The Second-Order Bias and Mean Squared Error of Estimators in Time Series Models*

Yong Bao†
Department of Economics
University of California, Riverside

Aman Ullah‡
Department of Economics
University of California, Riverside

August 14, 2003

Abstract

We develop the analytical results on the second-order bias and mean squared error (MSE) of estimators in time series. These results provide a unified approach to developing the properties of a large class of estimators in the linear and nonlinear time series models and they are valid for both the normal and non-normal sample of observations, and where the regressors are stochastic. The estimators included are the generalized method of moments, maximum likelihood, least squares, and other extremum estimators. Our general results are applied to a wide variety of econometric models. Numerical results for some of these models are presented.

Keywords: Higher-order moments; Stochastic expansion; Time series; Quadratic form

JEL Classification: C10, C22, C32

*We would like to thank Gloria González-Rivera, Emma Iglesias and Victoria Zinde-Walsh for helpful comments. Bao thanks the UCR Chancellor Fellowship and Ullah thanks the UCR Academic Senate for research support. All remaining errors are our own.

†Department of Economics, University of California, Riverside, CA, 92521, (909) 788-7998, e-mail: yongbao@mail.ucr.edu.

‡Corresponding Author: Department of Economics, University of California, Riverside, CA, 92521, (909) 827-1591, Fax: (909) 787-5685, e-mail: aman.ullah@ucr.edu.

1 Introduction

There is an extensive literature on the analytical finite sample properties of econometric estimators and test statistics in linear models, see Nagar (1959), Anderson and Sawa (1973, 1979), Basmann (1974), Sargan (1974, 1976), Phillips (1977), Rothenberg (1984), Ullah and Srivastava (1994), and Ullah (2002), among others. In contrast, not much work has been done on the finite sample properties of the nonlinear statistics, although see Roberston and Fryer (1970), Amemiya (1980), Chesher and Spady (1989), Cordeiro and McCullagh (1991), Koenker *et al.* (1992), and Newy and Smith (2001). However, most of these works are for some specific estimators and there is little with the dependent observations for nonlinear cases, although see Cordeiro and Klein (1994) and Linton (1997). Recently, Rilstone *et al.* (1996) developed the large-n second-order bias and mean squared error (MSE) of a class of nonlinear estimators. Nevertheless, their results are for the i.i.d. sample so they are not applicable to the models with dependent observations, for example, the time series models.

In this paper, we extend the second-order bias and MSE results of Rilstone *et al.* (1996) for the time series dependent observations. These results provide a unified way of developing the properties of a given class of estimators in the linear and nonlinear time series models. The estimators included are the generalized method of moments (GMM), maximum likelihood (ML), least squares (LS) and other extremum estimators, and the two step estimators which involve a nuisance parameter. Our results are also valid for both the normal and non-normal sample of observations, and where the regressors are stochastic. Next, in a special case of the ML estimators (MLE) we also show that our bias result reduces to that of the bias of MLE in Cox and Snell (1968) for the i.i.d. case and its extension in Cordeiro and Klein (1994) for the dependent observations. However, we note that our bias result is for a general class of estimators including the ML as a special case and that Cox and Snell's (1968) approach does not provide the MSE of estimators. Furthermore, as an application of our general results, we develop the second-order bias and MSE for some time series models. These include the AR(1) model, structural model with AR(1) errors, VAR model, MA(1) model, partial adjustment model, and absolute regression model.

The plan of the paper is as follows. In Section 2, we present the estimators identified by some moment condition and their bias and MSE results. Then in Section 3 we develop the results on the moments of estimators in six time series models. Section 4 concludes. Some useful results on the expectations of quadratic forms in a normal vector and proofs are given in the appendix.

2 Second-Order Bias and MSE

Consider a class of estimators in a time series model as

$$\hat{\beta} = \arg\left\{\psi_T(\beta) = 0\right\} = \arg\left\{\frac{1}{T}\sum_{t=1}^{T} q_t(Z_t;\beta) = 0\right\}, \tag{2.1}$$

where $q_t(\beta) = q_t(Z_t;\beta)$ is a known $k \times 1$ vector-valued function of the observable data Z_t and a parameter vector β of k elements such that $E\left[q_t(\beta)\right] = 0$. We consider Z_t in (2.1) to be a sequence of m-dimensional non-i.i.d. random vectors. Rilstone *et al.* (1996) considered the case where Z_t are i.i.d. random vectors. The class of estimators in (2.1) include many estimators based on nonlinear and linear time series models, which include the ML, LS, and GMM estimators. For example, the ML estimator of the parameters of an ARMA process.

To obtain the second-order bias and MSE of $\hat{\beta}$, the assumption of A to C in Rilstone *et al.* (1996) are assumed to hold along with the consistency of $\hat{\beta}$ and we follow their notations. Then the stochastic expansion of $\hat{\beta}$, up to $O_P\left(T^{-3/2}\right)$, in Rilstone *et al.* (1996) derived for the i.i.d. case can be extended to the non-i.i.d. case as

$$\hat{\beta} - \beta = a_{-1/2} + a_{-1} + a_{-2/3} \tag{2.2}$$

where $a_{-s/2}$ represents terms of order $O_P\left(T^{-s/2}\right)$ for $s = 1, 2, 3$ and they are

$$
\begin{aligned}
a_{-1/2} &= -Q\psi_T, \\
a_{-1} &= -QVa_{-1/2} - \frac{1}{2}Q\overline{H_2}[a_{-1/2} \otimes a_{-1/2}], \\
a_{-3/2} &= -QVa_{-1} - \frac{1}{2}QW[a_{-1/2} \otimes a_{-1/2}] - \frac{1}{2}Q\overline{H_2}\left\{\left[a_{-1/2} \otimes a_{-1}\right] + \left[a_{-1} \otimes a_{-1/2}\right]\right\} \\
&\quad -\frac{1}{6}Q\overline{H_3}[a_{-1/2} \otimes a_{-1/2} \otimes a_{-1/2}],
\end{aligned} \tag{2.3}
$$

in which $\overline{X} = E(X)$ denotes the expectation of a random vector X, $H_i = \nabla^i \psi_T$, $i = 1, 2, 3$, $Q = \overline{H_1}^{-1}$, $V = H_1 - \overline{H_1} = \frac{1}{T}\sum_{t=1}^{T} V_t$, $W_t = H_2 - \overline{H_2} = \frac{1}{T}\sum_{t=1}^{T} W_t$, $\otimes$ represents the Kronecker product and $\nabla^s A(\beta)$ are the matrix of recursive s-th order partial derivatives of a matrix $A(\beta)$.

It follows from (2.2) that the second-order bias of $\hat{\beta}$, $B\left(\hat{\beta}\right)$, up to $O\left(T^{-1}\right)$, and the MSE of $\hat{\beta}$, up to $O\left(T^{-2}\right)$, respectively, are

$$
\begin{aligned}
B\left(\hat{\beta}\right) &= E\left(a_{-1/2}\right) + E\left(a_{-1}\right), \\
M(\hat{\beta}) &= E(a_{-1/2}a'_{-1/2}) + E(a_{-1}a'_{-1/2} + a_{-1/2}a'_{-1}) \\
&\quad + E(a_{-1}a'_{-1} + a_{-3/2}a'_{-1/2} + a_{-1/2}a'_{-3/2}),
\end{aligned} \tag{2.4}
$$

where $E\left(a_{-1/2}\right) = 0$.

Using (2.3) in (2.4) and denoting $d = Q\psi_T$, one can write (2.4) as

$$
B\left(\hat{\beta}\right) = Q\left[\overline{Vd} - \frac{1}{2}\overline{H_2}\left(\overline{d \otimes d}\right)\right], \tag{2.5}
$$

and

$$
M\left(\hat{\beta}\right) = A_{-1} + A_{-3/2} + A_{-2} \tag{2.6}
$$

where $A_{-s/2} = O\left(T^{-s/2}\right)$, $s = 2, 3, 4$ are

$A_{-1} = \overline{dd'}$,

$A_{-3/2} = -Q\left[\overline{Vdd'} + \frac{1}{2}\overline{H_2}\,\overline{(d \otimes d)\, d'}\right] - \left[\overline{dd'V} + \frac{1}{2}\overline{d\,(d' \otimes d')}\,\overline{H_2}'\right] Q$,

$A_{-2} = Q\left[\overline{Vdd'V'} + \frac{1}{4}\overline{H_2}\,\overline{(d \otimes d)\,(d' \otimes d')}\,\overline{H_2}' - \frac{1}{2}\overline{Vd\,(d' \otimes d')}\,\overline{H_2}' - \frac{1}{2}\overline{H_2}\,\overline{(d \otimes d)d'V'}\right] Q$

$$+Q[\overline{VQVdd'} - \tfrac{1}{2}\overline{VQ\overline{H_2}(d\otimes d)d'} + \tfrac{1}{2}\overline{W(d\otimes d)d'} - \tfrac{1}{2}\overline{H_2}\,\overline{(d\otimes(QVd))\,d'}$$
$$+\tfrac{1}{4}\overline{H_2}\,\overline{\left(\left(Q\overline{H_2}(d\otimes d)\right)\otimes d\right)d'} - \tfrac{1}{2}\overline{H_2}\,\overline{((QVd)\otimes d)\,d'}$$
$$+\tfrac{1}{4}\overline{H_2}\,\overline{\left(d\otimes\left(Q\overline{H_2}(d\otimes d)\right)\right)d'} - \tfrac{1}{6}\overline{H_3}\,\overline{(d\otimes d\otimes d)d'}]$$
$$+[\overline{dd'VQV} - \tfrac{1}{2}\overline{d\,(d'\otimes d')\,\overline{H_2}'QV} + \tfrac{1}{2}\overline{d\,(d'\otimes d')\,W'} - \tfrac{1}{2}\overline{d\,((d'VQ)\otimes d')}\,\overline{H_2}'$$
$$+\tfrac{1}{4}\overline{d\left(d'\otimes\left((d'\otimes d')\,\overline{H_2}'Q\right)\right)}\,\overline{H_2}' - \tfrac{1}{2}\overline{d\,(d'\otimes(d'VQ))}\,\overline{H_2}'$$
$$+\tfrac{1}{4}\overline{d\left(\left((d'\otimes d')\,\overline{H_2}'Q\right)\otimes d'\right)}\,\overline{H_2}' - \tfrac{1}{6}\overline{d(d'\otimes d'\otimes d')}\;\overline{H_3}']Q.$$

When the observations are i.i.d. the bias and MSE results in (2.4) or (2.5) and (2.6) reduce to the following results:

$$B(\hat{\beta}) = \frac{1}{T}Q\left\{\overline{V_1d_1} - \frac{1}{2}\overline{H_2}\,\overline{[d_1\otimes d_1]}\right\}, \qquad (2.7)$$

$$M(\hat{\beta}) = \frac{1}{T}\mathcal{V}_1 + \frac{1}{T^2}\left(\mathcal{V}_2 + \mathcal{V}_2'\right) + \frac{1}{T^2}\left(\mathcal{V}_3 + \mathcal{V}_4 + \mathcal{V}_4'\right), \qquad (2.8)$$

where $\mathcal{V}_1 = E(\nabla q_1)^{-1}E(q_1q_1')E(\nabla q_1)^{-1\prime}$,

$$\mathcal{V}_2 = Q\left\{-\overline{V_1d_1d_1'} + \tfrac{1}{2}\overline{H_2}\,\overline{[d_1\otimes d_1]\,d_1'}\right\},$$
$$\mathcal{V}_3 = Q\left\{\overline{V_1d_1}\,\overline{d_2'V_2'} + \overline{V_1d_2d_1'V_2'} + \overline{V_1d_2d_2'V_1'}\right\}Q$$
$$+\tfrac{1}{4}Q\overline{H_2}\left\{\overline{d_1\otimes d_1}\,\overline{d_2'\otimes d_2'} + \overline{[d_1\otimes d_2]\,[d_1'\otimes d_2']} + \overline{[d_1\otimes d_2]\,[d_2'\otimes d_1']}\right\}\overline{H_2'}Q$$
$$-\tfrac{1}{2}Q\left\{\overline{V_1d_1}\,\overline{d_2'\otimes d_2'} + \overline{V_1d_2\,[d_1'\otimes d_2']} + \overline{V_1d_2\,[d_2'\otimes d_1']}\right\}\overline{H_2'}Q$$
$$-\tfrac{1}{2}Q\overline{H_2'}\left\{\overline{d_1\otimes d_1}\,\overline{d_2'V_2'} + \overline{[d_1\otimes d_2]\,d_1'V_2'} + \overline{[d_1\otimes d_2]\,d_2'V_1'}\right\}Q,$$
$$\mathcal{V}_4 = Q\left\{\overline{V_1QV_1}\,\overline{d_2d_2'} + \overline{V_1QV_2d_1d_2'} + \overline{V_1QV_2d_2d_1'}\right\}$$
$$-\tfrac{1}{2}Q\left\{\overline{V_1Q\overline{H_2}\,[d_1\otimes d_2]\,d_2'} + \overline{V_1Q\overline{H_2}\,[d_2\otimes d_1]\,d_2'} + \overline{V_1Q\overline{H_2}\,[d_2\otimes d_2]\,d_1'}\right\}$$
$$+\tfrac{1}{2}Q\left\{\overline{W_1\,[d_1\otimes d_2]\,d_2'} + \overline{W_1\,[d_2\otimes d_1]\,d_2'} + \overline{W_1\,[d_2\otimes d_2]\,d_1'}\right\}$$
$$-\tfrac{1}{2}Q\overline{H_2}\left\{\overline{[d_1\otimes QV_1d_2]\,d_2'} + \overline{[d_1\otimes QV_2d_1]\,d_2'} + \overline{[d_1\otimes QV_2d_2]\,d_1'}\right\}$$
$$+\tfrac{1}{4}Q\overline{H_2}\left\{\overline{\left[d_1\otimes Q\overline{H_2}\,[d_1\otimes d_2]\right]d_2'} + \overline{\left[d_1\otimes Q\overline{H_2}\,[d_2\otimes d_1]\right]d_2'} + \overline{\left[d_1\otimes Q\overline{H_2}\,[d_2\otimes d_2]\right]d_1'}\right\}$$
$$-\tfrac{1}{2}Q\overline{H_2}\left\{\overline{[QV_1d_1\otimes d_2]\,d_2'} + \overline{[QV_1d_2\otimes d_1]\,d_2'} + \overline{[QV_1d_2\otimes d_2]\,d_1'}\right\}$$
$$+\tfrac{1}{4}Q\overline{H_2}\left\{\overline{\left[Q\overline{H_2}\,[d_1\otimes d_1]\otimes d_2\right]d_2'} + \overline{\left[Q\overline{H_2}\,[d_1\otimes d_2]\otimes d_1\right]d_2'} + \overline{\left[Q\overline{H_2}\,[d_1\otimes d_2]\otimes d_2\right]d_1'}\right\}$$
$$-\tfrac{1}{6}Q\overline{H_3}\left\{\overline{[d_1\otimes d_1\otimes d_2]\,d_2'} + \overline{[d_1\otimes d_2\otimes d_1]\,d_2'} + \overline{[d_1\otimes d_2\otimes d_2]\,d_1'}\right\}.$$

The result in (2.7) is the same as in Rilstone *et al.* (1996). The result in (2.8) is the corrected version of Rilstone *et al.* (1996) result in Proposition 3.4.

The stochastic expansion in (2.2) and the bias and MSE expressions in (2.4) are valid for both non-i.i.d. and i.i.d. observations. In practical applications the bias and MSE expressions may often be easily obtained directly from (2.4), especially for dependent observations.

Note that if ψ_T is the score function for the MLE of the parameters, then the Rilstone *et al.* (1996) second-order bias for the i.i.d. case in (2.7) and our bias result for dependent observations in (2.5) will be essentially the same as the Cox and Snell (1968) result, since both use a Taylor series expansion of the score function. Cordeiro and Klein (1994) rewrote conveniently the bias vector from the Cox and Snell (1968) expansion of the $p \times 1$ MLE $\hat{\beta}$ as

$$E\left(\hat{\beta}-\beta\right) = K^{-1}Avec(K^{-1}), \tag{2.9}$$

and for a single parameter estimator $\hat{\beta}_s$, $1 \le s \le p$,

$$E\left(\hat{\beta}_s-\beta_s\right) = \sum_{i=1}^{p}\sum_{j=1}^{p}\sum_{l=1}^{p} k^{sj}\left(k_{ij}^{(l)} - \frac{1}{2}k_{ijl}\right)k^{jl} \tag{2.9'}$$

where $K^{-1} = \{-k^{ij}\}$ is the inverse of the information matrix, which is equal to $-Q$ in our notations, $A = \left(A^{(1)}, A^{(2)}, \cdots, A^{(p)}\right)$ with $A^{(l)} = \left\{a_{ij}^{(l)}\right\}, a_{ij}^{(l)} = k_{ij}^{(l)} - \frac{1}{2}k_{ijl}$, $k_{ij} = E\left(\partial^2 L/\partial\beta_i\partial\beta_j\right)$, $k_{ijl} = E\left(\partial^3 L/\partial\beta_i\partial\beta_j\partial\beta_l\right)$, $k_{ij}^{(l)} = \partial k_{ij}/\partial\beta_l$, for $i, j, l, s \le p$, and L is the log likelihood function. Apparently, k_{ijl} corresponds to the element of $\overline{H_2}$. The Taylor series expansion in Rilstone *et al.* (1996) is carried out with respect to the whole parameter vector β whereas in Cox and Snell (1968) the expansion is carried out with respect to each element of β and then a set of simultaneous equations are solved to arrive at $E\left(\hat{\beta}-\beta\right)$. Following Rilstone *et al.* (1996), the expansion of ψ_T, up to order $T^{-1/2}$ and T^{-1}, respectively, is

$$\psi_T\left(\hat{\beta}\right) = \psi_T(\beta) + \nabla\psi_T(\bar{\beta})\left(\hat{\beta}-\beta\right), \tag{2.10}$$

$$\psi_T\left(\hat{\beta}\right) = \psi_T(\beta) + \nabla\psi_T(\bar{\beta})\left(\hat{\beta}-\beta\right) + \frac{1}{2}\nabla^2\psi_T(\bar{\beta})\left[\left(\hat{\beta}-\beta\right)\otimes\left(\hat{\beta}-\beta\right)\right], \tag{2.11}$$

where $\bar{\beta}$ lies between $\hat{\beta}$ and β. From (2.10) and (2.11) we can get equation (18) and (20) of Cox and Snell (1968) by taking expectation on both sides of (2.10) and (2.11), and utilizing the relation,

$$E\left(\nabla\psi_T\left(\bar{\beta}\right)\left(\hat{\beta}-\beta\right)\right) = E\left(\nabla\psi_T\left(\bar{\beta}\right)\right)E\left(\hat{\beta}-\beta\right) + Cov\left(\nabla\psi_T\left(\bar{\beta}\right), \hat{\beta}-\beta\right)$$
$$= E\left(\nabla\psi_T\left(\bar{\beta}\right)\right)E\left(\hat{\beta}-\beta\right) + \left(-\nabla\psi_T^{-1}\right)Cov\left(\psi_T, \nabla\psi_T\left(\bar{\beta}\right)\right) + o\left(T^{-1}\right)$$
$$= E\left(\nabla\psi_T\right)E\left(\hat{\beta}-\beta\right) + \left(-\nabla\psi_T^{-1}\right)Cov\left(\psi_T, \nabla\psi_T\right) + o\left(T^{-1}\right), \text{ and}$$
$$E\left(\nabla^2\psi_T\left(\bar{\beta}\right)\left[\left(\hat{\beta}-\beta\right)\otimes\left(\hat{\beta}-\beta\right)\right]\right) = E\left(\nabla^2\psi_T\left(\bar{\beta}\right)\right)E\left(\left[\left(\hat{\beta}-\beta\right)\otimes\left(\hat{\beta}-\beta\right)\right]\right) + o\left(T^{-1}\right)$$
$$= E\left(\nabla^2\psi_T\right)E\left(\left[\left(\hat{\beta}-\beta\right)\otimes\left(\hat{\beta}-\beta\right)\right]\right) + o\left(T^{-1}\right),$$

where $\nabla\psi_T = \nabla\psi_T\left(\beta\right)$, $Cov\left(\psi_T, \nabla\psi_T\left(\beta\right)\right)$ corresponds to J in Cox and Snell (1968). Then $E\left(\nabla\psi_T\left(\bar{\beta}\right)\left(\hat{\beta}-\beta\right)\right)$ and $E\left(\nabla^2\psi_T\left(\bar{\beta}\right)\left[\left(\hat{\beta}-\beta\right)\otimes\left(\hat{\beta}-\beta\right)\right]\right)$ are plugged in the expectations of (2.10) and (2.11) to derive the second-order bias. As can be seen clearly, Rilstone *et al.* (1996) solved for $\hat{\beta}-\beta$ first from the Taylor series expansion and then take expectations to express $E\left(\hat{\beta}-\beta\right)$ in terms of $E\left(a_{-1}\right)$, while Cox and Snell (1968) first took the expectation of the Taylor series expansion and then solved for $E\left(\hat{\beta}-\beta\right)$.

In fact, if we examine (2.5) and (2.9) or (2.9′) carefully, we find

$$\begin{aligned}
\left\{Q\overline{Vd}\right\}_s &= \left\{Q\overline{H_1 Q\psi_T}\right\}_s = \sum_{i=1}^{p}\sum_{j=1}^{p}\sum_{l=1}^{p} k^{sj}k_{ij,l}k^{jl}, \\
\frac{1}{2}\left\{Q\overline{H_2}\left(\overline{d\otimes d}\right)\right\}_s &= \frac{1}{2}\left\{Q\,\overline{H_2}\left(Q\otimes Q\right)\overline{\left(\psi_T\otimes\psi_T\right)}\right\}_s \\
&= -\frac{1}{2}\sum_{i=1}^{p}\sum_{j=1}^{p}\sum_{l=1}^{p} k^{sj}k_{ijl}k^{jl}, \qquad (2.12)
\end{aligned}$$

for the single parameter estimator $\hat{\beta}_s$, $1 \le s \le p$, where we use the information equality $E\left(\partial^2 L/\partial\beta_i\partial\beta_j\right) = -E\left(\partial L/\partial\beta_i\right)E\left(\partial L/\partial\beta_j\right)$. If we replace $k_{ijl} = k_{ij}^{(l)} - k_{ij,l}$, then we immediately get

$$B\left(\hat{\beta}\right) = Q\left[\overline{Vd} - \frac{1}{2}\overline{H_2}\left(\overline{d\otimes d}\right)\right] = K^{-1}Avec(K^{-1}). \qquad (2.13)$$

This establishes the equality of (2.9) and (2.5). However, we should point out that (2.5) is more general since it includes the MLE as a special case. Furthermore, Cox and Snell (1968) did not give us the second-order MSE by using their approach, but we develop this for the dependent observations here and Rilstone *et al.* (1996) for the i.i.d. case.

Of course, we should bear in mind that all the expectations involved in (2.4) or (2.5) and (2.6) are in general much complicated. However, it is observed that a lot of econometric estimators are derived from some quadratic moment conditions. Then some well known results on the expectations of quadratic forms in a normal or nonnormal vector can be used for our purpose. In the following section, we use these results extensively for most of the examples.

3 Illustrations

In this section, we give the application of our second-order bias and MSE results to some time series models. These include the AR(1) model, structural model with AR(1) errors, VAR model, MA(1) model, partial adjustment model, and absolute regression model. Here we do not attempt to give an exhaustive list of all interesting econometric models, for example, the ARCH model of Engle (1982).[1] The approach, however, is unified, and is practically applicable as long as we can take expectations on the derivatives (up to third order) of the moment function used for estimation. We note that some results (e.g., AR(1)) are readily available through other methods in the literature and are consistent with our results or degenerate as our special cases, but most results are new through our method and are easy to implement numerically.

3.1 AR Model

Consider an AR(1) model

$$y_t = \beta y_{t-1} + \varepsilon_t, \tag{3.1}$$

where $|\beta| < 1$, ε_t is i.i.d. $N(0, \sigma_\varepsilon^2)$ and $t = 1, 2, ..., T$. Denote $y = (y_1, y_2, \cdots, y_T)'$, then $y \sim N(0, \Sigma)$, where $\Sigma = \{\sigma_{tt'}\}$ is a $T \times T$ matrix with $\sigma_{tt'} = \sigma_\varepsilon^2 \beta^{|t-t'|} / \left(1 - \beta^2\right)$ for t, $t' = 1, 2, ..., T$.

[1] We find that the second-order bias for ML estimator in ARCH(1) model is $Q\left\{\overline{V_1 d_1} - \frac{1}{2}\overline{H_2}\,\overline{[d_1 \otimes d_1]}\right\}/T - Q\left(\sum_{i>j} V_i d_j\right)/T$, which is of course equivalent to the result using (2.9). Iglesias and Phillips (2001, 2002) derived the result using (2.9) directly. We got the same result in an early version of this paper. To save space, we do not repeat the result in this paper. The MSE result is more involved and is in our future research agenda.

It is well known that the OLS estimator of β is equivalent to the (conditional) ML estimator. That is, to estimate, we use the following moment condition

$$\psi_T = \frac{1}{T-1}\sum_{t=2}^{T} y_{t-1}\varepsilon_t = \frac{1}{T-1}y'Cy = 0, \tag{3.2}$$

where $C = \{c_{tt'}\}$ is a $T \times T$ matrix with $c_{tt'} = -\beta$ for $t = t' = 1, 2, \cdots, T-1$, $c_{tt'} = 1/2$ if $t = t' + 1$ or $t = t' - 1$ and it is 0 otherwise. Therefore, we have

$$\begin{aligned} H_1 &= \nabla\psi_T = \frac{y'C_1y}{T-1},\ H_2 = H_3 = 0, \\ Q &= \overline{H_1}^{-1} = \left(\frac{trC_1\Sigma}{T-1}\right)^{-1},\ V = H_1 - \overline{H_1},\ W = 0, \end{aligned} \tag{3.3}$$

where $C_1 = \nabla C = \{\nabla c_{tt'}\}$ with $\nabla c_{tt'} = -1$ for $t = t' = 1, 2, \cdots, T-1$ and otherwise it is 0. Using (3.3) in (2.4) or (2.5) and (2.6) and noting that $tr\,(C\Sigma) = 0$ we obtain the second-order bias of $\hat{\beta}$, up to $O\left(T^{-1}\right)$, and MSE, up to $O\left(T^{-2}\right)$, as

$$\begin{aligned} B\left(\hat{\beta}\right) &= \frac{1}{(T-1)^2}Q^2\lambda_{11}, \\ M\left(\hat{\beta}\right) &= \frac{6Q^2}{(T-1)^2}\lambda_{20} + \frac{2Q^3}{(T-1)^3}\left(1+3Q^2\right)\lambda_{21} + \frac{3Q^4}{(T-1)^4}\lambda_{22}, \end{aligned} \tag{3.4}$$

where $\lambda_{rs} = E\left[(y'Cy)^r \cdot (y'C_1y)^s\right]$ for $r, s = 0, 1, 2$ are given in Appendix A.2. Note that in Appendix A.2 we further simplify, up to $o\left(T^{-1}\right)$,

$$B\left(\hat{\beta}\right) = \frac{-2\beta}{(T-1)}, \tag{3.5}$$

which is consistent with Phillips (1977), for instance.

Remark 1: If we include some nonstochastic exogenous regressors, for example, X, in the AR(1) model, then the extension is straightforward. Also, we can generalize the case to an AR(p) model with some exogenous regressors, as long as we can rewrite

$$\frac{1}{T-j}\sum_{t=j+1}^{T} y_t y_{t-j} = y'N_j y,\ j = 0, 1, \cdots, p. \tag{3.6}$$

See, for example, Kiviet and Phillips (1993).

Remark 2: The exact log likelihood function (excluding a constant term) when we normalize $\sigma_\varepsilon^2 = 1$ is

$$L = \frac{1}{T}\left\{-\frac{1}{2}\log\left(\frac{1}{1-\beta^2}\right) - \frac{1}{2}y_1^2\left(1-\beta^2\right) - \frac{1}{2}\sum_{t=2}^{T}\left(y_t - \beta y_{t-1}\right)^2\right\} \tag{3.7}$$

and the score function is

$$\psi_T = \frac{1}{T}\left\{\frac{-\beta}{1-\beta^2} + \beta y_1^2 + \sum_{t=2}^{T}\left(y_t - \beta y_{t-1}\right)y_{t-1}\right\} = 0. \tag{3.8}$$

Define a $T \times T$ diagonal matrix $D = diag\,(1, 0, \cdots, 0)$. Then we can rewrite the score function as

$$\psi_T = \frac{1}{T}y'C^*y - \frac{1}{T}\frac{\beta}{1-\beta^2} = 0, \tag{3.9}$$

where $C^* = \beta D + C$. Then immediately we have

$$\begin{aligned}
\nabla\psi_T &= \frac{y'Dy}{T} - \frac{y'C_1y}{T} - \frac{1}{T}\frac{1+\beta^2}{\left(1-\beta^2\right)^2},\ Q = \frac{T\left(1-\beta^2\right)^2}{(T-3)\,\beta^2 + 1 - T}, \\
H_2 &= -\frac{1}{T}\frac{6\beta + 2\beta^3}{\left(1-\beta^2\right)^3},\ H_3 = -\frac{1}{T}\frac{6 + 36\beta^2 + 6\beta^4}{\left(1-\beta^2\right)^4},\ W = 0.
\end{aligned} \tag{3.10}$$

Then it is straightforward to follow the procedure as in the conditional ML estimator case. This can be easily generalized to AR(p) model with some nonstochastic exogenous regressors.

3.2 Simultaneous Equation Model

Consider the model

$$y_{1t} = \beta y_{2t} + \varepsilon_t,\ y_{2t} = \pi x_t + v_{2t}, \tag{3.11}$$

where $\varepsilon_t = \rho\varepsilon_{t-1} + u_t$, x_t is a scalar of nonstochastic regressor, y_{1t} and y_{2t} are endogenous variables, $|\rho| < 1$, and ε_t, u_t and v_{2t} are the disturbances. The reduced form for y_{1t} is

$$y_{1t} = x_t\pi\beta + v_{1t},\ v_{1t} = \varepsilon_t + \beta v_{2t}. \tag{3.12}$$

We assume that $Eu_t = Ev_{2t} = 0$, $Ev_{2t}^2 = \omega_{22}$, $Eu_t^2 = \sigma_u^2$, $E\left(u_tv_{2t}\right) = E\left(\varepsilon_tv_{2t}\right) = \omega_{u2}$, $E\varepsilon_t^2 = \sigma_\varepsilon^2 = \sigma_u^2/\left(1-\rho^2\right)$, and all the third (cross) moments are zero. Also we assume u and

v_2 are independent at all lags and leads so that $E\left(\varepsilon_t v_{2t'}\right) = \omega_{u2}\rho^{t-t'}$. Then, assuming joint normality of $v_t = \left(v_{1t},\ v_{2t}\right)'$ and denoting $y_t = \left(y_{1t},\ y_{2t}\right)'$, we note that

$$Ey = \mu_y,\ V\left(y\right) = \Omega_y, \tag{3.13}$$

where y is a $2T \times 1$ vector, $\mu_y = \left(\left(x\pi\beta\right)',\ \left(x\pi\right)'\right)'$ and Ω_y is a $2T \times 2T$ matrix

$$\Omega_y = \begin{bmatrix} \Omega_{11} & \Omega_{12} \\ \Omega_{21} & \Omega_{22} \end{bmatrix}; \tag{3.14}$$

$\Omega_{11} = Ev_1v_1' = \left(\frac{\sigma_u^2}{1-\rho^2} + 2\beta\omega_{u2}\right)\Sigma + \beta^2\omega_{22}I_T$, $\Omega_{12} = Ev_1v_2' = \omega_{u2}\Sigma + \beta\omega_{22}I_T$, $\Omega_{22} = Ev_2v_2' = \omega_{22}I_T$ and $\Sigma = \{\sigma_{tt'}\}$ with $\sigma_{tt'} = \rho^{t-t'}$ for $t,\ t' = 1, 2, ..., T$.

The moment condition for estimating β *neglecting* the dependence structure in the disturbances is

$$\psi_T = \frac{1}{T}\sum_{t=1}^{T}\left(y_{1t} - \beta y_{2t}\right)x_t = \frac{1}{T}x'Cy = 0, \tag{3.15}$$

where $C = \begin{pmatrix} 1 & -\beta \end{pmatrix} \otimes I_T$ and $Ex'Cy = 0$. Thus

$$\begin{aligned} H_1 &= \nabla\psi_T = \frac{1}{T}x'C_1y,\ H_2 = H_3 = 0, \\ Q &= \left(\overline{H_1}\right)^{-1} = \left(\frac{1}{T}x'C_1\mu_y\right)^{-1} = \left(-\frac{x'x\pi}{T}\right)^{-1},\ V = H_1 - \overline{H_1},\ W = 0, \end{aligned} \tag{3.16}$$

where $C_1 = \nabla C = \begin{pmatrix} 0 & -1 \end{pmatrix} \otimes I_T$. Given all these conditions, we have Theorem 1 on the second-order bias and MSE of $\hat{\beta}$ estimated from (3.15).

Theorem 1: Under normality assumption of v_t, the bias, up to $O\left(T^{-1}\right)$, and the MSE, up to $O\left(T^{-2}\right)$, respectively, of $\hat{\beta}$ estimated from (3.15) when x_t is nonstochastic, are

$$\begin{aligned} B\left(\hat{\beta}\right) &= -\frac{\omega_{u2}x'\Sigma x}{\pi^2\left(x'x\right)^2}, \\ M\left(\hat{\beta}\right) &= \frac{\sigma_u^2}{1-\rho^2}\frac{x'\Sigma x}{\pi^2\left(x'x\right)^2} + \frac{6\omega_{u2}^2\left(x'\Sigma x\right)^2}{\pi^4\left(x'x\right)^4} + \frac{3\sigma_u^2\omega_{22}}{1-\rho^2}\frac{\left(x'\Sigma x\right)}{\pi^4\left(x'x\right)^3}. \end{aligned} \tag{3.17}$$

Proof: See Appendix A.3. ■

Nagar's (1959) result apparently follows from Theorem 1 by setting $\rho = 0$, which is given in Corollary 1.

Corollary 1: When $\rho = 0$, that is, ε_t is i.i.d., then the bias result in (3.17) reduces to

$$B\left(\hat{\beta}\right) = -\frac{\omega_{u_2}}{\pi^2 \sum_{t=1}^{T} x_t^2}. \tag{3.18}$$

If x is a stochastic variable with $Ex = \mu_x$ and $V(x) = \Sigma_x$ but independent of v_1 and v_2, then the bias and MSE can again be derived from (2.4).

For this, we write

$$\psi_T = \frac{1}{T} x^{*\prime} C y = \frac{1}{T} Z' C^* Z = 0 \tag{3.19}$$

where we redefine

$$C = \begin{bmatrix} I & 0 \\ 0 & -\beta I \end{bmatrix}, \; C^* = \frac{1}{2} \begin{bmatrix} 0 & C \\ C & 0 \end{bmatrix}, \tag{3.20}$$

and $x^* = (x', x')'$, $Z = (x^{*\prime}, y')'$. Here x^* is $2T \times 1$, C is $2T \times 2T$, Z is $4T \times 1$, and C^* is $4T \times 4T$. Note that $Z \sim N(\mu_Z, \Sigma_Z)$ with $\mu_Z = \left(\mu'_{x^*}, \mu'_y\right)'$ and

$$\Sigma_Z = \begin{bmatrix} \Omega_{x^*} & \Omega_{x^* y} \\ \Omega_{y x^*} & \Omega_y \end{bmatrix}, \tag{3.21}$$

where $\mu_{x^*} = (\mu'_x, \mu'_x)'$, $\mu_y = \left((\mu_x \pi \beta)', \; (\mu_x \pi)'\right)'$, and Ω_{x^*} is a $2T \times 2T$ block matrix with each block element being Σ_x, $\Omega_{x^* y} = \Omega'_{y x^*}$, and

$$\Omega_{y x^*} = \begin{bmatrix} \pi \beta \Sigma_x & \pi \beta \Sigma_x \\ \pi \Sigma_x & \pi \Sigma_x \end{bmatrix}. \tag{3.22}$$

Then $H_2 = H_3 = 0$, $W = 0$, and

$$\begin{aligned} H_1 &= \frac{1}{T} Z' C_1^* Z, \\ Q^{-1} &= \frac{1}{T} E\left(Z' C_1^* Z\right) = -\frac{1}{T} \pi \mu'_x \mu_x - \frac{1}{T} \pi tr\left(\Sigma_x\right), \end{aligned} \tag{3.23}$$

where C_1^* is C^* with C replaced by

$$C_1 = \begin{bmatrix} 0 & 0 \\ 0 & -I_T \end{bmatrix}. \tag{3.24}$$

Note that both C^* and C_1^* are symmetric.

Theorem 2: Under normality assumption of v_t, the bias, up to $O\left(T^{-1}\right)$, and the MSE, up to $O\left(T^{-2}\right)$, respectively, of $\hat{\beta}$ estimated from (3.15) when x_t is stochastic, are

$$\begin{aligned} B\left(\hat{\beta}\right) &= \frac{\left(-\omega_{u2} tr\left(\Sigma\Sigma_x\right) - \omega_{u2}\mu_x'\Sigma\mu_x\right)}{\pi^2\left[\mu_x'\mu_x + tr\left(\Sigma_x\right)\right]^2}, \\ M\left(\hat{\beta}\right) &= 3\frac{Q^4}{T^4}\lambda_{11} - 2\frac{Q^2}{T^2}\lambda_{10}, \end{aligned} \tag{3.25}$$

where $\lambda_{ij} = E\left[\left(Z'C^*Z\right)^{2i}\left(Z'C_1^*Z\right)^{2j}\right]$ for $i, j = 0, 1$.

Proof: See Appendix A.3. ■

Note that (3.17) is just a special case of (3.25). Corollary 2 and 3 given below give the second-order bias results for some specific cases of x_t.

Corollary 2: If x_t follows an AR(1) process as $x_t = \rho_x x_{t-1} + \eta_t$, $|\rho_x| < 1$, $\eta_t \sim$ i.i.d. $\left(0,\ \sigma_\eta^2\right)$, $\sigma_x^2 = \sigma_\eta^2/\left(1-\rho_x^2\right)$, then the bias result in (3.25) reduces to

$$B\left(\hat{\beta}\right) = \frac{-\omega_{u2}\left[T\sigma_x^2 + \sum_{\substack{t=1 \\ t\neq t'}}^{T}\sum_{t'=1}^{T}\left(\rho\rho_x\sigma_x^2\right)^{t-t'}\right]}{\pi^2\left(T\sigma_x^2\right)^2}. \tag{3.26}$$

Corollary 3: If x_t or ε_t is i.i.d. with mean zero, then the bias result in (3.26) reduces to

$$B\left(\hat{\beta}\right) = -\frac{\omega_{u2}}{T\pi^2 E x_1^2}. \tag{3.27}$$

Note that (3.27) encompasses the result in Rilstone *et al.* (1996) in that even though ε_t is an AR(1) process the bias result in Rilstone *et al.* (1996) will hold as long as x_t is i.i.d. with mean zero.

3.3 VAR Model

Consider the following VAR(1) model

$$\begin{cases} y_{1t} = \beta_{11} y_{1,t-1} + \beta_{12} y_{2,t-1} + \varepsilon_{1t} \\ y_{2t} = \beta_{21} y_{1,t-1} + \beta_{22} y_{2,t-1} + \varepsilon_{2t} \end{cases}, \quad \begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{pmatrix} \sim N(0, \Omega), \tag{3.28}$$

where $\Omega = \{\omega_{ij}\}$, i, $j = 1, 2$.

Define $y_t = (y_{1t}, y_{2t})'$, $B = \left((\beta_{11}, \beta_{12})', (\beta_{21}, \beta_{22})'\right)'$, $\varepsilon_t = (\varepsilon_{1t}, \varepsilon_{2t})'$, then we can compactly write

$$y_t = B y_{t-1} + \varepsilon_t. \tag{3.29}$$

We assume that the process is covariance stationary, that is, for all λ satisfying $|I_2\lambda - B| = 0$, $|\lambda| < 1$. For nonstationary VAR models, see Phillips (1987). In particular, this *implies* that $|\beta_{11}| < 1$, $|\beta_{22}| < 1$, $|\beta_{12}\beta_{21}| < 1$. Denote $\Gamma_0 = E(y_t y_t')$, $\Gamma_j = E\left(y_t y_{t-j}'\right)$, $\Gamma_{-j} = E(y_{t-j} y_t') = \Gamma_j'$. Let $vec(A)$ be a column vector of all the columns of A stacked from the first one to the last one. We know that $vec(\Gamma_0) = (I_4 - B \otimes B)^{-1} vec(\Omega)$ and $\Gamma_j = B^j \Gamma_0$. Therefore, $y_t \sim N\left(0, vec^{-1}\left[(I_4 - B \otimes B)^{-1} vec(\Omega)\right]\right)$. If we stack all observations as $y = (y_1', y_2', \cdots, y_T')'$, then we have $y \sim N(0, \Sigma)$, where Σ is the $2T \times 2T$ variance-variance matrix with the tt'th element being $B^{t-t'}\Gamma_0$ if $t \geq t'$ and $B^{t'-t}\Gamma_0'$ if $t < t'$ for t, $t' = 1, 2, ..., T$.

It is well known that the maximum likelihood estimators of each single equation is the same as the OLS estimators. As in Section 3.1, if we properly define C_{ij}, i, $j = 1, 2$, then we have the following moment condition for $\beta \equiv vec(B)$

$$\psi_T = \frac{1}{T-1} \sum_{t=2}^{T} q_t = \frac{1}{T-1} \sum_{t=2}^{T} vec\left[(y_t - B y_{t-1}) y_t'\right] = \frac{1}{T-1} \begin{pmatrix} y' C_{11} y \\ y' C_{21} y \\ y' C_{12} y \\ y' C_{22} y \end{pmatrix} = 0, \tag{3.30}$$

where

$$
\begin{aligned}
C_{11} &= \begin{pmatrix} 0_{2(T-1)\times 2} & AA \\ 0_{2\times 2} & 0_{2\times 2(T-1)} \end{pmatrix} - \beta_{11} \begin{pmatrix} AA & 0_{2(T-1)\times 2} \\ 0_{2\times 2(T-1)} & 0_{2\times 2} \end{pmatrix} - \beta_{12} \begin{pmatrix} BB & 0_{2(T-1)\times 2} \\ 0_{2\times 2(T-1)} & 0_{2\times 2} \end{pmatrix}, \\
C_{21} &= \begin{pmatrix} 0_{2(T-1)\times 2} & BB \\ 0_{2\times 2} & 0_{2\times 2(T-1)} \end{pmatrix} - \beta_{21} \begin{pmatrix} AA & 0_{2(T-1)\times 2} \\ 0_{2\times 2(T-1)} & 0_{2\times 2} \end{pmatrix} - \beta_{22} \begin{pmatrix} BB & 0_{2(T-1)\times 2} \\ 0_{2\times 2(T-1)} & 0_{2\times 2} \end{pmatrix}, \\
C_{12} &= \begin{pmatrix} 0_{2(T-1)\times 2} & CC \\ 0_{2\times 2} & 0_{2\times 2(T-1)} \end{pmatrix} - \beta_{11} \begin{pmatrix} CC & 0_{2(T-1)\times 2} \\ 0_{2\times 2(T-1)} & 0_{2\times 2} \end{pmatrix} - \beta_{12} \begin{pmatrix} DD & 0_{2(T-1)\times 2} \\ 0_{2\times 2(T-1)} & 0_{2\times 2} \end{pmatrix}, \\
C_{22} &= \begin{pmatrix} 0_{2(T-1)\times 2} & DD \\ 0_{2\times 2} & 0_{2\times 2(T-1)} \end{pmatrix} - \beta_{21} \begin{pmatrix} CC & 0_{2(T-1)\times 2} \\ 0_{2\times 2(T-1)} & 0_{2\times 2} \end{pmatrix} - \beta_{22} \begin{pmatrix} DD & 0_{2(T-1)\times 2} \\ 0_{2\times 2(T-1)} & 0_{2\times 2} \end{pmatrix},
\end{aligned}
$$

where the $2(T-1) \times 2(T-1)$ matrices AA, BB, CC, and DD are block diagonal with the block element $aa = \left((1,0)', (0,0)'\right)'$, $bb = \left((0,1)', (0,0)'\right)'$, $cc = \left((0,0)', (1,0)'\right)'$, and $dd = \left((0,0), (0,1)'\right)'$, respectively.

Define $C_{11} = C_{11}^{(1)} - \beta_{11} C_{11}^{(2)} - \beta_{12} C_{11}^{(3)}$, where $C_{11}^{(l)}$, $l = 1,2,3$, represents the lth matrix part in the definition of C_{11} in (3.30). Similar definitions are used for C_{21}, C_{12}, and C_{22}. Note that $y'C_{11}^{(3)}y = y'C_{12}^{(2)}y$, $y'C_{21}^{(3)}y = y'C_{22}^{(2)}y$.

Immediately we have

$$
\begin{aligned}
\nabla \psi_T &= H_1 = -\frac{1}{T-1} \begin{pmatrix} y'C_{11}^{(2)}y & 0 & y'C_{11}^{(3)}y & 0 \\ 0 & y'C_{21}^{(2)}y & 0 & y'C_{21}^{(3)}y \\ y'C_{12}^{(2)}y & 0 & y'C_{12}^{(3)}y & 0 \\ 0 & y'C_{22}^{(2)}y & 0 & y'C_{22}^{(3)}y \end{pmatrix}, \\
H_2 &= H_3 = W = 0. \qquad (3.31)
\end{aligned}
$$

The results on quadratic forms in Appendix A.1 then can be used here to evaluate the second-order bias and MSE of β. Denote $E\left(y'C_{ij}^{(l)}y\right) = tr\left[\left(C_{ij}^{(l)} + C_{ij}^{(l)\prime}\right)\Sigma/2\right] = \lambda_{ij}^{(l)}$, $E\left(y'C_{ij}y \cdot y'C_{mn}^{(l)}y\right) = 2tr\left[\left(C_{ij} + C_{ij}'\right)\Sigma\left(C_{mn}^{(l)} + C_{mn}^{(l)\prime}\right)\Sigma/4\right] + tr\left[\left(C_{ij} + C_{ij}'\right)\Sigma/2\right] \times tr\left[\left(C_{mn}^{(l)} + C_{mn}^{(l)\prime}\right)\Sigma/2\right] = \mu_{ijmn}^{(l)}$ for $i,j,m,n = 1,2$ and $\lambda_{i1}^{(3)}\lambda_{i2}^{(2)} - \lambda_{i1}^{(2)}\lambda_{i2}^{(3)} = \phi_i$ for $i = 1,2$, then it is easy to verify the bias result as given in Theorem 3.

Theorem 3: The bias , up to $O\left(T^{-1}\right)$, of $\hat{\beta}$ estimated from (3.30) in VAR(1) model (3.28) is

$$
B\left(\hat{\beta}_{i1}\right) = \frac{\xi_i}{\phi_i^2},\ B\left(\hat{\beta}_{i2}\right) = \frac{\eta_i}{\phi_i^2},\ i = 1,2, \qquad (3.32)
$$

where

$$\xi_i = \left(\mu^{(3)}_{i2i2} - \mu^{(3)}_{i2i1}\right)\lambda^{(2)}_{i1}\lambda^{(3)}_{i1} + \left(\mu^{(2)}_{i2i1} - \mu^{(3)}_{i1i2}\right)\lambda^{(3)}_{i1}\lambda^{(2)}_{i2} + \mu^{(2)}_{i1i2}\lambda^{(3)}_{i2}\lambda^{(3)}_{i1} + \mu^{(3)}_{i1i1}\lambda^{(2)}_{i2}\lambda^{(3)}_{i2} \\ -\mu^{(2)}_{i1i1}\left(\lambda^{(3)}_{i2}\right)^2 - \mu^{(2)}_{i2i2}\left(\lambda^{(3)}_{i1}\right)^2,$$

$$\eta_i = \left(\mu^{(3)}_{i1i2} - \mu^{(3)}_{i2i1}\right)\lambda^{(2)}_{i1}\lambda^{(2)}_{i2} + \left(\mu^{(2)}_{i2i2} - \mu^{(2)}_{i1i2}\right)\lambda^{(2)}_{i1}\lambda^{(3)}_{i1} + \mu^{(2)}_{i1i1}\lambda^{(2)}_{i2}\lambda^{(3)}_{i2} + \mu^{(2)}_{i2i1}\lambda^{(3)}_{i1}\lambda^{(2)}_{i2} \\ -\mu^{(3)}_{i1i1}\left(\lambda^{(2)}_{i2}\right)^2 - \mu^{(3)}_{i2i2}\left(\lambda^{(2)}_{i1}\right)^2.$$

Proof: Substitute (3.30) and (3.31) into (2.5) and use the results on quadratic forms in Appendix A.1. ■

Here we do not give explicitly expressions for the second-order MSE. However, it is straightforward to write a computer program to do the evaluation numerically.

Table 1 gives some numerical results for different sample sizes when $B = \left((0.8,\ 0.3)',\ \ (0.2,\ 0.5)'\right)'$ and $\left((0.1,\ 0.4)',\ \ (0.4,\ 0.3)'\right)'$, where eigenvalues refer to those of the parameter matrix B and we normalize $\Omega = \left((1,\rho)', (\rho,1)'\right)'$. We tried many other different B's, whose results are available upon request. Three outstanding points worthy of mentioning are that i) increase of sample size will significantly reduce the second-order bias and MSE; ii) the higher ρ, the larger the second-order MSE, which implies that even though the OLS estimators are consistent, the correlation between disturbance terms across equations will cause imprecision of the OLS estimators; iii) the difference between the second-order and first-order results is not insignificant for small samples we have considered, and it seems that the first-order variance, which is equal to the first-order MSE, usually underestimates the second-order variance, $\left(M\left(\hat{\beta}\right) - B\left(\hat{\beta}\right)^2\right)$, as well as the seond-order MSE. Of course, the first-order variance can be calculated from $\Omega \otimes \Gamma_0$ (see Hamilton, 1994, p. 299), which is equal to A_{-1} is our expression. Since we normalize $\Omega = \left((1,\rho)', (\rho,1)'\right)'$, the first-order variance is the same for the parameter estimators in the same row of the variance matrix of $vec\left(\hat{B}\right)$. Also, we should emphasize that our results are only approximations. For high $|\rho|$ around 0.9, we do encounter some second-order MSE with negative values (represented by "/" in Table 1) or very large unrealistic values around two hundred in small samples, which though is consistent with the fact that our results are *asymptotic* second-order results.

Remark 3: We can easily generalize to VAR(p) model of n variables, as long as we can write

$$\sum_{k=l+1}^{T} y_{i,k}y_{j,k-l} = y'N_{ijk}y,\ \ i,j \leq n,\ \ l \leq p. \tag{3.33}$$

3.4 MA Model

Consider the simplest case

$$y_t = \varepsilon_t - \beta\varepsilon_{t-1},\ \ |\beta| < 1, \tag{3.34}$$

where we normalize $\varepsilon_t \sim i.i.d.\ N(0,1)$. Define $y = (y_1, \cdots, y_t, y_{t+1}, \cdots, y_T)'$. Following Hamilton (1994), we have the averaged sample (conditional) log likelihood function, excluding a constant term, $L = -\frac{1}{T}\sum_{t=1}^{T}\frac{\varepsilon_t^2}{2}$, where $\varepsilon_t = y_t + \beta y_{t-1} + \cdots + \beta^{t-1}y_1 = \sum_{i=0}^{t-1} a_i y_{t-i} \equiv y'A_t$ with $a_i = \beta^i$ and

$$A_t = A_t(\beta) = \Bigg(a_{t-1}, \cdots, a_1, a_0, \underbrace{0, \cdots, 0}_{T-t \text{ zeros}}\Bigg)' \tag{3.35}$$

Then we have

$$\frac{\partial \varepsilon_t}{\partial \theta} = y'B_t,\ \ \frac{\partial^2 \varepsilon_t}{\partial \theta^2} = y'C_t,\ \ \frac{\partial^3 \varepsilon_t}{\partial \theta^3} = y'U_t,\ \ \frac{\partial^4 \varepsilon_t}{\partial \theta^4} = y'V_t, \tag{3.36}$$

where

$$B_t = \frac{\partial A_t}{\partial \beta},\ \ C_t = \frac{\partial^2 A_t}{\partial \beta^2},\ \ U_t = \frac{\partial^3 A_t}{\partial \beta^3},\ \ V_t = \frac{\partial^4 A_t}{\partial \beta^4}, \tag{3.37}$$

and

$$\begin{aligned}
q_t &= -\left(\frac{\partial \varepsilon_t}{\partial \beta}\right)\varepsilon_t = -y'B_tA_t'y, \\
\nabla q_t &= -\left(\frac{\partial^2 \varepsilon_t}{\partial \beta^2}\right)\varepsilon_t - \left(\frac{\partial \varepsilon_t}{\partial \beta}\right)^2 = -y'C_tA_t'y - y'B_tB_t'y, \\
\nabla^2 q_t &= -\left(\frac{\partial^3 \varepsilon_t}{\partial \beta^3}\right)\varepsilon_t - 3\left(\frac{\partial^2 \varepsilon_t}{\partial \beta^2}\right)\left(\frac{\partial \varepsilon_t}{\partial \beta}\right) = -y'U_tA_t'y - 3y'C_tB_t'y, \\
\nabla^3 q_t &= -\left(\frac{\partial^4 \varepsilon_t}{\partial \beta^4}\right)\varepsilon_t - 4\left(\frac{\partial^3 \varepsilon_t}{\partial \beta^3}\right)\left(\frac{\partial \varepsilon_t}{\partial \beta}\right) - 3\left(\frac{\partial^2 \varepsilon_t}{\partial \beta^2}\right)^2 \\
&= -y'V_tA_t'y - 4y'U_tB_t'y - 3y'C_tC_t'y.
\end{aligned} \tag{3.38}$$

In matrix notation, we have

$$\psi_T = \frac{1}{T} y' N y, \; H_i = \frac{1}{T} y' N_i y, \; i = 1, 2, 3, \tag{3.39}$$

where $N = -A'B$, $N_1 = -A'C - B'B$, $N_2 = -A'U - 3B'C$, $N_3 = -A'V - 4B'U - 3C'C$, $A = (A_1, A_2, \cdots, A_T)$ and similarly for B, C, U, V. Note that y is normally distributed with mean 0 and variance-covariance matrix Σ such that it has tt'th element $1 + \beta^2$ when $t = t'$, $-\beta$ when $|t - t'| = 1$, and 0 elsewhere, for t, $t' = 1, 2, \cdots, T$. With these notations, we use the results on quadratic forms in Appendix A.1 to derive the second-order bias and MSE of $\hat{\beta}$. However, we should bear in mind that if we write $\varepsilon_t = y' A_t$ in stead of $\varepsilon_t = \sum_{i=0}^{\infty} a_i y_{t-i}$, then it is not necessary that $E\left(a_{-1/2}\right) = 0$ in (2.4), but $\text{plim} a_{-1/2} = 0$. The asymptotic expansion in Rilstone *et al.* (1996) is nevertheless valid as long as we have $\sqrt{T}$-consistency of the estimator $\hat{\beta}$, together with their assumptions A to C.[2] Therefore, instead of (2.5), we use the bias expression in (2.4) directly with $E\left(a_{-1/2}\right) \neq 0$. The MSE expression given by (2.6), on the other hand, is still valid here. Theorem 4 gives a closed-form formula for the bias based on the above observation.

Theorem 4: The bias, up to $O\left(T^{-1}\right)$, of the conditional ML estimator $\hat{\beta}$ for model (3.34) is

$$\begin{aligned} B\left(\hat{\beta}\right) &= \frac{tr\left(N^*\Sigma\right) \cdot tr\left(N_1^*\Sigma\right) + 2tr\left(N^*\Sigma N_1^*\Sigma\right)}{\left[tr\left(N_1^*\Sigma\right)\right]^2} - 2\frac{tr\left(N^*\Sigma\right)}{tr\left(N_1^*\Sigma\right)} \\ &\quad - \frac{tr\left(N_2^*\Sigma\right) \cdot \left[\left(trN^*\Sigma\right)^2 + 2tr\left(\left(N^*\Sigma\right)^2\right)\right]}{2\left[tr\left(N_1^*\Sigma\right)\right]^3}, \end{aligned} \tag{3.40}$$

where $N_i^* = \frac{N_i + N_i'}{2}$, $i = 0, 1, 2$, $N_0 = N$, $N_0^* = N^*$.

Proof: Substitute (3.39) into (2.4) and use the results on quadratic forms in Appendix A.1. ■

[2] We can of course generalize for any $\sqrt[g]{T}$-consistent estimator, $g > 0$. But (2.3), (2.5) and (2.6) then have to be modified accordingly.

As for $M\left(\hat{\beta}\right)$, we do not report the explicit result here. But it is very easy to write a computer program to evaluate it. Figure 1 plots $B\left(\hat{\beta}\right)$ and Figure 2 plots $M\left(\hat{\beta}\right)$ against β for given sample size T. We see clearly the following patterns for $B\left(\hat{\beta}\right)$ and $M\left(\hat{\beta}\right)$: i) $B\left(\hat{\beta}\right)$ and $M\left(\hat{\beta}\right)$ reduces significantly as sample size increases and $\hat{\beta}$ is almost unbiased even for a sample size as small as 20; ii) the second-order bias behaves sinusoidally for β between (β_{B1}, β_{B2}) and the magnitude of it is approximately symmetric around the point $\beta_{B0} = 0$, at which the second-order bias is zero; iii)the absolute bias reaches two peaks at β_{B1} for $\beta < 0$ and at β_{B2} for $\beta > 0$; iv) the second-order MSE is approximately symmetric around the point where $\beta = 0$; v) starting from the origin, the second-order MSE monotonically increases as the magnitude of β increases until it reaches two peaks at β_{M1} for $\beta < 0$ and β_{M2} for $\beta > 0$, and then decreases. The exact values of β_{B1}, β_{B2}, β_{M1}, and β_{M2} can read from Figure 1 and Figure 2.

Of course, we should point out that as $|\beta|$ approaches one, the initial values of y for the conditional ML estimator are not negligible in small samples and hence we may cast some doubt on the second-order results built upon the conditional ML condition when $|\beta|$ is close to one.

Remark 4: In principle, this can be generalized to an MA(q) model. Instead of using quadratic forms in y, we use the normal vector ε directly. For simplicity, consider, for example, an MA(2) model, $y_t = \varepsilon_t + \theta_1\varepsilon_{t-1} + \theta_2\varepsilon_{t-2}$. Again, we have the log likelihood function $L = -\sum_{t=1}^{T} \varepsilon_t^2/2T$, with $\varepsilon_t = y_t - \theta_1\varepsilon_{t-1} - \theta_2\varepsilon_{t-2}$. Here we set initial values $\varepsilon_0 = \varepsilon_{-1} = 0$. Note that q_t is 2×1, ∇q_t is 2×2, $\nabla^2 q_t$ is 2×4, and $\nabla^3 q_t$ is 2×8. All the expectations of cross products will boil down to the form of cross products of $E\left(\partial^4\varepsilon/\partial\theta_i\partial\theta_j\partial\theta_k\partial\theta_l\right)$, $1 \leq i, j, k, l \leq 2$, or of lower order. For example, consider $E\left(\partial^4\varepsilon/\partial\theta_1^4\right)$. From $\varepsilon_t = y_t - \theta_1\varepsilon_{t-1} - \theta_2\varepsilon_{t-2}$ we have $\partial\varepsilon_t/\partial\theta_1 = -\varepsilon_{t-1} - \theta_1\partial\varepsilon_{t-1}/\partial\theta_1 - \theta_2\partial\varepsilon_{t-2}/\partial\theta_1$. Rewrite it as $\varepsilon_{t-1} = -\partial\varepsilon_t/\partial\theta_1 - \theta_1\partial\varepsilon_{t-1}/\partial\theta_1 - \theta_2\partial\varepsilon_{t-2}/\partial\theta_1$. Therefore, we have $\varepsilon = A_1\partial\varepsilon/\partial\theta_1$, where A_1 is $T \times T$ such that its tt'th element is equal to $-\theta_1$ if $t = t'$, -1 if $t' = t + 1$, $-\theta_2$ if $t = t' + 1$, and 0 elsewhere, for t, $t' = 1, 2, \cdots, T$. Then $\partial\varepsilon/\partial\theta_1 = A_1^{-1}\varepsilon = C_1\varepsilon$ for $C_1 = A_1^{-1}$. Invertibility will guarantee that A_1 is nonsingular. Next, from $\varepsilon_{t-1} = -\partial\varepsilon_t/\partial\theta_1 - \theta_1\partial\varepsilon_{t-1}/\partial\theta_1 - \theta_2\partial\varepsilon_{t-2}/\partial\theta_1$ we have $\partial\varepsilon_{t-1}/\partial\theta_1 = -\partial^2\varepsilon_t/\partial\theta_1^2 - \partial\varepsilon_{t-1}/\partial\theta_1 - \theta_1\partial^2\varepsilon_{t-1}/\partial\theta_1^2 - \theta_2\partial^2\varepsilon_{t-2}/\partial\theta_1^2$. Rewrite it as $\partial\varepsilon_{t-1}/\partial\theta_1 = -\frac{1}{2}\left(\partial^2\varepsilon_t/\partial\theta_1^2 + \theta_1\partial^2\varepsilon_{t-1}/\partial\theta_1^2 + \theta_2\partial^2\varepsilon_{t-2}/\partial\theta_1^2\right)$. Then we can find some matrix C_2 such that $\partial^2\varepsilon/\partial\theta_1^2 = C_2\partial\varepsilon/\partial\theta_1 = C_2C_1\varepsilon$. Carrying on this step we can get $\partial^3\varepsilon/\partial\theta_1^3 = C_3\partial^2\varepsilon/\partial\theta_1^2 = C_3C_2C_1\varepsilon$, and $\partial^4\varepsilon/\partial\theta_1^4 = C_4\partial^3\varepsilon/\partial\theta_1^3 = C_4C_3C_2C_1\varepsilon$. Therefore, all the expectations of cross products of $E\left(\partial^4\varepsilon/\partial\theta_i\partial\theta_j\partial\theta_k\partial\theta_l\right)$ will essentially take some quadratic form in the normal vector ε. As a result, the standard procedure in Section 3.1 is applicable here.

Remark 5: We can easily extend to the case of an MA(q) with mean $\mu_y = x_t'\beta$. If we combine with Section 3.1, we can generally evaluate a general ARMA process with mean $\mu_y = x_t'\beta$. In essence, we can express the relevant expectations of quadratic forms in the normal vector $(y', \varepsilon')'$.

Remark 6: We restrict $|\beta| < 1$ because our exercise is for the conditional ML estimator. It would be more interesting to extend the exercise to unconditional ML estimator, regardless of whether β is associated with an invertible representation or not. The difficulty of this exercise is that there is no appropriate way to handle the expectations of the score function and its derivatives up to third order. We defer this to our future research.

3.5 Partial Adjustment Model

Let $y_t^* = x_t\beta + \varepsilon_t$ be the desired level of some economic variable. The partial adjustment model describes an adjustment equation $y_t - y_{t-1} = (1-\gamma)\left(y_t^* - y_{t-1}\right)$. See Brown (1952) and Lovell (1961) for consumption model of habit persistence. The econometric regression equation is obtained by substituting the first equation into the adjustment equation and solving for y_t

$$y_t = \gamma y_{t-1} + (1-\gamma)\beta x_t + \varepsilon_t = \gamma y_{t-1} + \delta x_t + \varepsilon_t, \tag{3.41}$$

where $\delta = \beta(1-\gamma)$. For simplicity we assume x_t to be an i.i.d. normal scalar and $\varepsilon_t \sim N(0, \sigma_\varepsilon^2)$. Suppose $E(x_t) = \mu_x$, $V(x_t) = \sigma_x^2$, and x is uncorrelated with ε. Note that even though γ and δ can be estimated consistently and efficiently by OLS as this model is intrinsically linear, the inference about $\beta = \delta/(1-\gamma)$ is not so straightforward. Here we try to derive $B\left(\hat{\beta}\right)$ and $M\left(\hat{\beta}\right)$.

The OLS estimator needs the following moment condition

$$\psi_T = \frac{1}{T-1}\left(\begin{array}{c} \sum_{t=2}^{T}\varepsilon_t y_{t-1} \\ \sum_{t=2}^{T}\varepsilon_t x_t \end{array}\right) = \frac{1}{T-1}\left(\begin{array}{c} y'Cy-(1-\gamma)\beta x'Ay \\ x'Dy-(1-\gamma)\beta x'Bx \end{array}\right) = 0, \tag{3.42}$$

where A is $T\times T$ with tt'th element equal to 1 if $t-t'=1$ and 0 elsewhere, B is $T\times T$ with tt'th element equal to 1 if $t=t'\geq 2$ and 0 elsewhere, C is $T\times T$ with tt'th element equal to $-\gamma$ if $t=t'\leq T-1$, $1/2$ if $|t-t'|=1$, and 0 elsewhere, for $t,t'=1,2,\cdots,T$, and $D=B-\gamma A$.

Then we have

$$\begin{aligned}
\nabla\psi_T &= H_1 = \frac{1}{T-1}\left(\begin{array}{cc} -(1-\gamma)x'Ay & -y'C_1y+\beta x'Ay \\ -(1-\gamma)x'Bx & -x'Ay+\beta x'Bx \end{array}\right), \\
\nabla^2\psi_T &= H_2 = \frac{1}{T-1}\left(\begin{array}{cccc} 0 & x'Ay & x'Ay & 0 \\ 0 & x'Bx & x'Bx & 0 \end{array}\right), \\
\nabla^3\psi_T &= H_3 = 0.
\end{aligned} \tag{3.43}$$

where C_1 is $T\times T$ with tt'th element equal to 1 if $t=t'=1,2,\cdots,T-1$, and 0 elsewhere

Therefore, we can follow the same procedure as in Section 3.1. The only difference is that we will encounter some bilinear forms like $x'Ay$. For this we can define $y^*=(x',y')'$, then (3.42) and (3.43) can be rewritten as

$$\psi_T = \frac{1}{T-1}\left(\begin{array}{c} y^{*\prime}C^*y^*-(1-\gamma)\beta y^{*\prime}A^*y^* \\ y^{*\prime}D^*y^*-(1-\gamma)\beta y^{*\prime}B^*y^* \end{array}\right) = 0, \tag{3.42$'$}$$

and

$$\begin{aligned}
\nabla\psi_T &= H_1 = \frac{1}{T-1}\left(\begin{array}{cc} -(1-\gamma)y^{*\prime}A^*y^* & -y^{*\prime}C_1^*y^*+\beta y^{*\prime}A^*y^* \\ -(1-\gamma)y^{*\prime}B^*y^* & -y^{*\prime}A^*y^*+\beta y^{*\prime}B^*y^* \end{array}\right), \\
\nabla^2\psi_T &= H_2 = \frac{1}{T-1}\left(\begin{array}{cccc} 0 & y^{*\prime}A^*y^* & y^{*\prime}A^*y^* & 0 \\ 0 & y^{*\prime}B^*y^* & y^{*\prime}B^*y^* & 0 \end{array}\right),
\end{aligned} \tag{3.43$'$}$$

where

$$\begin{aligned}
A^* &= \frac{1}{2}\left(\begin{array}{cc} 0 & A \\ A & 0 \end{array}\right),\ B^* = \left(\begin{array}{cc} B & 0 \\ 0 & 0 \end{array}\right),\ C^* = \left(\begin{array}{cc} 0 & 0 \\ 0 & C \end{array}\right), \\
C_1^* &= \left(\begin{array}{cc} 0 & 0 \\ 0 & C_1 \end{array}\right),\ D^* = \frac{1}{2}\left(\begin{array}{cc} 0 & D \\ D & 0 \end{array}\right),
\end{aligned} \tag{3.44}$$

and $y^* \sim N(\mu, \Sigma)$, where

$$\mu = \begin{pmatrix} \mu_x \tau_T \\ \beta \mu_x \tau_T \end{pmatrix}, \ \Sigma = \begin{pmatrix} \sigma_x^2 I_T & \Sigma_{xy} \\ \Sigma_{xy} & \Sigma_y \end{pmatrix}, \tag{3.45}$$

where τ_T is a $T \times 1$ vector of ones, Σ_y is the $T \times T$ variance-covariance matrix of y with tt'th element equal to $\gamma^{|t-t'|}\left[\delta^2\left(\mu_x^2+\sigma_x^2\right)+\sigma_\varepsilon^2\right]/\left(1-\gamma^2\right)-\beta^2\mu_x^2$, Σ_{xy} is the $T \times T$ covariance matrix between x and y with tt'th element equal to $\delta\gamma^{t'-t}\left(\sigma_x^2+\mu_x^2\right)-\delta\mu_x^2\gamma^{j-i}/\left(1-\gamma\right)$ if $t' \geq t$ and 0 elsewhere, for $t,\ t' = 1, 2, \cdots, T$. Denote $trA^*\Sigma = a$, $trB^*\Sigma = b$, $trC_1^*\Sigma = c$, $\mu' A^* \mu = k$, $\mu' B^* \mu = m$, $\mu' C_1^* \mu = l$. With these notations, we have Theorem 5 for the second-order bias of $\hat{\beta}$.

Theorem 5: The bias, up to $O\left(T^{-1}\right)$, of the estimator $\hat{\beta}$ from (3.42) for model (3.41) is

$$B\left(\hat{\beta}\right) = \frac{\lambda_{12} - \lambda_{34} + \lambda_{56}}{\left(\gamma-1\right)^2\left[\left(a+k\right)^2 - \left(c+l\right)\left(b+m\right)\right]^2}, \tag{3.46}$$

where $\lambda_{ij} = E\left(y^{*\prime} N_i y^* \cdot y^{*\prime} N_j y^*\right),\ i,\ j = 1, 2, \cdots, 6,$

$N_1 = (\gamma-1)C^* + \beta(\gamma-1)^2 A^*,$

$N_2 = \left(a+k-2\beta b\right)\left(mC_1^* - kA^*\right) + \left(c+l\right)\left[\left(a+k\right)B^* - \left(m+b\right)A^*\right] - \beta(b^2+m^2)C_1^*$
$\quad +\beta k(2mA^* - kB^* - 2aB^*) + k(bC_1^* - aA^*) - a^2(A^* + \beta B^*) + 2a\beta(b+m)A^* + abC_1^*,$

$N_3 = (\gamma-1)D^* + \beta(\gamma-1)^2 B^*,$

$N_4 = (c+l)(2kA^* + \beta kB^* + 2aA^* + a\beta B^*) - (c+l)^2 B^* - (k+a)^2(\beta A^* + C_1^*)$
$\quad +\beta(aC_1^* + kC_1^* - lA^* - cA^*)(b+m),$

$N_5 = (b+m)C^* + \beta(\gamma-1)\left[(b+m)A^* - (a+k)B^*\right] - (a+l)D^*,$

$N_6 = \beta(C^* - \beta A^*)(b+m) + \beta B^*(\beta k - l) - (a+k)(C^* + \beta D^*) - \beta(aA^* + a\beta B^* - cB^*)(\gamma-1)$
$\quad +(c+l)D^* + \beta\gamma(lB^* - kA^*) + \beta^2\gamma(bA^* + mA^* - kB^*).$

Proof: Substitute (3.42′) and (3.43′) into (2.5) and use the results on quadratic forms in Appendix A.1. ■

As for the bias of γ, it is clear that we can use the results in Section 3.1 (including the exogenous variable x_t) instead. We do not give explicitly the expression for the second-order MSE. However, again it is easy to write a computer program to do the calculations numerically.

3.6 Absolute Regression Model

Consider the "absolute" regression model

$$y_t = -\beta \left| y_{t-1} \right| + \varepsilon_t, \tag{3.47}$$

where $0 < \beta < 1$, ε_t is i.i.d. $N(0,1)$ and $t = 1, 2, ..., T$. Note that (3.47) is a very special case of the self-exciting autoregressive (SETAR) model of Tong (1990) and it is in nature a nonlinear regression model. Tong (1990, p. 141) showed that the density function of y_t is

$$f(y) = \sqrt{\frac{2\left(1-\beta^2\right)}{\pi}} \exp\left[-\frac{1}{2}\left(1-\beta^2\right) y^2\right] \Phi(-\beta y), \tag{3.48}$$

where $\Phi(\cdot)$ is the cumulative distribution function of a standard normal variable. Andel *et al.* (1984) showed that the nth moments (about 0) of y_t are

$$m_n = \beta^{-n-1} \left[2\left(1-\beta^2\right)\pi\right]^{1/2} J_n, \tag{3.49}$$

where $J_n = \int_{-\infty}^{+\infty} x^n e^{-kx^2/2} \Phi(-x)\, dx$, $k = \beta^{-2} - 1$ (also see Tong, 1990, p. 209). In particular, we have

$$\begin{aligned} m_1 &= -\beta\sqrt{\frac{2}{\pi\left(1-\beta^2\right)}},\ m_2 = \frac{1}{1-\beta^2}, \\ m_3 &= \frac{\beta\left(\beta^2-3\right)}{1-\beta^2}\sqrt{\frac{2}{\pi\left(1-\beta^2\right)}},\ m_4 = \frac{3}{\left(1-\beta^2\right)^2}, \\ m_5 &= \frac{\beta\left(15-10\beta^2+3\beta^4\right)}{\left(1-\beta^2\right)^2}\sqrt{\frac{2}{\pi\left(1-\beta^2\right)}},\ m_6 = \frac{15}{\left(1-\beta^2\right)^3}. \end{aligned} \tag{3.50}$$

Note that for an AR(1) model (3.1), the even moments of y_t are the same as those for model (3.47) while all the odd moments are zero.

In most applications, β is estimated by LS. That is, the moment condition is

$$\psi_T = \frac{1}{T-1}\sum_{t=2}^{T} \frac{\partial \varepsilon_t}{\partial \beta}\varepsilon_t = \frac{1}{T-1}\sum_{t=2}^{T} \left|y_{t-1}\right| \left(y_t + \beta\left|y_{t-1}\right|\right) = 0. \tag{3.51}$$

Then following the notations as before, we have

$$H_1 = \frac{1}{T-1} \sum_{t=2}^{T} |y_{t-1}|^2, \ H_2 = H_3 = W = 0. \tag{3.52}$$

Further, $Q = 1/m_2$. Since we have a scalar case and higher derivatives of the moment condition are all zero, this will simplify our results a lot. From (2.5), the bias expression reduces to

$$B\left(\hat{\beta}\right) = \frac{Q^2}{(T-1)^2} E\left[\sum_{t=2}^{T} |y_{t-1}|^2 \sum_{t'=2}^{T} |y_{t'-1}| \left(y_{t'} + \beta |y_{t'-1}|\right)\right]. \tag{3.53}$$

Note that y here is no longer a normal vector so we can not use the expectation results in Appendix A.1. But in Appendix A.4, we prove the following

$$E\left[\sum_{t=2}^{T} |y_{t-1}|^2 \sum_{t'=2}^{T} |y_{t'-1}| \left(y_{t'} + \beta |y_{t'-1}|\right)\right] = -\frac{2m_2}{\beta} \sum_{\substack{t=2 \\ t>t'}}^{T} \sum_{t'=2}^{T} \beta^{2(t-t')}. \tag{3.54}$$

It is easy to verify that

$$\sum_{\substack{t=2 \\ t>t'}}^{T} \sum_{t'=2}^{T} \left[\beta^{2(t-t')}\right] = \left[\frac{\beta^2 (T-1)}{1-\beta^2} - \frac{\beta^2}{1-\beta^2} \frac{1-\beta^{2(T-1)}}{1-\beta^2}.\right] \tag{3.55}$$

By substitution, we have $B\left(\hat{\beta}\right) = -2\beta/(T-1)$, which is the same bias derived for an AR(1) model. Further, from (2.6), (3.51), and (3.52), we have the MSE expression

$$M\left(\hat{\beta}\right) = 6Q^2 \overline{\psi_T^2} - 8Q^3 \overline{H_1 \psi_T^2} + 3Q^4 \overline{H_1^2 \psi_T^2}, \tag{3.56}$$

where in Appendix A.4 we discuss how to evaluate $\overline{\psi_T^2}$, $\overline{H_1 \psi_T^2}$, and $\overline{H_1^2 \psi_T^2}$, and we show that $M\left(\hat{\beta}\right) = M(\beta,\ m_2,\ m_4,\ m_6)$. In addition, if we apply (2.6) to model (3.1) instead of using the quadratic form, the same steps will follow exactly as for model (3.47). That is, we will arrive at the same MSE for model (3.1) and (3.47). In summary, we have the following theorem.

Theorem 6: The bias and MSE, up to $O(T^{-1})$ and $O(T^{-2})$ respectively, of the LS estimator $\hat{\beta}$ for the absolute regression model (3.47) are the same as the bias and MSE of the LS estimator $\hat{\beta}$ for the AR(1) model (3.1).

The intuition here is that the nonlinearity imposed on the original AR(1) model will distort only the odd moments of the process and preserve all the even moments, but all the second-order bias and MSE results under LS estimation take the same functional form and involve only the even momets, and hence we have the same second-order bias and MSE. Of course, equality of the first two moments of $\hat{\beta}$ does not suggest the same distribution of $\hat{\beta}$ for model (3.47) and (3.1).

4 Conclusions

We have developed the analytical results on the properties of estimators in time series framework. General results on the second-order bias and MSE are given. The applications of these results to a wide variety of econometric models are also analyzed. We indicate that our general results are valid for both normal and non-normal observations. It would be desirable if we could approximate the distributions of the estimators since we may often need to know about the skewness and kurtosis of the estimators, construct confidence intervals, and investigate the power or size of tests. This will be the subject of a future study.

Appendix

A.1. Expectations of Quadratic Forms in a Normal Vector

Let trA be trace of any matrix A. For any symmetric matrix N_i, Magnus (1978, 1979), among others, derived the following results on the expectations of products of quadratic forms in a normal vector $y \sim N(0, I)$,

$$E(y'N_1y) = trN_1,$$

$$E(y'N_1y \cdot y'N_2y) = (trN_1)(trN_2) + 2tr(N_1N_2),$$

$$\begin{aligned}E(y'N_1y \cdot y'N_2y \cdot y'N_3y) &= (trN_1)(trN_2)(trN_3) + 8trN_1N_2N_3 \\ &\quad +2[(trN_1)(trN_2N_3) + (trN_2)(trN_1N_3) + (trN_3)(trN_1N_2)],\end{aligned}$$

$$\begin{aligned}E(y'N_1y \cdot y'N_2y \cdot y'N_3y \cdot y'N_4y) &= (trN_1)(trN_2)(trN_3)(trN_4) \\ &\quad +8[(trN_1)(trN_2N_3N_4) + (trN_2)(trN_1N_3N_4) + (trN_3)(trN_1N_2N_4) + (trN_4)(trN_1N_2N_3)] \\ &\quad +4[(trN_1N_2)(trN_3N_4) + (trN_1N_3)(trN_2N_4) + (trN_1N_4)(trN_2N_3)] \\ &\quad +2[(trN_1)(trN_2)(trN_3N_4) + (trN_1)(trN_3)(trN_2N_4) + (trN_1)(trN_4)(trN_2N_3) \\ &\quad + (trN_2)(trN_3)(trN_1N_4) + (trN_2)(trN_4)(trN_1N_3) + (trN_3)(trN_4)(trN_1N_2)] \\ &\quad +16[trN_1N_2N_3N_4 + trN_1N_2N_4N_3 + trN_1N_3N_2N_4].\end{aligned}$$

When $y \sim N(0, \Sigma)$, we replace N_i with $N_i\Sigma$ in the above formulae. Also we note that for a general matrix N, $y'Ny = y'\left(\frac{N+N'}{2}\right)y$ where $\frac{N+N'}{2}$ is always symmetric. In the following we assume that the matrices involved are symmetric.

When $y \sim N(\mu, \Sigma)$, we can write $E(y'N_1y) = E\left[(y-\mu+\mu)' N_1 (y-\mu+\mu)\right] = \mu'N_1\mu + tr(N_1\Sigma)$. Similarly, if we define $a_i = tr(N_i\Sigma)$, $a_{ij} = tr(N_i\Sigma N_j\Sigma)$, $a_{ijk} = tr(N_i\Sigma N_j\Sigma N_k\Sigma)$, $a_{ijkl} = tr(N_i\Sigma N_j\Sigma N_k\Sigma N_l\Sigma)$, $\theta_i = \mu'N_i\mu$, $\theta_{ij} = \mu'N_i\Sigma N_j\mu$, $\theta_{ijk} = \mu'N_i\Sigma N_j\Sigma N_k\mu$, $\theta_{ijkl} = \mu'N_i\Sigma N_j\Sigma N_k\Sigma N_l\mu$, then we can conveniently write

$$E(y'N_1y) = a_1 + \theta_1, \tag{A.1}$$

$$E(y'N_1y \cdot y'N_2y) = \prod_{i=1}^{2}(a_i + \theta_i) + \sum_{i=1}^{2}\sum_{\substack{j=1 \\ i\neq j}}^{2}(a_{ij} + 2\theta_{ij}), \tag{A.2}$$

$$\begin{aligned}E[y'N_1y \cdot y'N_2y \cdot y'N_3y] &= \prod_{i=1}^{3}(a_i + \theta_i) \\ &\quad + \sum_{i=1}^{3}\sum_{\substack{j=1 \\ i\neq j\neq k}}^{3}\sum_{k=1}^{3}[a_ia_{jk} + a_{ij}\theta_k + \tfrac{2^3}{3!}a_{ijk} + 4\theta_{ikl} + 2(a_i + \theta_i)\theta_{jk}],\end{aligned} \tag{A.3}$$

$$E\left(y'N_1y\cdot y'N_2y\cdot y'N_3y\cdot y'N_4y\right)=\prod_{i=1}^{4}\left(a_i+\theta_i\right)$$
$$+\sum_{i=1}^{4}\sum_{\substack{j=1\\ i\neq j\neq k\neq l}}^{4}\sum_{k=1}^{4}\sum_{l=1}^{4}[\left(2a_ia_j+2a_i\theta_j+2a_{ij}+\theta_i\theta_j+4\theta_{ij}\right)\theta_{kl}$$
$$+\left(\tfrac{1}{2}a_{ij}\theta_k+a_ia_{jk}+a_{ijk}\right)\theta_l+a_{kl}\left(a_ia_j+a_{ij}\right)+\tfrac{2^4}{4!}a_{ijkl}$$
$$+8\theta_{ijkl}+\tfrac{2^3}{3!}\left(\theta_i+\tfrac{1}{2}a_i\right)\theta_{jkl}+\tfrac{2^3}{3!}\left(a_{jkl}+\tfrac{1}{2}\theta_{jkl}\right)a_i]. \quad \text{(A.4)}$$

The above results also follow from Ullah (1990, 2002) and Mathai and Provost (1992), where a nonstochatic operator d and the moment generating function are used respectively to derive them.

A.2. AR(1) Model

First note that $tr\left(C_1\Sigma\right)=\left(T-1\right)\sigma^2/\left(1-\beta^2\right)$ and $tr\left(C\Sigma\right)=0$. Then

$$\lambda_{11}=E\left[\left(y'Cy\right)\cdot\left(y'C_1y\right)\right]=2tr\left(C_1\Sigma C\Sigma\right). \quad \text{(A.5)}$$

Also, we write

$$\begin{aligned} C_1 &= \begin{pmatrix} -I_{T-1} & 0 \\ 0 & 0 \end{pmatrix}, \\ C &= -\beta\begin{pmatrix} I_{T-1} & 0 \\ 0 & 0 \end{pmatrix}+\frac{1}{2}\begin{pmatrix} 0 & I_{T-1} \\ 0 & 0 \end{pmatrix}+\frac{1}{2}\begin{pmatrix} 0 & 0 \\ I_{T-1} & 0 \end{pmatrix}, \end{aligned} \quad \text{(A.6)}$$

and partition

$$\Sigma=\frac{\sigma_\varepsilon^2}{1-\beta^2}\begin{pmatrix} \Omega_1 \\ \Sigma_T^* \end{pmatrix}=\frac{\sigma_\varepsilon^2}{1-\beta^2}\begin{pmatrix} \Sigma_1^* \\ \Omega_2 \end{pmatrix}, \quad \text{(A.7)}$$

where Σ_i^* is the ith row of $\Sigma^*=\Sigma/\frac{\sigma_\varepsilon^2}{1-\beta^2}$, Ω_1 is the 1st to $(T-1)$th rows of Σ^* and Ω_2 is the 2nd to Tth rows of Σ^*, then

$$\begin{aligned} B\left(\hat{\beta}\right) &= \frac{2Q^2tr\left(C_1\Sigma C\Sigma\right)}{\left(T-1\right)^2} \\ &= \frac{2}{\left(T-1\right)^2}\left\{\beta tr\left[\begin{pmatrix} \Omega_1 \\ 0 \end{pmatrix}\begin{pmatrix} \Omega_1 \\ 0 \end{pmatrix}\right]-2tr\left[\begin{pmatrix} \Omega_2 \\ 0 \end{pmatrix}\begin{pmatrix} 0 \\ \Omega_1 \end{pmatrix}\right]\right\} \\ &= \frac{2}{\left(T-1\right)^2}\left\{\beta\left(T-1\right)-2\beta\left(T-1\right)\right\}+o\left(T^{-1}\right) \\ &= \frac{-2\beta}{\left(T-1\right)}+o\left(T^{-1}\right). \end{aligned} \quad \text{(A.8)}$$

Next, by using the results in Appendix A.1, it is straightforward to verify

$$\begin{aligned}
\lambda_{20} &= E\left[(y'Cy)^2\right] = 2tr\left[(C\Sigma)^2\right], \\
\lambda_{21} &= E\left[(y'C_1y)\cdot(y'Cy)^2\right] \\
&= 8tr\left[C_1\Sigma\,(C\Sigma)^2\right] + 2tr(C_1\Sigma)tr\left[(C\Sigma)^2\right], \\
\lambda_{22} &= E\left[(y'C_1y)^2\cdot(y'Cy)^2\right] \\
&= 16tr\,(C_1\Sigma)\,tr\left[C_1\Sigma\,(C\Sigma)^2\right] + 4tr\left[(C_1\Sigma)^2\right]tr\left[(C\Sigma)^2\right] \\
&\quad +8\left[tr\,(C_1\Sigma C\Sigma)\right]^2 + 2\left[tr\,(C_1\Sigma)\right]^2 tr\left[(C\Sigma)^2\right] \\
&\quad +32tr\left[(C_1\Sigma)^2(C\Sigma)^2\right] + 16tr\left[(C_1\Sigma C\Sigma)^2\right].
\end{aligned} \tag{A.9}$$

A.3. Simultaneous Equation Model

From (2.5) we have

$$\begin{aligned}
B\left(\hat{\beta}\right) &= Q\overline{Vd} = Q\overline{\left(H_1 - Q^{-1}\right)Q\psi_T} = Q^2\overline{H_1\psi_T} \\
&= \frac{Q^2}{T^2}E\left(x'C_1y\cdot x'Cy\right) = \frac{Q^2}{T^2}x'C\Omega_yC_1'x \\
&= -\frac{\omega_{u2}x'\Sigma x}{\pi^2\left(x'x\right)^2}.
\end{aligned} \tag{A.10}$$

Similarly

$$\begin{aligned}
A_{-1} &= Q^2\overline{\psi_T^2} = \frac{Q^2}{T^2}x'C\Omega_yC'x \\
&= \frac{\sigma_u^2}{1-\rho^2}\frac{x'\Sigma x}{\pi^2\left(x'x\right)^2}.
\end{aligned} \tag{A.11}$$

By symmetry, it is straightforward to prove that $A_{-3/2} = 0$. Furthermore, we can write $A_{-2} = \frac{3Q^4}{T^4}E\left[(x'C_1y)^2(x'Cy)^2\right] - 3A_{-1}$. Note that $E\left[(x'C_1y)^2(x'Cy)^2\right] = E\left[y'N_1y\cdot y'N_2y\right]$, where $N_1 = C_1'xx'C_1$, $N_2 = C'xx'C$, and both are symmetric. Then we can use the results on quadratic forms to derive the expectation. Specifically,

$$\begin{aligned}
tr\left(N_1\Omega_y\right) &= tr\left(xx'\Omega_{22}\right) = \omega_{22}x'x, \\
tr\left(N_2\Omega_y\right) &= tr\left[xx'\left(\Omega_{11} - 2\beta\Omega_{12} + \beta^2\Omega_{22}\right)\right] = \frac{\sigma_u^2}{1-\rho^2}x'\Sigma x, \\
tr\left(N_1\Omega_y N_2\Omega_y\right) &= tr\left[xx'\left(\Omega_{21} - \beta\Omega_{22}\right)xx'\left(\Omega_{12} - \beta\Omega_{22}\right)\right] = \omega_{u2}^2\left(x'\Sigma x\right)^2, \\
\mu_y' N_1 \mu_y &= \pi^2\left(x'x\right)^2, \\
\mu_y' N_2 \mu_y &= 0, \\
\mu_y' N_1 \Omega_y N_2 \mu_y &= 0. \qquad\qquad (A.12)
\end{aligned}$$

Therefore, by substitution we have

$$\begin{aligned}
A_{-2} &= \frac{3Q^4}{T^4}\left(x'\Sigma x\right)\left[\frac{\omega_{22}\sigma_u^2}{1-\rho^2}\left(x'x\right) + 2\omega_{u2}^2\left(x'\Sigma x\right) + \frac{\pi^2\sigma_u^2}{1-\rho^2}\left(x'x\right)^2\right] - 3\frac{\sigma_u^2}{1-\rho^2}\frac{x'\Sigma x}{\pi^2\left(x'x\right)^2} \\
&= \frac{6\omega_{u2}^2\left(x'\Sigma x\right)^2}{\pi^4\left(x'x\right)^4} + \frac{3\sigma_u^2\omega_{22}}{1-\rho^2}\frac{\left(x'\Sigma x\right)\left(x'x\right)}{\pi^4\left(x'x\right)^4}. \qquad\qquad (A.13)
\end{aligned}$$

Then the expression for $M\left(\hat{\beta}\right)$ follows immediately.

When x is stochastic, we have

$$\begin{aligned}
B\left(\hat{\beta}\right) &= \frac{Q^2}{T^2}[tr\left(C_1\Omega_y C\Omega_{x^*}\right) + \mu_{x^*}' C_1\Omega_y C\mu_{x^*} + \mu_y' C_1\Omega_{x^*} C\mu_y \\
&\quad + \mu_{x^*}' C_1\Omega_{yx^*} C\mu_y + \mu_y' C_1\Omega_{x^*y} C\mu_{x^*}] \\
&= \frac{Q^2}{T^2}\left(-\omega_{u2} tr\left(\Sigma\Sigma_x\right) - \omega_{u2}\mu_x'\Sigma\mu_x\right) \\
&= \frac{\left(-\omega_{u2} tr\left(\Sigma\Sigma_x\right) - \omega_{u2}\mu_x'\Sigma\mu_x\right)}{\pi^2\left[\mu_x'\mu_x + tr\left(\Sigma_x\right)\right]^2}, \qquad\qquad (A.14)
\end{aligned}$$

and

$$\begin{aligned}
A_{-1} &= \frac{Q^2}{T^2}\lambda_{10} = \frac{Q^2}{T^2}E\left(Z'C^*Z \cdot Z'C^*C\right) \\
&= \frac{Q^2}{T^2}\left[2tr\left(C^*\Sigma_Z C^*\Sigma_Z\right) + 4\mu_z' C^*\Sigma_Z C^*\mu_Z\right] = \frac{Q^2}{T^2}\frac{\sigma_u^2}{1-\rho^2}\left[tr\left(\Sigma\Sigma_x\right) + \mu_x'\Sigma\mu_x\right] \\
&= \frac{\sigma_u^2}{1-\rho^2}\frac{tr\left(\Sigma\Sigma_x\right) + \mu_x'\Sigma\mu_x}{\pi^2\left[\mu_x'\mu_x + tr\left(\Sigma_x\right)\right]^2}, \\
A_{-3/2} &= 0, \\
A_{-2} &= 3\frac{Q^4}{T^4}\lambda_{11} - 3\frac{Q^2}{T^2}\lambda_{10} \\
&= 3\frac{Q^4}{T^4}E\left(Z'C_1^*Z \cdot Z'C_1^*Z \cdot Z'C^*Z \cdot Z'C^*Z\right) - 3A_{-1}. \qquad\qquad (A.15)
\end{aligned}$$

The exact expression for λ_{11} can be derived using the results on quadratic forms. Note that if x is nonstochastic $(\Sigma_x = 0)$, then (A.14) or (A.15) will degenerate to (A.10) or (A.11) and (A.13).

A.4. Absoulte Regression Model

First note that

$$E\left[|y_{t-1}|^2 |y_{t'-1}| (y_{t'} + \beta |y_{t'-1}|)\right] = E\left(|y_{t-1}|^2 |y_{t'-1}| \varepsilon_{t'}\right) := U. \tag{A.16}$$

If $t \le t'$, $U = 0$. If $t > t'$, U can also be derived recursively from

$$\begin{aligned} E\left(|y_{t-1}|^2 |y_{t'-1}| \varepsilon_{t'}\right) &= \beta^2 E\left(|y_{t-2}|^2 |y_{t'-1}| \varepsilon_{t'}\right) \text{ for } t > t'+1, \text{ and} \\ E\left(|y_{t'}|^2 |y_{t'-1}| \varepsilon_{t'}\right) &= -2m_2\beta. \end{aligned} \tag{A.17}$$

Hence for $t > t'$

$$U = -2m_2\beta^{2(t-t')-1}. \tag{A.18}$$

Next, we try to derive the three terms for the MSE:

$$\begin{aligned} E\left(\psi_T^2\right) &= \frac{1}{T-1} E\left\{\left[|y_{t-1}| (y_t + \beta |y_{t-1}|)\right]^2\right\} = \frac{m_2}{T-1}, \\ E\left(H_1\psi_T^2\right) &= \frac{1}{(T-1)^3} E\left\{\left(\sum_{t=2}^{T} |y_{t-1}|^2\right)\left[\sum_{t'=2}^{T} |y_{t'-1}| (y_t' + \beta |y_{t'-1}|)\right]^2\right\}, \\ E\left(H_1^2\psi_T^2\right) &= \frac{1}{(T-1)^4} E\left\{\left[\sum_{t=2}^{T} |y_{t-1}|^2\right]^2 \left[\sum_{t'=2}^{T} |y_{t'-1}| (y_t' + \beta |y_{t'-1}|)\right]^2\right\}. \end{aligned} \tag{A.19}$$

Note that for the second term $E\left(|y_{t-1}|^2 |y_{t'-1}| \varepsilon_{t'} |y_{t''-1}| \varepsilon_{t''}\right) := V_{tt't''} \neq 0$ only if (the subscript is suppressed for V in the following)

i) $t' = t'' > t$, $V = E\left(|y_{t-1}|^2 |y_{t'-1}|^2 \varepsilon_{t'}^2\right) = E\left(|y_{t-1}|^2 |y_{t'-1}|^2\right)$ can be derived recursively from

$$\begin{aligned} E\left(|y_{t-1}|^2 |y_{t'-1}|^2\right) &= \beta^2 E\left(|y_{t-1}|^2 |y_{t'-2}|^2\right) + m_2, \text{ for } t' > t+1, \text{ and} \\ E\left(|y_{t-1}|^2 |y_t|^2\right) &= \beta^2 m_4 + m_2. \end{aligned} \tag{A.20}$$

Hence for $t' = t'' > t$

$$V = \beta^{2(t'-t)} m_4 + m_2 \frac{1-\beta^{2(t'-t)}}{1-\beta^2}. \tag{A.21}$$

ii) $t = t' = t''$, $V = m_4$.

iii) $t > t' = t''$, $V = E\left(|y_{t-1}|^2 |y_{t'-1}|^2 \varepsilon_{t'}^2\right)$ can be derived recursively from

$$\begin{aligned} E\left(|y_{t-1}|^2 |y_{t'-1}|^2 \varepsilon_{t'}^2\right) &= \beta^2 E\left(|y_{t-2}|^2 |y_{t'-1}|^2 \varepsilon_{t'}^2\right) + m_2 \text{ for } t > t'+1, \text{ and} \\ E\left(|y_{t'}|^2 |y_{t'-1}|^2 \varepsilon_{t'}^2\right) &= \beta^2 m_4 + 3m_2. \end{aligned} \tag{A.22}$$

Hence for $t > t' = t''$

$$V = \beta^{2(t-t')} m_4 + m_2 \frac{1-\beta^{2(t-t'-1)}}{1-\beta^2} + 3m_2. \tag{A.23}$$

iv) $t > t' > t''$, V can be derived recursively from

$$\begin{aligned} E\left(|y_{t-1}|^2 |y_{t'-1}| \varepsilon_{t'} |y_{t''-1}| \varepsilon_{t''}\right) &= \beta^2 E\left(|y_{t-2}|^2 |y_{t'-1}| \varepsilon_{t'} |y_{t''-1}| \varepsilon_{t''}\right) \text{ for } t > t'+1, \text{ and} \\ E\left(|y_{t'}|^2 |y_{t'-1}| \varepsilon_{t'} |y_{t''-1}| \varepsilon_{t''}\right) &= -2\beta E\left(|y_{t'-1}|^2 |y_{t''-1}| \varepsilon_{t''}\right) \\ &= 4m_2 \beta^{2(t'-t'')} \text{ (substitute (A.18))}. \end{aligned} \tag{A.24}$$

Hence for $t > t' > t''$

$$V = 4m_2 \beta^{2(t-t''-1)}. \tag{A.25}$$

Also, for the third term $E\left(|y_{t-1}|^2 |y_{t'-1}|^2 |y_{t''-1}| \varepsilon_{t''} |y_{t'''-1}| \varepsilon_{t'''}\right) := W_{tt't''t'''} \neq 0$ only if (the subscript is suppressed for W in the following)

i) $t = t' = t'' = t'''$, $W = m_6$

ii) $t = t'' = t''' > t'$ (or $t' = t'' = t''' > t$ due to symmetry; the result follows by exchanging the index t' with t), $W = E\left(|y_{t-1}|^4 \varepsilon_t^2 |y_{t'-1}|^2\right) = E\left(|y_{t-1}|^4 |y_{t'-1}|^2\right)$ can be derived recursively from

$$\begin{aligned} E\left(|y_{t-1}|^4 |y_{t'-1}|^2\right) &= \beta^4 E\left(|y_{t-2}|^4 |y_{t'-1}|^2\right) + 6\beta^2 E\left(|y_{t-2}|^2 |y_{t'-1}|^2\right) + 3m_2 \\ &= \beta^4 E\left(|y_{t-2}|^4 |y_{t'-1}|^2\right) + 6\left[\beta^{2(t-t')} m_4 + m_2 \frac{1-\beta^{2(t-t')}}{1-\beta^2}\right] + 3m_2 \\ \text{for } t &> t'+1 \text{ (substitute (A.21))}, \text{ and} \\ E\left(|y_{t'}|^4 |y_{t'-1}|^2\right) &= \beta^4 m_6 + 6\beta^2 m_4 + 3m_2. \end{aligned} \tag{A.26}$$

Hence for $t = t'' = t''' > t'$

$$\begin{aligned} W &= \beta^{4(t-t')}m_6 + 6\left[\beta^{2(t-t')}m_4 + m_2\frac{1-\beta^{2(t-t')}}{1-\beta^2}\right]\frac{1-\beta^{2(t-t'-1)}}{1-\beta^2} \\ &\quad +6\beta^2 m_4 + 3m_2\frac{1-\beta^{4(t-t')}}{1-\beta^4} \end{aligned} \tag{A.27}$$

iii) $t = t' > t'' = t'''$, $W = E\left(|y_{t-1}|^4 |y_{t''-1}|^2 \varepsilon_{t''}^2\right)$ can be derived recursively from

$$\begin{aligned} E\left(|y_{t-1}|^4 |y_{t''-1}|^2 \varepsilon_{t''}^2\right) &= \beta^4 E\left(|y_{t-2}|^4 |y_{t''-1}|^2 \varepsilon_{t''}^2\right) + 3m_2 + 6\beta^2 E\left(|y_{t-2}|^2 |y_{t''-1}|^2 \varepsilon_{t''}^2\right) \\ &= \beta^4 E\left(|y_{t-2}|^4 |y_{t''-1}|^2 \varepsilon_{t''}^2\right) + 3m_2 \\ &\quad +6\left[\beta^{2(t-t')}m_4 + m_2\frac{1-\beta^{2(t-t'-1)}}{1-\beta^2} + 3m_2\right] \\ \text{for } t &> t''+1 \text{ (substitute (A.23)), and} \\ E\left(|y_{t''}|^4 |y_{t''-1}|^2 \varepsilon_{t''}^2\right) &= 15m_2 + \beta^4 m_6 + 18\beta^2 m_4. \end{aligned} \tag{A.28}$$

Hence for $t = t' > t'' = t'''$

$$\begin{aligned} W &= \beta^{4(t-t')}m_6 + 6\left[\beta^{2(t-t')}m_4 + m_2\frac{1-\beta^{2(t-t'-1)}}{1-\beta^2} + 3m_2\right]\frac{1-\beta^{2(t-t'-1)}}{1-\beta^2} \\ &\quad +18\beta^2 m_4 + 3m_2\frac{1-\beta^{4(t-t'-1)}}{1-\beta^4} + 15m_2. \end{aligned} \tag{A.29}$$

iv) $t'' = t''' > t = t'$, $W = E\left(|y_{t-1}|^4 |y_{t''-1}|^2\right)$ (note that it is different from ii)) can be derived recursively from

$$\begin{aligned} E\left(|y_{t-1}|^4 |y_{t''-1}|^2\right) &= \beta^2 E\left(|y_{t-1}|^4 |y_{t''-2}|^2\right) + m^4 \text{ for } t'' > t+1, \text{ and} \\ E\left(|y_{t-1}|^4 |y_t|^2\right) &= \beta^2 m_6 + m_4. \end{aligned} \tag{A.30}$$

Hence for $t'' = t''' > t = t'$

$$W = \beta^{2(t''-t)}m_6 + m_4\frac{1-\beta^{2(t''-t)}}{1-\beta^2} \tag{A.31}$$

v) $t = t' > t'' > t'''$, $W = E\left(|y_{t-1}|^4 |y_{t''-1}| \varepsilon_{t''} |y_{t'''-1}| \varepsilon_{t'''}\right)$ can be derived recursively from

$$\begin{aligned} E\left(|y_{t-1}|^4 |y_{t''-1}| \varepsilon_{t''} |y_{t'''-1}| \varepsilon_{t'''}\right) &= \beta^4 E\left(|y_{t-2}|^4 |y_{t''-1}| \varepsilon_{t''} |y_{t'''-1}| \varepsilon_{t'''}\right) \\ &\quad +6\beta^2 E\left(|y_{t-2}|^2 |y_{t''-1}| \varepsilon_{t''} |y_{t'''-1}| \varepsilon_{t'''}\right) \\ &= \beta^4 E\left(|y_{t-2}|^4 |y_{t''-1}| \varepsilon_{t''} |y_{t'''-1}| \varepsilon_{t'''}\right) + 24 m_2 \beta^{2(t-t'''-1)} \\ \text{for } t &> t''+1 \text{ (substitute (A.25)), and} \\ E\left(|y_{t''}|^4 |y_{t''-1}| \varepsilon_{t''} |y_{t'''-1}| \varepsilon_{t'''}\right) &= -4\beta^3 E\left(|y_{t''-1}|^4 |y_{t'''-1}| \varepsilon_{t'''}\right) \\ &\quad -12\beta E\left(|y_{t''-1}|^2 |y_{t'''-1}| \varepsilon_{t'''}\right). \end{aligned} \tag{A.32}$$

Apparently, $E\left(|y_{t''-1}|^2 |y_{t'''-1}| \varepsilon_{t'''}\right) = -2m_2\beta^{2(t''-t''')-1}$ from (A.18) and $E\left(|y_{t''-1}|^4 |y_{t'''-1}| \varepsilon_{t'''}\right)$ follows a recursion

$$\begin{aligned} E\left(|y_{t''-1}|^4 |y_{t'''-1}| \varepsilon_{t'''}\right) &= \beta^4 E\left(|y_{t''-2}|^4 |y_{t'''-1}| \varepsilon_{t'''}\right) + 6\beta^2 E\left(|y_{t''-2}|^2 |y_{t'''-1}| \varepsilon_{t'''}\right) \\ &= \beta^4 E\left(|y_{t''-2}|^4 |y_{t'''-1}| \varepsilon_{t'''}\right) - 12 m_2 \beta^{2(t''-t''')-1} \\ \text{for } t'' &> t'''+1 \text{ (substitute (A.18)), and} \\ E\left(|y_{t'''}|^4 |y_{t'''-1}| \varepsilon_{t'''}\right) &= -4\beta^4 m_4 - 12\beta m_2. \end{aligned} \tag{A.33}$$

Hence it is straightforward to evaluate W numerically for $t = t' > t'' > t'''$.

vi) $t'' = t''' > t > t'$ (or $t'' = t''' > t' > t$ due to symmetry; the result follows by exchanging the index t' with t), $W = E\left(|y_{t''-1}|^2 |y_{t-1}|^2 |y_{t'-1}|^2\right)$ follows a recursion

$$\begin{aligned} E\left(|y_{t''-1}|^2 |y_{t-1}|^2 |y_{t'-1}|^2\right) &= \beta^2 E\left(|y_{t''-2}|^2 |y_{t-1}|^2 |y_{t'-1}|^2\right) + E\left(|y_{t-1}|^2 |y_{t'-1}|^2\right) \\ &= \beta^2 E\left(|y_{t''-2}|^2 |y_{t-1}|^2 |y_{t'-1}|^2\right) + \beta^{2(t-t')} m_4 + m_2 \frac{1-\beta^{2(t-t')}}{1-\beta^2} \\ \text{for } t'' &> t+1 \text{ (substitute (A.21)), and} \\ E\left(|y_t|^2 |y_{t-1}|^2 |y_{t'-1}|^2\right) &= \beta^2 E\left(|y_{t-1}|^4 |y_{t'-1}|^2\right) + E\left(|y_{t-1}|^2 |y_{t'-1}|^2\right) \\ &= 6\beta^2 \left[\beta^{2(t-t')} m_4 + m_2 \frac{1-\beta^{2(t-t')}}{1-\beta^2}\right] \frac{1-\beta^{2(t-t'-1)}}{1-\beta^2} \\ &\quad \beta^{4(t-t')+2} m_6 + 6\beta^4 m_4 + 3\beta^2 m_2 \frac{1-\beta^{4(t-t')}}{1-\beta^4} + \beta^{2(t-t')} m_4 \\ &\quad + m_2 \frac{1-\beta^{2(t-t')}}{1-\beta^2} \text{ (from (A.21) and (A.27)).} \end{aligned} \tag{A.34}$$

Hence it is straightforward to evaluate W numerically for $t'' = t''' > t > t'$.

vii) $t > t' > t'' > t'''$ (or $t' > t > t'' > t'''$, $t > t' > t''' > t''$, $t' > t > t''' > t''$; the results follow by exchanging the indices), $W = E\left(|y_{t-1}|^2 |y_{t'-1}|^2 |y_{t''-1}| \varepsilon_{t''} |y_{t'''-1}| \varepsilon_{t'''}\right)$ can be derived recursively from

$$\begin{aligned}
E\left(|y_{t-1}|^2 |y_{t'-1}|^2 |y_{t''-1}| \varepsilon_{t''} |y_{t'''-1}| \varepsilon_{t'''}\right) &= \beta^2 E\left(|y_{t-2}|^2 |y_{t'-1}|^2 |y_{t''-1}| \varepsilon_{t''} |y_{t'''-1}| \varepsilon_{t'''}\right) \\
&\quad + E\left(|y_{t'-1}|^2 |y_{t''-1}| \varepsilon_{t''} |y_{t'''-1}| \varepsilon_{t'''}\right) \\
\text{for } t &> t' + 1, \text{ and} \\
E\left(|y_{t'}|^2 |y_{t'-1}|^2 |y_{t''-1}| \varepsilon_{t''} |y_{t'''-1}| \varepsilon_{t'''}\right) &= \beta^2 E\left(|y_{t'-1}|^4 |y_{t''-1}| \varepsilon_{t''} |y_{t'''-1}| \varepsilon_{t'''}\right) \\
&\quad + E\left(|y_{t'-1}|^2 |y_{t''-1}| \varepsilon_{t''} |y_{t'''-1}| \varepsilon_{t'''}\right), \qquad \text{(A.35)}
\end{aligned}$$

where $E\left(|y_{t'-1}|^4 |y_{t''-1}| \varepsilon_{t''} |y_{t'''-1}| \varepsilon_{t'''}\right)$ follows from v) and $E\left(|y_{t'-1}|^2 |y_{t''-1}| \varepsilon_{t''} |y_{t'''-1}| \varepsilon_{t'''}\right)$ follows from (A.25). Hence it is straightforward to evaluate W numerically for $t > t' > t'' > t'''$.

Therefore, $E\left(H_1 \psi_T^2\right) = \frac{1}{(T-1)^3} \sum \sum \sum V_{tt't''}$ and $E\left(H_1^2 \psi_T^2\right) = \frac{1}{(T-1)^4} \sum \sum \sum \sum W_{tt't''t'''}$ can easily be evaluated numerically.

References

Amemiya, T., 1980, The n^2-order mean squared errors of the maximum likelihood and the minimum logit chi-square estimators, Annals of Mathematical Statistics 8, 488-505.

Andel, J., I. Netuka and K. Svara, 1984, On threshold autoregressive processes, Kybernetika 20, 89-106.

Anderson, T. W. and T. Sawa, 1973, Distributions of estimator of coeifficients of a single equation in a simultaneous system and their asymptotic expansions, Econometrica 41, 683-714.

Anderson, T. W. and T. Sawa, 1979, Evaluation of the distribution function of the two-stage least square estimate, Econometrica 47, 163-182.

Basmann, R. L., 1974, Exact finite sample distribution for some econometric estimators and test statistics: a survey and appraisal, in: M.D. Intriligator and D. A. Kendrick, eds., Frontiers of quantitative economics, Vol. 2 (North-Holland, Amsterdam) 209-288.

Brown, T. M., 1952, Habit persistence and lags in consumer behavior, Econometrica 20, 355-371.

Chesher, A. D. and R. Spady, 1989, Asymptotic expansions of the information matrix test statistic, Econometrica 59, 787-816.

Cordeiro, G. M. and R. Klein, 1994, Bias correction in ARMA models, Statistics and Probabilities Letters 19, 169-796.

Cordeiro, G. M. and P. McCullagh, 1991, Bias correction in generalized linear models, Journal of the Royal Statistical Society B 53, 629–643.

Cox, D. R. and J. Snell, 1968, A general definition of residuals, Journal of the Royal Statistical Society B 30, 248-275.

Engle, R. F., 1982, Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation, Econometrica 50, 987-1007.

Hamilton, J. D., 1994, Time series analysis (Princeton University Press, Princeton).

Iglesias, E. and G. D. A. Phillips, 2001, Small sample properties of ML estimators in AR-ARCH Models, Mimeo. (Cardiff Business School, Cardiff University, Cardiff, Wales, UK).

Iglesias, E. and G. D. A. Phillips, 2002, Small sample estimation bias in ARCH Models, Mimeo. (Cardiff Business School, Cardiff University, Cardiff, Wales, UK).

Kiviet, J. F. and Phillips, G. D. A., 1993, Alternative bias approximations in regressions with a lagged dependent variable, Econometric Theory 9, 62-80.

Koenker, R., J. A. F. Machado, C. L. Skeels, and A. H. Welsh, 1992, Momentary lapses: moment expansions and the robustness of minimum distance estimation, Mimeo. (University of Illinois, Evanston, IL).

Linton, O., 1997, An asymptotic expansion in the GARCH(1,1) model, Econometric Theory 13, 558-581.

Lovell, M. C., 1961, Manufacturers' inventories, sales expectations and the acceleration principle, Econometrica 29, 293-314.

Magnus, J. R., 1978, The moments of products of quadratic forms in normal variables, Statistica Neerlandica 32, 201-210.

Magnus, J. R., 1979, The expectation of products of quadratic forms in normal variables: the practice, Statistica Neerlandica 33, 131-136.

Mathai, A. M. and S. B. Provost, 1992, Quadratic forms in random variables (Marcel Dekker, New York).

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.

Newey, W. K. and R. J. Smith, 2001, Higher order properties of GMM and generalized empirical likelihood estimators, Working paper. (MIT and University of Bristol).

Phillips, P. C. B., 1977, Approximations to some finite sample distributions associated with a first-order stochastic difference equations, Econometrica 45, 463-485.

Phillips, P. C. B., 1987, Asymptotic expansions in nonstationary vector autoregression, Econometric Theory 3, 45-68.

Rilstone, P., V. K. Srivastara, and A. Ullah, 1996, The second-order bias and mean squared error of nonlinear estimators, Journal of Econometrics 75, 369-385.

Roberston, C. A. and J. G. Fryer, 1970, The bias and accuracy of moment estimators, Biometrika 57, 57-65.

Rothenberg, T. J., 1984, Approximating the distribution of econometric estimators and test statistics, in: M. L. Intriligator and Z. Griliches, eds., Handbook of econometrics, Vol. 2, (North-Holland, Amsterdam).

Sargan, J. D., 1974, The validity of Nagar's expansion for the moments of econometric estimators, Econometrica 42, 169-176.

Sargan, J. D., 1976, Econometric estimators and the Edgeworth approximations, Econometrica 44, 421-448.

Tong, H., 1990, Non-linear time series: a dynamical system approcah (Oxford University Press, New York).

Ullah, A., 1990, Finite sample econometrics: a unified approach, in: R. A. L. Carter, J. Dutta and A. Ullah, eds., Contributions to econometric theory and application, Volume in honor of A. L. Nagar (Springer-Verlag, New York).

Ullah, A. and V. K. Srivastava, 1994, Moments of the ratio of quadratic forms in non-normal variables with econometric examples, Journal of Econometrics 62, 129-141.

Ullah, A., 2002, Finite sample econometrics, Manuscript. (Department of Economics, University of California, Riverside).

Table 1: Second-order Bias (row 1) and MSE (row 2) and First-order MSE (row 3) of OLS Estimators of $vec(\hat{B})$ in VAR(1) Model with True Parameter *B*= ((0.8, 0.3)', (0.2, 0.5)')'

	$T = 20$				$T = 30$				$T = 50$			
$r = -0.9$	-0.1027	0.0827	-0.0720	0.0450	-0.0779	0.0622	-0.0600	0.0390	-0.0518	0.0412	-0.0426	0.0283
	1.4543	0.7966	1.0595	0.8410	0.9661	0.6158	0.8813	0.7541	0.6295	0.4863	0.7393	0.6870
	0.3458	0.3458	0.6063	0.6063	0.3458	0.3458	0.6063	0.6063	0.3458	0.3458	0.6063	0.6063
$r = -0.5$	-0.0698	0.0259	-0.0315	-0.0305	-0.0520	0.0178	-0.0282	-0.0202	-0.0342	0.0109	-0.0208	-0.0120
	0.9467	0.5539	0.8807	0.8166	0.6532	0.4482	0.7393	0.7015	0.4592	0.3747	0.6349	0.6185
	0.2879	0.2879	0.5131	0.5131	0.2879	0.2879	0.5131	0.5131	0.2879	0.2879	0.5131	0.5131
$r = -0.2$	-0.0610	0.0130	-0.0231	-0.0493	-0.0452	0.0076	-0.0215	-0.0349	-0.0295	0.0040	-0.0162	-0.0220
	0.9882	0.5739	0.9735	0.9599	0.6856	0.4734	0.8172	0.8117	0.4886	0.4023	0.7054	0.7032
	0.3136	0.3136	0.5625	0.5625	0.3136	0.3136	0.5625	0.5625	0.3136	0.3136	0.5625	0.5625
$r = 0$	-0.0569	0.0084	-0.0207	-0.0573	-0.0420	0.0040	-0.0195	-0.0412	-0.0274	0.0015	-0.0148	-0.0263
	1.1817	0.6427	1.1174	1.1374	0.8077	0.5333	0.9338	0.9522	0.5651	0.4546	0.8064	0.8152
	0.3528	0.3528	0.6348	0.6348	0.3528	0.3528	0.6348	0.6348	0.3528	0.3528	0.6348	0.6348
$r = 0.2$	-0.0531	0.0055	-0.0200	-0.0638	-0.0391	0.0018	-0.0189	-0.0463	-0.0254	-0.0000	-0.0144	-0.0298
	1.6586	0.7784	1.3660	1.4345	1.0944	0.6451	1.1300	1.1863	0.7276	0.5475	0.9743	1.0008
	0.4175	0.4175	0.7531	0.7531	0.4175	0.4175	0.7531	0.7531	0.4175	0.4175	0.7531	0.7531
$r = 0.5$	-0.0462	0.0044	-0.0233	-0.0733	-0.0338	0.0009	-0.0212	-0.0537	-0.0219	-0.0006	-0.0158	-0.0348
	4.4723	1.2958	2.2662	2.4698	2.7021	1.0504	1.7896	1.9818	1.5278	0.8668	1.5172	1.6133
	0.6147	0.6147	1.1119	1.1119	0.6147	0.6147	1.1119	1.1119	0.6147	0.6147	1.1119	1.1119
$r = 0.9$	-0.0095	0.0234	-0.0673	-0.1086	-0.0065	0.0157	-0.0537	-0.0813	-0.0040	0.0094	-0.0370	-0.0535
	216.1311	9.9156	71.6586	18.4018	120.0730	7.1224	42.2908	13.1765	50.8611	4.9998	20.6674	9.2209
	2.1356	2.1356	3.8738	3.8738	2.1356	2.1356	3.8738	3.8738	2.1356	2.1356	3.8738	3.8738

Eignevalues: 0.9372, 0.3628.

Table 1 (Continued): Second-order Bias (row 1) and MSE (row 2) and First-order MSE (row 3) of OLS Estimators of $vec(\hat{B})$ in VAR(1) Model with True Parameter B = ((0.1, 0.4)', (0.4, 0.3)')'

	$T = 20$				$T = 30$				$T = 50$			
$r = -0.9$	-0.0235	-0.0210	-0.0285	-0.0181	-0.0152	-0.0149	-0.0186	-0.0130	-0.0089	-0.0093	-0.0109	-0.0082
	7.9322	/	/	/	5.4407	0.8221	/	/	4.0841	1.6959	0.6082	0.7966
	2.9278	2.9278	3.0711	3.0711	2.9278	2.9278	3.0711	3.0711	2.9278	2.9278	3.0711	3.0711
$r = -0.5$	-0.0230	-0.0259	-0.0291	-0.0243	-0.0155	-0.0176	-0.0196	-0.0166	-0.0093	-0.0107	-0.0118	-0.0101
	1.9363	1.6420	1.0409	1.0338	1.5348	1.3804	0.9881	0.9864	1.2683	1.1936	0.9566	0.9564
	0.9490	0.9490	0.9247	0.9247	0.9490	0.9490	0.9247	0.9247	0.9490	0.9490	0.9247	0.9247
$r = -0.2$	-0.0223	-0.0257	-0.0298	-0.0256	-0.0151	-0.0175	-0.0201	-0.0174	-0.0091	-0.0106	-0.0121	-0.0106
	1.7294	1.4596	1.0303	1.0170	1.3634	1.2200	0.9212	0.9167	1.1173	1.0470	0.8482	0.8469
	0.8174	0.8174	0.7598	0.7598	0.8174	0.8174	0.7598	0.7598	0.8174	0.8174	0.7598	0.7598
$r = 0$	-0.0216	-0.0252	-0.0304	-0.0264	-0.0146	-0.0171	-0.0205	-0.0179	-0.0089	-0.0104	-0.0124	-0.0109
	1.7272	1.4613	1.0691	1.0532	1.3683	1.2271	0.9426	0.9369	1.1262	1.0572	0.8565	0.8548
	0.8300	0.8300	0.7500	0.7500	0.8300	0.8300	0.7500	0.7500	0.8300	0.8300	0.7500	0.7500
$r = 0.2$	-0.0206	-0.0243	-0.0313	-0.0275	-0.0139	-0.0165	-0.0211	-0.0186	-0.0085	-0.0100	-0.0128	-0.0113
	1.8326	1.5613	1.1725	1.1512	1.4642	1.3208	1.0268	1.0186	1.2151	1.1455	0.9265	0.9239
	0.9096	0.9096	0.8007	0.8007	0.9096	0.9096	0.8007	0.8007	0.9096	0.9096	0.8007	0.8007
$r = 0.5$	-0.0175	-0.0213	-0.0341	-0.0305	-0.0119	-0.0145	-0.0230	-0.0206	-0.0072	-0.0088	-0.0139	-0.0125
	2.3308	2.0199	1.5644	1.5219	1.8963	1.7351	1.3666	1.3492	1.6026	1.5260	1.2292	1.2233
	1.2414	1.2414	1.0546	1.0546	1.2414	1.2414	1.0546	1.0546	1.2414	1.2414	1.0546	1.0546
$r = 0.9$	0.0115	0.0081	-0.0602	-0.0571	0.0074	0.0051	-0.0404	-0.0383	0.0044	0.0029	-0.0243	-0.0231
	7.1723	6.2716	4.8100	4.7357	6.1245	5.7231	4.4899	4.4397	5.4763	5.3228	4.2380	4.2146
	4.7797	4.7797	3.8935	3.8935	4.7797	4.7797	3.8935	3.8935	4.7797	4.7797	3.8935	3.8935

Eignevalues: -0.2123, 0.6123.

Figure 1: Second−order Bias of Conditional ML Estimator of β in MA(1) Model

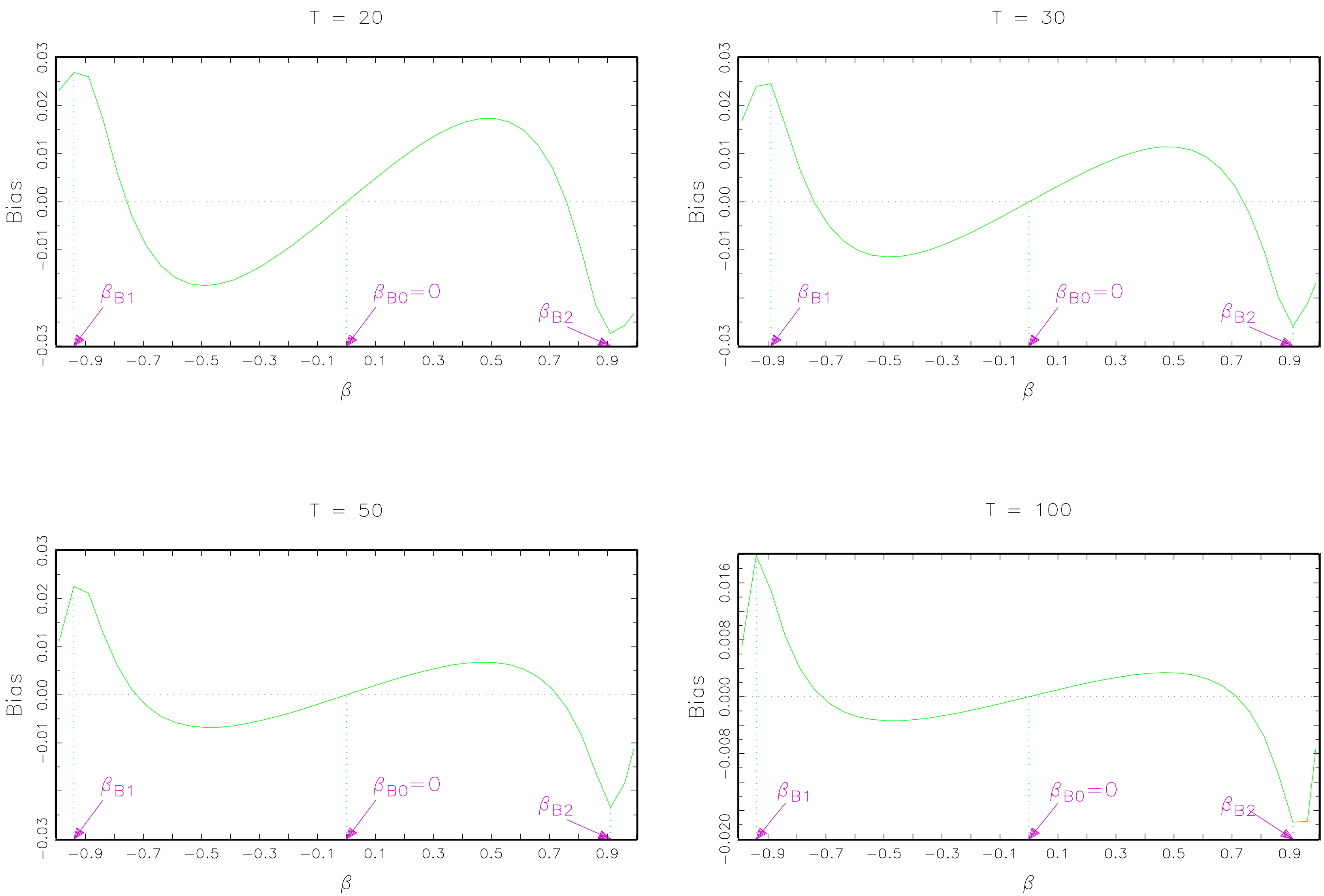

Figure 2: Second-order MSE of Conditional ML Estimator of β in MA(1) Model