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

†Department of Economics, University of California, Riverside, CA, 92521, (909) 788-7998, e-mail: yong-bao@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 Robertson 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

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

(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 $\beta$ of $k$ elements such that $E[q_t(\beta)] = 0$. 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 $O_P(T^{-3/2})$, in Rilstone et al. (1996) derived for the i.i.d. case can be extended to the non-i.i.d. case as

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

(2.2)
where \( a_{-s/2} \) represents terms of order \( O_p(T^{-s/2}) \) for \( s = 1, 2, 3 \) and they are

\[
\begin{align*}
a_{-1/2} &= -Q\psi_T, \\
a_{-1} &= -QV a_{-1/2} - \frac{1}{2}QH_2[a_{-1/2} \otimes a_{-1/2}], \\
a_{-3/2} &= -QV a_{-1} - \frac{1}{2}QW[a_{-1/2} \otimes a_{-1/2}] - \frac{1}{2}QH_2\left\{[a_{-1/2} \otimes a_{-1}] + [a_{-1} \otimes a_{-1/2}]\right\} \\
&\quad - \frac{1}{6}QH_3[a_{-1/2} \otimes a_{-1/2} \otimes a_{-1/2}],
\end{align*}
\]

(2.3)
in which \( X = E(X) \) denotes the expectation of a random vector \( X \), \( H_i = \nabla^i \psi_T, i = 1, 2, 3, \)
\( Q = H_1^{-1}, V = H_1 - H_1 = \frac{1}{T} \sum_{t=1}^T V_t, W_t = H_2 - H_2 = \frac{1}{T} \sum_{t=1}^T W_t, \otimes \) 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 \( \hat{\beta} \), \( B(\hat{\beta}) \), up to \( O(T^{-1}) \), and the MSE
of \( \hat{\beta} \), up to \( O(T^{-2}) \), respectively, are

\[
\begin{align*}
B(\hat{\beta}) &= E(a_{-1/2}) + E(a_{-1}), \\
M(\hat{\beta}) &= E(a_{-1/2}a'_{-1/2}) + E(a_{-1}a'_{-1} + a_{-1/2}a'_{-1}) \\
&\quad + E(a_{-1}a'_{-1} + a_{-3/2}a'_{-1/2} + a_{-1/2}a'_{-3/2}),
\end{align*}
\]

(2.4)
where \( E(a_{-1/2}) = 0. \)

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

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

(2.5)
and

\[
M(\hat{\beta}) = A_{-1} + A_{-3/2} + A_{-2}
\]

(2.6)
where \( A_{-s/2} = O(T^{-s/2}) \), \( s = 2, 3, 4 \) are

\[
A_{-1} = dd', \\
A_{-3/2} = -Q \left[ Vdd' + \frac{1}{2}H_2(d \otimes d) d' \right] - \left[ dd'V + \frac{1}{2}d(d' \otimes d')H_2' \right] Q, \\
A_{-2} = Q \left[ VdV' + \frac{1}{4}H_2 (d \otimes d')(d' \otimes d')H_2' - \frac{1}{2}Vd(d' \otimes d')H_2' - \frac{1}{2}H_2 (d \otimes d')d'V' \right] Q
\]
reduce to the following results:

\[ B(\hat{\beta}) = \frac{1}{T} Q \left\{ V_1 d_1 - \frac{1}{2} H_2 [d_1 \otimes d_1] \right\}, \]  
(2.7)

\[ M(\hat{\beta}) = \frac{1}{T} \nu_1 + \frac{1}{T^2} (\nu_2 + \nu_2') + \frac{1}{T^2} (\nu_3 + \nu_4 + \nu_4'), \]  
(2.8)

where \( \nu_1 = E(\nabla q_1)^{-1} E(q_1 q_1') E(\nabla q_1)^{-1}', \)

\[ \nu_2 = Q \left\{ -V_1 d_1 d_1' + \frac{1}{2} H_2 [d_1 \otimes d_1] d_1' \right\}, \]

\[ \nu_3 = Q \left\{ V_1 d_1 d_2' V_2 + V_1 d_2 d_1' V_2 + V_1 d_2 d_2' V_1 \right\} \]

\[ +\frac{1}{4} Q H_2 \left\{ d_1 \otimes d_1 d_2 \otimes d_2' + [d_1 \otimes d_2] [d_1' \otimes d_1''] + [d_1 \otimes d_2] [d_2' \otimes d_1'] \right\} H_2 Q, \]

\[ -\frac{1}{2} Q \left\{ V_1 d_1 d_2' d_2 + V_1 d_2 d_1' d_2 + V_1 d_2 d_2' d_1 \right\} H_2 Q, \]

\[ -\frac{1}{2} Q H_2 \left\{ d_1 \otimes d_1 d_2' V_1 + [d_1 \otimes d_2] d_1' V_1 + [d_1 \otimes d_2] d_2' V_1 \right\} Q, \]

\[ \nu_4 = Q \left\{ V_1 Q V_1 d_2 d_2' + V_1 Q V_2 d_1 d_2' + V_1 Q V_2 d_2 d_1' \right\} \]

\[ -\frac{1}{2} Q \left\{ V_1 Q H_2 [d_1 \otimes d_2] d_2' + V_1 Q H_2 [d_2 \otimes d_1] d_2' + V_1 Q H_2 [d_2 \otimes d_2] d_1' \right\}, \]

\[ +\frac{1}{2} Q \left\{ W_1 [d_1 \otimes d_2] d_2' + W_1 [d_1 \otimes d_1] d_2' + W_1 [d_2 \otimes d_2] d_1' \right\}, \]

\[ -\frac{1}{2} Q H_2 \left\{ [d_1 \otimes Q V_1 d_2] d_2' + [d_1 \otimes Q V_2 d_1] d_2' + [d_1 \otimes Q V_2 d_2] d_1' \right\}, \]

\[ +\frac{1}{2} Q H_2 \left\{ [d_1 \otimes Q H_2 [d_1 \otimes d_2]] d_2' + [d_1 \otimes Q H_2 [d_2 \otimes d_1]] d_2' + [d_1 \otimes Q H_2 [d_2 \otimes d_2]] d_1' \right\}, \]

\[ -\frac{1}{2} Q H_2 \left\{ [Q V_1 d_1 \otimes d_2] d_2' + [Q V_1 d_2 \otimes d_1] d_2' + [Q V_1 d_2 \otimes d_2] d_1' \right\}, \]

\[ +\frac{1}{2} Q H_2 \left\{ [Q H_2 [d_1 \otimes d_1] d_2] d_2' + [Q H_2 [d_1 \otimes d_2] d_1] d_2' + [Q H_2 [d_1 \otimes d_2] d_2] d_1' \right\} \]

\[ -\frac{1}{6} Q H_3 \left\{ [d_1 \otimes d_1 \otimes d_2] d_2' + [d_1 \otimes d_2 \otimes d_1] d_2' + [d_1 \otimes d_2 \otimes d_2] d_1' \right\}. \]

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 $\psi_T$ 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

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

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

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

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

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

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

where $\bar{\beta}$ lies between $\hat{\beta}$ and $\beta$. 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,
\[ E \left( \nabla \psi_T (\hat{\beta}) \left( \hat{\beta} - \beta \right) \right) = E \left( \nabla \psi_T (\hat{\beta}) \right) E \left( \hat{\beta} - \beta \right) + Cov \left( \nabla \psi_T (\hat{\beta}), \hat{\beta} - \beta \right) \]
\[ \quad = E \left( \nabla \psi_T (\hat{\beta}) \right) E \left( \hat{\beta} - \beta \right) + (-\nabla \psi_T^{-1}) Cov (\psi_T, \nabla \psi_T (\hat{\beta})) + o(T^{-1}) \]
\[ \quad = E \left( \nabla \psi_T (\hat{\beta}) \right) E \left( \hat{\beta} - \beta \right) + (-\nabla \psi_T^{-1}) Cov (\psi_T, \nabla \psi_T) + o(T^{-1}), \]
and
\[ E \left( \nabla^2 \psi_T (\hat{\beta}) \left[ \left( \hat{\beta} - \beta \right) \otimes \left( \hat{\beta} - \beta \right) \right] \right) = E \left( \nabla^2 \psi_T (\hat{\beta}) \right) E \left( \left[ \left( \hat{\beta} - \beta \right) \otimes \left( \hat{\beta} - \beta \right) \right] \right) + o(T^{-1}) \]
\[ \quad = E \left( \nabla^2 \psi_T (\hat{\beta}) \right) E \left( \left[ \left( \hat{\beta} - \beta \right) \otimes \left( \hat{\beta} - \beta \right) \right] \right) + o(T^{-1}), \]
where \( \nabla \psi_T = \nabla \psi_T (\beta) \), \( Cov (\psi_T, \nabla \psi_T (\hat{\beta})) \) corresponds to \( J \) in Cox and Snell (1968). Then \( E \left( \nabla \psi_T (\hat{\beta}) \left( \hat{\beta} - \beta \right) \right) \) and \( E \left( \nabla^2 \psi_T (\hat{\beta}) \left[ \left( \hat{\beta} - \beta \right) \otimes \left( \hat{\beta} - \beta \right) \right] \right) \) 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 \( E (a_{-1}) \), 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'), we find
\[ \{QVd\}_s = \{QH_1Q\psi_T\}_s = \sum_{i=1}^{p} \sum_{j=1}^{p} \sum_{l=1}^{p} k^{ijl} k^{ijl}, \]

\[ \frac{1}{2} \{QH_2 (d \otimes d)\}_s = \frac{1}{2} \left\{ Q \left( Q \otimes Q \right) \left( \psi_T \otimes \psi_T \right) \right\}_s \]
\[ \quad = -\frac{1}{2} \sum_{i=1}^{p} \sum_{j=1}^{p} \sum_{l=1}^{p} k^{ijl} k^{ijl}, \quad (2.12) \]
for the single parameter estimator \( \hat{\beta}_s \), \( 1 \leq s \leq p \), 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
\[ B \left( \hat{\beta} \right) = Q \left[ \nabla d - \frac{1}{2} H_2 (d \otimes d) \right] = K^{-1} \text{Avec}(K^{-1}). \quad (2.13) \]
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).

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

\[ y_t = \beta y_{t-1} + \varepsilon_t, \quad (3.1) \]

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

\(^1\)We find that the second-order bias for ML estimator in ARCH(1) model is

\[ Q \left\{ V_1 d_1 - \frac{1}{2} H_2 [d_1 \otimes d_1] \right\} / T - Q \left( \sum_{i>j} V_i d_j \right) / T, \]

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 $\beta$ is equivalent to the (conditional) ML estimator. That is, to estimate, we use the following moment condition

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

(3.2)

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

$$
H_1 = \nabla \psi_T = \frac{y'C_1 y}{T-1}, \quad H_2 = H_3 = 0,
$$

$$
Q = H_1^{-1} = \left( \frac{tr C_1 \Sigma}{T-1} \right)^{-1}, \quad V = H_1 - H_1, \quad W = 0,
$$

(3.3)

where $C_1 = \nabla C = \{\nabla c_{tt'}\}$ with $\nabla c_{tt'} = -1$ for $t = t' = 1, 2, \ldots, T - 1$ 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 $O(T^{-1})$, and MSE, up to $O(T^{-2})$, as

$$
B(\hat{\beta}) = \frac{1}{(T - 1)^2} Q^2 \lambda_{11},
$$

$$
M(\hat{\beta}) = \frac{6 Q^2}{(T - 1)^2} \lambda_{20} + \frac{2 Q^3}{(T - 1)^3} (1 + 3 Q^2) \lambda_{21} + \frac{3 Q^4}{(T - 1)^4} \lambda_{22},
$$

(3.4)

where $\lambda_{rs} = E[(y' Cy)^r \cdot (y' C_1 y)^s]$ for $r, s = 0, 1, 2$ are given in Appendix A.2. Note that in Appendix A.2 we further simplify, up to $o(T^{-1})$,

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

(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

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

(3.6)

See, for example, Kiviet and Phillips (1993).
Remark 2: The exact log likelihood function (excluding a constant term) when we normalize 
\( \sigma^2 \epsilon = 1 \) is

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

(3.7)

and the score function is

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

(3.8)

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

\[
\psi_T = \frac{1}{T} y' C^* y - \frac{1}{T} \frac{\beta}{1 - \beta^2} = 0,
\]

(3.9)

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

\[
\nabla \psi_T = \frac{y' D y}{T} - \frac{y'C_1 y}{T} - \frac{1}{T} \frac{1 + \beta^2}{(1 - \beta^2)^2}, \quad Q = \frac{T (1 - \beta^2)^2}{(T - 3) \beta^2 + 1 - T},
\]

\[
H_2 = -\frac{1}{T} \frac{6 \beta + 2 \beta^3}{(1 - \beta^2)^3}, \quad H_3 = -\frac{1}{T} \frac{6 + 36 \beta^2 + 6 \beta^4}{(1 - \beta^2)^4}, \quad W = 0.
\]

(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

\[
y_{1t} = \beta y_{2t} + \epsilon_t, \quad y_{2t} = \pi x_t + v_{2t},
\]

(3.11)

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

\[
y_{1t} = x_t \pi \beta + v_{1t}, \quad v_{1t} = \epsilon_t + \beta v_{2t}.
\]

(3.12)

We assume that \( E u_t = E v_{2t} = 0, E v^2_{2t} = \omega_{22}, E u^2_t = \sigma^2_u, E (u_t v_{2t}) = E (\epsilon_t v_{2t}) = \omega_{u2}, E \epsilon_t^2 = \sigma^2_\epsilon = \sigma^2_u / (1 - \rho^2) \), and all the third (cross) moments are zero. Also we assume \( u \) and
\( v_2 \) are independent at all lags and leads so that \( E(\varepsilon_t v_{2t}) = \omega_{u2}\rho^{t-t'} \). Then, assuming joint normality of \( v_t = (v_{1t}, v_{2t})' \) and denoting \( y_t = (y_{1t}, y_{2t})' \), we note that

\[
Ey = \mu_y, \quad V(y) = \Omega_y, \quad (3.13)
\]

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

\[
\Omega_y = \begin{bmatrix} \Omega_{11} & \Omega_{12} \\ \Omega_{12} & \Omega_{22} \end{bmatrix}; \quad (3.14)
\]

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

The moment condition for estimating \( \beta \) neglecting the dependence structure in the disturbances is

\[
\psi_T = \frac{1}{T} \sum_{t=1}^{T} (y_{1t} - \beta y_{2t}) x_t = \frac{1}{T} x' C y = 0, \quad (3.15)
\]

where \( C = \begin{pmatrix} 1 & -\beta \end{pmatrix} \otimes I_T \) and \( E x' C y = 0 \). Thus

\[
H_1 = \nabla \psi_T = \frac{1}{T} x' C_1 y, \quad H_2 = H_3 = 0, \quad Q = (H_1)^{-1} = \left( \frac{1}{T} x' C_1 \mu_y \right)^{-1} = \left( -\frac{x' x_\pi}{T} \right)^{-1}, \quad V = H_1 - H_1, \quad W = 0, \quad (3.16)
\]

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 \( O(T^{-1}) \), and the MSE, up to \( O(T^{-2}) \), respectively, of \( \hat{\beta} \) estimated from (3.15) when \( x_t \) is nonstochastic, are

\[
B(\hat{\beta}) = \frac{-\omega_{u2} x' \Sigma x}{\pi^2 (x' x)^2}, \quad (3.17)
\]

\[
M(\hat{\beta}) = \frac{\sigma^2_u}{1-\rho^2} \frac{x' \Sigma x}{\pi^2 (x' x)^2} + \frac{6\omega_{u2}^2 (x' \Sigma x)^2}{\pi^4 (x' x)^4} + \frac{3\sigma^2_u \omega_{22}}{1-\rho^2} \frac{(x' \Sigma x)}{\pi^4 (x' x)^3}. \quad (3.17)
\]

**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, \( \varepsilon_t \) is i.i.d., then the bias result in (3.17) reduces to

\[
B \left( \hat{\beta} \right) = -\frac{\omega u_2}{\pi^2 \sum_{t=1}^{T} x_t^2}.
\]

(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

\[
\psi_T = \frac{1}{T} x' C y = \frac{1}{T} Z'C'Z = 0
\]

(3.19)

where we redefine

\[
C = \begin{bmatrix} I & 0 \\ 0 & -\beta I \end{bmatrix}, \quad C^* = \frac{1}{2} \begin{bmatrix} 0 & C \\ C & 0 \end{bmatrix},
\]

(3.20)

and \( x^* = (x', x')', \quad Z = (x'', y')' \). 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 \( \mu_Z = (\mu'_{x'}, \mu'y)' \) and

\[
\Sigma_Z = \begin{bmatrix} \Omega_{x^x} & \Omega_{x^y} \\ \Omega_{y^x} & \Omega_{y^y} \end{bmatrix},
\]

(3.21)

where \( \mu_{x^x} = (\mu'_{x'}, \mu'_{x'})', \quad \mu_y = ((\mu_x \pi \beta)', (\mu_x \pi))' \), and \( \Omega_{x^*} \) is a \( 2T \times 2T \) block matrix with each block element being \( \Sigma_x, \Omega_{x^y} = \Omega'_{y^x} \), and

\[
\Omega_{y^x} = \begin{bmatrix} \pi \beta \Sigma_x & \pi \beta \Sigma_x \\ \pi \Sigma_x & \pi \Sigma_x \end{bmatrix}.
\]

(3.22)

Then \( H_2 = H_3 = 0, \quad W = 0 \), and

\[
H_1 = \frac{1}{T} Z' C_1^* Z,
\]

\[
Q^{-1} = \frac{1}{T} E (Z'C_1^* Z) = -\frac{1}{T} \pi \mu'_{x'} \mu_x - \frac{1}{T} \pi \text{tr} (\Sigma_x),
\]

(3.23)
where $C_1^*$ is $C^*$ with $C$ replaced by

$$C_1 = \begin{bmatrix} 0 & 0 \\ 0 & -I_T \end{bmatrix}. \quad (3.24)$$

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

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

$$B(\hat{\beta}) = \frac{(-\omega u_2 tr (\Sigma x) - \omega u_2 \mu_x' \Sigma x)}{\pi^2 [\mu_x' \mu_x + tr (\Sigma x)]^2},$$

$$M(\hat{\beta}) = 3 \frac{Q^4}{T^4} \lambda_{11} - 2 \frac{Q^2}{T^2} \lambda_{10}. \quad (3.25)$$

where $\lambda_{ij} = E[(Z' C^* Z)^{2i} (Z' C_1^* Z)^{2j}]$ 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. $(0, \sigma_\eta^2)$, $\sigma_x^2 = \sigma_\eta^2 / (1 - \rho_x^2)$, then the bias result in (3.25) reduces to

$$B(\hat{\beta}) = \frac{-\omega u_2 \left[ T \sigma_x^2 + \sum_{t=1}^T \sum_{t' = 1, t' \neq t} (\rho \rho_x \sigma_x^2)^{t-t'} \right]}{\pi^2 (T \sigma_x^2)^2}. \quad (3.26)$$

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

$$B(\hat{\beta}) = -\frac{\omega u_2}{T \pi^2 E x_1^2}. \quad (3.27)$$
Note that (3.27) encompasses the result in Rilstone et al. (1996) in that even though \( \varepsilon_t \) 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

\[
\begin{align*}
    y_{1t} &= \beta_{11} y_{1,t-1} + \beta_{12} y_{2,t-1} + \varepsilon_{1t}, \\
    y_{2t} &= \beta_{21} y_{1,t-1} + \beta_{22} y_{2,t-1} + \varepsilon_{2t}, \\
\end{align*}
\]

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

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

\[
y_t = By_{t-1} + \varepsilon_t.
\]

We assume that the process is covariance stationary, that is, for all \( \lambda \) 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 \( \Gamma_0 = E(y_t y_t'), \Gamma_j = E(y_t y_{t-j}'), \Gamma_{-j} = E(y_{t-j} y_t') = \Gamma_j \). 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, \( y_t \sim N(0, vec^{-1}[(I_4 - B \otimes B)^{-1} vec(\Omega)]) \). If we stack all observations as \( y = (y_1', y_2', \ldots, y_T') \), then we have \( y \sim N(0, \Sigma) \), where \( \Sigma \) 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 \( t, t' = 1, 2, \ldots, T \).

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 \( \beta \equiv vec(B) \)

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

(3.30)
where

\[
C_{11} = \begin{pmatrix}
0_{2(T-1) \times 2} & AA \\
0_{2 \times 2} & 0_{2 \times 2(T-1)}
\end{pmatrix} - \beta_{11} \begin{pmatrix}
AA & 0_{2(T-1) \times 2} \\
0_{2 \times 2(T-1)} & 0_{2 \times 2}
\end{pmatrix} - \beta_{12} \begin{pmatrix}
BB & 0_{2(T-1) \times 2} \\
0_{2 \times 2(T-1)} & 0_{2 \times 2}
\end{pmatrix},
\]

\[
C_{21} = \begin{pmatrix}
0_{2(T-1) \times 2} & BB \\
0_{2 \times 2} & 0_{2 \times 2(T-1)}
\end{pmatrix} - \beta_{21} \begin{pmatrix}
AA & 0_{2(T-1) \times 2} \\
0_{2 \times 2(T-1)} & 0_{2 \times 2}
\end{pmatrix} - \beta_{22} \begin{pmatrix}
BB & 0_{2(T-1) \times 2} \\
0_{2 \times 2(T-1)} & 0_{2 \times 2}
\end{pmatrix},
\]

\[
C_{12} = \begin{pmatrix}
0_{2(T-1) \times 2} & CC \\
0_{2 \times 2} & 0_{2 \times 2(T-1)}
\end{pmatrix} - \beta_{11} \begin{pmatrix}
CC & 0_{2(T-1) \times 2} \\
0_{2 \times 2(T-1)} & 0_{2 \times 2}
\end{pmatrix} - \beta_{12} \begin{pmatrix}
DD & 0_{2(T-1) \times 2} \\
0_{2 \times 2(T-1)} & 0_{2 \times 2}
\end{pmatrix},
\]

\[
C_{22} = \begin{pmatrix}
0_{2(T-1) \times 2} & DD \\
0_{2 \times 2} & 0_{2 \times 2(T-1)}
\end{pmatrix} - \beta_{21} \begin{pmatrix}
CC & 0_{2(T-1) \times 2} \\
0_{2 \times 2(T-1)} & 0_{2 \times 2}
\end{pmatrix} - \beta_{22} \begin{pmatrix}
DD & 0_{2(T-1) \times 2} \\
0_{2 \times 2(T-1)} & 0_{2 \times 2}
\end{pmatrix},
\]

where the \(2(T-1) \times 2(T-1)\) matrices \(AA, BB, CC,\) and \(DD\) are block diagonal with the block element \(aa = ((1,0)^T, (0,0)^T)^T, bb = ((0,1)^T, (0,0)^T)^T, cc = ((0,0)^T, (1,0)^T)^T,\) and \(dd = ((0,0,0)^T)\), 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 \(l\)th 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

\[
\nabla \psi_T = H_1 = -\frac{1}{T-1} \begin{pmatrix}
y'C_{11}^{(2)}y & 0 & y'C_{11}^{(3)}y & 0 \\
0 & y'C_{21}^{(2)}y & 0 & y'C_{21}^{(3)}y \\
y'C_{12}^{(2)}y & 0 & y'C_{12}^{(3)}y & 0 \\
0 & y'C_{22}^{(2)}y & 0 & y'C_{22}^{(3)}y
\end{pmatrix},
\]

\[
H_2 = H_3 = W = 0. \tag{3.31}
\]

The results on quadratic forms in Appendix A.1 then can be used here to evaluate the second-order bias and MSE of \(\hat{\beta}.\) Denote \(E\left(y'C_{ij}^{(l)}y\right) = tr \left[(C_{ij}^{(l)} + C_{ij}^{(l)})\Sigma/2\right] = \chi_{ij}^{(l)},\)

\[
E\left(y'C_{ij}y \cdot y'C_{mn}^{(l)}y\right) = 2tr \left[(C_{ij} + C_{ij})\Sigma (C_{mn}^{(l)} + C_{mn}^{(l)})\Sigma/4\right] + tr \left[(C_{ij} + C_{ij})\Sigma/2\right] \times tr \left[(C_{mn}^{(l)} + C_{mn}^{(l)})\Sigma/2\right] = \mu_{ijmn}^{(l)} \text{ for } i, j, m, n = 1, 2 \text{ and } \lambda_{i1}^{(3)}\lambda_{i2}^{(2)} - \lambda_{i1}^{(2)}\lambda_{i2}^{(3)} = \phi_i \text{ for } i = 1, 2,
\]

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

\[
B \left(\hat{\beta}_{i1}\right) = \frac{\xi_i}{\phi_i}, \quad B \left(\hat{\beta}_{i2}\right) = \frac{\eta_i}{\phi_i}, \quad i = 1, 2, \tag{3.32}
\]
where
\[
\xi_i = \left( \mu^{(3)}_{12i2} - \mu^{(3)}_{i12i} \right) \lambda^{(2)}_{i1} \lambda^{(3)}_{i1} + \left( \mu^{(2)}_{i12i} - \mu^{(3)}_{i12i} \right) \lambda^{(3)}_{i1} \lambda^{(3)}_{i2} + \mu^{(2)}_{i1i12} \lambda^{(3)}_{i1} \lambda^{(3)}_{i2} + \mu^{(3)}_{i1i1i1} \lambda^{(2)}_{i1} \lambda^{(3)}_{i2} \\
- \mu^{(2)}_{i2i1} \lambda^{(3)}_{i2} \lambda^{(3)}_{i1} - \mu^{(2)}_{i2i2} \lambda^{(3)}_{i1} \lambda^{(3)}_{i2} ,
\]
\[
\eta_i = \left( \mu^{(3)}_{i1i2} - \mu^{(3)}_{i2i1} \right) \lambda^{(2)}_{i2} \lambda^{(3)}_{i2} + \left( \mu^{(2)}_{i2i2} - \mu^{(3)}_{i2i2} \right) \lambda^{(3)}_{i1} \lambda^{(3)}_{i2} + \mu^{(2)}_{i1i1i1} \lambda^{(3)}_{i1} \lambda^{(3)}_{i2} + \mu^{(3)}_{i2i1i1} \lambda^{(2)}_{i2} \lambda^{(3)}_{i1} \\
- \mu^{(3)}_{i2i1} \lambda^{(2)}_{i2} \lambda^{(3)}_{i1} - \mu^{(3)}_{i2i2} \lambda^{(2)}_{i1} \lambda^{(3)}_{i2} .
\]

**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 \( B = ((0.8, 0.3)', (0.2, 0.5)') \) and \(((0.1, 0.4)', (0.4, 0.3)')\), where eigenvalues refer to those of the parameter matrix \( B \) and we normalize \( \Omega = ((1, \rho)', (\rho, 1)') \). 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 \( \rho \), 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 second-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 \( \Omega = ((1, \rho)', (\rho, 1)') \), 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

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

### 3.4 MA Model

Consider the simplest case

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

where we normalize \(\varepsilon_t \sim \text{i.i.d.} \ N(0,1)\). 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

\[
A_t = A_t(\beta) = \left( a_{t-1}, \cdots, a_1, a_0, 0, \cdots, 0 \right)' \tag{3.35}
\]

Then we have

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

where

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

and

\[
q_t = -\left( \frac{\partial \varepsilon_t}{\partial \beta} \right) \varepsilon_t = -y'B_t A_t' y,
\]

\[
\nabla q_t = -\left( \frac{\partial^2 \varepsilon_t}{\partial \beta^2} \right) \varepsilon_t - \left( \frac{\partial \varepsilon_t}{\partial \beta} \right)^2 = -y'C_t A_t' y - y'B_t B_t' y,
\]

\[
\nabla^2 q_t = -\left( \frac{\partial^3 \varepsilon_t}{\partial \beta^3} \right) \varepsilon_t - 3 \left( \frac{\partial^2 \varepsilon_t}{\partial \beta^2} \right) \left( \frac{\partial \varepsilon_t}{\partial \beta} \right) = -y'U_t A_t' y - 3y'C_t B_t' y,
\]

\[
\nabla^3 q_t = -\left( \frac{\partial^4 \varepsilon_t}{\partial \beta^4} \right) \varepsilon_t - 4 \left( \frac{\partial^3 \varepsilon_t}{\partial \beta^3} \right) \left( \frac{\partial \varepsilon_t}{\partial \beta} \right) - 3 \left( \frac{\partial^2 \varepsilon_t}{\partial \beta^2} \right)^2 = -y'V_t A_t' y - 4y'U_t B_t' y - 3y'C_t C_t' y. \tag{3.38}
\]
In matrix notation, we have
\[ \psi_T = \frac{1}{T} y'y N, \quad H_i = \frac{1}{T} y'_i N_i y, \quad i = 1, 2, 3, \]
(3.39)
where \( N = -A'B, \quad N_1 = -A'C - B'B, \quad N_2 = -A'U - 3B'C, \quad N_3 = -A'V - 4B'U - 3C'C, \)
\( A = (A_1, A_2, \cdots, A_T) \) and similarly for \( B, C, U, V. \) Note that \( y \) is normally distributed with
mean 0 and variance-covariance matrix \( \Sigma \) 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 \( E(a_{-1/2}) = 0 \) 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 \( C. \)
Therefore, instead of (2.5), we use the bias expression in (2.4) directly with \( E(a_{-1/2}) \neq 0. \) 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 \( O(T^{-1}) \), of the conditional ML estimator \( \hat{\beta} \) for model (3.34)
is
\[
B \left( \hat{\beta} \right) = \frac{\text{tr} (N^* \Sigma) \cdot \text{tr} (N_1^* \Sigma) + 2 \text{tr} (N^* \Sigma N_1^* \Sigma)}{[\text{tr} (N_1^* \Sigma)]^2} - 2 \frac{\text{tr} (N^* \Sigma)}{\text{tr} (N_1^* \Sigma)} \cdot \frac{\text{tr} (N_2^* \Sigma) \cdot [\text{tr} (N^* \Sigma)^2 + 2 \text{tr} ((N^* \Sigma)^2)]}{2 [\text{tr} (N_1^* \Sigma)]^3},
\]
(3.40)
where \( N_i^* = \frac{N_i + N'_i}{2}, \quad i = 0, 1, 2, \quad N_0 = N, \quad N_0^* = N^*. \)

**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 \( \sqrt{T} \)-consistent estimator, \( g > 0. \) But (2.3), (2.5) and (2.6) then have
to be modified accordingly.
As for $M(\hat{\beta})$, we do not report the explicit result here. But it is very easy to write a computer program to evaluate it. Figure 1 plots $B(\hat{\beta})$ and Figure 2 plots $M(\hat{\beta})$ against $\beta$ for given sample size $T$. We see clearly the following patterns for $B(\hat{\beta})$ and $M(\hat{\beta})$: i) $B(\hat{\beta})$ and $M(\hat{\beta})$ 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 $\beta$ between $(\beta_{B1}, \beta_{B2})$ 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 $\beta_{B1}$ for $\beta < 0$ and at $\beta_{B2}$ 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 $\beta$ increases until it reaches two peaks at $\beta_{M1}$ for $\beta < 0$ and $\beta_{M2}$ for $\beta > 0$, and then decreases. The exact values of $\beta_{B1}$, $\beta_{B2}$, $\beta_{M1}$, and $\beta_{M2}$ 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 $\varepsilon$ 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 $2 \times 1$, $\nabla q_t$ is $2 \times 2$, $\nabla^2 q_t$ is $2 \times 4$, and $\nabla^3 q_t$ is $2 \times 8$. All the expectations of cross products will boil down to the form of cross products of $E(\partial^4 \varepsilon / \partial \theta_i \partial \theta_j \partial \theta_k \partial \theta_l)$, $1 \leq i, j, k, l \leq 2$, or of lower order. For example, consider $E(\partial^4 \varepsilon / \partial \theta_1^4)$. From $\varepsilon_t = y_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2}$ we have $\partial \varepsilon_t / \partial \theta_1 = -\varepsilon_{t-1} - \theta_1 \varepsilon_{t-1} / \partial \theta_1 - \theta_2 \varepsilon_{t-2} / \partial \theta_1$. Rewrite it as $\varepsilon_{t-1} = -\partial \varepsilon_t / \partial \theta_1 - \theta_1 \varepsilon_{t-1} / \partial \theta_1 - \theta_2 \varepsilon_{t-2} / \partial \theta_1$. Therefore, we have $\varepsilon = A_1 \varepsilon / \partial \theta_1$, 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 $t$, $t' = 1, 2, \ldots, T$. Then $\partial \varepsilon / \partial \theta_1 = A_1^{-1} \varepsilon = C_1 \varepsilon$ for $C_1 = A_1^{-1}$. Invertibility will guarantee
Remark 5: We can easily extend to the case of an MA(q) with mean $x_t' \beta$. 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 $\beta$ 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 $y_t - y_{t-1} = (1 - \gamma) (y_t^* - y_{t-1})$. 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 $y_t$

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

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 $\varepsilon$. Note that even though $\gamma$ and $\delta$ 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 $B(\hat{\beta})$ and $M(\hat{\beta})$.

The OLS estimator needs the following moment condition

$$
\psi_T = \frac{1}{T - 1} \left( \frac{1}{T - 1} \right) = \frac{1}{T - 1} \left( y' C y - (1 - \gamma) A \beta^* A y ight) = 0, \quad (3.42)
$$

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 $t, t' = 1, 2, \ldots, T$, and $D = B - \gamma A$.

Then we have

$$
\nabla \psi_T = H_1 = \frac{1}{T - 1} \left( \begin{array}{cc}
- (1 - \gamma) x' A y & y' C y + \beta x'A y \\
- (1 - \gamma) x'B x & - x'A y + \beta x'B x
\end{array} \right),
$$

$$
\nabla^2 \psi_T = H_2 = \frac{1}{T - 1} \left( \begin{array}{cc}
0 & x' A y \\
0 & x'B x
\end{array} \right),
$$

and

$$
\nabla^3 \psi_T = H_3 = 0. \quad (3.43)
$$

where $C_1$ is $T \times T$ with $tt'$th element equal to 1 if $t = t' = 1, 2, \ldots, T - 1$, 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 $x'A y$. For this we can define $y^* = (x', y')'$, then (3.42) and (3.43) can be rewritten as

$$
\psi_T = \frac{1}{T - 1} \left( y'^* C^* y^* - (1 - \gamma) \beta y'^* A y^* \\
y'^* D^* y^* - (1 - \gamma) \beta y'^* B y^*
\right) = 0, \quad (3.42')
$$

and

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

$$
\nabla^2 \psi_T = H_2 = \frac{1}{T - 1} \left( \begin{array}{cc}
y'^* A y^* & y'^* A y^* \\
y'^* B y^* & y'^* B y^*
\end{array} \right), \quad (3.43')
$$

where

$$
A^* = \frac{1}{2} \begin{pmatrix} 0 & A \\ A & 0 \end{pmatrix}, \quad B^* = \begin{pmatrix} B & 0 \\ 0 & 0 \end{pmatrix}, \quad C^* = \begin{pmatrix} 0 & 0 \\ 0 & C \end{pmatrix},
$$

$$
C_1^* = \begin{pmatrix} 0 & 0 \\ 0 & C_1 \end{pmatrix}, \quad D^* = \frac{1}{2} \begin{pmatrix} 0 & D \\ D & 0 \end{pmatrix}, \quad (3.44)
$$

20
and \( y^* \sim N(\mu, \Sigma) \), where
\[
\mu = \begin{pmatrix} \mu_x \tau_T \\ \beta \mu_x \tau_T \end{pmatrix}, \quad \Sigma = \begin{pmatrix} \sigma^2 \Sigma_T & \Sigma_{xy} \\ \Sigma_{xy} & \Sigma_y \end{pmatrix},
\]
(3.45)
where \( \tau_T \) is a \( T \times 1 \) vector of ones, \( \Sigma_y \) is the \( T \times T \) variance-covariance matrix of \( y \) with \( tt' \)th element equal to \( \gamma^{t-t'} \left[ \sigma^2 (\mu_x^2 + \sigma^2_x) + \sigma^2_y \right] / (1 - \gamma^2) - \beta^2 \mu_x^2 \), \( \Sigma_{xy} \) is the \( T \times T \) covariance matrix between \( x \) and \( y \) with \( tt' \)th element equal to \( \delta \gamma^{t-t'} (\sigma^2_x + \mu_x^2) - \delta \mu_x^2 \gamma^{t-t'} / (1 - \gamma) \) if \( t' \geq t \) and 0 elsewhere, for \( t, t' = 1, 2, \cdots, T \). Denote \( tr A^* \Sigma = a, tr B^* \Sigma = b, tr C^*_1 \Sigma = c, \mu' A^* \mu = k, \mu' B^* \mu = m, \mu' C^*_1 \mu = l \). With these notations, we have Theorem 5 for the second-order bias of \( \hat{\beta} \).

**Theorem 5:** The bias, up to \( O(T^{-1}) \), of the estimator \( \hat{\beta} \) from (3.42) for model (3.41) is
\[
B(\hat{\beta}) = \frac{\lambda_{12} - \lambda_{34} + \lambda_{56}}{(\gamma - 1)^2 [(a + k)^2 - (c + l)(b + m)]^2},
\]
(3.46)
where \( \lambda_{ij} = E (y^* N_i y^* \cdot y^* N_j y^*) \), \( i, j = 1, 2, \cdots, 6 \),
\[
N_1 = (\gamma - 1)C^* + \beta (\gamma - 1)^2 A^*,
\]
\[
N_2 = (a + k - 2 \beta b) (m C^*_1 - k A^*) + (c + l) [(a + k) B^* - (m + b) A^*] - \beta (b^2 + m^2) C^*_1
+ \beta k (2m A^* -kB^* - 2a B^*) + k (b C^*_1 - a A^*) - a^2 (A^* + B^*) + 2a \beta (b + m) A^* + ab C^*_1,
\]
\[
N_3 = (\gamma - 1)D^* + \beta (\gamma - 1)^2 B^*,
\]
\[
N_4 = (c + l)(2k A^* + \beta k B^* + 2a A^* + a \beta B^*) - (c + l)^2 B^* - (k + a)(\beta A^* + C^*_1)
+ \beta (a C^*_1 + k C^*_1 - l A^* - c A^*)(b + m),
\]
\[
N_5 = (b + m) C^* + \beta (\gamma - 1) [(b + m) A^* - (a + k) B^*] - (a + l) D^*,
\]
\[
N_6 = \beta (C^* - \beta A^*)(b + m) + \beta B^*(\beta k - l) - (a + k)(C^* + \beta D^*) - \beta (a A^* + a \beta B^* - c B^*)(\gamma - 1)
+ (c + l) D^* + \beta \gamma (l B^* - k A^*) + \beta^2 \gamma (b A^* + m A^* - k B^*).
\]
**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 \( \gamma \), 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

\[ y_t = -\beta |y_{t-1}| + \varepsilon_t, \]  

(3.47)

where \(0 < \beta < 1\), \(\varepsilon_t\) 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

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

(3.48)

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

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

(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

\[
\begin{align*}
m_1 &= -\beta \sqrt{\frac{2}{\pi (1 - \beta^2)}}, \quad m_2 = \frac{1}{1 - \beta^2}, \\
m_3 &= \frac{\beta (\beta^2 - 3)}{1 - \beta^2} \sqrt{\frac{2}{\pi (1 - \beta^2)}}, \quad m_4 = \frac{3}{(1 - \beta^2)^2}, \\
m_5 &= \frac{\beta (15 - 10\beta^2 + 3\beta^4)}{(1 - \beta^2)^2} \sqrt{\frac{2}{\pi (1 - \beta^2)}}, \quad m_6 = \frac{15}{(1 - \beta^2)^3}.
\end{align*}
\]  

(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, \(\beta\) is estimated by LS. That is, the moment condition is

\[
\psi_T = \frac{1}{T-1} \sum_{t=2}^{T} \frac{\partial \varepsilon_t}{\partial \beta} \varepsilon_t = \frac{1}{T-1} \sum_{t=2}^{T} |y_{t-1}| (y_t + \beta |y_{t-1}|) = 0.
\]  

(3.51)
Then following the notations as before, we have

\[ H_1 = \frac{1}{T-1} \sum_{t=2}^{T} |y_{t-1}|^2, \quad H_2 = H_3 = W = 0. \]  

(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

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

(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

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

(3.54)

It is easy to verify that

\[ \sum_{t=2}^{T} \sum_{t'=t+1}^{T} \beta^{2(t-t')} = \left[ \frac{\beta^2 (T-1)}{1-\beta^2} - \frac{\beta^2}{1-\beta^2} \frac{1-\beta^{2(T-1)}}{1-\beta^2} \right]. \]  

(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

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

(3.56)

where in Appendix A.4 we discuss how to evaluate \( \psi_T^2, H_1 \psi_T^2, \) 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 moments, 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 tr$A$ 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 $y \sim N(0, I)$,

\[
E (y'N_1y) = trN_1,
\]

\[
E (y'N_1y \cdot y'N_2y) = (tr N_1)(tr N_2) + 2tr(N_1N_2),
\]

\[
E (y'N_1y \cdot y'N_2y \cdot y'N_3y) = (tr N_1)(tr N_2)(tr N_3) + 8tr N_1N_2N_3 + 2[(tr N_1)(tr N_2N_3) + (tr N_2)(tr N_1N_3) + (tr N_3)(tr N_1N_2)],
\]

\[
E (y'N_1y \cdot y'N_2y \cdot y'N_3y \cdot y'N_4y) = (tr N_1)(tr N_2)(tr N_3)(tr N_4) + 2[(tr N_1)(tr N_2N_3N_4) + (tr N_2)(tr N_1N_3N_4) + (tr N_3)(tr N_1N_2N_4) + (tr N_4)(tr N_1N_2N_3)]
\]

\[
+ 4[(tr N_1N_2)(tr N_3N_4) + (tr N_1N_3)(tr N_2N_4) + (tr N_1N_4)(tr N_2N_3) + (tr N_2)(tr N_3N_4) + (tr N_3)(tr N_1N_4) + (tr N_4)(tr N_1N_3)]
\]

\[
+ 2[(tr N_1)(tr N_2)(tr N_3N_4) + (tr N_1)(tr N_2N_3)(tr N_4) + (tr N_1)(tr N_3N_4)(tr N_2) + (tr N_2)(tr N_3)(tr N_4) + (tr N_3)(tr N_2)(tr N_4) + (tr N_4)(tr N_1)(tr N_2)]
\]

\[
+ 16 [tr N_1N_2N_3N_4 + tr N_1N_2N_4N_3 + tr N_1N_3N_2N_4].
\]

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

\[
E (y'N_1y) = a_1 + \theta_1, \quad (A.1)
\]

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

\[
E [y'N_1y \cdot y'N_2y \cdot y'N_3y] = \prod_{i=1}^{3} (a_i + \theta_i)
\]

\[
+ \sum_{i=1}^{3} \sum_{i \neq j \neq k}^{3} [a_i a_{jk} + a_{ij} \theta_k + a_{ik} \theta_j + 3 \theta_{ijkl} + 4 \theta_{ijk} + 2(a_i + \theta_i) \theta_{jk}], \quad (A.3)
\]
\[ E (y'N_1 y \cdot y'N_2 y \cdot y'N_3 y \cdot y'N_4 y) = \prod_{i=1}^{4} (a_i + \theta_i) \]
\[ + \sum_{i=1}^{4} \sum_{j=1}^{4} \sum_{k=1}^{4} \sum_{l=1}^{4} \left( 2a_i a_j + 2a_i \theta_j + 2a_{ij} + \theta_i \theta_j + 4\theta_{ij} \right) \theta_{kl} \]
\[ + \left( \frac{1}{2} a_{ij} \theta_k + a_i a_{jk} + a_{ijk} \right) \theta_i + a_{kl} (a_i a_j + a_{ij}) + \frac{2}{\pi} a_{ijkl} \]
\[ + 8\theta_{ijkl} + \frac{2}{\pi} \left( \theta_i + \frac{1}{2} a_i \right) \theta_{jkl} + \frac{2}{\pi} \left( a_{jkl} + \frac{1}{2} \theta_{jkl} \right) a_i. \] 
(A.4)

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

A.2. AR(1) Model

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

\[ \lambda_{11} = E [(y'yC y) \cdot (y'yC_1 y)] = 2tr (C_1 \Sigma C \Sigma). \] 
(A.5)

Also, we write

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

and partition

\[ \Sigma = \frac{\sigma^2}{1 - \beta^2} \begin{pmatrix} \Omega_1 \\ \Sigma^*_T \end{pmatrix} = \frac{\sigma^2}{1 - \beta^2} \begin{pmatrix} \Sigma^*_1 \\ \Omega_2 \end{pmatrix}, \] 
(A.7)

where \( \Sigma^*_i \) is the \( i \)th row of \( \Sigma^* = \Sigma / \frac{\sigma^2}{1 - \beta^2} \), \( \Omega_1 \) is the 1st to \( (T - 1) \)th rows of \( \Sigma^* \) and \( \Omega_2 \) is the 2nd to \( T \)th rows of \( \Sigma^* \), then

\[ B(\beta) = \frac{2Q^2 tr (C_1 \Sigma C \Sigma)}{(T - 1)^2} \]
\[ = \frac{2}{(T - 1)^2} \left\{ \beta tr \left[ \begin{pmatrix} \Omega_1 \\ 0 \end{pmatrix} \begin{pmatrix} \Omega_1 \\ 0 \end{pmatrix} \right] - 2tr \left[ \begin{pmatrix} \Omega_2 \\ 0 \end{pmatrix} \begin{pmatrix} 0 \\ \Omega_1 \end{pmatrix} \right] \right\} \]
\[ = \frac{2}{(T - 1)^2} \left\{ \beta (T - 1) - 2\beta (T - 1) \right\} + o(T^{-1}) \]
\[ = \frac{-2\beta}{(T - 1)} + o(T^{-1}). \] 
(A.8)
Next, by using the results in Appendix A.1, it is straightforward to verify

\[
\lambda_{20} = E \left[ (y'Cy)^2 \right] = 2tr \left[ (C\Sigma)^2 \right],
\]

\[
\lambda_{21} = E \left[ (y'C_1y) \cdot (y'Cy)^2 \right]
= 8tr \left[ C_1\Sigma (C\Sigma)^2 \right] + 2tr(C_1\Sigma)tr \left[ (C\Sigma)^2 \right],
\]

\[
\lambda_{22} = E \left[ (y'C_1y)^2 \cdot (y'Cy)^2 \right]
= 16tr \left( C_1\Sigma \right) tr \left[ C_1\Sigma (C\Sigma)^2 \right] + 4tr \left[ (C_1\Sigma)^2 \right] tr \left[ (C\Sigma)^2 \right]
+ 8 [tr \left( C_1\Sigma C\Sigma \right)]^2 + 2 [tr \left( C_1\Sigma \right)]^2 tr \left[ (C\Sigma)^2 \right]
+ 32tr \left[ (C_1\Sigma)^2 \right] (C\Sigma)^2 + 16tr \left[ (C_1\Sigma C\Sigma)^2 \right]. \tag{A.9}
\]

### A.3. Simultaneous Equation Model

From (2.5) we have

\[
B \left( \hat{\beta} \right) = Q^\top d = Q(H_1 - Q^{-1}) Q\psi_T = Q^2 H_1 \psi_T
= \frac{Q^2}{T^2} E \left( x' C_1 y \cdot x' C y \right) = \frac{Q^2}{T^2} x' C \Omega_y C_x'
= -\frac{\omega_{12} x' \Sigma_x}{\pi^2 \left( x' x \right)^2}. \tag{A.10}
\]

Similarly

\[
A_{-1} = Q^2 \psi_T^2 = \frac{Q^2}{T^2} x' C \Omega_y C_x'
= \frac{\sigma_u^2}{1 - \rho^2} \frac{x' \Sigma_x}{\pi^2 \left( x' x \right)^2}. \tag{A.11}
\]

By symmetry, it is straightforward to prove that \( A_{-3/2} = 0 \). Furthermore, we can write

\[
A_{-2} = \frac{3Q^4}{T^2} E \left[ (x'C_1y)^2 \times (x'Cy)^2 \right] - 3A_{-1}.
\]

Note that \( E \left[ (x'C_1y)^2 \times (x'Cy)^2 \right] = E \left[ y'N_1y \cdot y'N_2y \right] \),
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,

\[ tr(N_1\Omega_y) = tr(xx'\Omega_{22}) = \omega_{22}x'x, \]

\[ tr(N_2\Omega_y) = tr[xx'(\Omega_{11} - 2\beta\Omega_{12} + \beta^2\Omega_{22})] = \frac{\sigma_u^2}{1 - \rho^2} x'Sx, \]

\[ tr(N_1\Omega_y N_2\Omega_y) = tr[xx'(\Omega_{21} - \beta\Omega_{22}) xx'(\Omega_{12} - \beta\Omega_{22})] = \omega_{u2}^2 (x'Sx)^2, \]

\[ \mu'_{yN_1\mu_y} = \pi^2 (x'x)^2, \]

\[ \mu'_{yN_2\mu_y} = 0, \]

\[ \mu'_{yN_1\Omega_y N_2\mu_y} = 0. \]  \hspace{1cm} (A.12)

Therefore, by substitution we have

\[ A_{-2} = \frac{3Q^4}{T^4} (x'Sx) \left[ \frac{\omega_{22}\sigma_u^2}{1 - \rho^2} (x'x) + 2\omega_{u2}^2 (x'Sx) + \frac{\pi^2\sigma_u^2}{1 - \rho^2} (x'x)^2 \right] - 3\frac{\sigma_u^2}{1 - \rho^2} x'Sx \]

\[ = \frac{6\omega_{u2}^2 (x'Sx)^2}{\pi^4 (x'x)^4} + \frac{3\sigma_u^2\omega_{22} (x'Sx) (x'x)}{1 - \rho^2 \pi^4 (x'x)^4}. \]  \hspace{1cm} (A.13)

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

When \( x \) is stochastic, we have

\[ B(\hat{\beta}) = \frac{Q^2}{T^2} [tr(C_1'\Omega_y C_1 \mu_x) + \mu_{x'}C_1'\Omega_y C_1 \mu_x + \mu_{y'}C_1'\Omega_y C_1 \mu_y + \mu_{x'}C_1'\Omega_y C_1 \mu_{x'} + \mu_{y'}C_1'\Omega_y C_1 \mu_{y'}] \]

\[ = \frac{Q^2}{T^2} (-\omega_{u2} tr(\Sigma \Sigma_x) - \omega_{u2} \mu_x' \Sigma \mu_x) \]

\[ = \frac{(-\omega_{u2} tr(\Sigma \Sigma_x) - \omega_{u2} \mu_x' \Sigma \mu_x)}{\pi^2 [\mu_x' \mu_x + tr(\Sigma \Sigma_x)]^2}, \]  \hspace{1cm} (A.14)

and

\[ A_{-1} = \frac{Q^2}{T^2} \lambda_{10} = \frac{Q^2}{T^2} E(Z'C^*Z \cdot Z'C^*C) \]

\[ = \frac{Q^2}{T^2} [2tr(C^*\Sigma Z' C^* Z) + 4\mu_x' C^* \Sigma Z' C^* C] = \frac{Q^2}{T^2} \frac{\sigma_u^2}{1 - \rho^2} [tr(\Sigma \Sigma_x) + \mu_x' \Sigma \mu_x] \]

\[ = \frac{\sigma_u^2}{1 - \rho^2 \pi^2 [\mu_x' \mu_x + tr(\Sigma \Sigma_x)]^2}, \]

\[ A_{-3/2} = 0, \]

\[ A_{-2} = 3\frac{Q^4}{T^4} \lambda_{11} - 3\frac{Q^2}{T^2} \lambda_{10} \]

\[ = 3\frac{Q^4}{T^4} E(Z'C^*Z \cdot Z'C^*Z \cdot Z'C^*Z \cdot Z'C^*Z) - 3A_{-1}. \]  \hspace{1cm} (A.15)
The exact expression for $\lambda_{11}$ 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. Absoluute Regression Model

First note that

$$E \left[ |y_{t-1}|^2 |y_{t-1}^e (y_t + \beta |y_{t-1}|) \right] = E \left( |y_{t-1}|^2 |y_{t-1}^e \varepsilon_t \right) := U. \quad (A.16)$$

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

$$E \left( |y_{t-1}|^2 |y_{t-1}^e \varepsilon_t \right) = \beta^2 E \left( |y_{t-2}|^2 |y_{t-1}^e \varepsilon_t \right) \quad \text{for } t > t' + 1, \quad \text{and}$$

$$E \left( |y_t|^2 |y_{t-1}^e \varepsilon_t \right) = -2m_2\beta. \quad (A.17)$$

Hence for $t > t'$

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

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

$$E \left( \psi_T^2 \right) = \frac{1}{T-1} E \left\{ |y_{t-2}| (y_t + \beta |y_{t-1}|) \right\}^2 = \frac{m_2}{T-1}.$$

$$E \left( H_1 \psi_T^2 \right) = \frac{1}{(T-1)^2} E \left\{ \left( \sum_{t=2}^T |y_{t-1}| \right)^2 \left( \sum_{t' = 2}^T |y_{t-1}^e (y_t + \beta |y_{t-1}|) \right) \right\}. \quad (A.19)$$

$$E \left( H_1^2 \psi_T^2 \right) = \frac{1}{(T-1)^4} E \left\{ \left( \sum_{t=2}^T |y_{t-1}|^2 \right)^2 \left( \sum_{t' = 2}^T |y_{t-1}^e (y_t + \beta |y_{t-1}|) \right)^2 \right\}. \quad (A.19)$$

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

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

$$E \left( |y_{t-1}|^2 |y_{t-1}^e \varepsilon_t^2 \right) = \beta^2 E \left( |y_{t-2}|^2 |y_{t-2}^e \varepsilon_t^2 \right) + m_2, \quad \text{for } t' > t + 1, \quad \text{and}$$

$$E \left( |y_{t-1}|^2 |y_t^2 \right) = \beta^2 m_2 + m_2. \quad (A.20)$$
Hence for \( t' = t'' > t \)

\[
V = \beta^{2(t'-t)}m_4 + m_2 \frac{1 - \beta^{2(t'-t)}}{1 - \beta^2}.
\]  
(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

\[
E \left( |y_{t-1}|^2 |y_{t'-1}|^2 \varepsilon_{t'}^2 \right) = \beta^2 E \left( |y_{t-2}|^2 |y_{t'-1}|^2 \varepsilon_{t'}^2 \right) + m_2 \text{ for } t > t' + 1, \text{ and}
\]

\[
E \left( |y_{t'}|^2 |y_{t'-1}|^2 \varepsilon_{t'}^2 \right) = \beta^2 m_4 + 3m_2.
\]  
(A.22)

Hence for \( t > t' = t'' \)

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

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

\[
E \left( |y_{t-1}|^2 |y_{t'-1}| \varepsilon_{t'} |y_{t''-1}| \varepsilon_{t''} \right) = \beta^{2} E \left( |y_{t-2}|^2 |y_{t'-1}| \varepsilon_{t'} |y_{t''-1}| \varepsilon_{t''} \right) \text{ for } t > t' + 1, \text{ and}
\]

\[
E \left( |y_{t'}|^2 |y_{t'-1}| \varepsilon_{t'} |y_{t''-1}| \varepsilon_{t''} \right) = \beta^2 m_4 + 3m_2.
\]  
(A.24)

Hence for \( t > t' > t'' \)

\[
V = 4m_2 \beta^{2(t'-t'')}. \]  
(A.25)

Also, for the third term \( E \left( |y_{t-1}|^2 |y_{t'-1}|^2 |y_{t''-1}| \varepsilon_{t'} |y_{t''-1}| \varepsilon_{t''} \right) := W_{t't't''} \neq 0 \) only if

(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

\[
E \left( |y_{t-1}|^4 |y_{t'-1}|^2 \right) = \beta^4 E \left( |y_{t-2}|^4 |y_{t'-1}|^2 \right) + 6\beta^2 E \left( |y_{t-2}|^2 |y_{t'-1}|^2 \right) + 3m_2
\]

\[
= \beta^4 E \left( |y_{t-2}|^4 |y_{t'-1}|^2 \right) + 6 \left[ \beta^{2(t'-t')}m_4 + m_2 \frac{1 - \beta^{2(t'-t')}}{1 - \beta^2} \right] + 3m_2
\]

for \( t > t' + 1 \) (substitute \( \text{(A.21)} \)), and

\[
E \left( |y_{t'}|^4 |y_{t'-1}|^2 \right) = \beta^4 m_6 + 6\beta^2 m_4 + 3m_2.
\]  
(A.26)
Hence for \( t = t'' = t''' > t' \)

\[
W = \beta^4(t-t') m_6 + 6 \left[ \beta^{2(t-t')} m_4 + m_2 \frac{1 - \beta^{2(t-t')}}{1 - \beta^2} \right] \frac{1 - \beta^{2(t-t'-1)}}{1 - \beta^2} \\
+ 6\beta^2 m_4 + 3m_2 \frac{1 - \beta^4(t-t')}{1 - \beta^4}
\]  

(A.27)

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

\[
E(\|y_{t-1}\|^4 | y_{t''-1}|^2 \varepsilon_{t''}^2) = \beta^4 E(\|y_{t-2}\|^4 | y_{t''-1}|^2 \varepsilon_{t''}^2) + 3m_2 + 6\beta^2 E(\|y_{t-2}\|^2 | y_{t''-1}|^2 \varepsilon_{t''}^2) \\
= \beta^4 E(\|y_{t-2}\|^4 | y_{t''-1}|^2 \varepsilon_{t''}^2) + 3m_2 \\
+ 6 \left[ \beta^{2(t-t')} m_4 + m_2 \frac{1 - \beta^{2(t-t'-1)}}{1 - \beta^2} + 3m_2 \right]
\]

for \( t > t'' + 1 \) (substitute (A.31)), and

\[
E(\|y_{t''}\|^4 | y_{t''-1}|^2 \varepsilon_{t''}^2) = 15m_2 + \beta^4 m_6 + 18\beta^2 m_4.
\]  

(A.28)

Hence for \( t = t' > t'' = t''' \)

\[
W = \beta^4(t-t') m_6 + 6 \left[ \beta^{2(t-t')} m_4 + m_2 \frac{1 - \beta^{2(t-t'-1)}}{1 - \beta^2} + 3m_2 \right] \frac{1 - \beta^{2(t-t'-1)}}{1 - \beta^2} \\
+ 18\beta^2 m_4 + 3m_2 \frac{1 - \beta^4(t-t')}{1 - \beta^4} + 15m_2.
\]  

(A.29)

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

\[
E(\|y_{t-1}\|^4 | y_{t''-1}|^2) = \beta^2 E(\|y_{t-1}\|^4 | y_{t''-1}|^2) + m_4 \text{ for } t'' > t + 1, \text{ and}
\]

\[
E(\|y_{t-1}\|^4 | y_t|^2) = \beta^2 m_6 + m_4.
\]  

(A.30)

Hence for \( t'' = t''' > t = t' \)

\[
W = \beta^{2(t''-t)} m_6 + m_4 \frac{1 - \beta^{2(t''-t)}}{1 - \beta^2}
\]  

(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

\[
E \left( |y_{t-1}|^4 |y_{t''-1}| \varepsilon_{t'} |y_{t''-1}| \varepsilon_{t''} \right) = \beta^4 E \left( |y_{t-2}|^4 |y_{t''-1}| \varepsilon_{t'} |y_{t''-1}| \varepsilon_{t''} \right) \\
+ 6 \beta^2 E \left( |y_{t-2}|^2 |y_{t''-1}| \varepsilon_{t'} |y_{t''-1}| \varepsilon_{t''} \right) \\
= \beta^4 E \left( |y_{t-2}|^4 |y_{t''-1}| \varepsilon_{t'} |y_{t''-1}| \varepsilon_{t''} \right) + 24 m_2 \beta^2 (t-t''-1)
\]

for \( t > t'' + 1 \) (substitute (A.25)), and

\[
E \left( |y_{t''}|^4 |y_{t''-1}| \varepsilon_{t'} |y_{t''-1}| \varepsilon_{t''} \right) = -4 \beta^3 E \left( |y_{t''-1}|^4 |y_{t''-1}| \varepsilon_{t''} \right) \\
- 12 \beta E \left( |y_{t''-1}|^2 |y_{t''-1}| \varepsilon_{t''} \right).
\]

(App. 32)

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

\[
E \left( |y_{t''-1}|^4 |y_{t''-1}| \varepsilon_{t''} \right) = \beta^4 E \left( |y_{t''-2}|^4 |y_{t''-1}| \varepsilon_{t''} \right) + 6 \beta^2 E \left( |y_{t''-2}|^2 |y_{t''-1}| \varepsilon_{t''} \right) \\
= \beta^4 E \left( |y_{t''-2}|^4 |y_{t''-1}| \varepsilon_{t''} \right) - 12 m_2 \beta^2 (t''-t''')^{-1}
\]

for \( t'' > t''' + 1 \) (substitute (A.18)), and

\[
E \left( |y_{t''}|^4 |y_{t''-1}| \varepsilon_{t''} \right) = -4 \beta^4 m_4 - 12 \beta m_2. \tag{A.33}
\]

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

\[
E \left( |y_{t''-1}|^2 |y_{t-1}|^2 |y_{t-1}|^2 \right) = \beta^2 E \left( |y_{t''-2}|^2 |y_{t-1}|^2 |y_{t-1}|^2 \right) + E \left( |y_{t-1}|^2 |y_{t-1}|^2 \right) \\
= \beta^2 E \left( |y_{t''-2}|^2 |y_{t-1}|^2 |y_{t-1}|^2 \right) + \beta^2 (t-t') m_4 + m^2 \frac{1 - \beta^2 (t-t')}{1 - \beta^2}
\]

for \( t'' > t + 1 \) (substitute (A.21)), and

\[
E \left( |y_{t}|^2 |y_{t-1}|^2 |y_{t-1}|^2 \right) = \beta^2 E \left( |y_{t-1}|^4 |y_{t-1}|^2 \right) + E \left( |y_{t-1}|^2 |y_{t-1}|^2 \right) \\
= 6 \beta^2 \left[ \beta^2 (t-t') m_4 + m^2 \frac{1 - \beta^2 (t-t')}{1 - \beta^2} \right] \frac{1 - \beta^2 (t-t'-1)}{1 - \beta^2}
\]

\[
\beta^4 (t-t')^2 m_6 + 6 \beta^4 m_4 + 3 \beta^2 m_2 \frac{1 - \beta^4 (t-t')}{1 - \beta^4} + \beta^2 (t-t') m_4
\]

for \( m_2 \frac{1 - \beta^2 (t-t')}{1 - \beta^2} \) (from (A.21) and (A.27)), \( \tag{A.34} \)
Hence it is straightforward to evaluate $W$ numerically for $t^\prime = t^\prime\prime > t > t^\prime$. 

vii) $t > t^\prime > t^\prime\prime > t^\prime\prime\prime$ (or $t^\prime > t > t^\prime\prime > t^\prime\prime\prime$, $t > t^\prime > t^\prime\prime > t^\prime\prime\prime$, $t^\prime > t > t^\prime\prime > t^\prime\prime\prime$); the results follow by exchanging the indices), $W = E \left( |y_{t-1}|^2 |y_{t^\prime-1}|^2 |y_{t^\prime\prime-1}| \epsilon_t \mid y_{t^\prime\prime\prime-1}| \epsilon_{t^\prime\prime\prime} \right)$ can be derived recursively from

$$E \left( |y_{t-1}|^2 |y_{t^\prime-1}|^2 |y_{t^\prime\prime-1}| \epsilon_t \mid y_{t^\prime\prime\prime-1}| \epsilon_{t^\prime\prime\prime} \right) = \beta^2 E \left( |y_{t-1}|^2 |y_{t^\prime-1}|^2 |y_{t^\prime\prime-1}| \epsilon_t \mid y_{t^\prime\prime\prime-1}| \epsilon_{t^\prime\prime\prime} \right)$$

$$+ E \left( |y_{t^\prime-1}|^2 |y_{t^\prime\prime-1}| \epsilon_{t^\prime} \mid y_{t^\prime\prime\prime-1}| \epsilon_{t^\prime\prime\prime} \right)$$

for $t > t^\prime + 1$, and

$$E \left( |y_{t^\prime-1}|^2 |y_{t^\prime\prime-1}|^2 |y_{t^\prime\prime\prime-1}| \epsilon_{t^\prime} \mid y_{t^\prime\prime\prime-1}| \epsilon_{t^\prime\prime\prime} \right) = \beta^2 E \left( |y_{t^\prime-1}|^4 |y_{t^\prime\prime-1}| \epsilon_{t^\prime} \mid y_{t^\prime\prime\prime-1}| \epsilon_{t^\prime\prime\prime} \right)$$

$$+ E \left( |y_{t^\prime-1}|^2 |y_{t^\prime\prime-1}| \epsilon_{t^\prime} \mid y_{t^\prime\prime\prime-1}| \epsilon_{t^\prime\prime\prime} \right),$$

where $E \left( |y_{t^\prime-1}|^4 |y_{t^\prime\prime-1}| \epsilon_{t^\prime} \mid y_{t^\prime\prime\prime-1}| \epsilon_{t^\prime\prime\prime} \right)$ follows from v) and $E \left( |y_{t^\prime-1}|^2 |y_{t^\prime\prime-1}| \epsilon_{t^\prime} \mid y_{t^\prime\prime\prime-1}| \epsilon_{t^\prime\prime\prime} \right)$ follows from (A.25). Hence it is straightforward to evaluate $W$ numerically for $t > t^\prime > t^\prime\prime > t^\prime\prime\prime$.

Therefore, $E \left( H_1 \psi_T^2 \right) = \frac{1}{(T-1)^3} \sum \sum \sum V_{t^\prime t^\prime\prime}$ and $E \left( H_1^2 \psi_T^2 \right) = \frac{1}{(T-1)^3} \sum \sum \sum \sum W_{t^\prime t^\prime\prime t^\prime\prime\prime}$ can easily be evaluated numerically.
References


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

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

Eigenvalues: 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 \( \hat{vec}(B) \) in VAR(1) Model with True Parameter \( B = ((0.1, 0.4)', (0.4, 0.3)')' \)

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

Eigenevalues: -0.2123, 0.6123.
Figure 1: Second-order Bias of Conditional ML Estimator of $\beta$ in MA(1) Model
Figure 2: Second-order MSE of Conditional ML Estimator of $\beta$ in MA(1) Model

- $T = 20$
- $T = 30$
- $T = 50$
- $T = 100$