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

Kyoo il Kim

Department of Economics, Michigan State University, 486W Circle Dr, East Lansing, MI 48824, USA;
kyookim@msu.edu; Tel.: +1-517-353-9008

Abstract: This paper studies an alternative bias correction for the M-estimator, which is obtained by correcting the moment equations 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. This finding suggests that the comparison between the one-step bias correction and the method of correcting the moment equations or the fully-iterated bias correction should be based on the stochastic expansions higher than the third order.

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 (see, e.g., Quenouille (1956) [1], Hall (1992) [2], Shao and Tu (1995) [3], MacKinnon and Smith (1998) [4], Andrews (2002) [5], Hahn and Newey (2004) [6], Bun and Carree (2005) [7], Bao and Ullah (2007) [8,9], Bao (2013) [10] and Yang (2015) [11]). This variety of bias correction methods evokes the issue whether one method is preferable to others at least on asymptotic efficiency grounds (e.g., see Hahn et al. (2004) [12]). For the maximum likelihood (ML) estimation, they show that a method of bias correction does not affect the higher order efficiency of any estimator that is first-order efficient in parametric or semiparametric models. An ML estimator is a class of M-estimators, and this paper extends their intuition to a general class of M-estimators.[1]

Specifically, this paper considers an alternative bias correction for the M-estimator, which is achieved by correcting moment equations in the spirit of Firth (1993) [13]. 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 than comparing higher order variances, since it implies that these two estimators do not only have the same higher order variances, but would also agree upon more properties in terms of their stochastic expansions.[2] We do not consider other bias correction methods, such as bootstrap and jackknife methods, in this paper.

---

[1] This possible extension was noted in Hahn and Newey (2004) [6].
[2] This is subject to some caveats, such as the existence of moments and other negligible remainder terms in the stochastic expansions.

In the literature (see Hahn and Newey (2004) [6] and Fernandez-Val (2004) [14] for nonlinear panel data models), it has been discussed that removing the bias directly from the moment equations has an attractive feature that it does not use pre-estimated parameters that are not bias corrected, though this alternative approach requires more intensive computations.[3] Because the analytically-bias-corrected estimator is a two-step estimator, for which an initial estimator needs to be plugged in, while the bias-corrected moment equations estimator is a one-step estimator that does not need an initial estimator, the higher order asymptotic equivalence of these two estimators is not obvious. This paper, however, shows 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 in the context of M-estimators. This finding suggests that the comparison between the one-step bias correction and the method of correcting the moment equations should be based on the stochastic expansions higher than the third order.

Examples of the M-estimation include maximum likelihood estimation (MLE), least squares and instrumental variable (IV) estimation. Many other useful estimators can also fit into the M-estimation framework with the appropriate definition of the moment equations. It includes some cases of the generalized method of moments (GMM; see examples in Rilstone et al. (1996) [15]) and two-step estimators (Newey (1984) [16]). We note that the generalized empirical likelihood (GEL) can also fit into this framework. This suggests that Firth (1993)'s [13] correcting moment equations approach can be an alternative to Newey and Smith's approach to obtain the higher order bias and variance terms of GEL (2004) [17].

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 gathered in Appendix A and B.

2. Higher Order Expansion for the M-Estimator

Consider a moment condition:

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

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 includes 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 MLE, 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 $s\left(z_i, \theta\right) = \frac{\partial f(X_i; \theta)}{\partial \theta}\left(y_i - f(X_i; \theta)\right)$ and $z_i = (y_i \ X_i')'$ for a known function $f(\cdot)$. In the linear IV model, we have $s\left(z_i, \theta\right) = w_i(y_i - X_i'\theta)$ 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) [18]'s two-step estimator also fit into this framework (see Newey (1984) [16]). Rilstone et al. (1996) [15] provide some special cases of GMM estimators that can be put into the M-estimation, but the examples are not restricted to those. Partly motivated with this wide applicability, we study the stochastic expansion and the bias correction of the M-estimator.

---

[3] Note that the bias correction problem in nonlinear panel data models is the correction for the first-order bias due to the incidental parameters, while the bias correction in this paper is for the second-order bias.

We obtain the higher order stochastic expansion of the M-estimator using the iterative approach used in Rilstone et al. (1996) [15] up to a certain order. This approach is analytically convenient and straightforward to implement since the estimators are expressed as functions of the sums of random variables. Edgeworth expansion can be considered as an alternative whose validity has been derived in Bhattacharya and Ghosh (1978) [19], 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 that 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) [20] and Ghosh et al. (1980) [21] and as discussed in Rothenberg (1984) [22], it suffices to use the simple stochastic expansions for our purposes.

Here, we borrow Rilstone et al. (1996) [15]'s notation. We denote the matrix of v-th order partial derivatives of a matrix $A(\theta)$ as $\nabla^v A(\theta)$. 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 $A(\theta)$. $\nabla^v A(\theta)$ (a $k \times k^v$ matrix) is defined recursively, such that the j-th element of the l-th row of $\nabla^v A(\theta)$ is the $1 \times k$ vector $a_{lj}^v(\theta) = \partial a_{lj}^{v-1}(\theta)/\partial \theta'$, where a_{lj}^{v-1} is the l-th row and the j-th element of $\nabla^{v-1} A(\theta)$. We use $\otimes$ to denote a usual Kronecker product. Using this Kronecker product, we can express $\nabla^v A(\theta) = \underbrace{\frac{\partial^v A(\theta)}{\partial \theta' \otimes \partial \theta' \otimes \ldots \otimes \partial \theta'}}_{v \text{ Kronecker product of } \partial \theta'}$. Finally, we use a matrix norm $\|A\| = \sqrt{tr(A'A)}$ for a matrix A.

We first derive the higher order stochastic expansion of the M-estimator and consider the one-step bias correction here. In the next section, we introduce the bias-corrected moment equations estimator and derive its higher order stochastic expansion. Then, we compare these two approaches.

Before we derive the second-order expansion of the M-estimator to obtain the second-order bias analytically, we introduce simplifying notation. Let $H_1(\theta) = E[\nabla s(z_i, \theta)]$, $H_2(\theta) = E[\nabla^2 s(z_i, \theta)]$, $Q(\theta) = (-E[\nabla s(z_i, \theta)])^{-1}$, and write $H_1 = H_1(\theta_0)$, $H_2 = H_2(\theta_0)$, $Q = Q(\theta_0)$. Let $\widehat{H}_1(\theta) = \frac{1}{n}\sum_{i=1}^n \nabla s(z_i, \theta)$, $\widehat{H}_2(\theta) = \frac{1}{n}\sum_{i=1}^n \nabla^2 s(z_i, \theta)$, $\widehat{Q}(\theta) = (-\widehat{H}_1(\theta))^{-1}$, $\widehat{H}_1 = \widehat{H}_1(\theta_0)$, $\widehat{H}_2 = \widehat{H}_2(\theta_0)$ and $\widehat{Q} = \widehat{Q}(\theta_0)$. Furthermore, define $J \equiv \frac{1}{\sqrt{n}}\sum_{i=1}^n s(z_i, \theta_0)$, $V \equiv \frac{1}{\sqrt{n}}\sum_{i=1}^n (\nabla s(z_i, \theta_0) - E[\nabla s(z_i, \theta_0)])$, $W \equiv \frac{1}{\sqrt{n}}\sum_{i=1}^n (\nabla^2 s(z_i, \theta_0) - E[\nabla^2 s(z_i, \theta_0)])$.

Lemma 1. *(Rilstone et al. (1996) [15]) Suppose $\{z_i\}_{i=1}^n$ are i.i.d.; 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 $z \in \mathcal{Z} \equiv Support(z_i)$, $\kappa \geq 3$ with probability one; (iia) $\nabla^v s(z, \theta)$ is integrable for each fixed $\theta \in \Theta_0$, $v = \{0, 1, 2, \ldots \kappa\}$, $\kappa \geq 3$; and (iib) $E[\nabla^3 s(z, \theta)]$ is continuous and bounded at θ_0; (iii) $\left\| \frac{1}{n}\sum_{i=1}^n \nabla^v s(z_i, \overline{\theta}) - E[\nabla^v s(z_i, \theta_0)] \right\| = o_p(1)$ for $\overline{\theta} = \theta_0 + o_p(1)$ and $v = 1, 2$; (iv) $\frac{1}{\sqrt{n}}\sum_{i=1}^n (\nabla^2 s(z_i, \overline{\theta}) - H_2(\overline{\theta})) - \frac{1}{\sqrt{n}}\sum_{i=1}^n (\nabla^2 s(z_i, \theta_0) - H_2(\theta_0)) = o_p(1)$ for $\overline{\theta} = \theta_0 + o_p(1)$; (v) $Q(\theta_0)$ exists, i.e., $E[\nabla s(z_i, \theta_0)]$ is nonsingular; (vi) $J = O_p(1)$; (vii) $V = O_p(1)$; (viii) $W = O_p(1)$. Then, we have $\sqrt{n}(\widehat{\theta} - \theta_0) = QJ + O_p\left(\frac{1}{\sqrt{n}}\right)$, and moreover, $\sqrt{n}(\widehat{\theta} - \theta_0) = QJ + \frac{1}{\sqrt{n}}Q\left(VQJ + \frac{1}{2}H_2(QJ \otimes QJ)\right) + O_p(n^{-1})$.*

This lemma and the following Lemma 2 are available in Rilstone et al. (1996) [15], but we reproduce them since some of their results are useful to derive our new results. From Lemma 1, the higher order bias of $\widehat{\theta}$ is obtained as:

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

Defining $d_i(\theta) = Q(\theta)s(z_i, \theta)$ and $v_i(\theta) = \nabla s(z_i, \theta) - E[\nabla s(z_i, \theta)]$ and letting $d_i = d_i(\theta_0)$ and $v_i = v_i(\theta_0)$, it is not difficult to see $Q\left(E[VQJ] + \frac{1}{2}H_2 E[(QJ \otimes QJ)]\right) = Q\left(E[v_i d_i] + \frac{1}{2}H_2 E[(d_i \otimes d_i)]\right)$, as shown below. In this regard, we will write $B(\theta) \equiv Q(\theta)\left(E[v_i(\theta)d_i(\theta)] + \frac{1}{2}H_2(\theta)E[d_i(\theta) \otimes d_i(\theta)]\right)$.

Lemma 2. *(Rilstone et al. (1996) [15]) Suppose (1) holds and $\{z_i\}_{i=1}^{n}$ are i.i.d. Then, $E[VQJ] + \frac{1}{2}H_2 E[QJ \otimes QJ] = E[v_i d_i] + \frac{1}{2}H_2 E[d_i \otimes d_i]$, where $d_i = Qs(z_i, \theta_0)$ and $v_i = \nabla s(z_i, \theta_0) - E[\nabla s(z_i, \theta_0)]$.*

Thus, we can eliminate the second-order bias of the M-estimator $\widehat{\theta}$ by subtracting a consistent estimator of the bias.[4] Now, let $\widehat{\theta}_{bc}$ denote the bias-corrected estimator of this sort defined by:

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

where the function $\widehat{B}(\theta)$, a consistent estimator of $B(\theta)$, 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 $\widehat{v}_i(\theta) = \nabla s(z_i, \theta)$. In particular, we can replace $\widehat{\theta}$ in $\widehat{B}(\widehat{\theta})$ with any $\sqrt{n}$-consistent estimator of θ_0. 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-corrected estimators, we need to expand the M-estimator to the third order. We use some additional simplifying terms: $H_3(\theta) = E[\nabla^3 s(z, \theta)]$, $\widehat{H}_3(\theta) = \frac{1}{n}\sum_{i=1}^{n}\nabla^3 s(z_i, \theta)$, $H_3 = H_3(\theta_0)$, $W_3 \equiv \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(\nabla^3 s(z_i, \theta_0) - E[\nabla^3 s(z_i, \theta_0)]\right)$. Furthermore, we write:

$$a_{-1/2} = QJ, \quad a_{-1} = Q\left(Va_{-1/2} + \frac{1}{2}H_2(a_{-1/2} \otimes a_{-1/2})\right)$$

$$a_{-3/2} = QVa_{-1} + \frac{1}{2}QW(a_{-1/2} \otimes a_{-1/2}) + \frac{1}{2}QH_2(a_{-1/2} \otimes a_{-1} + a_{-1} \otimes a_{-1/2})$$

$$+ \frac{1}{6}QH_3(a_{-1/2} \otimes a_{-1/2} \otimes a_{-1/2})$$

for the ease of notation. We obtain:

Lemma 3. *Suppose $\{z_i\}_{i=1}^{n}$ are i.i.d., θ_0 is in the interior of Θ, 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) $\nabla^v s(z, \theta)$ is integrable for each fixed $\theta \in \Theta_0$, $v = \{0, 1, 2, \ldots \kappa\}$, $\kappa \geq 4$; (iib) $E[\nabla^4 s(z, \theta)]$ is continuous and bounded at θ_0; (iii) $\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(\nabla^3 s(z_i, \overline{\theta}) - H_3(\overline{\theta})\right) - \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(\nabla^3 s(z_i, \theta_0) - H_3(\theta_0)\right) = o_p(1)$ 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 $\sqrt{n}\left(\widehat{\theta} - \theta_0\right) = a_{-1/2} + \frac{1}{\sqrt{n}}a_{-1} + \frac{1}{n}a_{-3/2} + O_p(n^{-3/2})$.*

Note that the conditions in Lemma 3 are all standard regularity conditions.

In the following section, we propose an alternative one-step estimator that eliminates the second-order bias by adjusting the moment equations inspired by Firth (1993) [13].

---

[4] The fact that the estimating equation is the sum of n independent terms allows one to simply estimate the bias terms using their sample analogues. This approach does not have a direct extension to cases where the estimating function takes a more general form.

3. Bias-Corrected Moment Equation

Here, we consider an alternative higher order bias reduced estimator that solves bias-corrected moment equations. This idea was proposed in Firth (1993) [13] for the ML with a fixed number of parameters and exploited in Hahn and Newey (2004) [6] and Fernandez-Val (2004) [14] for the nonlinear panel data models with individual specific effects. We refer to this estimator as Firth's estimator.

To be precise, consider:

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

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

$$c(\theta) = Q(\theta)^{-1} B(\theta) = \frac{1}{2} H_2(\theta) E\left[Q(\theta)s(z_i, \theta) \otimes Q(\theta)s(z_i, \theta)\right] + E\left[\nabla s(z_i, \theta) Q(\theta)s(z_i, \theta)\right]. \quad (5)$$

This correction term $c(\theta)$ is obtained following Firth (1993) [13] and using the bias term for the M-estimator. In the ML context, Firth (1993) [13] shows that by adjusting the score function (he refers to this as a modified score function) with the correction term defined by the product of the Fisher information matrix and the bias term, one can obtain a bias-corrected ML estimator. $c(\theta)$ 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, (5) is a generalization of Firth (1993) [13]'s approach to the M-estimation . In general $c(\theta)$ contains population terms, and hence, to implement this alternative estimator, we need to estimate the function $c(\theta)$. We use a sample analogue of (5) as:

$$\begin{aligned}
\widehat{c}(\theta) &= \widehat{Q}(\theta)^{-1} \widehat{B}(\theta) & (6) \\
&= \frac{1}{2} \widehat{H}_2(\theta) \left(\frac{1}{n} \sum_{i=1}^{n} \left[\widehat{Q}(\theta)s(z_i, \theta) \otimes \widehat{Q}(\theta)s(z_i, \theta) \right] \right) + \frac{1}{n} \sum_{i=1}^{n} \left[\nabla s(z_i, \theta) \widehat{Q}(\theta)s(z_i, \theta) \right].
\end{aligned}$$

Now, we estimate θ_0 by solving:

$$0 = \frac{1}{n} \sum_{i=1}^{n} s(z_i, \theta) - \frac{1}{n} \widehat{c}(\theta), \quad (7)$$

and claim that the solution of this modified moment condition eliminates the second-order bias of $\widehat{\theta}$ that solves the original moment condition (2).

Assumption 1. *(i)* $\{z_i\}_{i=1}^{n}$ *are i.i.d.; (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)* $E\left[\sup_{\theta \in \Theta_0} \|\nabla^v s(z, \theta)\|^2\right] < \infty$, $v = \{0, 1, 2, \ldots \kappa\}$, $\kappa \geq 4$; *(iv)* Θ *is compact; (v)* θ_0 *is in the interior of Θ and is the only $\theta \in \Theta$ satisfying (1); (vi)* $E\left[\|\nabla^{\overline{v}} s(z, \theta_0)\|^4\right] < \infty$ *for* $\overline{v} = \{0, 1, 2, \ldots, \overline{\kappa}\}, \overline{\kappa} \geq 3$.

Assumption 2. *For $\theta \in \Theta_0$,* $E\left[\frac{\partial s(z_i, \theta)}{\partial \theta'}\right]$ *is nonsingular.*

Alternatively, we can assume the following instead of Assumption 1.

Assumption 3. *(i)* $\{z_i\}_{i=1}^{n}$ *are i.i.d.; (ii)* $\nabla^v s(z, \theta)$ *satisfies the Lipschitz condition in θ as:*

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

for some function $B_v(\cdot): \mathcal{Z} \to R$ and $E\left[B_v(\cdot)^{2t+\delta}\right] < \infty$, $v = \{0, 1, 2, \ldots \kappa\}$, *with positive integer $t \geq 2$ and for some $\delta > 0$ and $\kappa \geq 4$ in a neighborhood of θ_0; (iii)* $E\left[\sup_{\theta \in \Theta_0} \|\nabla^v s(z, \theta)\|^{2t+\delta}\right] < \infty$, $v = \{0, 1, 2, \ldots \kappa\}$,

$\kappa \geq 4$ *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 Assumptions 1 and 2 or Assumptions 3 and 2, the following three conditions are satisfied (see Lemma A.9 in Appendix A).

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

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

Note that these three conditions are required to control for the estimation error in $\widehat{c}(\theta)$ in the stochastic expansions. Now, we are ready to present one of our main results.

Proposition 1. *Suppose* θ^* *solves (7) where* $\widehat{c}(\theta)$ *is given by (6) and that* θ^* *is a consistent estimator of* θ_0. *Further, suppose that Conditions 1–3 and Conditions (i)–(viii) in Lemma 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 equations as (7), and it is a proper alternative to the analytic bias correction of (3).

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 bias correction, including analytical corrections that we focus on in this paper, the jackknife and bootstrap methods. These abundant ways of bias correction evoke the issue of which method is preferable to others at least on asymptotic efficiency grounds. For the ML estimation, Hahn et al. (2004) [12] show that the method of bias correction does not affect the higher order efficiency of any bias-corrected estimator that is first-order efficient. Although the ML estimator is a class of the M-estimator we consider, it is not trivial to conjecture that the same equivalence result will hold for a general class of M-estimators because the equivalence in the ML can hold due to some specific properties of the ML estimator. In this section, we formally extend the equivalence result to a general M-estimator.

We compare the higher order efficiency of several first-order efficient bias-corrected estimators by comparing the higher order variances, which are defined by the $O\left(\frac{1}{n}\right)$ variance in a third-order stochastic expansion of an estimator.

4.1. Third-Order Expansion of the One-Step Bias-Corrected Estimator

To compare with the estimator of interest θ^* in (7), first we consider a one-step 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 (6). 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 some $\widetilde{\theta}$ between $\widehat{\theta}$ and θ_0, a first-order Taylor expansion 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 because $\widehat{\theta} - \theta_0 = O_p\left(\frac{1}{\sqrt{n}}\right)$. Furthermore, we have:

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

by the triangle inequality, Lemma A.7 (in Appendix A) and the continuity of $c(\theta)$ at θ_0 (applying the Slutsky theorem), hence, both $B(\widehat{\theta})$ and $\widehat{B}(\widehat{\theta})$ are indeed consistent estimators of the higher order bias noting that $Q(\widehat{\theta}) = Q(\theta_0) + o_p(1)$ by the continuity of $Q(\theta)$ at θ_0 and $\widehat{Q}(\widehat{\theta}) = Q(\theta_0) + o_p(1)$.

Now, from the result of Lemma 3 and a second-order Taylor expansion of $B(\widehat{\theta})$, it follows that:

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

where $\widetilde{\theta}$ is a point between $\widehat{\theta}$ and θ_0, and hence:

$$
\sqrt{n}(\widehat{\theta}_b - \theta_0) = a_{-1/2} + \frac{1}{\sqrt{n}}\left(a_{-1} - B(\theta_0)\right) + \frac{1}{n}\left(a_{-3/2} - \nabla B(\theta_0)a_{-1/2}\right) + O_p(n^{-3/2}), \tag{8}
$$

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 conclude $\frac{1}{2\sqrt{n}}\nabla^2 B(\widetilde{\theta})((\widehat{\theta} - \theta_0) \otimes (\widehat{\theta} - \theta_0)) = O_p(n^{-3/2})$.

Now, similarly for $\widehat{\theta}_{bc}$, we obtain:

$$
\begin{aligned}
\sqrt{n}(\widehat{\theta}_{bc} - \theta_0) &= \sqrt{n}(\widehat{\theta} - \theta_0) - \frac{1}{\sqrt{n}}\widehat{B}(\widehat{\theta}) \\
&= \begin{array}{l} a_{-1/2} + \frac{1}{\sqrt{n}}a_{-1} + \frac{1}{n}a_{-3/2} + O_p(n^{-3/2}) \\ - \frac{1}{\sqrt{n}}\widehat{B}(\theta_0) - \frac{1}{\sqrt{n}}\nabla \widehat{B}(\theta_0)(\widehat{\theta} - \theta_0) - \frac{1}{2\sqrt{n}}\nabla^2 \widehat{B}(\widetilde{\theta})((\widehat{\theta} - \theta_0) \otimes (\widehat{\theta} - \theta_0)). \end{array}
\end{aligned} \tag{9}
$$

Then, applying the following three results (that hold under Assumptions 1 and 2 or 3 and 2, as shown in Lemma A.12 in Appendix A):

Condition 4. $\widehat{B}(\theta_0) = B(\theta_0) + O_p\left(1/\sqrt{n}\right)$,

Condition 5. $\nabla\widehat{B}(\theta_0) = \nabla B(\theta_0) + O_p\left(1/\sqrt{n}\right)$,

Condition 6. $\nabla^2\widehat{B}(\theta) = O_p(1)$ *in the neighborhood of* θ_0,

we obtain:

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

noting that $\frac{1}{\sqrt{n}}\nabla\widehat{B}(\theta_0)(\widehat{\theta} - \theta_0) = \frac{1}{n}\nabla\widehat{B}(\theta_0)\left(a_{-1/2} + O_p\left(\frac{1}{\sqrt{n}}\right)\right) = \frac{1}{n}\nabla B(\theta_0)a_{-1/2} + O_p(n^{-3/2})$ by Condition 5 and that $\frac{1}{2\sqrt{n}}\nabla^2\widehat{B}(\widetilde{\theta})((\widehat{\theta} - \theta_0) \otimes (\widehat{\theta} - \theta_0)) = O_p(n^{-3/2})$ by Condition 6 and the fact that $\widehat{\theta} - \theta_0 = O_p\left(1/\sqrt{n}\right)$.

4.2. Third-Order Expansion of the Bias-Corrected Moment Equations Estimator

Now, we derive the higher order expansion of the proposed bias-corrected estimator θ^* up to the third order. For this, we need to verify an additional condition below, which is satisfied under Assumptions 1 and 2 or 3 and 2 with $\kappa \geq 5$ as shown in Lemma A.10 and A.11 in Appendix A.

Condition 7. *(i)* $\nabla \widehat{c}(\theta_0) = \nabla c(\theta_0) + O_p\left(\frac{1}{\sqrt{n}}\right)$; *(ii)* $\nabla^3 \widehat{c}(\theta) = O_p(1)$ *in 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 2. *Suppose* θ^* *solves (7), where* $\widehat{c}(\theta)$ *is given in (6), and that* θ^* *is consistent. Further, suppose that Conditions 1–7 and Conditions (i)–(viii) in Lemma 3 are satisfied, and assume* $\sqrt{n}(\widehat{\theta} - \theta_0) = a_{-1/2} + \frac{1}{\sqrt{n}}(a_{-1} - B(\theta_0)) + O_p\left(\frac{1}{n}\right)$. *Then, we have:*

$$
\begin{aligned}
&\sqrt{n}\left(\theta^* - \theta_0\right) \\
&= a_{-1/2} + \frac{1}{\sqrt{n}}(a_{-1} - B(\theta_0)) + \frac{1}{n}\left(a_{-3/2} - \nabla B(\theta_0)a_{-1/2} - \sqrt{n}(\widehat{B}(\theta_0) - B(\theta_0))\right) + O_p(n^{-3/2}).
\end{aligned} \tag{11}
$$

Comparing (10) and (11), we therefore 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 implies that θ^* and $\widehat{\theta}_{bc}$ at least agree upon their higher order variances, as we discuss in the following section.

4.3. Higher Order Variances

For a three-term stochastic expansion of an estimator $\widecheck{\theta}$, such as:

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

the higher order variance is given by:

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

with $\Sigma = \mathrm{Var}[T_{-1/2}]$ and $\Xi = \mathrm{Var}[T_{-1}] + E\left[\left(\sqrt{n}T_{-1} + T_{-3/2}\right)T'_{-1/2}\right] + E\left[T_{-1/2}\left(\sqrt{n}T_{-1} + T_{-3/2}\right)'\right]$. Then, from the third-order stochastic expansions of the bias-corrected estimators derived in (8), (10) and (11), we can obtain the higher order variances of three alternative estimators, denoted by $\Lambda_{\widehat{\theta}_b}$, $\Lambda_{\widehat{\theta}_{bc}}$, and Λ_{θ^*}, respectively, as:[5]

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

$$
\Lambda_{\widehat{\theta}_{bc}} = \Lambda_{\widehat{\theta}_b} - \frac{1}{n}E\left[a_{-1/2}\sqrt{n}(\widehat{B}(\theta_0) - B(\theta_0))'\right] - \frac{1}{n}E\left[\sqrt{n}(\widehat{B}(\theta_0) - B(\theta_0))a'_{-1/2}\right] \tag{12}
$$

$$
\Lambda_{\theta^*} = \Lambda_{\widehat{\theta}_{bc}}. \tag{13}
$$

First note that the result of (12) reveals that the higher order variance of $\widehat{\theta}_{bc}$ has additional terms compared with $\widehat{\theta}_b$ (infeasible estimator) because we use the sample analogue of the second-order bias, unless $E\left[a_{-1/2}\sqrt{n}(\widehat{B}(\theta_0) - B(\theta_0))\right] = 0$. These additional terms contribute to the cost of estimating

---

5   The analytic forms of these variances are provided in Appendix C.

the bias term $\widehat{B}(\cdot)$ in the analytic bias correction approach. Now, the result of (13) tells that the higher order variances of two alternative bias-corrected estimators are the same, so on the grounds of this higher order variances' comparison, we find that the bias correction method by adjusting moment equations does not improve over the analytic bias correction.

Indeed, it is more remarkable that comparing the third-order expansions of (10) and (11), we further find:

$$n^{3/2}(\widehat{\theta}_{bc} - \theta^*) = o_p(1). \tag{14}$$

This is a stronger result than just comparing the higher order variances because it implies that these two estimators do not only have the same higher order variance, but also agree on more properties in terms of their stochastic expansions. In the literature, it has been argued that removing the bias directly from the moment equations has an attractive feature that it does not use pre-estimated parameters that are not bias corrected, though this alternative approach requires more intensive computations, since it requires solving some nonlinear equation. However, in view of the result (14), this paper concludes 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.

4.4. Further Comparison of Alternative Bias Corrections

To have a better understanding of the equivalence result (14), here we compare several versions of bias-corrected estimators that are infeasible by their nature. First, let θ_1^* be the solution of $0 = \frac{1}{n}\sum_{i=1}^n s(z_i, \theta) - \frac{1}{n}c(\theta)$ where $c(\theta)$ is known. We also define two other bias-corrected estimators:

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

so for $\widehat{\theta}_2$, $c(\theta)$ is known, but $Q(\theta)$ is estimated, while for $\widehat{\theta}_3$, $c(\theta)$ is estimated, but $Q(\theta)$ is known. For these estimators, we obtain the following results.[6]

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

$$\sqrt{n}(\theta_1^* - \theta_0) = \sqrt{n}(\widehat{\theta}_2 - \theta_0) + O_p(n^{-3/2}),$$

$$\sqrt{n}(\theta^* - \theta_0) = \sqrt{n}(\widehat{\theta}_3 - \theta_0) + \frac{1}{n}QVB(\theta_0) + O_p(n^{-3/2}),$$

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

The results illustrate that using $\widehat{Q}(\cdot)$ rather than $Q(\cdot)$ in the bias correction term plays a critical role for equating the stochastic expansions (up to the third order) of the one-step bias-corrected estimator and the bias-corrected moment equations estimator. To see the point compare θ_1^* and $\widehat{\theta}_2$ and compare θ^* and $\widehat{\theta}_{bc}$ where in the former $c(\cdot)$ is known, and in the latter, $c(\cdot)$ is also estimated.

Next, we consider the possible iteration of the bias correction. Hahn and Newey (2004) [6] discuss the relationship between the bias correction of moment equations and the iterated bias correction. The iteration idea is that one can update $\widehat{B}(\cdot)$ several times using the previous estimator of $\widehat{\theta}$. To be 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 $\widehat{\theta}_{bc}^k = \widehat{\theta} - \widehat{B}(\widehat{\theta}_{bc}^{k-1})/n$ for $k \geq 2$ where $\widehat{\theta}_{bc}^1 = \widehat{\theta}_{bc}$. If we iterate this procedure until convergence, we will obtain $\widehat{\theta}_{bc}^\infty = \widehat{\theta} - \widehat{B}(\widehat{\theta}_{bc}^\infty)/n$, which implies that $\widehat{\theta}_{bc}^\infty$ solves (note $\widehat{B}(\theta) = \widehat{Q}(\theta)\widehat{c}(\theta)$):

$$0 = \widehat{Q}(\theta)^{-1}(\widehat{\theta} - \theta) - \frac{1}{n}\widehat{c}(\theta) = \frac{1}{n}\sum_{i=1}^n s(z_i, \widehat{\theta}) + \widehat{Q}(\theta)^{-1}(\widehat{\theta} - \theta) - \frac{1}{n}\widehat{c}(\theta), \tag{15}$$

6　Detailed derivations are available upon request.

where the second equality is from the definition of $\widehat{\theta}$ in (2). Observing that $\widehat{Q}(\theta)^{-1} = -\frac{1}{n}\sum_{i=1}^{n}\nabla s(z_i, \theta)$, if $s(z_i, \theta)$ is linear in θ, then we find that the Equation (15) is the same as Equation (7) for the bias-corrected moment equations; hence, $\widehat{\theta}_{bc}^{\infty}$ is exactly the same with θ^*. Otherwise, (15) is an approximation of (7). 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 (7). Similarly to (9), for $\widetilde{\widetilde{\theta}}$ between $\widehat{\theta}_{bc}^{\infty}$ and θ_0, we can show that:

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

using Conditions 4, 5 and 6 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 $\widehat{\theta}_{bc}^k$ ($k \geq 2$).

From this equivalence of the higher order expansions for $\widehat{\theta}_{bc}^{\infty}$ and $\widehat{\theta}_{bc}$ at least up to the third order term, one would expect that the higher order expansion of θ^* will be equivalent to that of $\widehat{\theta}_{bc}$ at least up to the third order, and we have verified that this intuition is correct. However, as observed in some Monte Carlo examples of Hahn and Newey (2004) [6] and Fernandez-Val (2004) [14], the iterative bias correction can lower the 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 equations (or the fully-iterated bias correction) should be based on the stochastic expansions higher than the third order. The comparison of the two alternative bias-corrected estimators over the fourth order or even higher order stochastic expansions can be challenging and is beyond the scope of this paper.

5. Conclusions

This paper considers an alternative bias correction for the M-estimator, which is achieved by correcting the moment equations in the spirit of Firth (1993) [13]. 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 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. 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 the asymptotic properties in general; neither will the alternative estimator. We have verified this intuition in this paper.

Appendix A. Technical Lemmas and Proofs

Some preliminary Lemmas and their proofs are available in the Supplementary Material, which are useful to derive the main results presented in the paper.

Appendix B. Proofs of the Main Lemmas and Propositions

This section collects proofs for the main results in the paper.

Appendix B.1. Proposition 1

Proof. By the first-order Taylor series approximation of (7), we have:

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

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

$$
\begin{aligned}
&\sqrt{n}\left(\theta^* - \theta_0\right) \\
&= -\left(\frac{1}{n} \sum_{i=1}^{n} \nabla s\left(z_i, \widetilde{\theta}\right) - \frac{1}{n} \nabla \widehat{c}(\widetilde{\theta})\right)^{-1} \left(\frac{1}{\sqrt{n}} \sum_{i=1}^{n} s\left(z_i, \theta_0\right) - \frac{1}{\sqrt{n}} \widehat{c}(\theta_0)\right) \\
&= -\left(E\left[\nabla s\left(z_i, \theta_0\right)\right] + o_p(1) + O_p\left(\frac{1}{n}\right)\right)^{-1} \left(\frac{1}{\sqrt{n}} \sum_{i=1}^{n} s\left(z_i, \theta_0\right) + O_p\left(\frac{1}{\sqrt{n}}\right)\right) = QJ + o_p(1),
\end{aligned} \tag{B1}
$$

by Conditions 1(i), 2 and $\widetilde{\theta} = \theta_0 + o_p(1)$ provided that $\left\| \frac{1}{n} \sum_{i=1}^{n} \nabla s\left(z_i, \overline{\theta}\right) - E\left[\nabla s\left(z_i, \theta_0\right)\right] \right\| = o_p(1)$ for $\overline{\theta} = \theta_0 + o_p(1)$. This confirms that the estimator has the same first-order asymptotic distribution with $\sqrt{n}(\widehat{\theta} - \theta_0)$. Recalling $\widehat{H}_1(\theta) \equiv \frac{1}{n} \sum_{i=1}^{n} \nabla s\left(z_i, \theta\right)$ and $H_1(\theta_0)(= -Q^{-1}) \equiv E[\nabla s\left(z_i, \theta_0\right)]$, we can rewrite (B1) as:

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

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

$$
0 = \left\{
\begin{array}{l}
\frac{1}{n} \sum_{i=1}^{n} s\left(z_i, \theta_0\right) + \left(\frac{1}{n} \sum_{i=1}^{n} \nabla s\left(z_i, \theta_0\right)\right)\left(\theta^* - \theta_0\right) \\
+ \frac{1}{2}\left(\frac{1}{n} \sum_{i=1}^{n} \nabla^2 s\left(z_i, \widetilde{\theta}\right)\right)\left(\left(\theta^* - \theta_0\right) \otimes \left(\theta^* - \theta_0\right)\right) \\
- \frac{1}{n} \widehat{c}(\theta_0) - \frac{1}{n} \nabla \widehat{c}(\theta_0)\left(\theta^* - \theta_0\right) - \frac{1}{2n} \nabla^2 \widehat{c}\left(\widetilde{\theta}\right)\left(\left(\theta^* - \theta_0\right) \otimes \left(\theta^* - \theta_0\right)\right)
\end{array}
\right\}. \tag{B2}
$$

We rewrite (B2) as:

$$
\begin{aligned}
0 \;=\; & \left\{
\begin{array}{l}
\frac{1}{\sqrt{n}}J + \left(-Q^{-1} + \frac{1}{\sqrt{n}}V\right)\left(\theta^* - \theta_0\right) \\
+\frac{1}{2}\left(H_2 + \frac{1}{\sqrt{n}}W\right)\left(\left(\theta^* - \theta_0\right) \otimes \left(\theta^* - \theta_0\right)\right) - \frac{1}{n}\widehat{c}(\theta_0) - \frac{1}{n}\nabla\widehat{c}(\theta_0)\left(\theta^* - \theta_0\right) \\
-\frac{1}{2n}\nabla^2\widehat{c}\left(\widetilde{\theta}\right)\left(\left(\theta^* - \theta_0\right) \otimes \left(\theta^* - \theta_0\right)\right) + O_p(n^{-3/2})
\end{array}
\right\} \\
\;=\; & \left\{
\begin{array}{l}
\frac{1}{\sqrt{n}}J + \left(-Q^{-1} + \frac{1}{\sqrt{n}}V\right)\left(\theta^* - \theta_0\right) \\
+\frac{1}{2}\left(H_2 + \frac{1}{\sqrt{n}}W\right)\left(\left(\theta^* - \theta_0\right) \otimes \left(\theta^* - \theta_0\right)\right) - \frac{1}{n}\widehat{c}(\theta_0) + O_p(n^{-3/2})
\end{array}
\right\} \quad \text{(B3)}
\end{aligned}
$$

since (a) $\frac{1}{n}\nabla\widehat{c}(\theta_0)(\theta^* - \theta_0) = O_p(n^{-3/2})$ by Condition 2 and $\theta^* = \theta_0 + O_p(\frac{1}{\sqrt{n}})$ from (B1) noting $J = O_p(1)$ and since (b):

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

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

From (B3), 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 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 $H_2 \equiv H_2(\theta_0)$):

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

by the definition of $c(\theta)$ in (5) and Lemma 2. $\square$

Appendix B.2. Lemma 3

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

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

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:

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

and hence:

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

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 to 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:

$$
0 = \left\{
\begin{aligned}
& \tfrac{1}{n} \sum_{i=1}^{n} s \left(z_i, \theta_0 \right) + \left(\tfrac{1}{n} \sum_{i=1}^{n} \nabla s \left(z_i, \theta_0 \right) \right) \left(\widehat{\theta} - \theta_0 \right) \\
& + \tfrac{1}{2} \left(\tfrac{1}{n} \sum_{i=1}^{n} \nabla^2 s \left(z_i, \theta_0 \right) \right) \left(\left(\widehat{\theta} - \theta_0 \right) \otimes \left(\widehat{\theta} - \theta_0 \right) \right) \\
& + \tfrac{1}{6} \left(\tfrac{1}{n} \sum_{i=1}^{n} \nabla^3 s \left(z_i, \theta_0 \right) \right) \left(\left(\widehat{\theta} - \theta_0 \right) \otimes \left(\widehat{\theta} - \theta_0 \right) \otimes \left(\widehat{\theta} - \theta_0 \right) \right) + O_p(n^{-2})
\end{aligned}
\right\}. \tag{B4}
$$

Now, note:

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

We then rewrite (B4) as:

$$
0 = \left\{
\begin{aligned}
& \tfrac{1}{\sqrt{n}} J + \left(-Q^{-1} + \tfrac{1}{\sqrt{n}} V \right) \left(\widehat{\theta} - \theta_0 \right) + \tfrac{1}{2} \left(H_2 + \tfrac{1}{\sqrt{n}} W \right) \left(\left(\widehat{\theta} - \theta_0 \right) \otimes \left(\widehat{\theta} - \theta_0 \right) \right) \\
& + \tfrac{1}{6} \left(H_3 + \tfrac{1}{\sqrt{n}} W_3 \right) \left(\left(\widehat{\theta} - \theta_0 \right) \otimes \left(\widehat{\theta} - \theta_0 \right) \otimes \left(\widehat{\theta} - \theta_0 \right) \right) + O_p(n^{-2})
\end{aligned}
\right\}. \tag{B5}
$$

Now, note that we can expand:

$$
\left(-Q^{-1} + \tfrac{1}{\sqrt{n}} V \right)^{-1} = \left(I - Q \tfrac{1}{\sqrt{n}} V \right)^{-1} \left(-Q \right) \tag{B6}
$$
$$
= -Q + O_p(\tfrac{1}{\sqrt{n}}) = -Q - \tfrac{1}{\sqrt{n}} QVQ + O_p(n^{-1}) = -Q - \tfrac{1}{\sqrt{n}} QVQ - \tfrac{1}{n} QVQVQ + O_p(n^{-3/2})
$$

depending on the expansions we need. Plugging (B6) into (B5) (depending on the orders we need) and inspecting the orders, we have:

$$
\widehat{\theta} - \theta_0 \tag{B7}
$$

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

$$
= \left\{
\begin{aligned}
& -\left(-Q - \tfrac{1}{\sqrt{n}} QVQ - \tfrac{1}{n} QVQVQ + O_p(n^{-3/2}) \right) \tfrac{1}{\sqrt{n}} J \\
& -\tfrac{1}{2} \left(-Q - \tfrac{1}{\sqrt{n}} QVQ + O_p(n^{-1}) \right) \left(H_2 + \tfrac{1}{\sqrt{n}} W \right) \left(\left(\widehat{\theta} - \theta_0 \right) \otimes \left(\widehat{\theta} - \theta_0 \right) \right) \\
& -\tfrac{1}{6} \left(-Q + O_p \left(\tfrac{1}{\sqrt{n}} \right) \right) \left(H_3 + \tfrac{1}{\sqrt{n}} W_3 \right) \left(\left(\widehat{\theta} - \theta_0 \right) \otimes \left(\widehat{\theta} - \theta_0 \right) \otimes \left(\widehat{\theta} - \theta_0 \right) \right) + O_p(n^{-2})
\end{aligned}
\right\}
$$

$$
= \left\{
\begin{aligned}
& \left(\tfrac{1}{\sqrt{n}} QJ + \tfrac{1}{n} QVQJ + \tfrac{1}{n^{3/2}} QVQVQJ \right) \\
& + \tfrac{1}{2} \left(Q \left(H_2 + \tfrac{1}{\sqrt{n}} W \right) + \tfrac{1}{\sqrt{n}} QVQH_2 \right) \left(\left(\widehat{\theta} - \theta_0 \right) \otimes \left(\widehat{\theta} - \theta_0 \right) \right) \\
& + \tfrac{1}{6} QH_3 \left(\left(\widehat{\theta} - \theta_0 \right) \otimes \left(\widehat{\theta} - \theta_0 \right) \otimes \left(\widehat{\theta} - \theta_0 \right) \right) + O_p(n^{-2})
\end{aligned}
\right\}. \tag{B8}
$$

Now, plugging $\sqrt{n}(\widehat{\theta}-\theta_0)=a_{-1/2}+O_p\left(1/\sqrt{n}\right)$ or $\sqrt{n}(\widehat{\theta}-\theta_0)=a_{-1/2}+\frac{1}{\sqrt{n}}a_{-1}+O_p\left(1/n\right)$ into (B8) depending on the orders required, we obtain:

$$
\begin{aligned}
&\widehat{\theta}-\theta_0\\
&=\left\{
\begin{array}{l}
\left(\frac{1}{\sqrt{n}}QJ+\frac{1}{n}QVQJ+\frac{1}{n^{3/2}}QVQVQJ\right)\\
+\frac{1}{2}\frac{1}{n}\left(Q\left(H_2+\frac{W}{\sqrt{n}}\right)+\frac{QVQH_2}{\sqrt{n}}\right)\left(\left(a_{-1/2}+\frac{a_{-1}}{\sqrt{n}}+O_p\left(\frac{1}{n}\right)\right)\otimes\left(a_{-1/2}+\frac{a_{-1}}{\sqrt{n}}+O_p\left(\frac{1}{n}\right)\right)\right)\\
+\frac{1}{6}\frac{1}{n^{3/2}}QH_3\left(\left(a_{-1/2}+O_p\left(\frac{1}{\sqrt{n}}\right)\right)\otimes\left(a_{-1/2}+O_p\left(\frac{1}{\sqrt{n}}\right)\right)\otimes\left(a_{-1/2}+O_p\left(\frac{1}{\sqrt{n}}\right)\right)\right)\\
+O_p(n^{-2})
\end{array}
\right\}\\[2ex]
&=\left\{
\begin{array}{l}
\left(\frac{1}{\sqrt{n}}QJ+\frac{1}{n}QVQJ+\frac{1}{n^{3/2}}QVQVQJ\right)\\
+\frac{1}{2}\frac{1}{n}\left(Q\left(H_2+\frac{1}{\sqrt{n}}W\right)+\frac{1}{\sqrt{n}}QVQH_2\right)\left(\left(a_{-1/2}+\frac{1}{\sqrt{n}}a_{-1}\right)\otimes\left(a_{-1/2}+\frac{1}{\sqrt{n}}a_{-1}\right)\right)\\
+\frac{1}{6}\frac{1}{n^{3/2}}QH_3\left(a_{-1/2}\otimes a_{-1/2}\otimes a_{-1/2}\right)+O_p(n^{-2})
\end{array}
\right\}\quad\text{(B9)}\\[2ex]
&=\left\{
\begin{array}{l}
\left(\frac{1}{\sqrt{n}}QJ+\frac{1}{n}QVQJ+\frac{1}{n^{3/2}}QVQVQJ\right)+\frac{1}{2}\frac{1}{n}QH_2\left(a_{-1/2}\otimes a_{-1/2}\right)\\
+\frac{1}{2}\frac{1}{n^{3/2}}\left(QH_2\left(\left(a_{-1/2}\otimes a_{-1}\right)+\left(a_{-1}\otimes a_{-1/2}\right)\right)+\left(QW+QVQH_2\right)\left(a_{-1/2}\otimes a_{-1/2}\right)\right)\\
+\frac{1}{6}\frac{1}{n^{3/2}}QH_3\left(a_{-1/2}\otimes a_{-1/2}\otimes a_{-1/2}\right)+O_p(n^{-2})
\end{array}
\right\}.
\end{aligned}
$$

Finally, rearranging (B9) according to the orders, we have:

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

□

Appendix B.3. Proposition 2

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

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

From this, similarly as (B4) to (B5), 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 $\frac{1}{2}\frac{1}{n}\nabla^2\widehat{c}(\theta_0)\left(\left(\theta^*-\theta_0\right)\otimes\left(\theta^*-\theta_0\right)\right) = O_p(n^{-2})$ by Condition 3 and $\theta^* = \theta_0 + O_p(\frac{1}{\sqrt{n}})$ and since:

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

by Condition 7(ii) and $\theta^* = \theta_0 + O_p(\frac{1}{\sqrt{n}})$. Similarly to (B7), we obtain:

$$\theta^*-\theta_0 \tag{B10}$$
$$= \left\{\begin{array}{l} -\left(-Q-\frac{1}{\sqrt{n}}QVQ-\frac{1}{n}QVQVQ+O_p(n^{-3/2})\right)\frac{1}{\sqrt{n}}J \\ -\frac{1}{2}\left(-Q-\frac{1}{\sqrt{n}}QVQ+O_p(n^{-1})\right)\left(H_2+\frac{1}{\sqrt{n}}W\right)\left(\left(\theta^*-\theta_0\right)\otimes\left(\theta^*-\theta_0\right)\right) \\ -\frac{1}{6}\left(-Q+O_p\left(\frac{1}{\sqrt{n}}\right)\right)\left(H_3+\frac{1}{\sqrt{n}}W_3\right)\left(\left(\theta^*-\theta_0\right)\otimes\left(\theta^*-\theta_0\right)\otimes\left(\theta^*-\theta_0\right)\right) \\ +\frac{1}{n}\left(-Q-\frac{1}{\sqrt{n}}QVQ+O_p(n^{-1})\right)\widehat{c}(\theta_0)+\frac{1}{n}\left(-Q+O_p\left(\frac{1}{\sqrt{n}}\right)\right)\nabla\widehat{c}(\theta_0)\left(\theta^*-\theta_0\right) \\ +O_p(n^{-2}) \end{array}\right\}$$

$$= \left\{\begin{array}{l} \left(\frac{1}{\sqrt{n}}QJ+\frac{1}{n}QVQJ+\frac{1}{n^{3/2}}QVQVQJ\right) \\ +\frac{1}{2}\left(Q\left(H_2+\frac{1}{\sqrt{n}}W\right)+\frac{QVQH_2}{\sqrt{n}}\right)\left(\left(\theta^*-\theta_0\right)\otimes\left(\theta^*-\theta_0\right)\right) \\ +\frac{1}{6}QH_3\left(\left(\theta^*-\theta_0\right)\otimes\left(\theta^*-\theta_0\right)\otimes\left(\theta^*-\theta_0\right)\right) \\ -\frac{1}{n}\left(Q+\frac{1}{\sqrt{n}}QVQ\right)\widehat{c}(\theta_0)-\frac{1}{n}Q\nabla\widehat{c}(\theta_0)\left(\theta^*-\theta_0\right)+O_p(n^{-2}). \end{array}\right\} \tag{B11}$$

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 (B11) depending on the orders required, we obtain:

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

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

where we replaced $\nabla \widehat{c}(\theta_0)$ with $\nabla c(\theta_0) + O_p\left(\frac{1}{\sqrt{n}}\right)$ from Condition 7(i). Rearranging terms according to the orders, we have:

$$\theta^* - \theta_0 \tag{B12}$$
$$= \left\{ \begin{array}{l} \frac{1}{\sqrt{n}} QJ + \frac{1}{n}\left(QVQJ + \frac{1}{2}QH_2(a_{-1/2} \otimes a_{-1/2}) - Q\widehat{c}(\theta_0)\right) \\ + \frac{1}{n^{3/2}}\left(\begin{array}{l} QVQVQJ + \frac{1}{2}QH_2\left((a_{-1/2} \otimes (a_{-1} - Qc(\theta_0))) + ((a_{-1} - Qc(\theta_0)) \otimes a_{-1/2})\right) \\ + \frac{1}{2}(QW + QVQH_2)(a_{-1/2} \otimes a_{-1/2}) \\ + \frac{1}{6}QH_3(a_{-1/2} \otimes a_{-1/2} \otimes a_{-1/2}) - QVQ\widehat{c}(\theta_0) - Q\nabla c(\theta_0)a_{-1/2} \end{array} \right) \\ + O_p(n^{-2}) \end{array} \right\}$$
$$= \left\{ \begin{array}{l} \frac{1}{\sqrt{n}} QJ + \frac{1}{n}\left(QVQJ + \frac{1}{2}QH_2(a_{-1/2} \otimes a_{-1/2}) - Qc(\theta_0)\right) \\ + \frac{1}{n^{3/2}}\left(\begin{array}{l} QVa_{-1} + \frac{1}{2}QH_2\left((a_{-1/2} \otimes (a_{-1} - Qc(\theta_0))) + ((a_{-1} - Qc(\theta_0)) \otimes a_{-1/2})\right) \\ + \frac{1}{2}QW(a_{-1/2} \otimes a_{-1/2}) + \frac{1}{6}QH_3(a_{-1/2} \otimes a_{-1/2} \otimes a_{-1/2}) \\ - QVQ\left(c(\theta_0) + O_p\left(\frac{1}{\sqrt{n}}\right)\right) - Q\nabla c(\theta_0)a_{-1/2} - \sqrt{n}Q(\widehat{c}(\theta_0) - c(\theta_0)) \end{array} \right) \\ + O_p(n^{-2}) \end{array} \right\}$$
$$= \left\{ \begin{array}{l} \frac{1}{\sqrt{n}} QJ + \frac{1}{n}\left(a_{-1} - Qc(\theta_0)\right) \\ + \frac{1}{n^{3/2}}\left(\begin{array}{l} a_{-3/2} - \frac{1}{2}QH_2\left((a_{-1/2} \otimes Qc(\theta_0)) + (Qc(\theta_0) \otimes a_{-1/2})\right) \\ - QVQc(\theta_0) - Q\nabla c(\theta_0)a_{-1/2} - \sqrt{n}Q(\widehat{c}(\theta_0) - c(\theta_0)) \end{array} \right) + O_p(n^{-2}), \end{array} \right\}$$

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

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:

$$\nabla c(\theta) = Q(\theta)^{-1}\nabla B(\theta) - vec^*\left(B(\theta)'\nabla\left(H_1(\theta)'\right)\right) \tag{B13}$$

from Remark A.2 in Appendix A. From (B12), note:

$$\sqrt{n}\left(\theta^* - \theta_0\right) \tag{B14}$$
$$= \begin{array}{l} a_{-1/2} + \frac{1}{\sqrt{n}}\left(a_{-1} - Qc(\theta_0)\right) \\ + \frac{1}{n}\left(a_{-3/2} - \frac{1}{2}QH_2\left((a_{-1/2} \otimes Qc(\theta_0)) + (Qc(\theta_0) \otimes a_{-1/2})\right) - QVQc(\theta_0) - Q\nabla c(\theta_0)a_{-1/2}\right) \\ - \frac{1}{n}Q\sqrt{n}(\widehat{c}(\theta_0) - c(\theta_0)) + O_p(n^{-3/2}) \end{array}$$

from (11), and also note that:

$$\begin{array}{l} \frac{1}{2}QH_2\left((a_{-1/2} \otimes Qc(\theta_0)) + (Qc(\theta_0) \otimes a_{-1/2})\right) + QVQc(\theta_0) + Q\nabla c(\theta_0)a_{-1/2} \\ = \frac{1}{2}QH_2\left((a_{-1/2} \otimes B(\theta_0)) + (B(\theta_0) \otimes a_{-1/2})\right) + QVB(\theta_0) \\ + Q\left(Q(\theta_0)^{-1}\nabla B(\theta_0) - vec^*\left(B(\theta_0)'\nabla\left(H_1(\theta_0)'\right)\right)\right)a_{-1/2} \\ = \frac{1}{2}QH_2\left((a_{-1/2} \otimes Qc(\theta_0)) + (Qc(\theta_0) \otimes a_{-1/2})\right) - Qvec^*\left(B(\theta_0)'\nabla\left(H_1(\theta_0)'\right)\right)a_{-1/2} \\ + \nabla B(\theta_0)a_{-1/2} + QVB(\theta_0) \end{array} \tag{B15}$$

from (B13) and $B(\theta) = Q(\theta)c(\theta)$. We claim that:

$$\frac{1}{2}H_2\left((a_{-1/2} \otimes B(\theta_0)) + (B(\theta_0) \otimes a_{-1/2})\right) - vec^*\left(B(\theta_0)'\nabla\left(H_1(\theta_0)'\right)\right)a_{-1/2} = 0, \tag{B16}$$

which simplifies (B15) to $\nabla B(\theta_0)a_{-1/2} + QVB(\theta_0)$. This is obvious when $\dim(\theta_0) = 1$, since:

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

and $vec^*\left(B(\theta_0)'\nabla\left(H_1\left(\theta_0\right)'\right)\right)a_{-1/2} = B(\theta_0)H_2 a_{-1/2}$ 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:

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

Finally, after some tedious algebra, we find $\tfrac{1}{2}H_2\left(I \otimes B(\theta_0) + B(\theta_0) \otimes I\right) = vec^*\left(B(\theta_0)'\nabla\left(H_1\left(\theta_0\right)'\right)\right)$, which concludes (B16). Therefore, we can rewrite (B14) as:

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

Now, 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:

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

This completes the proof.　□

Appendix C. Higher Order Variances

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

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

Rilstone et al. (1996) [15] derive the second-order mean squared error (MSE) of the M-estimator that solves the moment condition (2). Proposition 3.4 in Rilstone et al. (1996) [15] 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):

$$\gamma_1 = \overline{d_1 d_1'}, \gamma_2 = Q\left\{\overline{v_1 d_1 d_1'} + \frac{1}{2}H_2\overline{(d_1 \otimes d_1)\,d_1'}\right\}$$

$$\gamma_3 = \left\{\begin{array}{l} Q\left\{\overline{v_1 d_1 d_2' V_2'} + \overline{v_1 d_2 d_1' v_2'} + \overline{v_1 d_2 d_2' v_1'}\right\}Q' \\ +QH_2\left\{\overline{(d_1 \otimes d_1)(d_2' \otimes d_2')} + \overline{(d_1 \otimes d_2)(d_1' \otimes d_2')} + \overline{(d_1 \otimes d_2)(d_2' \otimes d_1')}\right\}H_2'Q' \\ -Q\left\{\overline{v_1 d_1\,(d_2' \otimes d_2')} + \overline{v_1 d_2\,(d_1' \otimes d_2')} + \overline{v_1 d_2\,(d_2' \otimes d_1')}\right\}H_2'Q' \\ -QH_2\left\{\overline{d_1 \otimes d_1 d_2' v_2'} + \overline{(d_1 \otimes d_2)\,d_1' v_2'} + \overline{(d_1 \otimes d_2)\,d_2' v_1'}\right\}Q' \end{array}\right\}$$

$$\gamma_4 = \left\{\begin{array}{l} Q\left\{\overline{v_1 Q v_1 d_2 d_2'} + \overline{v_1 Q v_2 d_1 d_2'} + \overline{v_1 Q v_2 d_2 d_1'}\right\} \\ +\frac{1}{2}Q\left\{\overline{v_1 Q H_2\,(d_1 \otimes d_2)\,d_2'} + \overline{v_1 Q H_2\,(d_2 \otimes d_1)\,d_2'} + \overline{v_1 Q H_2\,(d_2 \otimes d_2)\,d_1'}\right\} \\ +\frac{1}{2}Q\left\{\overline{w_1\,(d_1 \otimes d_2)\,d_2'} + \overline{w_1\,(d_2 \otimes d_1)\,d_2'} + \overline{w_1\,(d_2 \otimes d_2)\,d_1'}\right\} \\ +\frac{1}{2}QH_2\left\{\overline{(d_1 \otimes Q v_1 d_2)\,d_2'} + \overline{(d_1 \otimes Q v_2 d_1)\,d_2'} + \overline{(d_1 \otimes Q v_2 d_2)\,d_1'}\right\} \\ +\frac{1}{4}QH_2\left\{\overline{d_1 \otimes Q H_2\,(d_1 \otimes d_2)\,d_2'} + \overline{d_1 \otimes Q H_2\,(d_2 \otimes d_1)\,d_2'} + \overline{d_1 \otimes Q H_2\,(d_2 \otimes d_2)\,d_1'}\right\} \\ +\frac{1}{2}QH_2\left\{\overline{(Q V_1 d_1 \otimes d_2)\,d_2'} + \overline{(Q V_1 d_2 \otimes d_1)\,d_2'} + \overline{(Q V_1 d_2 \otimes d_2)\,d_1'}\right\} \\ +\frac{1}{4}QH_2\left\{\overline{(Q H_2\,(d_1 \otimes d_1) \otimes d_2)\,d_2'} + \overline{(Q H_2\,(d_1 \otimes d_2) \otimes d_1)\,d_2'} + \overline{(Q H_2\,(d_1 \otimes d_2) \otimes d_2)\,d_1'}\right\} \\ +\frac{1}{6}QH_3\left\{\overline{[d_1 \otimes d_1 \otimes d_2]\,d_2'} + \overline{[d_1 \otimes d_2 \otimes d_1]\,d_2'} + \overline{[d_1 \otimes d_2 \otimes d_2]\,d_1'}\right\} \end{array}\right\}.$$

for $d_i = Qs(z_i, \theta_0)$, $v_i = \nabla s(z_i, \theta_0) - E[\nabla s(z_i, \theta_0)]$ and $w_i = \nabla^2 s(z_i, \theta_0) - E[\nabla^2 s(z_i, \theta_0)]$. 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. 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 A.5 in Appendix A, we can show:

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

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

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:

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

References

1. Quenouille, M. Notes on Bias in Estimation. *Biometrika* **1956**, *43*, 353–360.
2. Hall, P. *The Bootstrap and Edgeworth Expansion*; Springer: New York, NY, USA, 1992.
3. Shao. J.; Tu, D. *The Jackknife and Bootstrap*; Springer: New York, NY, USA, 1995.
4. MacKinnon, J.; Smith, A. Approximate Bias Correction in Econometrics. *J. Econom.* **1998**, *85*, 205–230.
5. Andrews, D.W.K. Higher-Order Improvements of a Computationally Attractive k-Step Bootstrap for Extremum Estimators. *Econometrica* **2002**, *70*, 119–162.
6. Hahn, J.; Newey, W.K. Jackknife and Analytical Bias Reduction for Nonlinear Panel Models. *Econometrica* **2004**, *72*, 1295–1319.
7. Bun, M.J.G.; Carree, M.A. Bias-corrected Estimation in Dynamic Panel Data Models. *J. Bus. Econ. Stat.* **2005**, *23*, 200–210.

8. Bao, Y.; Ullah, A. Finite Sample Properties of Maximum Likelihood Estimator in Spatial Models. *J. Econom.* **2007**, *137*, 396–413.

9. Bao, Y.; Ullah, A. The Second-order Bias and Mean Squared Error of Estimators in Time-series Models. *J. Econom.* **2007** , *140*, 650–669.

10. Bao, Y. Finite Sample Bias of the QMLE in Spatial Autoregressive Models. *Econom. Theory* **2013**, *29*, 68–88.

11. Yang, Z. A General Method for Third-order Bias and Variance Corrections on a Nonlinear Estimator. *J. Econom.* **2015**, *186*, 178–200.

12. Hahn, J.; Kuersteiner, G.; Newey, W. *Higher Order Efficiency of Bias Corrections*; Working Paper; 2004. Available online: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.421.1735&rep=rep1&type=pdf (accessed on 2 March 2016).

13. Firth, D. Bias Reduction of Maximum Likelihood Estimation. *Biometrika* **1993**, *80*, 27–38.

14. Fernandez-Val, I. Bias Correction in Panel Data Models with Individual Specific Parameters. Ph.D. Thesis, MIT, Cambridge, MA, USA, 2004.

15. Rilstone, P.; Srivastava, V.K.; Ullah, A. The Second-order Bias and Mean Squared Error of Nonlinear Estimators. *J. Econom.* **1996**, *75*, 369–395.

16. Newey, W.K. A Method of Moments Interpretation of Sequential Estimators. *Econom. Lett.* **1984**, *14*, 201–206.

17. Newey, W.K.; Smith, R. Higher Order Properties of GMM and Generalized Empirical Likelihood Estimators. *Econometrica* **2004**, *72*, 219–255.

18. Heckman, J. Sample Selection Bias as a Specification Error. *Econometrica* **1979**, *47*, 153–161.

19. Bhattacharya, R.N.; Ghosh, J.K. On the Validity of the Formal Edgeworth Expansion. *Ann. Math. Stat.* **1978**, *6*, 434–451.

20. Pfanzagl, J.; Wefelmeyer, W. A Third-Order Optimum Property of the Maximum Likelihood Estimator. *J. Multivar. Anal.* **1978**, *8*, 1–29.

21. Ghosh, J.K.; Sinha, B.K.; Wieand, H.S. Second Order Efficiency of the MLE with Respect to Any Bowl Shaped Loss Function. *Ann. Stat.* **1980**, *8*, 506–521.

22. Rothenberg, T.J. Approximating the Distributions of Econometric Estimators and Test Statistics. In *Handbook of Econometrics V2*; Griliches, Z., Intriligator, M.D., Eds.; North Holland Publishing Co.: Amsterdam, The Netherlands, 1984; pp. 881–935.