

HIGHER-ORDER BIAS AND MSE OF NONLINEAR ESTIMATORS

Yong Bao¹ and Aman Ullah²

¹ Department of Economics, Purdue University, West Lafayette, IN 47907, USA.
Email: ybao@purdue.edu

² Department of Economics, University of California, Riverside, CA 92521, USA.
Email: aman.ullah@ucr.edu

ABSTRACT

We derive the approximate results for the bias and mean squared error of a large class of estimators to orders $O(n^{-5/2})$ and $O(n^{-3})$, respectively, given a sample of n observations. The results are built on a stochastic expansion of the moment condition used to identify the econometric estimator and a Nagar-type expansion of the inverse of the gradient of the moment condition.

KEYWORDS

Bias, Mean Squared Error, Higher-Order Expansion.

2000 Mathematics Subject Classification: 41A60.

1 INTRODUCTION

Recently, Rilstone et. al. (1996) and Bao and Ullah (2007) considered the issue of approximating the bias and mean squared error (MSE) of a class of nonlinear estimators identified by the moment condition

$$\hat{\beta}_n = \arg \{ \psi_n(\beta) = 0 \}, \quad (1.1)$$

where $\psi_n(\beta) = \psi_n(Z; \beta)$ is a known $k \times 1$ vector-valued function of the observable data $Z = \{Z_i\}_{i=1}^n$, which may or may not be IID, and a parameter vector β , with true value β_0 , of k elements such that $\mathbb{E}[\psi_n(\beta)] = 0$ only happens at $\beta = \beta_0$.

Based on the condition (1.1), Rilstone et. al. (1996) and Bao and Ullah (2007) expanded $\hat{\beta}_n - \beta_0$ to order $O(n^{-3/2})$ and defined the second-order bias, up to $O(n^{-1})$, and MSE, up to $O(n^{-2})$, of $\hat{\beta}_n$. In this note, we expand $\hat{\beta}_n - \beta_0$ to a higher order and derive the results of approximate bias and MSE of $\hat{\beta}_n$ up to $O(n^{-5/2})$ and $O(n^{-3})$, respectively. These results would be useful in econometrics and statistics for obtaining higher-order results for a large class of estimators (see Bao and Ullah (2007)) in linear and non-linear models.

The plan of this paper is as follows. The main results are presented in Section 2. The Appendix contains four lemmas that are used for deriving the results.

2 MAIN RESULTS

In what follows, we use $\text{tr}A$ to denote the trace, $\|A\|$ to denote the usual norm $[\text{tr}(AA')]^{1/2}$, and $\nabla^s A(\beta)$ is the matrix of s -th order partial derivative of $A(\beta)$ and is obtained recursively (specifically, if $A(\beta)$ is a $k \times 1$ vector function, the j th element of the l th row of $\nabla^s A(\beta)$ (a $k \times k^s$ matrix) is the $1 \times k$ vector $a_{lj}^s(\beta) = \partial a_{lj}^{s-1}(\beta) / \partial \beta'$). Throughout, the following assumptions are made.

Assumption 1: $\hat{\beta}_n$ and its first two moments exist.

Assumption 2: The s -th order derivatives of $\psi_n(\beta)$ exist for β in a neighborhood of β_0 and for s up to 5, $\mathbb{E}(\|\nabla^s \psi_n(\beta_0)\|^2) < \infty$.

Assumption 3: For β in some neighborhood of β_0 , $[\nabla \psi_n(\beta)]^{-1} = O_P(1)$.

Assumption 4: $\|\nabla^s \psi_n(\beta) - \nabla^s \psi_n(\beta_0)\| \leq \|\beta - \beta_0\| M_n$ for β in some neighborhood of β_0 , where $\mathbb{E}(|M_n|) < C < \infty$ for some positive constant C , for s up to 5.

Assumptions 1-4 are fairly standard and are satisfied for a large class of estimators, though excluding nonstationary time-series models involving a unit root. One may lay out a set of primitive conditions to guarantee existence and consistency of $\hat{\beta}_n$. We should point out that moments of the instrumental estimator for a just identified structural equation do not exist. However, it does have a well-defined exact distribution and limiting distribution and the ‘‘approximate moments’’ can still be obtained. Assumptions 2-4 are similar to Rilstone et al. (1996) and Bao and Ullah (2007) to guarantee that the moment condition is smooth enough so that a stochastic expansion of $\psi_n(\hat{\beta})$ as well as a Nagar-type (Nagar, 1959) expansion of the inverse of the gradient of the moment function can be implemented around β_0 .

Following the notational conventions in Rilstone et al. (1996), $\psi_n = \psi_n(\beta_0)$ (we suppress the argument of a function when it is evaluated at β_0), $H_i = \nabla^i \psi_n$, $Q = [\mathbb{E}(H_1)]^{-1}$, $V_i = H_i - \mathbb{E}(H_i)$, \otimes represent the Kronecker product, $a_{-s/2}$ represent terms of order $O_P(n^{-s/2})$, and put

$$\begin{aligned} a_{-1/2} &= -Q\psi_n, \\ a_{-1} &= -QV_1a_{-1/2} - \frac{1}{2}QE(H_2)(a_{-1/2} \otimes a_{-1/2}), \\ a_{-3/2} &= -QV_1a_{-1} - \frac{1}{2}QV_2(a_{-1/2} \otimes a_{-1/2}) - \frac{1}{2}QE(H_2)(a_{-1/2} \otimes a_{-1} + a_{-1} \otimes a_{-1/2}) \\ &\quad - \frac{1}{6}QE(H_3)(a_{-1/2} \otimes a_{-1/2} \otimes a_{-1/2}), \\ a_{-2} &= -QV_1a_{-3/2} - \frac{1}{2}QV_2(a_{-1/2} \otimes a_{-1} + a_{-1} \otimes a_{-1/2}) - \frac{1}{2}QE(H_2)(a_{-1/2} \otimes a_{-3/2} \\ &\quad + a_{-3/2} \otimes a_{-1/2} + a_{-1} \otimes a_{-1}) - \frac{1}{6}QV_3(a_{-1/2} \otimes a_{-1/2} \otimes a_{-1/2}) \\ &\quad - \frac{1}{6}QE(H_3)(a_{-1/2} \otimes a_{-1/2} \otimes a_{-1} + a_{-1/2} \otimes a_{-1} \otimes a_{-1/2} + a_{-1} \otimes a_{-1/2} \otimes a_{-1/2}) \\ &\quad - \frac{1}{24}QE(H_4)(a_{-1/2} \otimes a_{-1/2} \otimes a_{-1/2} \otimes a_{-1/2}), \\ a_{-5/2} &= -QV_1a_{-2} - \frac{1}{2}QV_2(a_{-1/2} \otimes a_{-3/2} + a_{-3/2} \otimes a_{-1/2} + a_{-1} \otimes a_{-1}) \\ &\quad - \frac{1}{2}QE(H_2)(a_{-1/2} \otimes a_{-2} + a_{-1} \otimes a_{-3/2} + a_{-3/2} \otimes a_{-1} + a_{-2} \otimes a_{-1/2}) \\ &\quad - \frac{1}{6}QV_3(a_{-1/2} \otimes a_{-1/2} \otimes a_{-1} + a_{-1/2} \otimes a_{-1} \otimes a_{-1/2} + a_{-1} \otimes a_{-1/2} \otimes a_{-1/2}) \\ &\quad - \frac{1}{6}QE(H_3)(a_{-1/2} \otimes a_{-1/2} \otimes a_{-3/2} + a_{-1/2} \otimes a_{-1} \otimes a_{-1} + a_{-1/2} \otimes a_{-3/2} \otimes a_{-1/2} \\ &\quad + a_{-1} \otimes a_{-1/2} \otimes a_{-1} + a_{-1} \otimes a_{-1} \otimes a_{-1/2} + a_{-3/2} \otimes a_{-1/2} \otimes a_{-1/2}) \\ &\quad - \frac{1}{24}QV_4(a_{-1/2} \otimes a_{-1/2} \otimes a_{-1/2} \otimes a_{-1/2}) - \frac{1}{24}QE(H_4)(a_{-1/2} \otimes a_{-1/2} \otimes a_{-1/2} \otimes a_{-1} \\ &\quad + a_{-1/2} \otimes a_{-1/2} \otimes a_{-1} \otimes a_{-1/2} + a_{-1/2} \otimes a_{-1} \otimes a_{-1/2} \otimes a_{-1/2} \\ &\quad + a_{-1} \otimes a_{-1/2} \otimes a_{-1/2} \otimes a_{-1/2}) - \frac{1}{120}QE(H_5)(a_{-1/2} \otimes a_{-1/2} \otimes a_{-1/2} \otimes a_{-1/2} \otimes a_{-1/2}). \end{aligned}$$

From the appendix (see Lemma 3.5), one can write a stochastic expansion

$$\hat{\beta}_n - \beta_0 = a_{-1/2} + a_{-1} + a_{-3/2} + a_{-2} + a_{-5/2} + o_P(n^{-5/2}), \tag{2.1}$$

where the order $O(n^{-1/2})$ term $a_{-1/2}$ represents the the asymptotic behavior of $\hat{\beta}_n$ and $AsyV(\sqrt{n}(\hat{\beta}_n - \beta_0)) = n\mathbb{E}(a_{-1/2}a'_{-1/2})$. Note that $\mathbb{E}(a_{-1/2}) = 0$ since $\mathbb{E}(\psi_n) = 0$. Based on this and the expansion (2.1), one can immediately derive the second-order bias, up to $O(n^{-1})$, and MSE, up to $O(n^{-2})$, of $\hat{\beta}_n$, as done in Rilstone et al. (1996) for models with IID data and Bao and Ullah (2007) for models with non-IID data. Now we go to the higher order. For the bias, we can approximate up to order $O(n^{-5/2})$,

$$B(\hat{\beta}_n) = \mathbb{E}(a_{-1} + a_{-3/2} + a_{-2} + a_{-5/2}) + o(n^{-5/2}), \tag{2.2}$$

and for the MSE, we can approximate up to order $O(n^{-3})$,

$$\begin{aligned}
M(\hat{\beta}_n) &= \mathbb{E}[a_{-1/2}(a_{-1/2} + a_{-1} + a_{-3/2} + a_{-2} + a_{-5/2})' \\
&\quad + a_{-1}(a_{-1/2} + a_{-1} + a_{-3/2} + a_{-2})' \\
&\quad + a_{-3/2}(a_{-1/2} + a_{-1} + a_{-3/2})' \\
&\quad + a_{-2}(a_{-1/2} + a_{-1})' + a_{-5/2}a'_{-1/2}] + o(n^{-3}). \tag{2.3}
\end{aligned}$$

In practice, the expectations involved in (2.2) and (2.3) might be evaluated analytically given the distribution assumption on Z and the functional form of $\psi_n(\cdot)$. In cases when analytical evaluation is not possible, researchers might have to resort to Monte Carlo simulations to approximate these expectations. For example, Linton (1997), Pesaran and Timmermann (2005), and Iglesias and Phillips (2008) employed this strategy to study the second-order bias problem in time series models.

3 APPENDIX: LEMMAS

We follow the steps outlined in Rilstone et al. (1996) and similar notations are used. In particular, ε_r refers to a generic random sequence such that $\varepsilon_r = O_p(n^r)$.

Lemma 3.1. *Let Assumptions 1-4 hold for $s = 3$. Then*

$$\hat{\beta} - \beta_0 = a_{-1/2} + a_{-1} + a_{-3/2} + \varepsilon_{-2}. \tag{3.1}$$

Proof. See Rilstone et al. (1996).

Lemma 3.2. *Let Assumptions 1-4 hold for $s = 4$ and suppose that*

$$\begin{aligned}
\psi_n(\hat{\beta}) &= \psi_n + \nabla\psi_n(\hat{\beta} - \beta_0) + \frac{1}{2}\nabla^2\psi_n[(\hat{\beta} - \beta_0) \otimes (\hat{\beta} - \beta_0)] \\
&\quad + \frac{1}{6}\nabla^3\psi_n[(\hat{\beta} - \beta_0) \otimes (\hat{\beta} - \beta_0) \otimes (\hat{\beta} - \beta_0)] \\
&\quad + \frac{1}{24}\nabla^4\psi_n[(\hat{\beta} - \beta_0) \otimes (\hat{\beta} - \beta_0) \otimes (\hat{\beta} - \beta_0) \otimes (\hat{\beta} - \beta_0)] + \varepsilon_{-5/2}. \tag{3.2}
\end{aligned}$$

Then

$$\begin{aligned}
\hat{\beta} - \beta_0 &= -\nabla\psi_n^{-1}\psi_n - \frac{1}{2}\nabla\psi_n^{-1}\nabla^2\psi_n[(a_{-1/2} + a_{-1} + a_{-3/2}) \otimes (a_{-1/2} + a_{-1} + a_{-3/2})] \\
&\quad - \frac{1}{6}\nabla\psi_n^{-1}\nabla^3\psi_n[(a_{-1/2} + a_{-1}) \otimes (a_{-1/2} + a_{-1}) \otimes (a_{-1/2} + a_{-1})] \\
&\quad - \frac{1}{24}\nabla\psi_n^{-1}\nabla^4\psi_n(a_{-1/2} \otimes a_{-1/2} \otimes a_{-1/2} \otimes a_{-1/2}) + \varepsilon_{-5/2}. \tag{3.3}
\end{aligned}$$

Proof. Since $\psi_n(\hat{\beta}) = 0$, rearranging terms yields

$$\begin{aligned}\hat{\beta} - \beta_0 &= -\nabla\psi_n^{-1}\psi_n - \frac{1}{2}\nabla\psi_n^{-1}\nabla^2\psi_n[(\hat{\beta} - \beta_0) \otimes (\hat{\beta} - \beta_0)] \\ &\quad - \frac{1}{6}\nabla\psi_n^{-1}\nabla^3\psi_n[(\hat{\beta} - \beta_0) \otimes (\hat{\beta} - \beta_0) \otimes (\hat{\beta} - \beta_0)] \\ &\quad - \frac{1}{24}\nabla\psi_n^{-1}\nabla^4\psi_n[(\hat{\beta} - \beta_0) \otimes (\hat{\beta} - \beta_0) \otimes (\hat{\beta} - \beta_0) \otimes (\hat{\beta} - \beta_0)] - \nabla\psi_n^{-1}\varepsilon_{-5/2}.\end{aligned}$$

Substituting from Lemma 3.1,

$$\begin{aligned}\hat{\beta} - \beta_0 &= -\nabla\psi_n^{-1}\psi_n - \frac{1}{2}\nabla\psi_n^{-1}\nabla^2\psi_n[(a_{-1/2} + a_{-1} + a_{-3/2} + \varepsilon_{-2}) \\ &\quad \otimes (a_{-1/2} + a_{-1} + a_{-3/2} + \varepsilon_{-2})] - \frac{1}{6}\nabla\psi_n^{-1}\nabla^3\psi_n[(a_{-1/2} + a_{-1} + \varepsilon_{-3/2}) \\ &\quad \otimes (a_{-1/2} + a_{-1} + \varepsilon_{-3/2}) \otimes (a_{-1/2} + a_{-1} + \varepsilon_{-3/2})] \\ &\quad - \frac{1}{24}\nabla\psi_n^{-1}\nabla^4\psi_n[(a_{-1/2} + \varepsilon_{-1}) \otimes (a_{-1/2} + \varepsilon_{-1}) \otimes (a_{-1/2} + \varepsilon_{-1}) \\ &\quad \otimes (a_{-1/2} + \varepsilon_{-1})] - \nabla\psi_n^{-1}\varepsilon_{-5/2}.\end{aligned}$$

An inspection of the order of the terms inside the above equation produces the result.

Lemma 3.3. *Let Assumptions 1-4 hold for $s = 4$, then*

$$\hat{\beta} - \beta_0 = a_{-1/2} + a_{-1} + a_{-3/2} + a_{-2} + O_P(n^{-5/2}). \quad (3.4)$$

Proof. Implementing a fourth-order Taylor expansion of $\psi_n(\hat{\beta})$, we have

$$\begin{aligned}\psi_n(\hat{\beta}) &= \psi_n + \nabla\psi_n(\hat{\beta} - \beta_0) + \frac{1}{2}\nabla^2\psi_n[(\hat{\beta} - \beta_0) \otimes (\hat{\beta} - \beta_0)] \\ &\quad + \frac{1}{6}\nabla^3\psi_n[(\hat{\beta} - \beta_0) \otimes (\hat{\beta} - \beta_0) \otimes (\hat{\beta} - \beta_0)] \\ &\quad + \frac{1}{24}\nabla^4\psi_n[(\hat{\beta} - \beta_0) \otimes (\hat{\beta} - \beta_0) \otimes (\hat{\beta} - \beta_0) \otimes (\hat{\beta} - \beta_0)] \\ &\quad + \frac{1}{24}[\nabla^4\psi_n(\bar{\beta}) - \nabla^4\psi_n(\beta_0)][(\hat{\beta} - \beta_0) \otimes (\hat{\beta} - \beta_0) \otimes (\hat{\beta} - \beta_0) \otimes (\hat{\beta} - \beta_0)],\end{aligned}$$

where $\nabla^4\psi_n(\bar{\beta})$ is the matrix of fourth-order partial derivatives of $\psi_n(\beta)$ with each element evaluated at some point between $\hat{\beta}$ and β_0 . From Assumption 4, the last term in the above expression is of order $O_P(n^{-5/2})$, so the condition (3.2) in Lemma 3.2 is satisfied, and we can write $\hat{\beta} - \beta_0$ as in (3.3). Next, we can use a Nagar-type (Nagar, 1959) expansion on

$\nabla\psi_n^{-1}$:

$$\begin{aligned}\nabla\psi_n^{-1} &= (Q^{-1} + V_1)^{-1} = (I + QV_1)^{-1}Q \\ &= Q + \begin{cases} +\varepsilon_{-1/2} \\ -QV_1Q + \varepsilon_{-1} \\ -QV_1Q + QV_1QV_1Q + \varepsilon_{-3/2} \\ -QV_1Q + QV_1QV_1Q - QV_1QV_1QV_1Q + \varepsilon_{-2} \\ -QV_1Q + QV_1QV_1Q - QV_1QV_1QV_1Q + QV_1QV_1QV_1QV_1Q + \varepsilon_{-5/2} \end{cases} \quad (3.5)\end{aligned}$$

since by Assumption 4, $V_1 = O_P(n^{-1/2})$. Substituting (3.5) into (3.3), we write

$$\begin{aligned}\hat{\beta} - \beta_0 &= -(Q - QV_1Q + QV_1QV_1Q - QV_1QV_1QV_1Q + \varepsilon_{-2})\psi_n \\ &\quad - \frac{1}{2}(Q - QV_1Q + QV_1QV_1Q + \varepsilon_{-3/2})[\mathbb{E}(H_2) + V_2][(a_{-1/2} + a_{-1} + a_{-3/2}) \\ &\quad \otimes (a_{-1/2} + a_{-1} + a_{-3/2})] - \frac{1}{6}(Q - QV_1Q + \varepsilon_{-1})[\mathbb{E}(H_3) + V_3][(a_{-1/2} + a_{-1}) \\ &\quad \otimes (a_{-1/2} + a_{-1}) \otimes (a_{-1/2} + a_{-1})] - \frac{1}{24}(Q + \varepsilon_{-1/2})[\mathbb{E}(H_4) + V_4] \\ &\quad \times (a_{-1/2} \otimes a_{-1/2} \otimes a_{-1/2} \otimes a_{-1/2}) + \varepsilon_{-5/2},\end{aligned}$$

where $V_1 \sim V_4$ are $O_P(n^{-1/2})$ by Assumption 4. Collecting terms of different orders and ignoring terms of lower orders, the result follows.

Lemma 3.4. *Let Assumptions 1-4 hold for $s = 5$ and suppose that*

$$\begin{aligned}\psi_n(\hat{\beta}) &= \psi_n + \nabla\psi_n(\hat{\beta} - \beta_0) + \frac{1}{2}\nabla^2\psi_n[(\hat{\beta} - \beta_0) \otimes (\hat{\beta} - \beta_0)] \\ &\quad + \frac{1}{6}\nabla^3\psi_n[(\hat{\beta} - \beta_0) \otimes (\hat{\beta} - \beta_0) \otimes (\hat{\beta} - \beta_0)] \\ &\quad + \frac{1}{24}\nabla^4\psi_n[(\hat{\beta} - \beta_0) \otimes (\hat{\beta} - \beta_0) \otimes (\hat{\beta} - \beta_0) \otimes (\hat{\beta} - \beta_0)] \\ &\quad + \frac{1}{120}\nabla^5\psi_n[(\hat{\beta} - \beta_0) \otimes (\hat{\beta} - \beta_0) \otimes (\hat{\beta} - \beta_0) \otimes (\hat{\beta} - \beta_0) \otimes (\hat{\beta} - \beta_0)] + \varepsilon_{-3}.\end{aligned}$$

Then

$$\begin{aligned} \hat{\beta} - \beta_0 &= -\nabla\psi_n^{-1}\psi_n - \frac{1}{2}\nabla\psi_n^{-1}\nabla^2\psi_n[(a_{-1/2} + a_{-1} + a_{-3/2} + a_{-2}) \\ &\quad \otimes (a_{-1/2} + a_{-1} + a_{-3/2} + a_{-2})] - \frac{1}{6}\nabla\psi_n^{-1}\nabla^3\psi_n[(a_{-1/2} + a_{-1} + a_{-3/2}) \\ &\quad \otimes (a_{-1/2} + a_{-1} + a_{-3/2}) \otimes (a_{-1/2} + a_{-1} + a_{-3/2})] \\ &\quad - \frac{1}{24}\nabla\psi_n^{-1}\nabla^4\psi_n[(a_{-1/2} + a_{-1}) \otimes (a_{-1/2} + a_{-1}) \otimes (a_{-1/2} + a_{-1}) \\ &\quad \otimes (a_{-1/2} + a_{-1})] - \frac{1}{120}\nabla\psi_n^{-1}\nabla^5\psi_n(a_{-1/2} \otimes a_{-1/2} \otimes a_{-1/2} \otimes a_{-1/2} \otimes a_{-1/2}) \\ &\quad + \varepsilon_{-3}. \end{aligned}$$

Proof. Following the same steps as in the proof of Lemma 3.2 and substituting (3.1) and (3.4) yields the result.

Lemma 3.5. *Let Assumptions 1-4 hold for $s = 5$, then*

$$\hat{\beta} - \beta_0 = a_{-1/2} + a_{-1} + a_{-3/2} + a_{-2} + a_{-5/2} + O_p(n^{-3}). \tag{3.6}$$

Proof. Follow the same steps as in the proof of Lemma 3.3 by implementing a fifth-order Taylor expansion of $\psi_n(\hat{\beta})$, using Lemma 3.4, substituting (3.5), and noting that $V_1 \sim V_5$ are $O_p(n^{-1/2})$ by Assumption 4, then the result follows.

ACKNOWLEDGEMENTS

We would like to thank an anonymous referee and Ejaz Ahmed for their helpful comments.

REFERENCES

1. Bao, Y. and Ullah, A. (2007). The second-order bias and mean squared error of estimators in time series models. *Journal of Econometrics*, **140**, 650-669.
2. Igelsias, E.M. and Phillips, G.D.A. (2008). Finite sample theory of QMLEs in ARCH Models with dynamics in the mean equation. *Journal of Time Series Analysis*, **29**, 719-737.
3. Linton, O. (1997). An asymptotic expansion in the GARCH(1,1) model. *Econometric Theory*, **13**, 558-81.

4. Nagar, A.L. (1959). The bias and moment matrix of the general k -class estimators of the parameters in simultaneous equations. *Econometrica*, **27**, 575-595.
5. Pesaran, M.H. and Timmermann, A. (2005). Small sample properties of forecasts from autoregressive models under structural breaks. *Journal of Econometrics*, **129**, 183-217.
6. Rilstone, P., Srivastava, V.K. and Ullah, A. (1996). The second-order bias and mean squared error of nonlinear estimators. *Journal of Econometrics*, **75**, 369-395.