Available online at www.sciencedirect.com
ScienceDirect
Journal of Econometrics 140 (2007) 650–669
JOURNAL OF Econometrics
www.elsevier.com/locate/jeconom

The second-order bias and mean squared error of estimators in time-series models

Yong Bao[a], Aman Ullah[b,*]

[a]*Department of Economics, University of Texas, San Antonio, TX 78249, USA*
[b]*Department of Economics, University of California, Riverside, CA 92521, USA*

Available online 18 September 2006

Abstract

We develop analytical results on the second-order bias and mean squared error of estimators in time-series models. These results provide a unified approach to developing the properties of a large class of estimators in linear and nonlinear time-series models and they are valid for both normal and nonnormal samples 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 four time-series models. We investigate the effects of nonnormality on the second-order bias results for two of these models, while for all four models, the second-order bias and mean squared error results are given under normality. Numerical results for some of these models are also presented.

JEL classification: C10; C22; C32

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

1. Introduction

There is 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 (2004), among others. In contrast, not much work has been done on the finite sample properties of nonlinear statistics, although see Roberston and

*Corresponding author. Tel.: +1 951 827 1591; fax: +1 951 787 5685.
E-mail addresses: yong.bao@utsa.edu (Y. Bao), aman.ullah@ucr.edu (A. Ullah).

doi:10.1016/j.jeconom.2006.07.007

Fryer (1970), Amemiya (1980), Chesher and Spady (1989), Cordeiro and McCullagh (1991), Koenker et al. (1992), and Newey and Smith (2004). However, most of these works addressed some specific estimators and the series of interest were assumed to be identically and independently distributed (i.i.d.), 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 in models with i.i.d. samples.

In this paper, we extend the second-order bias and MSE results of Rilstone et al. (1996) for time-series dependent observations. These results provide a unified way of developing the properties of a given class of estimators in 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. Our results are also valid for models with both normal and nonnormal samples of observations, and where the weakly exogenous regressors are stochastic. Next, in a special case of the ML estimator (MLE) we show that our bias result reduces to the result in Cox and Snell (1968) for the i.i.d. case and also for its extension in Cordeiro and Klein (1994) for models with dependent observations. However, we note that our bias result is for a general class of estimators including the MLE as a special case and that Cox and Snell's (1968) approach does not provide the MSE of the MLE.

As an application of our results, we derive the second-order bias and MSE of parameter estimators in four time-series models. These include the AR(1) model with exogenous regressors (ARX(1)), VAR(1), MA(1), and absolute regression models. For expositional purposes, we develop the second-order bias results under nonnormality for the ARX(1) and MA(1) models and find that the bias results under normality are robust against nonnormality for the MA(1) model, and also for the ARX(1) model without exogenous regressors or just a constant. For the ARX(1) model, when the nonnormality parameter does play a role, our simulation results show that bias correction accommodating the effects of nonnormality works quite well.

The plan of the paper is as follows. In Section 2, we present the estimators identified by some moment conditions and their bias and MSE results. Then in Section 3 we derive the second-order bias and MSE of estimators in four time-series models. Section 4 contains our conclusion. All the proofs are collected in the Appendix.

2. Second-order bias and MSE

Given a sample of T observations, we consider a class of estimators $\hat{\beta}$ which can be written as the solution to a set of moment equations of the form

$$\psi_T(\hat{\beta}) = \frac{1}{T}\sum_{t=1}^{T} q_t(\hat{\beta}) = 0, \tag{1}$$

where $q_t(\beta) = q(Z_t; \beta)$ is a known $p \times 1$ vector-valued function of the observable m-dimensional non-i.i.d. random vectors Z_t and the parameter vector $\beta \in \mathbb{R}^p$ with true value β_0 such that $\mathbb{E}[\psi_T(\beta)] = 0$ (conditional on the stochastic regressors, if any) only at $\beta = \beta_0$. We can easily think of $\psi_T(\cdot)$ as the orthogonality condition between the regressors and the error terms, or the first-order condition of some optimization criterion. The class of estimators identified by (1) include many estimators in linear and nonlinear time-series models, which include the ML, LS, and GMM estimators.

To obtain the stochastic expansion of $\hat{\beta}$, Assumptions A–C in Rilstone et al. (1996) are assumed to hold along with the $\sqrt{T}$-consistency of $\hat{\beta}$ and we follow similar notations. Throughout, $\nabla^s A(\beta)$ denote the matrices of recursively defined sth order partial derivatives of matrix $A(\beta)$ with respect to β. We can now implement a Taylor series expansion of $\psi_T(\hat{\beta}) = 0$ around β_0 up to the third order and use a Nagar-type (Nagar, 1959) expansion of the inverse of the gradient of the moment function to solve for $\hat{\beta} - \beta_0$ recursively. For more detail, see the steps outlined in Lemmas A.1–A.3 in Rilstone et al. (1996). The stochastic expansion of $\hat{\beta}$ is

$$\hat{\beta} - \beta_0 = a_{-1/2} + a_{-1} + a_{-3/2} + \mathrm{o_P}(T^{-3/2}), \tag{2}$$

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

$$a_{-1/2} = -Q\psi_T, \quad a_{-1} = -QVa_{-1/2} - \tfrac{1}{2} Q\overline{H_2}[a_{-1/2} \otimes a_{-1/2}],$$

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

in which $\overline{X} = \mathbb{E}(X)$ denotes the expectation of a random vector X, $H_i = \nabla^i \psi_T$ (we suppress the argument of a function when it is evaluated at β_0), $i = 1, 2, 3$, $Q = \overline{H_1}^{-1}$, $V = H_1 - \overline{H_1}$, $W = H_2 - \overline{H_2}$, and $\otimes$ represents the Kronecker product. We should emphasize that the stochastic expansion (2) is valid for both non-i.i.d. and i.i.d. cases. In fact, the stochastic expansion (2) is the same as in Lemma 3.3 of Rilstone et al. (1996). However, in their case Q and $\overline{H_i}$ are evaluated for the i.i.d. case, for example, $\overline{H_1} = T^{-1}\mathbb{E}\left(\nabla \sum_{t=1}^{T} q_t\right) = \overline{\nabla q_1}$.

We observe that $\mathbb{E}(a_{-1/2}) = 0$ since $\mathbb{E}(\psi_T) = \mathbb{E}[\psi_T(\beta_0)] = 0$. It follows from (2) that the second-order bias of $\hat{\beta}$, $B(\hat{\beta})$, up to $\mathrm{O}(T^{-1})$, is

$$B(\hat{\beta}) = \mathbb{E}(a_{-1}), \tag{4}$$

and the MSE of $\hat{\beta}$, $M(\hat{\beta})$, up to $\mathrm{O}(T^{-2})$, is

$$\begin{aligned} M(\hat{\beta}) = & \mathbb{E}(a_{-1/2}a'_{-1/2}) + \mathbb{E}(a_{-1}a'_{-1/2} + a_{-1/2}a'_{-1}) \\ & + \mathbb{E}(a_{-1}a'_{-1} + a_{-3/2}a'_{-1/2} + a_{-1/2}a'_{-3/2}). \end{aligned} \tag{5}$$

If we substitute (3) into (4) and (5), one an write the second-order bias and MSE results in an alternate way, which e present in the Appendix (see (A.1) and (A.2)).[1]

We note that the bias formula (4) (also (A.1) in the Appendix) and the MSE formula (5) (also (A.2) in the Appendix) are valid for both non-i.i.d. and i.i.d. cases of the series $\{Z_t\}_{t=1}^T$. This was not observed in Rilstone et al. (1996), where these results were evaluated and presented by assuming $\{Z_t\}_{t=1}^T$ to be i.i.d., and hence the elements of ψ_T, V, W, and d (defined as $Q\psi_T$) to be i.i.d.(also see Rilstone and Ullah, 2005 for a correction of the MSE expression in Rilstone et al., 1996). Under the i.i.d. assumption, (A.1) and (A.2) can be further simplified since most of the cross-terms in the matrix multiplications drop out. For

[1]The authors are thankful to a referee for the question on the existence of the exact moments of $\hat{\beta}$, which remains to be explored in a future study. This issue is important since the "approximate moments"can be obtained even if the exact moments do not exist, also see Sargan (1974).

example, in the bias expression, $\overline{Vd} = T^{-2}\sum_i\sum_j \overline{V_i d_j}$ if we write $V = T^{-1}\sum V_i$ and $d = T^{-1}\sum d_i$, the summand is zero except for $i = j$. Such simplifications are not available for a general non-i.i.d. series, so we stay with the parsimonious presentation (4) and (5) (or (A.1) and (A.2)) in this paper since we focus on non-i.i.d. observations and thereby it is the general results (4) and (5), or (A.1) and (A.2), that have more interesting applications. We should also point out that the general expressions (4) and (5) do not necessarily lead to easy derivations for various models. The amount of work may vary for different models and in some cases massive derivations may be undertaken to work out the expectations in (4) and (5) for a specific model. For models with low-dimensional parameters, we may be able to simplify (4) and (5) and express them explicitly as functions of the parameters (e.g. ARX(1)). In general, however, we do not make such an attempt and we shall present some of our results in the following section without final expressions being presented explicitly in terms of model parameters (though for comparison with the existing results on the ARX(1) model, we do simplify our results). But all the expectations needed in (4) and (5) for deriving the second-order results for the estimators considered in this paper can be evaluated or programmed. Readers may regard (4) and (5) as a recipe to follow to obtain general results for quite general stable time-series models and consistent extremum estimators.

The expressions in (4) and (5), based on the stochastic expansion of the estimator in (2), provides a unified approach for both linear and nonlinear models and when observations are not necessarily i.i.d. The stochastic expansion in (2) is based on the Taylor expansion of the moment condition for a class of estimators and the Nagar expansion of the inverse of gradient of the moment condition. For some linear models the results for a specific estimator under consideration can also be obtained by a direct Nagar-type stochastic expansion, which amounts to a binomial expansion of an inverse matrix or scalar of stochastic elements, and then grouping terms by the order of magnitude; for example, see the important results on the bias under normality by Kiviet and Phillips (1993) for the stable ARX(1) model, Kiviet and Phillips (1994) for the higher-order stable ARX(p) model, and Kiviet et al. (1999) for the VAR(1) model. These models are also considered as illustrative examples (Section 3.1 under nonnormality and Section 3.3 under normality) to obtain the results by directly using the unified expressions in (4) and (5). Our results are identical with the existing results available under normality. But for nonlinear models a direct Nagar expansion is often unavailable, and the results can be analyzed by following our method; see the applications in Sections 3.2 and 3.4. The main results are then summarized as follows: (i) providing the procedure to follow to obtain further results for quite general stable time-series models and consistent extremum estimators, and (ii) producing some analytical results for nonlinear time-series models and for other estimators than the LS estimator.

Note that if $\psi_T(\beta)$ is the score function for the MLE of the parameters, then the second-order bias result will be the same as the Cox and Snell (1968) result. Cordeiro and Klein (1994) rewrote the bias vector from Cox and Snell (1968) of the $p \times 1$ MLE $\hat{\beta}$ as

$$\mathbb{E}(\hat{\beta} - \beta_0) = K^{-1}A \cdot \mathbf{vec}(K^{-1}), \tag{6}$$

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

$$\mathbb{E}(\hat{\beta}_s - \beta_{s0}) = \sum_{i=1}^{p}\sum_{j=1}^{p}\sum_{l=1}^{p} k^{sj}\left(k_{ij}^{(l)} - \frac{1}{2}k_{ijl}\right)k^{jl}, \tag{7}$$

where **vec** stands for the vectorizing (column by column) operator on a matrix, $K^{-1} = \{-k^{ij}\}$ is the inverse of the information matrix evaluated at β_0, which is equal to $-Q$ in our notation, $A = (A^{(1)}, A^{(2)}, \ldots, A^{(p)})$ with the $p \times p$ matrix $A^{(l)} = \{a_{ij}^{(l)}\}, a_{ij}^{(l)} = k_{ij}^{(l)} - \frac{1}{2}k_{ijl}$, $k_{ij} = \mathbb{E}(\partial^2 \mathscr{L}/\partial\beta_i\partial\beta_j)$, $k_{ijl} = \mathbb{E}(\partial^3 \mathscr{L}/\partial\beta_i\partial\beta_j\partial\beta_l)$, $k_{ij}^{(l)} = \partial k_{ij}/\partial\beta_l$, for $i, j, l, s \leqslant p$, and $\mathscr{L}$ is the log-likelihood function. Essentially, we implement a Taylor series expansion, as in Rilstone et al. (1996), with respect to the whole estimated parameter vector $\hat{\beta}$ whereas in Cox and Snell (1968) the expansion was carried out with respect to each element of $\hat{\beta}$ and then a set of simultaneous equations were solved to arrive at $\mathbb{E}(\hat{\beta} - \beta_0)$. In fact, if we examine (A.1) and (6) or (7) carefully and use the information equality $\mathbb{E}(\partial^2 \mathscr{L}/\partial\beta_i\partial\beta_j) = -\mathbb{E}[(\partial\mathscr{L}/\partial\beta_i)(\partial\mathscr{L}/\partial\beta_j)]$, we find $\{Q\overline{Vd}\}_s = \sum_{i=1}^{p}\sum_{j=1}^{p}\sum_{l=1}^{p} k^{sj}k_{ij,l}k^{jl}$, where $k_{ij,l} = \mathbb{E}[(\partial^2\mathscr{L}/\partial\beta_i\partial\beta_j)(\partial\mathscr{L}/\partial\beta_l)]$, and $\{Q\overline{H_2}(\overline{d \otimes d})\}_s = -\sum_{i=1}^{p}\sum_{j=1}^{p}\sum_{l=1}^{p} k^{sj}k_{ijl}k^{jl}$ for the single parameter estimator $\hat{\beta}_s, 1 \leqslant s \leqslant p$. Replacing k_{ijl} by $k_{ij}^{(l)} - k_{ij,l}$, we immediately have

$$B(\hat{\beta}) = Q[\overline{Vd} - \tfrac{1}{2}\overline{H_2}(\overline{d \otimes d})] = K^{-1}A \cdot \mathbf{vec}(K^{-1}). \tag{8}$$

Since (4) is equivalent to (A.1), this establishes the equality of (6) and (4). Our result (4) is more general since it includes the MLE as a special case. Furthermore, we provide the second-order MSE in (5) (or (A.2)), which is not covered in Cox and Snell (1968).

3. Illustrations

In this section, we give the application of our second-order bias and MSE results to four time-series models. These include the ARX(1), MA(1), VAR(1), and absolute regression models. 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. To characterize nonnormality, for $\varepsilon_t \sim$i.i.d. $(0, \sigma^2)$, we assume

$$\mathbb{E}(\varepsilon_t^3) = \sigma^3\gamma_1, \quad \mathbb{E}(\varepsilon_t^4) = \sigma^4(\gamma_2 + 3), \tag{9}$$

where γ_1 and γ_2 are the Pearson's measures of skewness and excess kurtosis of the distribution. Throughout, when we write down a model, the sample observation runs from $t = 1$ to T. We assume that the model is correctly specified and for notational convenience, we drop the subscript 0 for the true parameter.

3.1. AR(1) model with exogenous regressors

The first example we consider is the stable first-order autoregressive model with exogenous regressors

$$y_t = \rho y_{t-1} + x_t'\beta + \varepsilon_t, \tag{10}$$

where $\rho \in (-1, 1)$, x_t is $k \times 1$ fixed and bounded so that $X'X = \mathrm{O}(T)$, where $X = (x_1, \ldots, x_T)'$, β is $k \times 1$, and $\varepsilon_t \sim$i.i.d. $(0, \sigma^2)$ and follows (9). Denote $\varepsilon = (\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_T)'$, $y = (y_1, y_2, \ldots, y_T)'$, $y_{-1} = (y_0, y_1, \ldots, y_{T-1})'$, where we assume that the first observation y_0 has been observed and it can be either fixed or stochastic (and independent of ε). For convenience, we shall call model (10) with $|\rho| < 1$ a stable ARX(1) model. Following Kiviet and Phillips (1993), we define F to be a $T \times 1$ vector with the tth element being ρ^{t-1}, C to be a strictly lower triangular $T \times T$ matrix with the tt'th lower off-diagonal element being $\rho^{t-t'-1}$, $y_D = y_0 F + CX\beta$, $M = I - X(X'X)^{-1}X'$ ($I = I_T$ is the identity matrix of size T), $r_D = My_D$, and write $y_{-1} = y_D + C\varepsilon$.

The OLS estimator $\hat{\rho}$ for the autoregressive coefficient ρ is based on the moment condition $\psi_T = 0$, where

$$\psi_T = \frac{1}{T} y'_{-1} M\varepsilon = \frac{1}{T} y'_{-1} M(y - X\beta - \rho y_{-1}), \tag{11}$$

which gives $H_1 = -(y'_{-1} M y_{-1})/T$ and $H_2 = H_3 = W = 0$. Given this, (4) or (A.1) provides the bias result presented in the following theorem, where $\imath$ is a unit vector of size T, **tr** is the trace operator on a matrix, and $\odot$ stands for the Hadamard product.

Theorem 1. *In the stable ARX*(1) *model, the second-order bias, up to* $\mathrm{O}(T^{-1})$, *of the OLS estimator* $\hat{\rho}$, *is*

$$B(\hat{\rho}) = \frac{\sigma^2[\mathbf{tr}(MC) - 2\rho(1-\rho^2)^{-1}]}{\bar{D}} - \frac{2\sigma^2[r'_D C r_D - \rho(1-\rho^2)^{-1} r'_D r_D]}{\bar{D}^2} + \gamma_1 \xi,$$

where $\xi = -\sigma^3 \imath'\{[I \odot (C'MC)]r_D + 2[I \odot (MC)]C' r_D\}/\bar{D}^2$ *and* $\bar{D} = r'_D r_D + \sigma^2\, \mathbf{tr}(C'MC)$.

The proof of Theorem 1 is given in Appendix A.3. Note that under normality ($\gamma_1 = 0$), $B(\hat{\rho})$ degenerates into the bias result of Kiviet and Phillips (1993), also see Section 2. Theorem 1 also indicates that for symmetric nonnormal error distributions (e.g. Student-t), the bias result remains the same as in the case of normality. Moreover, for asymmetric nonnormal error distributions, the contribution to the second-order bias from γ_1 will disappear in a pure AR(1) model ($k = 0$) and the intercept model ($X = \imath$ and $\beta \neq 0$). These are given in the following corollaries.

Corollary 1. *In a stable pure AR*(1) *model, the second-order bias, up to* $\mathrm{O}(T^{-1})$, *of the OLS estimator* $\hat{\rho}$ *is* $-2\rho/T$.

Corollary 2. *If in the stable ARX*(1) *model* $X = \imath$ *and* $\beta \neq 0$, *the second-order bias, up to* $\mathrm{O}(T^{-1})$, *of the OLS estimator* $\hat{\rho}$, *is given by* $B(\hat{\rho}) = -(1+3\rho)/T$.

Regarding the general result as given by Theorem 1, it is interesting to investigate the effect of y_0. Since $r'_D r_D = \beta' X' C' MCX\beta + \mathrm{O}(1)$ and $r'_D C r_D = \beta' X' C' MCMCX\beta + \mathrm{O}(1)$, replacing $r'_D r_D$ (note that $\bar{D}$ has a term $r'_D r_D$ also) with $\beta' X' C' MCX\beta$ and $r'_D C r_D$ with $\beta' X' C' MCMCX\beta$ does not affect the second-order bias up to $\mathrm{O}(T^{-1})$. Hence we do not need to distinguish between the random and fixed start-up cases for calculating the second-order bias (though assumptions on y_0 will affect the second-order MSE).

For the pure AR(1) model, the bias result $-2\rho/T$ was derived by White (1961) under normality. For the intercept model, Kendall's (1954) approximate bias under normality gives exactly the same result as in Corollary 2, also see Sawa (1978). Under the assumption that the errors are martingale differences, Pope (1990) obtained the bias result, which can be shown (Nicholls and Pope, 1988) to reduce to the result in Corollary 2. The two corollaries above demonstrate that for both the pure AR(1) and intercept models, the bias results under normality are robust, up to the second order, against the distribution assumption on the error terms, and are independent of the error variance, and of the starting value y_0.

As for the second-order MSE of $\hat{\rho}$, since $H_2 = H_3 = W = 0$, from (5) or (A.2) we immediately have $M(\hat{\rho}) = 6Q^2\overline{\psi_T^2} - 8Q^3\overline{H_1\psi_T^2} + 3Q^4\overline{H_1^2\psi_T^2}$. Essentially, this result will involve the expectation of quadratic forms in ε up to four terms. Assuming normality and $y_0 = 0$, we show in Appendix A.3 that $M(\hat{\rho}) = (1-\rho^2)/T + (14\rho^2 - 1)/T^2$ for the pure AR(1) model, which was also given in White (1961), Shenton and Johnson (1965), and Phillips (1977). Interestingly, the $O(T^{-3/2})$ term disappears for the MSE in this case.

Now we give some numerical calculations to present the second-order bias result given by Theorem 1. We use a noncentral t distribution to simulate the error terms from the following model

$$y_t = \rho y_{t-1} + x_t\beta + \varepsilon_t, \quad x_t = \text{Uniform } [0, 1],$$

for $\eta = 10$ and $v = 5, 7, 9$, where η is the noncentrality parameter and v is the degrees of freedom of the noncentral t distribution.[2] We try with $T = 50, 80$, $\sigma^2 = 0.5, 1$, $\rho = 0.1, 0.3, 0.5, 0.9$, and to prevent the signal-to-noise ratio increasing as ρ goes up, we set $\beta = 1 - \rho$. The results in Table 1 represent the averaged values across 5,000 simulations (y_0, x_t, and ε_t vary for each simulation). For a given sample size, the first column $\hat{\rho}$ gives the OLS estimate, the second column $\tilde{\rho}_N$ represents the feasible bias-corrected estimate using the bias formula from Theorem 1 *ignoring* nonnormality ($\gamma_1 = 0$), and the third column $\tilde{\rho}_{NN}$ represents the bias-corrected estimate using the bias formula from Theorem 1 under nonnormality.[3] We observe that firstly as T increases from 50 to 80, the second-order bias decreases, as we expect. Secondly, when ρ is small, the bias contribution due to nonnormality is not negligible in the sense that the bias performance of $\tilde{\rho}_{NN}$ is better than that of $\tilde{\rho}_N$. For example, when $v = 5$, $\sigma^2 = 0.5$, and $\rho = 0.1$, the percentage absolute bias ($|\text{bias}|/\rho \times 100\%$) of $\tilde{\rho}_{NN}$ is 0.2% but the percentage absolute bias of $\tilde{\rho}_N$ is 2.9%; when $v = 5$, $\sigma^2 = 0.5$, and $\rho = 0.5$, the percentage absolute bias of $\tilde{\rho}_{NN}$ is 0.12% but the percentage absolute bias of $\tilde{\rho}_N$ is 0.78%. On the other hand, when ρ is large (around 0.9), even though the (absolute) bias of $\tilde{\rho}_{NN}$ is much smaller than that of $\hat{\rho}$, it is bigger than that of $\tilde{\rho}_N$, which indicates that the nonnormal term $\hat{\gamma}_1\hat{\xi}$ in $\tilde{\rho}_{NN}$ does not capture the effect of nonnormality accurately. In fact, when ρ is large, it is often the case $\hat{\gamma}_1$ is overestimated for the simulations we have experienced so overall the performance of $\tilde{\rho}_{NN}$ is even worse than that of $\tilde{\rho}_N$. Thirdly, as the error variance σ^2 goes up, $\hat{\rho}$ is more biased, but the performance of $\tilde{\rho}_{NN}$ is quite robust to the change in σ^2. Fourthly, we note that bias correction does not diminish the accuracy of estimates since the MSEs of the bias corrected ones (using estimates for all parameters) are not that different from those of the uncorrected ones (while also changing the value of σ^2 and the characteristics of x_t).

3.2. MA(1) model

Consider an invertible MA(1) model,

$$y_t = \varepsilon_t - \phi\varepsilon_{t-1}, \tag{12}$$

[2] When $\eta = 10$ and $v = 5, 7, 9$, we have $\gamma_1 = 3.0279$, 1.9223, 1.4959, and $\gamma_2 = 32.3206$, 9.0466, 5.0308. We simulate y_t with initial value 0 for a larger sample size $T + T_0$, say, and then use observations from $T_0 + 1$ to $T + T_0$ for y and observations T_0 to $T + T_0 - 1$ for y_{-1}, so the "y_0" is random in our simulations.

[3] A feasible bias-corrected estimator is defined as $\tilde{\rho} = \hat{\rho} - \hat{B}(\hat{\rho})$, where $\hat{B}(\hat{\rho})$ (with all the unknown parameters ρ, σ^2, β, and γ_1 replaced with their sample estimators) consistently estimates $B(\hat{\rho})$, see Rilstone et al. (1996). We estimate γ_1 by $T^{-1}\sum\hat{\varepsilon}_t^3/\hat{\sigma}^3$, where $\hat{\varepsilon}$ is the OLS residual and $\hat{\sigma} = \sqrt{T^{-1}\sum\hat{\varepsilon}_t^2}$.

Table 1
Bias correction for ARX(1) model

v	ρ	$T=50$						$T=80$					
		$\sigma^2=0.5$			$\sigma^2=1$			$\sigma^2=0.5$			$\sigma^2=1$		
		$\hat{\rho}$	$\tilde{\rho}_N$	$\tilde{\rho}_{NN}$	$\hat{\rho}$	$\tilde{\rho}_N$	$\tilde{\rho}_{NN}$	$\hat{\rho}$	$\tilde{\rho}_N$	$\tilde{\rho}_{NN}$	$\hat{\rho}$	$\tilde{\rho}_N$	$\tilde{\rho}_{NN}$
5	0.1	**0.0877**	**0.0971**	**0.1002**	**0.0798**	**0.0952**	**0.0980**	**0.0898**	**0.0958**	**0.0981**	**0.0887**	**0.0987**	**0.1008**
		0.1065	0.1100	0.1103	0.1249	0.1295	0.1295	0.0871	0.0888	0.0889	0.0991	0.1015	0.1016
	0.3	**0.2776**	**0.2953**	**0.2987**	**0.2689**	**0.2944**	**0.2971**	**0.2846**	**0.2960**	**0.2985**	**0.2812**	**0.2977**	**0.2997**
		0.1105	0.1135	0.1136	0.1237	0.1264	0.1263	0.0869	0.0882	0.0882	0.0980	0.0994	0.0994
	0.5	**0.4678**	**0.4961**	**0.4994**	**0.4614**	**0.4982**	**0.5005**	**0.4770**	**0.4951**	**0.4976**	**0.4737**	**0.4972**	**0.4989**
		0.1078	0.1091	0.1090	0.1196	0.1200	0.1198	0.0853	0.0853	0.0852	0.0930	0.0925	0.0924
	0.9	**0.8437**	**0.9034**	**0.9041**	**0.8404**	**0.9026**	**0.9029**	**0.8631**	**0.9011**	**0.9018**	**0.8637**	**0.9035**	**0.9039**
		0.0961	0.0933	0.0929	0.0993	0.0930	0.0929	0.0675	0.0612	0.0610	0.0687	0.0631	0.0630
7	0.1	**0.0841**	**0.0937**	**0.0962**	**0.0790**	**0.0945**	**0.0967**	**0.0907**	**0.0969**	**0.0987**	**0.0869**	**0.0969**	**0.0985**
		0.1118	0.1152	0.1153	0.1281	0.1329	0.1329	0.0880	0.0897	0.0898	0.1019	0.1043	0.1043
	0.3	**0.2779**	**0.2960**	**0.2987**	**0.2706**	**0.2964**	**0.2986**	**0.2850**	**0.2966**	**0.2985**	**0.2808**	**0.2973**	**0.2988**
		0.1118	0.1147	0.1148	0.1257	0.1290	0.1289	0.0902	0.0916	0.0917	0.1018	0.1034	0.1033
	0.5	**0.4672**	**0.4960**	**0.4986**	**0.4617**	**0.4987**	**0.5005**	**0.4791**	**0.4975**	**0.4993**	**0.4757**	**0.4992**	**0.5005**
		0.1128	0.1142	0.1141	0.1234	0.1244	0.1243	0.0852	0.0856	0.0855	0.0952	0.0954	0.0954
	0.9	**0.8427**	**0.9017**	**0.9023**	**0.8425**	**0.9045**	**0.9049**	**0.8629**	**0.9006**	**0.9012**	**0.8614**	**0.9008**	**0.9010**
		0.0972	0.1316	0.1314	0.0990	0.1018	0.1016	0.0686	0.0623	0.0621	0.0720	0.0654	0.0653
9	0.1	**0.0868**	**0.0966**	**0.0987**	**0.0805**	**0.0963**	**0.0981**	**0.0947**	**0.1010**	**0.1025**	**0.0893**	**0.0994**	**0.1007**
		0.1119	0.1156	0.1157	0.1279	0.1330	0.1330	0.0891	0.0911	0.0913	0.1020	0.1046	0.1047
	0.3	**0.2804**	**0.2987**	**0.3009**	**0.2721**	**0.2981**	**0.2998**	**0.2863**	**0.2980**	**0.2996**	**0.2812**	**0.2977**	**0.2990**
		0.1138	0.1174	0.1175	0.1280	0.1319	0.1318	0.0898	0.0913	0.0914	0.1007	0.1023	0.1023
	0.5	**0.4662**	**0.4951**	**0.4972**	**0.4618**	**0.4987**	**0.5003**	**0.4784**	**0.4968**	**0.4984**	**0.4748**	**0.4984**	**0.4995**
		0.1124	0.1134	0.1134	0.1236	0.1246	0.1245	0.0874	0.0877	0.0876	0.0952	0.0952	0.0951
	0.9	**0.8413**	**0.8999**	**0.9004**	**0.8406**	**0.9027**	**0.9029**	**0.8636**	**0.9004**	**0.9008**	**0.8615**	**0.9009**	**0.9012**
		0.1006	0.0951	0.0948	0.1006	0.0962	0.0961	0.0696	0.0880	0.0879	0.0701	0.0632	0.0631

Note: The numbers in **bold** font are the estimated parameters and bias-corrected ones. The row following in regular font represents the MSE (across simulations) of the corresponding estimates above. $\hat{\rho}$ is the OLS estimate; $\tilde{\rho}_N$ is the feasible bias-corrected estimate ignoring nonnormality; $\tilde{\rho}_{NN}$ is the feasible bias-corrected estimate accommodating nonnormality; v is the degrees of freedom for a standardized noncentral t distribution with noncentral parameter 10 used to simulate the errors.

where $|\phi|<1$, $\varepsilon_t \sim$i.i.d. $(0,\sigma^2)$ and follows (9). Define $y=(y_1,\ldots,y_T)'$, C to be a lower triangular $T\times T$ matrix with unit diagonal and its tt'th lower off-diagonal element being $-\phi$ if $t-t'=1$, and zero otherwise, $B=I-C$, and $B_1=\partial B/\partial\phi$. Then we can write $y=C\varepsilon$, $\varepsilon=y+B\varepsilon$, and the following derivatives:

$$\frac{\partial\varepsilon}{\partial\phi}=C^{-1}B_1\varepsilon,\quad \frac{\partial^2\varepsilon}{\partial\phi^2}=2(C^{-1}B_1)^2\varepsilon,\quad \frac{\partial^3\varepsilon}{\partial\phi^3}=6(C^{-1}B_1)^3\varepsilon,\quad \frac{\partial^4\varepsilon}{\partial\phi^4}=24(C^{-1}B_1)^4\varepsilon. \tag{13}$$

To estimate the parameter vector $(\phi,\sigma^2)'$, the standard approach is to maximize the conditional quasi-likelihood function (conditioning on $\varepsilon_0=0$)

$$\mathscr{L}(\phi,\sigma^2)=-\frac{1}{2}\log(2\pi)-\frac{1}{2}\log(\sigma^2)-\frac{1}{T}\frac{\varepsilon'\varepsilon}{2\sigma^2}, \tag{14}$$

where $\varepsilon = (\varepsilon_1, \ldots, \varepsilon_T)'$. Denoting $N_1 = C^{-1}B_1$, $N_2 = 2(C^{-1}B_1)^2 + (C^{-1}B_1)'(C^{-1}B_1)$, $N_3 = 6(C^{-1}B_1)^3 + 6(C^{-1}B_1)'(C^{-1}B_1)^2$, and $N_4 = 24(C^{-1}B_1)^4 + 24(C^{-1}B_1)'(C^{-1}B_1)^3 + 12[(C^{-1}B_1)^2]'\,(C^{-1}B_1)^2$, we can conveniently write score function as

$$\psi_T = \begin{pmatrix} -\dfrac{\varepsilon' N_1 \varepsilon}{T\sigma^2} \\ -\dfrac{1}{2\sigma^2} + \dfrac{\varepsilon'\varepsilon}{2T\sigma^4} \end{pmatrix}, \tag{15}$$

as well as the Hessian and higher-order derivative matrices as follows:

$$H_1 = \begin{pmatrix} -\dfrac{\varepsilon' N_2 \varepsilon}{T\sigma^2} & \dfrac{\varepsilon' N_1 \varepsilon}{T\sigma^4} \\ \dfrac{\varepsilon' N_1 \varepsilon}{T\sigma^4} & \dfrac{1}{2\sigma^4} - \dfrac{\varepsilon'\varepsilon}{T\sigma^6} \end{pmatrix}, \quad H_2 = \begin{pmatrix} -\dfrac{\varepsilon' N_3 \varepsilon}{T\sigma^2} & \dfrac{\varepsilon' N_2 \varepsilon}{T\sigma^4} & \dfrac{\varepsilon' N_2 \varepsilon}{T\sigma^4} & -\dfrac{2\varepsilon' N_1 \varepsilon}{T\sigma^6} \\ \dfrac{\varepsilon' N_2 \varepsilon}{T\sigma^4} & -\dfrac{2\varepsilon' N_1 \varepsilon}{T\sigma^6} & -\dfrac{2\varepsilon' N_1 \varepsilon}{T\sigma^6} & -\dfrac{1}{\sigma^6} + \dfrac{3\varepsilon'\varepsilon}{T\sigma^8} \end{pmatrix},$$

$$H_3 = \begin{pmatrix} -\dfrac{\varepsilon' N_4 \varepsilon}{T\sigma^2} & \dfrac{\varepsilon' N_3 \varepsilon}{T\sigma^4} & \dfrac{\varepsilon' N_3 \varepsilon}{T\sigma^4} & -\dfrac{2\varepsilon' N_2 \varepsilon}{T\sigma^6} & \dfrac{\varepsilon' N_3 \varepsilon}{T\sigma^4} & -\dfrac{2\varepsilon' N_2 \varepsilon}{T\sigma^6} & -\dfrac{2\varepsilon' N_2 \varepsilon}{T\sigma^6} & \dfrac{6\varepsilon' N_1 \varepsilon}{T\sigma^8} \\ \dfrac{\varepsilon' N_3 \varepsilon}{T\sigma^4} & -\dfrac{2\varepsilon' N_2 \varepsilon}{T\sigma^6} & -\dfrac{2\varepsilon' N_2 \varepsilon}{T\sigma^6} & \dfrac{6\varepsilon' N_1 \varepsilon}{T\sigma^8} & -\dfrac{2\varepsilon' N_2 \varepsilon}{T\sigma^6} & \dfrac{6\varepsilon' N_1 \varepsilon}{T\sigma^8} & \dfrac{6\varepsilon' N_1 \varepsilon}{T\sigma^8} & \dfrac{3}{\sigma^8} - \dfrac{12\varepsilon'\varepsilon}{T\sigma^{10}} \end{pmatrix}. \tag{16}$$

Since $\mathbf{tr}(N_1) = \mathbf{tr}(C^{-1}B) = 0$, we note that $\mathbb{E}(\psi_T) = 0$, as we expect, and that $Q = [\mathbb{E}(H_1)]^{-1}$ is diagonal

$$Q = \begin{pmatrix} \dfrac{-T}{\mathbf{tr}(N_2)} & 0 \\ 0 & -2\sigma^4 \end{pmatrix}. \tag{17}$$

Upon substitution, the bias and MSE results can now be obtained from (4) and (5) or (A.1) and (A.2), and they are given in the following theorem.

Theorem 2. *In the moving average model* (12), *the second-order bias, up to* $O(T^{-1})$, *of the conditional QMLE* $(\hat{\phi}, \hat{\sigma}^2)'$ *is*

$$B\begin{pmatrix} \hat{\phi} \\ \hat{\sigma}^2 \end{pmatrix} = \frac{1}{T}\begin{pmatrix} \phi \\ -\sigma^2 \end{pmatrix}$$

and the second-order MSE, up to $O(T^{-2})$, *is* $((m_{11}, m_{12})', (m_{12}, m_{22})')'$, *where*

$$m_{11} = \frac{6\lambda_{0200}}{n_2^2} - \frac{8\lambda_{0210}}{n_2^3} + \frac{3\lambda_{0220} + \lambda_{0301} + 4n_3\lambda_{0300}}{n_2^4} - \frac{12n_3\lambda_{0310} + n_4\lambda_{0400}}{3n_2^5} + \frac{5n_3^2\lambda_{0400}}{4n_2^6},$$

$$m_{22} = \sigma^4\left(\frac{\lambda_{2000}}{T^2} + \frac{4\lambda_{0200}}{Tn_2} - \frac{4\lambda_{1200}}{T^2 n_2} - \frac{2\lambda_{0210}}{Tn_2^2} + \frac{\lambda_{0400} + 2\lambda_{1210}}{T^2 n_2^2} + \frac{2n_3\lambda_{0300}}{3Tn_2^3} - \frac{2n_3\lambda_{1300}}{3T^2 n_2^3} - 1\right),$$

$$m_{12} = \sigma^2 \left(-\frac{3\lambda_{1100}}{Tn_2} - \frac{3\lambda_{0110}}{n_2^2} + \frac{3\lambda_{0300} + 3\lambda_{1110}}{Tn_2^2} + \frac{2\lambda_{0120} + \lambda_{0201} + 3n_3\lambda_{0200}}{2n_2^3} \right.$$
$$- \frac{4\lambda_{0310} + 2\lambda_{1120} + 2\lambda_{1201} + 3n_3\lambda_{1200}}{2Tn_2^3} - \frac{9n_3\lambda_{0210} + n_4\lambda_{0300}}{6n_2^4}$$
$$\left. + \frac{5n_3\lambda_{0400} + 9n_3\lambda_{1210} + n_4\lambda_{1300}}{6Tn_2^4} + \frac{n_3^2\lambda_{0300}}{2n_2^5} - \frac{n_3^2\lambda_{1300}}{2Tn_2^5} \right),$$

in which $\lambda_{ijkl} = \mathbb{E}[(\varepsilon'\varepsilon)^i(\varepsilon' N_1 \varepsilon)^j \cdot (\varepsilon' N_2 \varepsilon)^k \cdot (\varepsilon' N_3 \varepsilon)^l / \sigma^{2(i+j+k+l)}]$ *and* $n_i = \mathbf{tr}(N_i)$.

The proof of Theorem 2 is given in Appendix A.3. According to this theorem, the second-order bias of $\hat{\phi}$ is robust against nonnormality. We notice that Cordeiro and Klein (1994) derived the same result using the Cox and Snell (1968) expansion *and* assuming normality. Theorem 2 also indicates that the bias and MSE of $\hat{\phi}$ are independent of σ^2, while the bias of $\hat{\sigma}^2$ is independent of ϕ, but its second-order MSE depends on both ϕ and σ^2.

In Table 2, we report our simulation results by assuming ε_t following different distributions (normal, uniform, exponential, mixture of two normals $\mathrm{N}(-3, 1)$ and $\mathrm{N}(3, 1)$ with half probability for each normal, and Student-t with 5 degrees of freedom). Clearly, our simulation results confirm that the theoretical bias result is robust to nonnormality for the MA(1) model. The feasible bias-corrected estimates are almost unbiased for the normal and the four nonnormal distributions. We note that for very small $T = 30$, when $\phi = 0.9$, the QMLE may do better than the bias-corrected estimators, which indicates that a sample size of 30 may not be large enough for the asymptotics to work in this case.

As for the MSE, evaluations under normality are quite straightforward by using the results in Appendix A.2. Table 3 gives the theoretical second-order MSE results, the asymptotic variances ($(1-\phi^2)/T$ for $\hat{\phi}$, $2\sigma^4/T$ for $\hat{\sigma}^2$), which are also the first-order MSE, as well as their ratios, when ε_t is normally distributed. Regarding the theoretical results, we observe the following: firstly, the asymptotic variances appear to underestimate the sampling variations of $\hat{\phi}$ and $\hat{\sigma}^2$ compared with the second-order MSEs and the degree of underestimation increases as the moving average parameter increases. An immediate implication is that the standard t ratio constructed using the asymptotic standard deviation may give misleading results in small samples as it tends to over-reject. Secondly, comparing the degree of underestimation of the asymptotic variance for the sampling variation of the estimator, we find that it is more severe for $\hat{\phi}$ than $\hat{\sigma}^2$. For example, for a sample size of 100, m_{22} is almost the same as the asymptotic variance of $\hat{\sigma}^2$, but the difference between m_{11} and the asymptotic variance of $\hat{\phi}$ is still quite significant. Thirdly, even though second-order MSE of $\hat{\sigma}^2$ depends on both ϕ and σ^2, it seems that it is more sensitive to the change of σ^2 than that of ϕ. In fact, for a sample size as small as 50, the effects of ϕ are quite marginal. This is consistent with our second-order theoretical MSE result, since the terms involving σ^2 in m_{22} appear in the terms of order $\mathrm{O}(T^{-1})$ and of lower order, while the terms involving ϕ appear only in the terms of order lower than $\mathrm{O}(T^{-1})$. Lastly, as the sample size increases, the asymptotic variances and the second-order MSEs decrease, as we expect. Moreover, the gap between the first-order and the second-order MSEs tends to decrease. It seems that we need the sample size to be as large as 300 to close this gap.

Table 2
Bias correction for MA(1) model

σ^2	ϕ	$T=30$						$T=100$					
		$\hat{\phi}$	$\ddot{\phi}$	$\tilde{\phi}$	$\hat{\sigma}^2$	$\ddot{\sigma}^2$	$\tilde{\sigma}^2$	$\hat{\phi}$	$\ddot{\phi}$	$\tilde{\phi}$	$\hat{\sigma}^2$	$\ddot{\sigma}^2$	$\tilde{\sigma}^2$
1	0.2	0.2092	0.2025	0.2022	0.9641	0.9974	0.9962	0.2007	0.1987	0.1987	0.9866	0.9966	0.9965
		0.2082	0.2015	0.2012	0.9666	0.9999	0.9988	0.2039	0.2019	0.2019	0.9885	0.9985	0.9983
		0.2017	0.1950	0.1950	0.9686	1.0019	1.0009	0.2024	0.2004	0.2004	1.0019	1.0119	1.0119
		0.2074	0.2008	0.2005	0.9631	0.9964	0.9952	0.2021	0.2001	0.2001	0.9901	1.0001	1.0000
		0.2025	0.1958	0.1957	0.9711	1.0045	1.0035	0.2037	0.2017	0.2016	0.9955	1.0055	1.0054
	0.5	0.5240	0.5073	0.5065	0.9667	1.0000	0.9989	0.5065	0.5015	0.5015	0.9894	0.9994	0.9993
		0.5160	0.4994	0.4988	0.9633	0.9966	0.9954	0.5052	0.5002	0.5002	0.9914	1.0014	1.0013
		0.5134	0.4967	0.4963	0.9652	0.9986	0.9974	0.5030	0.4980	0.4980	0.9833	0.9933	0.9931
		0.5186	0.5020	0.5013	0.9598	0.9932	0.9918	0.5040	0.4990	0.4989	0.9894	0.9994	0.9993
		0.5171	0.5005	0.4999	0.9754	1.0087	1.0079	0.5046	0.4996	0.4996	0.9923	1.0023	1.0023
	0.9	0.9089	0.8789	0.8786	0.9713	1.0046	1.0036	0.9092	0.9002	0.9002	0.9894	0.9994	0.9992
		0.9113	0.8813	0.8809	0.9626	0.9960	0.9947	0.9089	0.8999	0.8998	0.9877	0.9977	0.9976
		0.9068	0.8768	0.8766	0.9717	1.0050	1.0041	0.9084	0.8994	0.8993	0.9901	1.0001	1.0000
		0.9102	0.8802	0.8799	0.9659	0.9993	0.9981	0.9095	0.9005	0.9004	0.9879	0.9979	0.9978
		0.9113	0.8813	0.8809	0.9493	0.9826	0.9809	0.9089	0.8999	0.8999	0.9871	0.9971	0.9969
2	0.2	0.2138	0.2072	0.2067	1.9253	1.9920	1.9895	0.2011	0.1991	0.1991	1.9844	2.0044	2.0042
		0.2139	0.2072	0.2067	1.9311	1.9978	1.9955	0.2023	0.2003	0.2003	1.9819	2.0019	2.0017
		0.1997	0.1931	0.1931	1.9208	1.9875	1.9849	0.2000	0.1980	0.1980	1.9984	2.0184	2.0183
		0.2073	0.2006	0.2004	1.9252	1.9919	1.9894	0.2002	0.1982	0.1982	1.9806	2.0006	2.0004
		0.2055	0.1988	0.1986	1.9484	2.0151	2.0134	0.2019	0.1999	0.1999	1.9874	2.0074	2.0073
	0.5	0.5209	0.5042	0.5035	1.9287	1.9953	1.9930	0.5038	0.4988	0.4988	1.9800	2.0000	1.9998
		0.5197	0.5031	0.5024	1.9214	1.9881	1.9854	0.5042	0.4992	0.4992	1.9800	2.0000	1.9998
		0.5097	0.4930	0.4927	1.9305	1.9972	1.9948	0.5014	0.4964	0.4964	1.9856	2.0056	2.0054
		0.5170	0.5003	0.4998	1.9209	1.9876	1.9849	0.5067	0.5017	0.5017	1.9784	1.9984	1.9982
		0.5236	0.5070	0.5062	1.9327	1.9993	1.9971	0.5051	0.5001	0.5000	1.9808	2.0008	2.0006
	0.9	0.9122	0.8822	0.8818	1.9444	2.0110	2.0092	0.9093	0.9003	0.9002	1.9799	1.9999	1.9997
		0.9089	0.8789	0.8786	1.9397	2.0064	2.0044	0.9096	0.9006	0.9005	1.9737	1.9937	1.9935
		0.9073	0.8773	0.8771	1.9179	1.9846	1.9818	0.9087	0.8997	0.8997	1.9585	1.9785	1.9781
		0.9109	0.8809	0.8806	1.9340	2.0007	1.9985	0.9095	0.9005	0.9004	1.9758	1.9958	1.9956
		0.9118	0.8818	0.8814	1.9668	2.0334	2.0323	0.9094	0.9004	0.9003	1.9771	1.9971	1.9968

Note: $\hat{\phi}$ is the QMLE of ϕ, $\ddot{\phi}$ is the bias-corrected estimate, $\tilde{\phi}$ is the feasible bias-corrected estimate. $\hat{\sigma}^2$, $\ddot{\sigma}^2$, and $\tilde{\sigma}^2$ are interpreted similarly. All the estimates refer to the averaged ones over 5,000 simulations. The standard normal, uniform, exponential, mixture of two normals, and Student-t with 5 degrees of freedom are used to simulate the error terms. For each parameter value of (ϕ, σ^2), the five rows in each column correspond to the estimates under the five distributions.

3.3. VAR(1) model

Let $y_t = (y_{1t}, y_{2t}, \ldots, y_{pt})'$ denote a $p \times 1$ vector containing the values that the p variables assume at time t. Suppose that the dynamics of y_t is described by a first-order vector autoregression

$$y_t = \Phi y_{t-1} + u_t, \tag{18}$$

Table 3
Theoretical asymptotic variance and second-order MSE in MA(1) model

ϕ	$T=30$			$T=50$			$T=100$			$T=300$		
Panel A: results for $\hat{\phi}$												
0.0	0.0333	0.0606	1.8194	0.0200	0.0273	1.3630	0.0100	0.0114	1.1350	0.0033	0.0034	1.0349
0.1	0.0330	0.0608	1.8428	0.0198	0.0272	1.3720	0.0099	0.0113	1.1377	0.0033	0.0034	1.0354
0.2	0.0320	0.0614	1.9180	0.0192	0.0269	1.4009	0.0096	0.0110	1.1461	0.0032	0.0033	1.0369
0.3	0.0303	0.0625	2.0615	0.0182	0.0265	1.4558	0.0091	0.0106	1.1620	0.0030	0.0032	1.0397
0.4	0.0280	0.0647	2.3107	0.0168	0.0261	1.5508	0.0084	0.0100	1.1892	0.0028	0.0029	1.0443
0.5	0.0250	0.0687	2.7460	0.0150	0.0257	1.7166	0.0075	0.0093	1.2361	0.0025	0.0026	1.0521
0.6	0.0213	0.0759	3.5580	0.0128	0.0259	2.0264	0.0064	0.0085	1.3226	0.0021	0.0023	1.0657
0.7	0.0170	0.0896	5.2683	0.0102	0.0274	2.6901	0.0051	0.0077	1.5055	0.0017	0.0019	1.0930
0.8	0.0120	0.1149	9.5768	0.0072	0.0325	4.5090	0.0036	0.0072	2.0082	0.0012	0.0014	1.1636
0.9	0.0063	0.1282	20.2440	0.0038	0.0463	12.1748	0.0019	0.0086	4.5105	0.0006	0.0009	1.4971

Panel B: results for $\hat{\sigma}^2$

σ^2	ϕ	$T=30$			$T=50$			$T=100$			$T=300$		
1	0.0	0.0667	0.0698	1.0472	0.0400	0.0405	1.0129	0.0200	0.0200	1.0007	0.0067	0.0067	0.9990
	0.1	0.0667	0.0699	1.0486	0.0400	0.0405	1.0134	0.0200	0.0200	1.0008	0.0067	0.0067	0.9990
	0.2	0.0667	0.0702	1.0528	0.0400	0.0406	1.0150	0.0200	0.0200	1.0012	0.0067	0.0067	0.9990
	0.3	0.0667	0.0707	1.0606	0.0400	0.0407	1.0178	0.0200	0.0200	1.0019	0.0067	0.0067	0.9991
	0.4	0.0667	0.0715	1.0729	0.0400	0.0409	1.0223	0.0200	0.0201	1.0031	0.0067	0.0067	0.9992
	0.5	0.0667	0.0728	1.0918	0.0400	0.0412	1.0293	0.0200	0.0201	1.0049	0.0067	0.0067	0.9994
	0.6	0.0667	0.0748	1.1218	0.0400	0.0416	1.0404	0.0200	0.0202	1.0077	0.0067	0.0067	0.9998
	0.7	0.0667	0.0781	1.1722	0.0400	0.0424	1.0595	0.0200	0.0203	1.0126	0.0067	0.0067	1.0003
	0.8	0.0667	0.0845	1.2679	0.0400	0.0439	1.0973	0.0200	0.0205	1.0225	0.0067	0.0067	1.0014
	0.9	0.0667	0.0985	1.4774	0.0400	0.0479	1.1985	0.0200	0.0210	1.0514	0.0067	0.0067	1.0048
2	0	0.2667	0.2792	1.0472	0.1600	0.1621	1.0129	0.0800	0.0801	1.0007	0.0267	0.0266	0.9990
	0.1	0.2667	0.2796	1.0486	0.1600	0.1621	1.0134	0.0800	0.0801	1.0008	0.0267	0.0266	0.9990
	0.2	0.2667	0.2808	1.0528	0.1600	0.1624	1.0150	0.0800	0.0801	1.0012	0.0267	0.0266	0.9990
	0.3	0.2667	0.2828	1.0606	0.1600	0.1628	1.0178	0.0800	0.0802	1.0019	0.0267	0.0266	0.9991
	0.4	0.2667	0.2861	1.0729	0.1600	0.1636	1.0223	0.0800	0.0802	1.0031	0.0267	0.0266	0.9992
	0.5	0.2667	0.2912	1.0918	0.1600	0.1647	1.0293	0.0800	0.0804	1.0049	0.0267	0.0267	0.9994
	0.6	0.2667	0.2991	1.1218	0.1600	0.1665	1.0404	0.0800	0.0806	1.0077	0.0267	0.0267	0.9998
	0.7	0.2667	0.3126	1.1722	0.1600	0.1695	1.0595	0.0800	0.0810	1.0126	0.0267	0.0267	1.0003
	0.8	0.2667	0.3381	1.2679	0.1600	0.1756	1.0973	0.0800	0.0818	1.0225	0.0267	0.0267	1.0014
	0.9	0.2667	0.3940	1.4774	0.1600	0.1918	1.1985	0.0800	0.0841	1.0514	0.0267	0.0268	1.0048

Note: For each sample size, the first column gives the asymptotic variance, the second column gives the second-order MSE, and the third column gives the ratio (column 2/column 1).

where Φ is the $p \times p$ matrix of autoregression parameters, $u_t = (u_{1t}, u_{2t}, \ldots, u_{pt})'$ is i.i.d. normal with zero mean and covariance matrix Ω. We assume that the process is stationary, i.e. for all λ satisfying $|I_p\lambda - \Phi| = 0$, $|\lambda| < 1$. Denote the $Tp \times 1$ vectors $y = (y_1', y_2', \ldots, y_T')'$, $y_{-1} = (y_0', y_1', \ldots, y_{T-1}')'$, where y_0 is y_t at $t = 0$ and is assumed to be fixed, and $u = (u_1', u_2', \ldots, u_T')'$. We also use the notation $\Phi_i = (\phi_{i1}, \ldots, \phi_{ip})'$, $U_i = (u_{i1}, u_{i2}, \ldots, u_{iT})'$, $Y_i = (y_{i1}, y_{i2}, \ldots, y_{iT})'$, $Y_{i,-1} = (y_{i0}, y_{i1}, \ldots, y_{iT-1})'$, $Y_{-1} = (Y_{1,-1}, Y_{2,-1}, \ldots, Y_{p,-1})$.

Since $y_t = \Phi^t y_0 + \sum_{s=1}^{t} \Phi^{t-s} u_s$, we can decompose y_{-1} into a nonstochastic part and a stochastic part: $y_{-1} = \bar{y}_{-1} + Cu$, where the nonstochastic part $\bar{y}_{-1} = (y_0', (\Phi y_0)', \ldots, (\Phi^{T-1} y_0)')'$ and the $Tp \times Tp$ matrix C has its tt'th block ($p \times p$) being $\Phi^{t-t'-1}$ if $t > t'$, and zero otherwise. Conditioning on y_0, the MLE of Φ is the same as the OLS estimator, equation by equation. Following Kiviet et al. (1999), we focus on the first equation

$$y_{1t} = \Phi_1' y_{t-1} + u_{1t}. \tag{19}$$

The OLS estimator of Φ_1 in (19) is $\hat{\Phi}_1 = (Y_{-1}'Y_{-1})^{-1}Y_{-1}'Y_1$, which is based on the moment condition $\psi_T = 0$, where

$$\psi_T = \frac{1}{T}(Y_{1,-1}'U_1, Y_{2,-1}'U_1, \ldots, Y_{p,-1}'U_1)', \tag{20}$$

which gives the $p \times p$ matrix H_1 with its ijth element being $-Y_{i,-1}'Y_{j,-1}/T$, and $H_2 = H_3 = W = 0$. If we define the $T \times Tp$ matrix $A_i = (A_i^{(1)}, A_i^{(2)}, \ldots, A_i^{(T)})$, $1 \leqslant i \leqslant p$, and the $T \times p$ matrix $A_i^{(t)}$, $1 \leqslant t \leqslant T$, having all elements being zero except the tith element being one, then we can write $Y_{i,-1} = A_i y_{-1} = A_i \bar{y}_{-1} + A_i C u$ and $U_1 = A_1 u$, which leads to $Y_{i,-1}'U_1 = \bar{y}_{-1}'A_i'A_1 u + u'C'A_i'A_1 u$ and $Y_{i,-1}'Y_{j,-1} = \bar{y}_{-1}'A_i'A_j\bar{y}_{-1} + u'C'A_i'A_j\bar{y}_{-1} + \bar{y}_{-1}'A_i'A_jCu + u'C'A_i'A_jCu$.

Note that since $u_t \sim$i.i.d. $(0, \Omega)$ and Gaussian, we have $u \sim \mathrm{N}(0, \Sigma)$, where the $Tp \times Tp$ covariance matrix Σ is block diagonal with each block being Ω. Using ψ_T and H_1 in (4) and (5) or (A.1) and (A.2) gives the following theorem.

Theorem 3. *In the VAR model* (18), *the second-order bias*, *up to* $\mathrm{O}(T^{-1})$, *of the conditional MLE* $\hat{\Phi}_1$ *has its ith component*

$$\begin{aligned} B(\hat{\phi}_{1i}) = & -\frac{1}{T^2}\sum_{j=1}^{p}\sum_{k=1}^{p}\sum_{l=1}^{p} Q_{ij}Q_{kl}[(\bar{y}_{-1}'A_j'A_k\bar{y}_{-1} + \mathbf{tr}(C'A_j'A_kC\Sigma))\mathbf{tr}(C'A_l'A_1\Sigma) \\ & + \bar{y}_{-1}'A_l'A_1\Sigma C'(A_j'A_k + A_k'A_j)\bar{y}_{-1} + 2\,\mathbf{tr}(C'A_j'A_kC\Sigma C'A_l'A_1\Sigma)], \end{aligned}$$

where Q_{ij} *is the ijth element of* $Q = \overline{H_1}^{-1}$ *and* $\overline{H_1}$ *has its ijth component* $-[\bar{y}_{-1}'A_i'A_j\bar{y}_{-1} + \mathbf{tr}(C'A_i'A_jC\Sigma)]/T$, *and the second-order MSE*, *up to* $\mathrm{O}(T^{-2})$, *of* $\hat{\Phi}_1$ *is given by*

$$\begin{aligned} M(\hat{\Phi}_1) = & \, 6Q[\overline{\psi\psi'} - 4(\overline{H_1Q\psi\psi'} + \overline{\psi\psi'QH_1}) \\ & + \overline{H_1Q\psi\psi'QH_1} + \overline{H_1QH_1Q\psi\psi'} + \overline{\psi\psi'QH_1QH_1}]Q, \end{aligned}$$

in which the ijth elements of $\overline{\psi\psi'}$, $\overline{H_1Q\psi\psi'}$, $\overline{\psi\psi'QH_1}$, $\overline{H_1Q\psi\psi'QH_1}$, $\overline{H_1QH_1Q\psi\psi'}$, *and* $\overline{\psi\psi'QH_1QH_1}$ *can be evaluated using the results in Appendix* A.2 *since* ψ *and* H_1 *involve quadratic forms in u.*

The proof of Theorem 3 is given in Appendix A.3 We note that our bias result will be consistent with the result in Nicholls and Pope (1988), Pope (1990), and the result in Kiviet et al. (1999) by setting the exogenous regressors equal to zero in their result. However, they did not consider the MSE result. Table 4 gives the theoretical second-order bias and the MSE, as well as the asymptotic variance (first-order MSE) results for $\hat{\Phi}_1$ when $\Phi = ((0.125, 0.125)', (-0.5, 0.5)')'$. We follow Kiviet et al. (1999) by setting $y_0 = 0$ and $\Omega = ((\omega, \rho_{12}\omega)', (\rho_{12}\omega, 1)')'$.[4] Though not directly comparable with the results by Kiviet et al. (1999), where some exogenous X is included in the regression, we note that the inclusion of exogenous regressors makes the OLS estimation more imprecise in that $\hat{\Phi}_1$ always gives

[4]We tried with many other different Φ's, whose results are available upon request.

Table 4
Theoretical second-order bias, second-order MSE, and asymptotic variance of OLS estimator of Φ_1 in VAR(1) model

		$T=20$		$T=50$	
$\rho_{12}=0.6$	$\omega=0.006$	−0.002400	−0.000711	−0.000920	−0.000274
ω varies		0.082677	0.000863	0.011659	0.000120
		0.005169	0.000022	0.001893	0.000009
	$\omega=0.012$	−0.006648	−0.000985	−0.002564	−0.000381
		0.153591	0.001981	0.022861	0.000283
		0.013156	0.000065	0.004878	0.000026
	$\omega=0.018$	−0.010611	−0.001105	−0.004119	−0.000428
		0.188343	0.002902	0.029341	0.000421
		0.020499	0.000117	0.007661	0.000046
	$\omega=0.024$	−0.013982	−0.001161	−0.005453	−0.000449
		0.203288	0.003650	0.032960	0.000538
		0.026747	0.000175	0.010054	0.000069
$\omega=0.0075$	$\rho_{12}=0.24$	−0.007956	0.000185	−0.003103	0.000080
ρ_{12} varies		0.077388	0.000829	0.011054	0.000115
		0.005432	0.000023	0.001997	0.000009
	$\rho_{12}=0.48$	−0.004982	−0.000466	−0.001929	−0.000177
		0.093072	0.001016	0.013316	0.000142
		0.006412	0.000028	0.002357	0.000011
	$\rho_{12}=0.72$	−0.001847	−0.001147	−0.000692	−0.000446
		0.118767	0.001326	0.016988	0.000187
		0.008033	0.000037	0.002953	0.000014
	$\rho_{12}=0.96$	0.001460	−0.001860	0.000612	−0.000728
		0.155366	0.001777	0.022191	0.000251
		0.010327	0.000049	0.003797	0.000019

Note: The true parameters are $\Phi = ((0.125, 0.125)', (-0.5, 0.5)')'$ and the covariance matrix $\Omega = (((\omega, \rho_{12}\omega)', \rho_{12}\omega, 1)')'$. Each cell contains the second-order bias (first row), the second-order MSE (second row), and the asymptotic variance (third row), of $\hat{\Phi}_1$.

larger bias in Kiviet et al. (1999) than here. In addition, we observe the following. Firstly, there is significant difference between the second-order and the first-order MSEs in small samples, and it seems that the first-order MSE usually underestimates the variation of $\hat{\Phi}_1$ compared with the second-order MSE. Again, this implies that the standard t ratio constructed using the asymptotic standard deviation may give misleading results in small samples. Secondly, the higher the covariance between u_{1t} and u_{2t} (by increasing ω or ρ_{12}), the larger the second-order MSE (as well as the asymptotic variance), which implies that even though the OLS estimators are consistent, the correlation between disturbance terms across equations will make the OLS estimators less precise. Lastly, as the sample size increases, the second-order bias and the MSE decrease, and the difference between the asymptotic and the second-order MSEs becomes smaller, as we expect.

3.4. Absolute regression model

Consider the "absolute"regression model

$$y_t = -\rho|y_{t-1}| + \varepsilon_t, \tag{21}$$

where $0<\rho<1$, ε_t is normal i.i.d. (0, 1). We assume that y_0 is fixed and bounded. Note that (21) is a very special case of the self-exciting autoregressive (SETAR) model of Tong (1990) and it is in nature a nonlinear regression model. Using Andel et al. (1984), we can show that the even moments of y_t are the same as those of y_t following a pure AR(1) process. In particular, $m_2 = (1-\rho^2)^{-1}$, $m_4 = 3(1-\rho^2)^{-2}$, $m_6 = 15(1-\rho^2)^{-3}$, where $m_i = \mathbb{E}(y_t^i)$.

In most applications, ρ is estimated by the LS estimator $\hat{\rho}$, which is a solution of $\psi_T = 0$, where

$$\psi_T = \frac{1}{T}\sum_{t=1}^{T} |y_{t-1}|(y_t + \rho|y_{t-1}|), \tag{22}$$

which gives $H_1 = \sum_{t=1}^{T}|y_{t-1}|^2/T$, and $H_2 = H_3 = W = 0$. Further, $Q = 1/m_2$. Using these in (4) and (5) or (A.1) and (A.2), we can obtain the results in the following theorem.

Theorem 4. *The bias and MSE, up to* $\mathrm{O}(T^{-1})$ *and* $\mathrm{O}(T^{-2})$, *respectively*, *of the LS estimator* $\hat{\rho}$ *in the absolute regression model* (21) *are*

$$B(\hat{\rho}) = -\frac{2\rho}{T}, \quad M(\hat{\rho}) = \frac{1-\rho^2}{T} + \frac{14\rho^2 - 1}{T^2}.$$

In obtaining these results we also note that y here is no longer a normal vector, and the expectations of terms needed are not in quadratic forms so we cannot use the expectation results in Appendix A.2, see the proof of Theorem 4 as given in Appendix A.3. We note that the results for the absolute regression model in Theorem 4 are the same as those for the pure AR(1) model given in Corollary 1 and in Section 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 the second-order bias and MSE results under LS estimation take the same functional form and involve only the even moments, and thereby we have the same second-order bias and MSE. Of course, equality of the first two moments of $\hat{\rho}$ does not suggest the same distribution of $\hat{\rho}$ for the two models.

4. Conclusions

We have developed analytical results on the properties of estimators in time-series models. General results on the second-order bias and MSE are given and applications of these results to some time-series models are also analyzed. We indicate that our general results are valid for both normal and nonnormal observations. We investigate the effects of nonnormality on the second-order bias results for two of these models, while for all the four models the second-order bias and MSE results are derived under normality. 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.

Acknowledgments

We thank Peter Robinson, two anonymous referees, and an associate editor for their extensive comments and constructive suggestions that have led to improvement of our paper. We would also like to thank Robin Carter, David Findley, Gloria González-Rivera, Bernard Gress, Emma Iglesias, Ivana Komunjer, Hashem Pesaran, Garry Phillips, Qiwei Yao, Victoria Zinde-Walsh, as well as conference participants at the 2003 NBER/NSF Time Series Conference (Chicago) and seminar participants at UCR, UC Irvine, and Cambridge University for very helpful comments. Ullah thanks the UCR Academic Senate for research support. All remaining errors are our own.

Appendix A

A.1. Alternative formulation of the bias and the MSE

Substituting (3) into (4) and (5) and denoting $d = Q\psi_T$, one can write (4) as

$$B(\hat{\beta}) = Q[\overline{Vd} - \tfrac{1}{2}\overline{H_2}(\overline{d \otimes d})], \tag{A.1}$$

and (5) as

$$M(\hat{\beta}) = A_{-1} + A_{-3/2} + A_{-2}, \tag{A.2}$$

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

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

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

$$\begin{aligned}
A_{-2} = {} & Q\Big[\overline{Vdd'V'} + \tfrac{1}{4}\overline{H_2}\overline{(d \otimes d)(d' \otimes d')}\overline{H_2}' - \tfrac{1}{2}\overline{Vd(d' \otimes d')}\overline{H_2}' - \tfrac{1}{2}\overline{H_2}\overline{(d \otimes d)d'V'}\Big]Q \\
& + Q\Big[\overline{VQVdd'} - \tfrac{1}{2}\overline{VQ\overline{H_2}(d \otimes d)d'} + \tfrac{1}{2}\overline{W(d \otimes d)d'} - \tfrac{1}{2}\overline{H_2}\overline{(d \otimes (QVd))d'} \\
& + \tfrac{1}{4}\overline{H_2}\overline{((Q\overline{H_2}(d \otimes d)) \otimes d)d'} - \tfrac{1}{2}\overline{H_2}\overline{((QVd) \otimes d)d'} \\
& + \tfrac{1}{4}\overline{H_2}\overline{(d \otimes (Q\overline{H_2}(d \otimes d)))d'} - \tfrac{1}{6}\overline{H_3}\overline{(d \otimes d \otimes d)d'}\Big] \\
& + \Big[\overline{dd'VQV} - \tfrac{1}{2}\overline{d(d' \otimes d')\overline{H_2}'QV} + \tfrac{1}{2}\overline{d(d' \otimes d')W'} - \tfrac{1}{2}\overline{d((d'VQ) \otimes d')}\overline{H_2}' \\
& + \tfrac{1}{4}\overline{d(d' \otimes ((d' \otimes d')\overline{H_2}'Q))}\overline{H_2}' - \tfrac{1}{2}\overline{d(d' \otimes (d'VQ))}\overline{H_2}' \\
& + \tfrac{1}{4}\overline{d(((d' \otimes d')\overline{H_2}'Q) \otimes d')}\overline{H_2}' - \tfrac{1}{6}\overline{d(d' \otimes d' \otimes d')}\,\overline{H_3}'\Big]Q.
\end{aligned}$$

When the observations are i.i.d., the bias result (A.1) reduces to Proposition 3.2 in Rilstone et al. (1996), and the MSE result (A.2) reduces to the corrected version of Proposition 3.4 in Rilstone et al. (1996). See Rilstone and Ullah (2005) for the corrected expression of the MSE of Rilstone et al. (1996) and its impact on other results in their paper.

A.2. Expectations of quadratic forms

For any symmetric matrix N_i, Magnus (1978, 1979), among others, derived the following results for $\varepsilon \sim \mathrm{N}(0, I)$:

$$\mathbb{E}(\varepsilon' N_1 \varepsilon) = \mathbf{tr}(N_1),$$

$$\mathbb{E}(\varepsilon' N_1 \varepsilon \cdot \varepsilon' N_2 \varepsilon) = \mathbf{tr}(N_1)\,\mathbf{tr}(N_2) + 2\,\mathbf{tr}(N_1 N_2),$$

$$\begin{aligned}\mathbb{E}(\varepsilon' N_1 \varepsilon \cdot \varepsilon' N_2 \varepsilon \cdot \varepsilon' N_3 \varepsilon) = {} & \mathbf{tr}(N_1)\,\mathbf{tr}(N_2)\,\mathbf{tr}(N_3) + 8\,\mathbf{tr}(N_1 N_2 N_3) \\ & + 2[\mathbf{tr}(N_1)\,\mathbf{tr}(N_2 N_3) + \mathbf{tr}(N_2)\,\mathbf{tr}(N_1 N_3) + \mathbf{tr}(N_3)\,\mathbf{tr}(N_1 N_2)],\end{aligned}$$

$$\begin{aligned}&\mathbb{E}(\varepsilon' N_1 \varepsilon \cdot \varepsilon' N_2 \varepsilon \cdot \varepsilon' N_3 \varepsilon \cdot \varepsilon' N_4 \varepsilon) \\ &\quad = \mathbf{tr}(N_1)\,\mathbf{tr}(N_2)\,\mathbf{tr}(N_3)\,\mathbf{tr}(N_4) + 8[\mathbf{tr}(N_1)\,\mathbf{tr}(N_2 N_3 N_4) \\ &\qquad + \mathbf{tr}(N_2)\,\mathbf{tr}(N_1 N_3 N_4) + \mathbf{tr}(N_3)\,\mathbf{tr}(N_1 N_2 N_4) + \mathbf{tr}(N_4)\,\mathbf{tr}(N_1 N_2 N_3)] \\ &\qquad + 4[\mathbf{tr}(N_1 N_2)\,\mathbf{tr}(N_3 N_4) + \mathbf{tr}(N_1 N_3)\,\mathbf{tr}(N_2 N_4) + \mathbf{tr}(N_1 N_4)\,\mathbf{tr}(N_2 N_3)] \\ &\qquad + 2[\mathbf{tr}(N_1)\,\mathbf{tr}(N_2)\,\mathbf{tr}(N_3 N_4) + \mathbf{tr}(N_1)\,\mathbf{tr}(N_3)\,\mathbf{tr}(N_2 N_4) + \mathbf{tr}(N_1)\,\mathbf{tr}(N_4)\,\mathbf{tr}(N_2 N_3) \\ &\qquad + \mathbf{tr}(N_2)\,\mathbf{tr}(N_3)\,\mathbf{tr}(N_1 N_4) + \mathbf{tr}(N_2)\,\mathbf{tr}(N_4)\,\mathbf{tr}(N_1 N_3) + \mathbf{tr}(N_3)\,\mathbf{tr}(N_4)\,\mathbf{tr}(N_1 N_2)] \\ &\qquad + 16[[\mathbf{tr}(N_1 N_2 N_3 N_4) + \mathbf{tr}(N_1 N_2 N_4 N_3) + \mathbf{tr}(N_1 N_3 N_2 N_4)].\end{aligned}$$

For a nonnormal vector ε with each element ε_t i.i.d. and following (9), we have $\mathbb{E}(\varepsilon' N_1 \varepsilon \cdot \varepsilon) = \sigma^3 \gamma_1 (I \odot N_1)\iota$, and $\mathbb{E}(\varepsilon' N_1 \varepsilon \cdot \varepsilon' N_2 \varepsilon) = \sigma^4[\gamma_2\,\mathbf{tr}(N_1 \odot N_2) + \mathbf{tr}(N_1)\,\mathbf{tr}(N_2) + 2\,\mathbf{tr}(N_1 N_2)]$, see Ullah (2004).

A.3. Proof

Proof of Theorem 1. Using (11) in (4) or (A.1), we get $B(\hat\rho) = Q^2\overline{H_1\psi_T} - 2Q\overline{\psi_T} = 2\bar N/\bar D - \overline{ND}/\bar D^2$, where $N = y_{-1}'M\varepsilon = y_D'M\varepsilon + \varepsilon' MC\varepsilon$ and $D = y_{-1}'My_{-1} = y_D'My_D + 2\varepsilon'C'My_D + \varepsilon'C'MC\varepsilon$. Now using the results from Appendix A.2, we have $\bar N = \sigma^2\,\mathbf{tr}(MC)$, $\bar D = r_D'r_D + \sigma^2\,\mathbf{tr}(C'MC)$, $\overline{ND} = \bar N\bar D + 2\sigma^2[r_D'Cr_D + \sigma^2\,\mathbf{tr}(C'MCMC)] + \gamma_1\sigma^3\iota'[I\odot (C'MC)]r_D + 2\gamma_1\sigma^3\iota'[I\odot(MC)]C'r_D + \gamma_2\sigma^4\,\mathbf{tr}[(MC)\odot(C'MC)]$. Grubb and Symons (1987) showed that $\mathbf{tr}(C'MC) = T/(1-\rho^2) + \mathrm{O}(T)$. Similarly, we can verify that $\mathbf{tr}(MC) = \mathrm{O}(1)$, $\iota'[I\odot(MC)]C'r_D = \mathrm{O}(T)$, $\iota'[I\odot(C'MC)]r_D = \mathrm{O}(T)$, $\mathbf{tr}[(MC)\odot(C'MC)] = \mathrm{O}(1)$, as long as y_0 is bounded or bounded a.s., and $\bar D = \mathrm{O}(T)$. Theorem 1 now follows. □

Proof of Corollary 1. For the pure AR(1) model, $M = I$, $r_D = y_D = y_0 F$. On substituting $y_{-1} = y_D + C\varepsilon$, we have $\mathbb{E}(y_{-1}'\varepsilon) = \mathbf{E}(y_0F'\varepsilon + \varepsilon'C'\varepsilon) = \mathbf{tr}(C') = 0$ since y_0 is independent of ε. Similarly, we can verify that

$$\bar D = \mathbb{E}(y_{-1}'y_{-1}) = \mathbb{E}(\varepsilon'C'C\varepsilon) + \mathrm{O}(1) = \sigma^2\,\mathbf{tr}(C'C) + \mathrm{O}(1) = \sigma^2 T/(1-\rho^2) + \mathrm{O}(1),$$

$$\begin{aligned}\overline{ND} &= \mathbb{E}(y_{-1}'y_{-1}\cdot y_{-1}'\varepsilon) = \mathbb{E}(\varepsilon'C'C\varepsilon\cdot\varepsilon'C\varepsilon) + \mathrm{O}(1) = 2\sigma^4\,\mathbf{tr}(C'CC) + \mathrm{O}(1) \\ &= 2\sigma^4 T\rho/(1-\rho^2)^2 + \mathrm{O}(1).\end{aligned}$$

Then the second-order bias result $-2\rho/T$ follows. □

Proof of Corollary 2. Since $M = I - \iota\iota'/T$, we can simplify

$$r_D'r_D = \beta^2\iota'C'C\iota - \beta^2(\iota'C\iota)^2/T + \mathrm{O}(1),$$

$$r_D'Cr_D = \beta^2\iota'C'CC\iota - \beta^2\iota'C'\iota\iota'CC\iota/T - \beta^2\iota'C'C\iota\iota'C\iota/T + \beta^2(\iota'C\iota)^3/T^2 + \mathrm{O}(1),$$

$$\imath'[I \odot (C'MC)]r_D = \imath'[I \odot (C'C)]C\imath\beta - \imath'[I \odot (C'C)]\imath\imath'C\imath\beta/T - \imath'[I \odot (C'\imath\imath'C)]C\imath\beta/T + \imath'[I \odot (C'\imath\imath'C)]\imath\imath'C\imath\beta/T^2 + \mathrm{O}(1),$$

$$\imath'[I \odot (MC)]C'r_D = \imath'[I \odot C]C'C\imath\beta - \imath'[I \odot C]C'\imath\imath'C'C\imath\beta/T - \imath'[I \odot (\imath\imath'C)]C'C\imath\beta/T + \imath'[I \odot (\imath\imath'C)]C'\imath\imath'C\imath\beta/T^2 + \mathrm{O}(1),$$

where

$$\imath'C\imath = \imath'C'\imath = T/(1-\rho) + \mathrm{O}(1), \quad \imath'C'C\imath = \imath'CC\imath + \mathrm{O}(1) = T/(1-\rho)^2 + \mathrm{O}(1),$$

$$\imath'C'CC\imath = T/(1-\rho)^3 + \mathrm{O}(1), \quad \imath'[I \odot (C'C)]C\imath = T/(1-\rho^2)(1-\rho) + \mathrm{O}(1),$$

$$\imath'[I \odot (\imath\imath'C)]C'C\imath = T/(1-\rho^2)(1-\rho)^2 + \mathrm{O}(1), \quad \imath'[I \odot C]C'C\imath\beta = \mathrm{O}(1),$$

$$\imath'[I \odot (C'\imath\imath'C)]\imath\imath'C\imath = T^2/(1-\rho)^3 + \mathrm{O}(T), \quad \imath'[I \odot C]C'\imath\imath'C'C\imath\beta = \mathrm{O}(1),$$

$$\imath'[I \odot (C'C)]\imath\imath'C\imath = T^2/(1-\rho^2)(1-\rho) + \mathrm{O}(T),$$

$$\imath'[I \odot (C'\imath\imath'C)]C\imath = T/(1-\rho)^2(1-\rho) + \mathrm{O}(1),$$

$$\imath'[I \odot (\imath\imath'C)]C'\imath\imath'C\imath = T^2/(1-\rho)^3 + \mathrm{O}(T).$$

Substituting the above results yields $r'_D r_D = \mathrm{O}(1)$, $r'_D C r_D = \mathrm{O}(1)$, $\imath'[I \odot (C'MC)]r_D = \mathrm{O}(1)$, and $\imath'[I \odot (MC)]C'r_D = \mathrm{O}(1)$. Then Corollary 2 follows. $\square$

Proof of MSE of $\hat{\rho}$ in Pure AR(1) Model under Normality. From Grubb and Symons (1987), $CC' = (I + \rho C + \rho C' - FF')/(1-\rho^2)$, where FF' always leads to terms of smaller order. Putting $y_0 = 0$, substituting CC' and noting $\mathbf{tr}(C^i) = 0$ for any integer i, we have the following, in which $C^* = (C + C')/2$:

$$\mathbb{E}(\varepsilon'C\varepsilon \cdot \varepsilon'C\varepsilon) = \sigma^4\,\mathbf{tr}(C'C) = T/(1-\rho^2) - 1/(1-\rho^2)^2 + \mathrm{o}(1),$$

$$\begin{aligned}\mathbb{E}(\varepsilon'C'C\varepsilon \cdot \varepsilon'C\varepsilon \cdot \varepsilon'C\varepsilon) &= \sigma^6[8\,\mathbf{tr}(C'C \cdot C^{*2}) + 2\,\mathbf{tr}(C'C)\,\mathbf{tr}(C^{*2})] \\ &= T^2/(1-\rho^2)^2 + (8\rho^2+2)/(1-\rho^2)^3 + \mathrm{o}(T),\end{aligned}$$

$$\begin{aligned}&\mathbb{E}(\varepsilon'C'C\varepsilon \cdot \varepsilon'C'C\varepsilon \cdot \varepsilon'C\varepsilon \cdot \varepsilon'C\varepsilon) \\ &\quad = \sigma^8\{16\,\mathbf{tr}(C'C)\,\mathbf{tr}(C'C \cdot C^{*2}) + 4\,\mathbf{tr}[(C'C)^2]\,\mathbf{tr}(C^{*2}) \\ &\qquad + 8[\mathbf{tr}(C'C \cdot C^*)]^2 + 2[\mathbf{tr}(C'C)]^2\,\mathbf{tr}(C^{*2}) + 32\,\mathbf{tr}[(C'C)^2 C^{*2}] + 16\,\mathbf{tr}[(C'C \cdot C^*)^2]\} \\ &\quad = T^3/(1-\rho^2)^3 + T^2(7+26\rho^2)/(1-\rho^2)^4 + \mathrm{o}(T^2).\end{aligned}$$

Since $M(\hat{\rho}) = 6Q^2\overline{\psi_T^2} - 8Q^3\overline{H_1\psi_T^2} + 3Q^4\overline{H_1^2\psi_T^2}$ and $Q = -T[\mathbb{E}(\varepsilon'C'C\varepsilon)]^{-1} = -(1-\rho^2)[1 + 1/\ T(1-\rho^2)]/\sigma^2 + \mathrm{o}(T^{-1})$, $\overline{\psi_T^2} = \mathbb{E}(\varepsilon'C\varepsilon \cdot \varepsilon'C\varepsilon)/T^2 + \mathrm{o}(T^{-2})$, $\overline{H_1\psi_T^2} = -\mathbb{E}(\varepsilon'C'C\varepsilon \cdot \varepsilon'C\varepsilon \cdot \varepsilon'C\varepsilon)/\ T^3 + \mathrm{o}(T^{-2})$, $\overline{H_1^2\psi_T^2} = \mathbb{E}(\varepsilon'C'C\varepsilon \cdot \varepsilon'C'C\varepsilon \cdot \varepsilon'C\varepsilon \cdot \varepsilon'C\varepsilon)/T^4 + \mathrm{o}(T^{-2})$, the second-order MSE of $\hat{\rho}$, up to $\mathrm{O}(T^{-2})$, is then given by $(1-\rho^2)/T + (14\rho^2 - 1)/T^2$.

Proof of Theorem 2. On plugging (15)–(17) into (4) or (A.1) and observing $\mathbf{tr}(N_1) = 0$, we have the bias vector

$$B\begin{pmatrix} \hat{\phi} \\ \hat{\sigma}^2 \end{pmatrix} = \begin{pmatrix} \dfrac{\mathbb{E}(\varepsilon' N_1 \varepsilon \cdot \varepsilon' N_2 \varepsilon)}{[\sigma^2\,\mathbf{tr}(N_2)]^2} - \dfrac{\mathbf{tr}(N_3)\mathbb{E}(\varepsilon' N_1 \varepsilon \cdot \varepsilon' N_1 \varepsilon)}{2\sigma^4[\mathbf{tr}(N_2)]^3} \\ -\dfrac{\mathbb{E}(\varepsilon' N_1 \varepsilon \cdot \varepsilon' N_1 \varepsilon)}{T\sigma^2\,\mathbf{tr}(N_2)} \end{pmatrix},$$

where $\mathbb{E}(\varepsilon' N_1 \varepsilon \cdot \varepsilon' N_1 \varepsilon) = \sigma^4\,\mathbf{tr}(N_1' N_1)$, $\mathbb{E}(\varepsilon' N_1 \varepsilon \cdot \varepsilon' N_2 \varepsilon) = \sigma^4\,\mathbf{tr}[(N_1 + N_1')N_2]$ by using A.2 and observing $\mathbf{tr}(N_1 \odot N_1) = \mathbf{tr}(N_1 \odot N_2) = 0$ (so the terms involving γ_2 in $\mathbb{E}(\varepsilon' N_1 \varepsilon \cdot \varepsilon' N_2 \varepsilon)$ and $\mathbb{E}(\varepsilon' N_1 \varepsilon \cdot \varepsilon' N_1 \varepsilon)$ disappear). Some algebra leads to $\mathbf{tr}(N_2) = \mathbf{tr}(N_1' N_1) = T/(1-\phi^2) + \mathrm{O}(1)$, $\mathbf{tr}(N_3) = 6T\phi/(1-\phi^2)^2 + \mathrm{O}(1)$, $\mathbf{tr}(N_1 N_2) = \mathbf{tr}(N_1' N_1 N_1) = T\phi/(1-\phi^2)^2 + \mathrm{O}(1)$, $\mathbf{tr}[N_1' N_2] = 3\,\mathbf{tr}(N_1 N_2)$. Upon substitution, $B(\hat{\phi}) = \phi/T$ and $B(\hat{\sigma}^2) = -\sigma^2/T$. The MSE expression also follows by substitution. □

Proof of Theorem 3. Substitute ψ_T and H_1 into (4) and (5) or (A.1) and (A.2), and use the results from Appendix A.2. Note that $H_2 = H_3 = W = 0$. □

Proof of Theorem 4. Using (22) in (4) or (A.1) we can write $B(\hat{\rho}) = Q^2\mathbb{E}[\sum_{t=1}^{T}|y_{t-1}|^2 \sum_{s=1}^{T}|y_{s-1}|(y_s + \rho|y_{s-1}|)]/T^2$. Now note that $\mathbb{E}[|y_{t-1}|^2|y_{s-1}|(y_s + \rho|y_{s-1}|)] = \mathbb{E}(|y_{t-1}|^2 |y_{s-1}|\varepsilon_s)$, which is equal to zero if $t \leqslant s$, and when $t > s$, it can be derived from the recursion $\mathbb{E}(|y_{t-1}|^2|y_{s-1}|\varepsilon_s) = \rho^2\mathbb{E}(|y_{t-2}|^2|y_{s-1}|\varepsilon_s)$ for $t > s+1$, and $\mathbb{E}(|y_{t-1}|^2|y_{s-1}|\varepsilon_s) = -2m_2\rho$ for $t = s+1$, which gives $\mathbb{E}(|y_{t-1}|^2|y_{s-1}|\varepsilon_s) = -2m_2\rho^{2(t-s)-1}$ for $t > s$. It is easy to verify that $\sum_{t>s}^{T}\sum_{s=1}^{T}\rho^{2(t-s)} = [T\rho^2(1-\rho^2) - \rho^2(1-\rho^{2T})]/(1-\rho^2)^2$. By substitution, we have $B(\hat{\rho}) = -2\rho/T$. Now we try to derive the three terms for the MSE: $\mathbb{E}(\psi_T^2)$, $\mathbb{E}(H_1\psi_T^2)$, and $\mathbb{E}(H_1^2\psi_T^2)$. We can easily show that $\mathbb{E}(\psi_T^2) = m_2/T$, $\mathbb{E}(H_1\psi_T^2) = \sum\sum\sum\mathbb{E}(|y_{t-1}|^2 |y_{t'-1}|\varepsilon_{t'}|y_{t''-1}|\varepsilon_{t''})/T^3$, $\mathbb{E}(H_1^2\psi_T^2) = \sum\sum\sum\sum\mathbb{E}(|y_{t-1}|^2|y_{t'-1}|^2|y_{t''-1}|\varepsilon_{t''}|y_{t'''-1}|\varepsilon_{t'''})/T^4$. Note that $\mathbb{E}(|y_{t-1}|^2|y_{t'-1}|\varepsilon_{t'}\ |y_{t''-1}|\varepsilon_{t''}) \neq 0$ only if (i) $t' = t'' > t$, (ii) $t = t' = t''$, (iii) $t > t' = t''$, (iv) $t > t' > t''$, and $\mathbb{E}(|y_{t-1}|^2|y_{t'-1}|^2|y_{t''-1}|\ \varepsilon_{t''}|y_{t'''-1}|\varepsilon_{t'''}) \neq 0$ only if (i) $t = t' = t'' = t'''$, (ii) $t = t'' = t''' > t'$ (or $t' = t'' = t''' > t$ due to symmetry), (iii) $t = t' > t'' = t'''$, (iv) $t'' = t''' > t = t'$, (v) $t = t' > t'' > t'''$, (vi) $t'' = t''' > t > t'$ (or $t'' = t''' > t' > t$ due to symmetry), (vii) $t > t' > t'' > t'''$ (or $t' > t > t'' > t'''$, $t > t' > t''' > t''$, $t' > t > t''' > t''$ due to symmetry). The exact expressions for these different cases can be derived following similar techniques as when we derive $\mathbb{E}[|y_{t-1}|^2|y_{s-1}|(y_s + \rho|y_{s-1}|)]$, i.e. by using appropriate recursions, also see Bao and Ullah (2002) for these expressions. The MSE result then follows. □

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., Netuka, I., Svara, K., 1984. On threshold autoregressive processes. Kybernetika 20, 89–106.

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

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

Bao, Y., Ullah, A., 2002. The second-order bias and mean squared error of estimators in time series models. Working paper, University of California, Riverside.

Basmann, R.L., 1974. Exact finite sample distribution for some econometric estimators and test statistics: a survey and appraisal. In: Intriligator, M.D., Kendrick, D.A. (Eds.), Frontiers of Quantitative Economics, vol. 2. North-Holland, Amsterdam, pp. 209–288.

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

Cordeiro, G.M., Klein, R., 1994. Bias correction in ARMA models. Statistics and Probability Letters 19, 169–176.

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

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

Grubb, D., Symons, J., 1987. Bias in regression with a lagged dependent variable. Econometric Theory 3, 371–386.

Kendall, M.G., 1954. Note on bias in estimation of autocorrelation. Biometrika 41, 403–404.

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

Kiviet, J.F., Phillips, G.D.A., 1994. Bias assessment and reduction in linear error-correction models. Journal of Econometrics 63, 215–243.

Kiviet, J.F., Phillips, G.D.A., Schipp, B., 1999. Alternative bias approximations in first-order dynamic reduced form models. Journal of Economic Dynamics and Control 23, 909–928.

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

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

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.

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., Smith, R.J., 2004. Higher order properties of GMM and generalized empirical likelihood estimators. Econometrica 72, 219–255.

Nicholls, D.F., Pope, A.L., 1988. Bias in the estimation of multivariate autoregressions. Australian Journal of Statistics 30A, 296–309.

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

Pope, A.L., 1990. Biases of estimators in multivariate non-Gaussian autoregressions. Journal of Time Series Analysis 11, 249–258.

Rilstone, P., Ullah, A., 2005. Corrigendum to "The second-order bias and mean squared error of nonlinear estimators": (Journal of Econometrics 75(2) (1996) 369–395). Journal of Econometrics 124, 203–204.

Rilstone, P., Srivastava, V.K., Ullah, A., 1996. The second-order bias and mean squared error of nonlinear estimators. Journal of Econometrics 75, 369–395.

Roberston, C.A., Fryer, J.G., 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: Intriligator, M.L., Griliches, Z. (Eds.), Handbook of Econometrics, vol. 2. North-Holland, Amsterdam, pp. 881–935.

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 approximation. Econometrica 44, 421–448.

Sawa, T., 1978. The exact moments of the least squares estimator for the autoregressive model. Journal of Econometrics 8, 159–172.

Shenton, L.R., Johnson, W.L., 1965. Moments of a serial correlation coefficient. Journal of the Royal Statistical Society, Series B 27, 308–320.

Tong, H., 1990. Non-linear Time Series: A Dynamical System Approach. Oxford University Press, New York.

Ullah, A., 2004. Finite Sample Econometrics. Oxford University Press, New York.

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

White, J.S., 1961. Asymptotic expansions for the mean and variance of the serial correlation coefficient. Biometrika 48, 85–94.