theta_0)$ as follows. Noting $vec^*\left(s(z_i,\theta_0)'\nabla\left(Q(\theta_0)'\right)\right) = vec^*\left(s(z_i,\theta_0)'Q'QH_2\right)$ from Remark 5, similarly with (38), we can show

$$\nabla c(\theta_0) = \frac{1}{2}vec^*\left(\left(\overline{d_1 \otimes d_1}\right)' \nabla\left(H_2(\theta)'\right)|_{\theta=\theta_0}\right) + \frac{1}{2}H_2\left(\overline{e_1 \otimes^* d_1} + \overline{vec^*\left(d_1'QH_2\right) \otimes^* d_1}\right)$$
$$+\frac{1}{2}H_2\left(\overline{d_1 \otimes e_1} + \overline{d_1 \otimes vec^*\left(d_1'QH_2\right)}\right) + \overline{\nabla s_1\left(e_1 + vec^*\left(d_1'QH_2\right)\right)} + \overline{vec^*\left(d_1' \nabla\left(\left(\nabla s_1(\theta)\right)'\right)\Big|_{\theta=\theta_0}\right)}$$

by inspecting (59) where $e_1 = Q\nabla s(z_1,\theta_0)$, $\nabla s_1(\theta) = \nabla s(z_1,\theta)$ and $\nabla s_1 = \nabla s(z_1,\theta_0)$. Combining this result with $\nabla B(\theta_0) = Q(\theta_0)\nabla c(\theta_0) + vec^*\left(c(\theta_0)'\nabla\left(Q(\theta_0)'\right)\right)$ and $B(\theta_0) = Q\left\{\overline{v_1 d_1} + \frac{1}{2}H_2\overline{d_1 \otimes d_1}\right\}$, we obtain

$$\begin{aligned}
\nabla B(\theta_0) &= Q(\theta_0)\nabla c(\theta_0) + vec^*\left(c(\theta_0)'\nabla\left(Q(\theta_0)'\right)\right) \\
&= Q(\theta_0)\nabla c(\theta_0) + vec^*\left(c(\theta_0)'Q'QH_2\right) = Q(\theta_0)\nabla c(\theta_0) + vec^*\left(B(\theta_0)'QH_2\right) \\
&= \left\{\begin{array}{l}
\frac{1}{2}Qvec^*\left(\left(\overline{d_1 \otimes d_1}\right)' \nabla\left(H_2(\theta)'\right)|_{\theta=\theta_0}\right) + \frac{1}{2}QH_2\left(\overline{e_1 \otimes^* d_1} + \overline{vec^*\left(d_1'QH_2\right) \otimes^* d_1}\right) \\
+\frac{1}{2}QH_2\left(\overline{d_1 \otimes e_1} + \overline{d_1 \otimes vec^*\left(d_1'QH_2\right)}\right) + \overline{e_1\left(e_1 + vec^*\left(d_1'QH_2\right)\right)} \\
+\overline{Qvec^*\left(d_1' \nabla\left(\left(\nabla s_1(\theta)\right)'\right)\Big|_{\theta=\theta_0}\right)} + vec^*\left(\left\{\overline{d_1'v_1'} + \frac{1}{2}\overline{\left(d_1 \otimes d_1\right)'}H_2'\right\}Q'QH_2\right)
\end{array}\right\}.
\end{aligned}$$

References

[1] Andrews, D.W.K. (1994), "Empirical Process Methods in Econometrics," *Handbook of Econometrics* IV, edited by R.F. Engle and D.L. McFadden, 2247-2294.

[2] Bhattacharya, R.N. and J.K. Ghosh (1978), "On the Validity of the Formal Edgeworth Expansion," *Annals of Mathematical Statistics*, 6, 434-451.

[3] Fernádez-Val, I. (2004), "Bias Correction in Panel Data Models with Individual Specific Parameters," working paper, MIT.

[4] Firth, D. (1993), "Bias Reduction of Maximum Likelihood Estimation," *Biometrika*, 80, 27-38.

[5] Ghosh, J.K, B.K. Sinha, and H.S. Wieand (1980), "Second Order Efficiency of the MLE with Respect to Any Bowl Shaped Loss Function," Annals of Statistics, 8, 506-521.

[6] Hahn, J., G. Kuersteiner, and W. Newey (2004), "Higher Order Efficiency of Bias Reduction," Working paper, UCLA and MIT.

[7] Hahn, J and W.K. Newey (2004), "Jackknife and Analytical Bias Reduction for Nonlinear Panel Models," *Econometrica*, 72, 1295-1319.

[8] Heckman, J. (1979), "Sample Selection Bias as a Specification Error," *Econometrica*, 47(1), 153-61.

[9] Kitamura, Y. and M. Stutzer (1997), "An Information-theoretic Alternative to Generalized Method of Moments Estimation," *Econometrica*, 69, 1661-1672.

[10] Newey, W.K. (1984), "A Method of Moments Interpretation of Sequential Estimators," *Economics Letters*, 14(2/3), 201-206.

[11] Newey, W.K. (1991), "Uniform Convergence in Probability and Stochastic Equicontinuity," *Econometrica*, 58, 809-837.

[12] Newey, W.K. and D. McFadden (1994), "Large Sample Estimation and Hypothesis Testing," *Handbook of Econometrics* IV, edited by R.F. Engle and D.L. McFadden, 2247-2294.

[13] Newey, W.K. and R. Smith (2004), "Higher Order Properties of GMM and Generalized Empirical Likelihood Estimators," *Econometrica*, 72, 219-255.

[14] Pfanzagl, J. and W. Wefelmeyer (1978), "A Third-Order Optimum Property of the Maximum Likelihood Estimator," *Journal of Multivariate Analysis*, 8, 1-29.

[15] Rilstone, P., V.K. Srivastava, and A. Ullah (1996), "The Second-order Bias and Mean Squared Error of Nonlinear Estimators," *Journal of Econometrics*, 75, 369-395.

[16] Rothenberg, T.J. (1984), "Approximating the Distributions of Econometric Estimators and Test Statistics," Ch. 15 in *Handbook of Econometrics*, V2, edited by Z. Griliches and M.D. Intriligator.

[17] Tauchen, G.E. (1985), "Diagnostic Testing and Evaluation of Maximum Likelihood Models," *Journal of Econometrics*, 30, 415-443.

[18] Woutersen, T.M. (2002), "Robustness Againt Incidental Parameters," unpublished manuscript, University of Western Ontario