Expectation of Quadratic Forms in Normal and Nonnormal Variables with Econometric Applications*

Yong Bao†
Department of Economics
Purdue University

Aman Ullah‡
Department of Economics
University of California, Riverside

June 2009

ABSTRACT

We derive some new results on the expectation of quadratic forms in normal and nonnormal variables. Using a nonstochastic operator, we show that the expectation of the product of an arbitrary number of quadratic forms in noncentral normal variables follows a recurrence formula. This formula includes the existing result for central normal variables as a special case. For nonnormal variables, while the existing results are available only for quadratic forms of limited order (up to 3), we derive analytical results to a higher order 4. We use the nonnormal results to study the effects of nonnormality on the finite sample mean squared error of the OLS estimator in an AR(1) model and the QMLE in an MA(1) model.

Keywords: Expectation; Quadratic form; Nonnormality

JEL Classification: C10; C19

*We are grateful to Peter Phillips for his comments on an earlier version of this paper. We are also thankful to Jason Abrevaya, Fathali Firoozi, seminar participants at Purdue University, and conference participants at the Midwest Econometrics Group Meeting (2007, St. Louis) and the Conference in Honor of Peter Phillips (2008, Singapore) for their comments and suggestions. All remaining errors are our own.

†Corresponding Author: Department of Economics, Purdue University, West Lafayette, IN 47907, (765) 496-2313, e-mail: ybao@purdue.edu.

‡Department of Economics, University of California, Riverside, CA 92521, e-mail: aman.ullah@ucr.edu.

1 Introduction

In evaluating statistical properties of a large class of econometric estimators and test statistics we often come across the problem of deriving the expectation of the product of an arbitrary number of quadratic forms in random variables. For example, see White (1957), Nagar (1959), Theil and Nagar (1961), Kadane (1971), Ullah, Srivastava, and Chandra (1983), Dufour (1984), Magee (1985), Hoque, Magnus, and Pesaran (1988), Kiviet and Phillips (1993), Smith (1993), Lieberman (1994), Srivastava and Maekawa (1995), Zivot, Startz, and Nelson (1998), and Pesaran and Yamagata (2005); also see the book by Ullah (2004). Econometric examples of the situations where the expectation of the product of quadratic forms can arise are: obtaining the moments of the residual variance; obtaining the moments of the statistics where the expectation of the ratio of quadratic forms is the ratio of the expectations of the quadratic forms, for example, the moments of the Durbin-Watson statistic; and obtaining the moments of a large class of estimators in linear and nonlinear econometric models (see Bao and Ullah, 2007a, 2007b), among others. In view of this econometricians and statisticians have long been interested in deriving $E\left(\prod_{i=1}^{n} Q_i\right)$, where $Q_i = y'A_iy$, A_i are nonstochastic symmetric matrices of dimension $m \times m$ (for asymmetric A_i we can always put $(A_i + A_i')/2$ in place of A_i), and y is an $m \times 1$ random vector with mean vector μ and identity covariance matrix.[1] We consider the cases where y is distributed as normal or nonnormal.

When y is normally distributed, the results were developed by various authors, see, for example, Mishra (1972), Kumar (1973), Srivastava and Tiwari(1976), Magnus (1978, 1979), Don (1979), Magnus and Neudecker (1979), Mathai and Provost (1992), Ghazal (1996), and Ullah (2004), where both the derivations and econometric applications were extensively studied. Loosely speaking, many of these works employed the moment generating function (m.g.f.) approach, while the commutation matrix (Magnus and Neudecker, 1979) and a recursive nonstochastic operator (Ullah, 2004) were also used to derive the results. When y is not normally distributed, however, the results are quite limited and are available only for quadratic forms of low order. In some applications (for example, see Section 4), the expectation of $E\left(y\prod_{i=1}^{n-1} y'A_iy\right)$, which is the product of linear function and quadratic forms, is also needed. (When y is a normal vector with zero mean, it is trivially equal to zero.) Under nonnormality, a general recursive procedure does not exist for deriving $E\left(\prod_{i=1}^{n} y'A_iy\right)$ or $E\left(y\prod_{i=1}^{n-1} y'A_iy\right)$, though for some special nonnormal distributions, including mixtures of normal, we may invoke some recursive algorithms (see Section 2.3 of Ullah, 2004).

The major purpose of this paper is two-fold. First, we try to derive a recursive algorithm for the expectation of an arbitrary number of products of quadratic forms in the random vector y when it is

[1] For the case of a general covariance matrix Ω, we can write $Q_i = (\Omega^{-1/2}y)'\Omega^{1/2}A_i\Omega^{1/2}(\Omega^{-1/2}y)$. Now $\Omega^{-1/2}y$ has mean vector $\Omega^{-1/2}\mu$ and covariance matrix I. For a normal vector y with mean μ and covariance matrix Ω, this normalization is innocuous for deriving the results for moments of quadratic forms in normal variables. This is also the case for the results on the first two moments of quadratic forms in nonnormal variables. However such a normalization may invalidate the assumption (see (6) in Section 3) on moments of the elements of the normalized vector. For tractability and simplicity, in Section 3 we assume that the nonnormal elements are i.i.d., as is usually the case considered for the error vector in a standard regression model.

normally distributed. We are going to utilize a nonstochastic operator proposed by Ullah (2004) to facilitate the derivation. When y has zero mean, the recursive result degenerates to the result given in Ghazal (1996). Secondly, we try to derive analytical results for $E\left(\prod_{i=1}^{n} y'A_iy\right)$ and $E\left(y\prod_{i=1}^{n-1} y'A_iy\right)$ for $n=4$ when y is nonnormally distributed. We express the nonnormal results explicitly as functions of the cumulants of the underlying nonnormal distribution of y.

The organization of this paper is as follows. In Section 2, we discuss the normal case and in Section 3 we derive the nonnormal results. Section 4 presents examples of using the nonnormal results to study the effects of nonnormality on the finite sample mean squared error (MSE) of econometric estimators in two time series models. We consider the ordinary least squares (OLS) estimator in an AR(1) model and the quasi maximum likelihood estimator (QMLE) in an MA(1) model. Section 5 concludes. Appendix A contains all the proofs and Appendix B contains the expressions needed for deriving the MSE result of QMLE in the MA(1) model.

2 The Normal Case

In this section, we focus on the case of normal variables. Before we are going to derive the main results, it would be helpful for us to discuss briefly a nonstochastic operator first introduced by Ullah (1990) and later in his monograph, Ullah (2004).

2.1 A Nonstochastic Operator

Ullah (1990, 2004) discussed an approach to deriving the moments of *any* analytic function involving a normal random vector by using a nonstochastic operator. More formally, he showed that for the $m\times 1$ vector $x\sim N(\mu,\Omega)$,

$$E[h(x)] = h(d)\cdot 1 = h(d) \text{ and } E[h(x)\,g(x)] = h(d)\,E[g(x)] \tag{1}$$

for the real-valued analytic functions $h(\cdot)$ and $g(\cdot)$, where $d=\mu+\Omega\partial_\mu$, $\partial_\mu=(\partial/\partial\mu_1,\cdots,\partial/\partial\mu_m)'$, is a nonstochastic $m\times 1$ vector operator. The transpose of this operator is defined as $d'=\mu'+\partial_\mu'\Omega$, where $\partial_\mu'=(\partial/\partial\mu_1,\cdots,\partial/\partial\mu_m)$. This result essentially follows from the fact that the density of x is an exponential function and for analytic functions, differentiation under the integral sign is allowed. Note that to use the result (1) correctly, readers must be cautious to the usual caveats, as pointed out by Ullah (2004, p. 12). Most importantly, power of d should be interpreted as a *recursive* operation. For example, we can easily derive $E(x'Ax)=d'Ad\cdot 1=d'A\mu=\mu'A\mu+\text{tr}(A\Omega)$ (where tr denotes the trace operator) by applying d' to $Ad\cdot 1=A\mu$. If one ignores this, $d'Ad\cdot 1=d'A\mu=\text{tr}(A\mu d')=\text{tr}(A\mu d'\cdot 1)=\mu'A\mu$, which is incorrect. This is so because in $d'A\mu$ the operator d' should be applied to $A\mu$, and thus $d'A\mu\neq\text{tr}(A\mu d')$.

Given (1), a recursive procedure $E\left(\prod_{i=1}^{n} Q_i\right)=d'A_1d\cdot E\left(\prod_{i=2}^{n} Q_i\right)$ immediately follows for $Q_i=y'A_iy$, $y\sim N(\mu,I)$. Apparently, we need to expand the operation $d'A_1d$ on $E(\prod_{i=2}^{n} Q_i)$, which may become demanding when n gets larger. Ullah (2004) gave the expressions for n up to 4 using this operator d. In this paper we take a step further to expand $d'A_1d \cdot E\left(\prod_{i=2}^n Q_i\right)$ explicitly. As a consequence, a recursive formula is derived for $E\left(\prod_{i=1}^n Q_i\right)$ for any n.

2.2 A Recursive Procedure

For the starting case when $n = 1$, $E(Q_1) = \mu' A_1 \mu + \mathrm{tr}(A_1)$, as shown before.

Theorem 1: The expectation of $\prod_{i=1}^n Q_i$ is given by the following recursion

$$E\left(\prod_{i=1}^n Q_i\right) = \sum_{i=0}^{n-1} 2^i \sum_{j_1=2}^{n} \cdots \sum_{j_i=2}^{n} \left[g_{j_1 \cdots j_i} E\left(\frac{Q_2 \cdots Q_n}{Q_{j_1} \cdots Q_{j_i}}\right)\right], \tag{2}$$

where for $i > 0$, $g_{j_1 \cdots j_i} = \mu'(A_1 A_{j_1} \cdots A_{j_i} + A_{j_1} A_1 A_{j_2} \cdots A_{j_i} + \cdots + A_{j_1} \cdots A_{j_i} A_1)\mu + \mathrm{tr}(A_1 A_{j_1} \cdots A_{j_i})$, and for $i = 0$, $g = \mu' A_1 \mu + \mathrm{tr}(A_1) = E(Q_1)$. ■

Note that in (2) an empty product in $g_{j_1 \cdots j_i}$ and $\frac{Q_2 \cdots Q_n}{Q_{j_1} \cdots Q_{j_i}}$ is to be interpreted as one. Now applying the theorem, we have the following results for n up to 4,

$E\left(\prod_{i=1}^1 Q_i\right) = \mu' A_1 \mu + \mathrm{tr}(A_1)$,

$E\left(\prod_{i=1}^2 Q_i\right) = E(Q_1)E(Q_2) + 4\mu' A_1 A_2 \mu + 2\mathrm{tr}(A_1 A_2)$,

$E\left(\prod_{i=1}^3 Q_i\right) = E(Q_1)E(Q_2Q_3) + [4\mu' A_1 A_2 \mu + 2\mathrm{tr}(A_1 A_2)]E(Q_3) + [4\mu' A_1 A_3 \mu + 2\mathrm{tr}(A_1 A_3)]E(Q_2) + 8\mu' A_1 A_2 A_3 \mu$
$+ 8\mu' A_1 A_3 A_2 \mu + 8\mu' A_2 A_1 A_3 \mu + 8\mathrm{tr}(A_1 A_2 A_3)$,

$E\left(\prod_{i=1}^4 Q_i\right) = E(Q_1)E(Q_2Q_3Q_4) + [4\mu' A_1 A_2 \mu + 2\mathrm{tr}(A_1 A_2)]E(Q_3Q_4) + [4\mu' A_1 A_3 \mu + 2\mathrm{tr}(A_1 A_3)]E(Q_2Q_4)$
$+ 4[\mu' A_1 A_4 \mu + 2\mathrm{tr}(A_1 A_4)]E(Q_2Q_3) + (8\mu' A_1 A_2 A_3 \mu + 8\mu' A_1 A_3 A_2 \mu)E(Q_4) + (8\mu' A_1 A_2 A_4 \mu + 8\mu' A_1 A_4 A_2 \mu)E(Q_3)$
$+ (8\mu' A_1 A_3 A_4 \mu + 8\mu' A_1 A_4 A_3 \mu)E(Q_2) + 16\mu' A_1 A_2 A_3 A_4 \mu + 16\mu' A_1 A_2 A_4 A_3 \mu + 16\mu' A_1 A_3 A_2 A_4 \mu + 16\mu' A_1 A_3 A_4 A_2 \mu$
$+ 16\mu' A_1 A_4 A_2 A_3 \mu + 16\mu' A_1 A_4 A_3 A_2 \mu + 16\mu' A_2 A_1 A_3 A_4 \mu + 16\mu' A_3 A_1 A_2 A_4 \mu + 16\mu' A_2 A_1 A_4 A_3 \mu + 16\mu' A_4 A_1 A_2 A_3 \mu$
$+ 16\mu' A_3 A_1 A_4 A_2 \mu + 16\mu' A_4 A_1 A_3 A_2 \mu + 16\mathrm{tr}(A_1 A_2 A_3 A_4) + 16\mathrm{tr}(A_1 A_2 A_4 A_3) + 16\mathrm{tr}(A_1 A_3 A_2 A_4)$.

Upon successive substitution, the above results for $n = 1, 2, 3, 4$ can be simplified to the expressions as given in Ullah (2004). When $A \equiv A_1 = A_2 = \cdots = A_n$, the result in Theorem 1 gives the nth moment (about zero) of the quadratic form $y'Ay$, as indicated in the following corollary, which obviously follows from (2).

Corollary 1: The nth moment of $y'Ay$ for $y \sim N(\mu, I)$ is given by the following recursion

$$E[(y'Ay)^n] = \sum_{i=0}^{n-1} g_i E\left[(y'Ay)^{n-i-1}\right], \tag{3}$$

where $E\left[(y'Ay)^0\right] = 1$, $g_i = \left(\begin{array}{c} n-1 \\ i \end{array}\right) 2^i i! [\mathrm{tr}(A^{i+1}) + (i+1)\mu' A^{i+1} \mu]$. ■

When $\mu = 0$, the above theorem degenerates to the result of Ghazal (1996), as given in Corollary 2.

Corollary 2: The expectation of $\prod_{i=1}^{n} Q_i$ when $y \sim N(0, I)$ is given by the following recursion

$$E\left(\prod_{i=1}^{n} Q_i\right) = E(Q_1) \cdot E\left(\prod_{i=2}^{n} Q_i\right) + 2\sum_{j=2}^{n} E\left(y' A_j A_1 y \cdot \frac{Q_2 \cdots Q_n}{Q_j}\right). \quad \blacksquare \tag{4}$$

3 The Nonnormal Case

Now suppose $y = (y_1, \cdots, y_m)'$ follows a general error distribution with an identity covariance matrix. In general, we can write

$$\begin{aligned} E\left(\prod_{i=1}^{n} Q_i\right) &= E(\otimes_{i=1}^{n} Q_i) \\ &= \mathrm{tr}\{E[\otimes_{i=1}^{n}(yy'A_i)]\} \\ &= \mathrm{tr}\{E[\otimes_{i=1}^{n}(yy')](\otimes_{i=1}^{n} A_i)\} \\ &= \mathrm{tr}\{E[(y^{\otimes})(y^{\otimes})']A^{\otimes}\}, \end{aligned} \tag{5}$$

where $\otimes$ denotes the Kronecker product symbol and $y^{\otimes} = y \otimes y \otimes \cdots \otimes y$ (n terms) and $A^{\otimes} = \otimes_{i=1}^{n} A_i$. Thus, under a general error distribution, the key determinant of the expectation required is the set of product moments of the y_i's that appear in the $m^n \times m^n$ matrix $P = (y^{\otimes})(y^{\otimes})'$, the elements of which are products of the type $\prod_{i=1}^{m} y_i^{\alpha(i)}$, where the nonnegative integers satisfy $\sum_{i=1}^{m} \alpha(i) = 2n$, i.e. they are a composition of $2n$ with m parts. Numerically, on can always write a computer program to calculate the expectation. However, when A_i are of high dimension, it may not be practically possible to store an $m^n \times m^n$ matrix P without torturing the computer. So it may be more promising if we can work out some explicit analytical expressions. More importantly, analytical expressions can help us understand explicitly the effects of nonnormality on the finite sample properties of many econometric estimators, examples of which are provided in the next section. Note that sometimes we may also be interested in the expectation of $y \prod_{i=1}^{n-1} y' A_i y$, which is the product of linear function and quadratic forms.

To facilitate our derivation, suppose now that the mean vector $\mu = 0$ and y_i is i.i.d. As it turns out, as n goes up, the work is more demanding. In the literature, results are only available for n up to 3, see Chandra (1983), Ullah, Srivastava, and Chandra (1983), and Ullah (2004). So now we take one step further to derive the analytical result for $n = 4$. From the analysis in the previous paragraph, when $n = 4$, the highest power of y_i in P is 8. So we assume that under a general error distribution, y_i has finite moments $m_j = E(y_i^j)$ up to the eighth order:

$$\begin{aligned}
m_1 &= 0, \quad m_2 = 1, \quad m_3 = \gamma_1, \quad m_4 = \gamma_2 + 3, \\
m_5 &= \gamma_3 + 10\gamma_1, \quad m_6 = \gamma_4 + 15\gamma_2 + 10\gamma_1^2 + 15, \\
m_7 &= \gamma_5 + 21\gamma_3 + 35\gamma_2\gamma_1 + 105\gamma_1, \\
m_8 &= \gamma_6 + 28\gamma_4 + 56\gamma_3\gamma_1 + 35\gamma_2^2 + 210\gamma_2 + 280\gamma_1^2 + 105, \qquad (6)
\end{aligned}$$

where γ_1 and γ_2 are the Pearson's measures of skewness and kurtosis of the distribution and these and $\gamma_3, \cdots, \gamma_6$ can be regarded as measures for deviation from normality. For a normal distribution, the parameters $\gamma_1, \cdots, \gamma_6$ are all zero. Note that these γs can also be expressed as cumulants of y_i, e.g., γ_1 and γ_2 represent the third and fourth cumulants.

Theorem 2: When y has mean zero and its elemetns are i.i.d. and have moments as specified in (6), then

$$\begin{aligned}
E\left(\prod_{i=1}^{4} Q_i\right) &= \mathrm{tr}(A_1)\mathrm{tr}\,(A_2)\,\mathrm{tr}\,(A_3)\,\mathrm{tr}\,(A_4) + 2[\mathrm{tr}\,(A_1)\,\mathrm{tr}\,(A_2)\,\mathrm{tr}\,(A_3A_4) \\
&\quad +\mathrm{tr}\,(A_1)\,\mathrm{tr}\,(A_3)\,\mathrm{tr}\,(A_2A_4) + \mathrm{tr}\,(A_1)\,\mathrm{tr}\,(A_4)\,\mathrm{tr}\,(A_2A_3) \\
&\quad +\mathrm{tr}\,(A_2)\,\mathrm{tr}\,(A_3)\,\mathrm{tr}\,(A_1A_4) + \mathrm{tr}\,(A_2)\,\mathrm{tr}\,(A_4)\,\mathrm{tr}\,(A_1A_3) \\
&\quad +\mathrm{tr}\,(A_3)\,\mathrm{tr}\,(A_4)\,\mathrm{tr}\,(A_1A_2)] + 4[\mathrm{tr}\,(A_1A_2)\,\mathrm{tr}\,(A_3A_4) + \mathrm{tr}\,(A_1A_3)\,\mathrm{tr}\,(A_2A_4) \\
&\quad +\mathrm{tr}\,(A_1A_4)\,\mathrm{tr}\,(A_2A_3)] + 8[\mathrm{tr}\,(A_1)\,\mathrm{tr}\,(A_2A_3A_4) + \mathrm{tr}\,(A_2)\,\mathrm{tr}\,(A_1A_3A_4) \\
&\quad +\mathrm{tr}\,(A_3)\,\mathrm{tr}\,(A_1A_2A_4) + \mathrm{tr}\,(A_4)\,\mathrm{tr}\,(A_1A_2A_3)] \\
&\quad +16[\mathrm{tr}\,(A_1A_3A_4A_2) + \mathrm{tr}\,(A_1A_4A_2A_3) + \mathrm{tr}\,(A_1A_4A_3A_2)] \\
&\quad +\gamma_2 f_{\gamma_2} + \gamma_4 f_{\gamma_4} + \gamma_6 f_{\gamma_6} + \gamma_1^2 f_{\gamma_1^2} + \gamma_2^2 f_{\gamma_2^2} + \gamma_1\gamma_3 f_{\gamma_1\gamma_3}, \qquad (7)
\end{aligned}$$

and

$$E\left(y\prod_{i=1}^{3} Q_i\right) = \gamma_5 h_{\gamma_5} + \gamma_3 h_{\gamma_3} + \gamma_1 h_{\gamma_1} + \gamma_1\gamma_2 h_{\gamma_1\gamma_2}, \qquad (8)$$

where f_γ's and h_γ's denote the contributions to $E\left(\prod_{i=1}^4 Q_i\right)$ and $E\left(y\prod_{i=1}^3 Q_i\right)$, respetively, due to nonnormality of y and they are given in Appendix A.

Note that setting all the γ's equal to zero, then (7) degenerates to the result for normal quadratic form of order 4 as given in Ullah (2004), and trivially in (8), $E(y\prod_{i=1}^3 Q_i) = 0$ for $y \sim N(0, I)$.

4 Effects of Nonnormality

Now we give applications of the nonnormal results from the previous section. We study the effects of nonnormality on the finite sample MSE of the OLS estimator in an AR(1) model and the QMLE in an

MA(1) model. For notational convenience, in the following we suppress the subscript 0 in the true parameter vector in a model.

4.1 AR(1) Model

Consider the following AR(1) model with exogenous regressors

$$y_t = \rho y_{t-1} + x_t'\beta + \sigma\varepsilon_t,\ t = 1, \cdots, T, \tag{9}$$

where $\rho \in (-1, 1)$, x_t is $k \times 1$ fixed and bounded so that $X'X = O(T)$, where $X = (x_1, \cdots, x_T)'$, β is $k \times 1$, $\sigma > 0$, and the error term ε_t is i.i.d. and ε_t/σ follows some nonnormal distribution with moments (6). Denote $\varepsilon = (\varepsilon_1, \varepsilon_2, \cdots, \varepsilon_T)'$, $y = (y_1, y_2, \cdots, y_T)'$, $M = I - X(X'X)^{-1}X'$. Also define F to be a $T \times 1$ vector with the t-th 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 $\rho^{t-t'-1}$, $A = MC$, $r = M(y_0F + CX\beta)$, where we assume that the first observation y_0 has been observed.

For the OLS estimator $\hat{\rho}$ of ρ, Bao and Ullah (2007b) found that nonnormality affects its approximate bias, up to $O(T^{-1})$, through the skewness coefficient γ_1. Consequently, for nonnormal symmetric distribution (e.g. Student-t), the bias is robust against nonnormality. Setting $\gamma_1 = 0$, the result degenerates into the bias result of Kiviet and Phillips (1993) under normality. The approximate MSE under nonnormality, up to $O(T^{-2})$, however, was not available due to the absence of the results (7) and (8) given in Section 3. Now we are ready to study the effects of nonnormality on the MSE. Following the lines of Bao and Ullah (2007b), we can derive the following result for the approximate MSE of $\hat{\rho}$:

$$\begin{aligned} M(\hat{\rho}) &= \frac{6(\lambda_{10000} + \lambda_{02000} + 2r'\omega_{01000})}{(r'r + \lambda_{00100})^2} \\ &- \frac{8[r'r(\lambda_{10000}+\lambda_{02000}+4\lambda_{01010})+\lambda_{10100}+\lambda_{02100}+2r'(rr'\omega_{01000}+A\omega_{10000}+A\omega_{02000}+\omega_{01100})]}{(r'r + \lambda_{00100})^3} \\ &+ \frac{3[(r'r)^2(\lambda_{10000} + \lambda_{02000}) + 2r'r(\lambda_{10100} + 4\lambda_{01010} + \lambda_{02100}) + 4\lambda_{10001} + \lambda_{10200} + 4\lambda_{02001}]}{(r'r + \lambda_{00100})^4} \\ &+ \frac{3[8\lambda_{01110} + \lambda_{02200} + 4r'rr'(A\omega_{10000} + A\omega_{02000} + \omega_{01100}) + 2(r'r)^2r'\omega_{01000}]}{(r'r + \lambda_{00100})^4} \\ &+ \frac{6r'(\omega_{01200} + 4\omega_{01001} + 2A\omega_{10100} + 2A\omega_{02100})}{(r'r + \lambda_{00100})^4} + o(T^{-2}), \end{aligned} \tag{10}$$

where $\lambda_{ijklm} = \sigma^{2(i+j+k+l+m)}E[(\varepsilon'rr'\varepsilon)^i \cdot (\varepsilon'A\varepsilon)^j \cdot (\varepsilon'A'A\varepsilon)^k \cdot (\varepsilon'A'rr'\varepsilon)^l \cdot (\varepsilon'A'rr'A\varepsilon)^m]$ and $\omega_{ijklm} = \sigma^{2(i+j+k+l+m)+1}E[\varepsilon \cdot (\varepsilon'rr'\varepsilon)^i \cdot (\varepsilon'A\varepsilon)^j \cdot (\varepsilon'A'A\varepsilon)^k \cdot (\varepsilon'A'rr'\varepsilon)^l \cdot (\varepsilon'A'rr'A\varepsilon)^m]$. Note that the result (10) holds for both normal and nonnormal ε. Under nonnormality, (7) is needed to evaluate λ_{ijklm} when $i+j+k+l+m = 4$ and (8) is needed for ω_{ijklm} when $i+j+k+l+m = 3$ (note that A and $A'rr'$ need to be symmetrized in (7) and (8)). So given this new result (10), together with (7) and (8), it is possible for us to investigate explicitly the effects of nonnormality on $M(\hat{\rho})$.

For the special case when x_t is a scalar constant, i.e., when we have the so-called intercept model, upon simplifying (10) and ignoring terms of lower orders, we can derive the following anlytical result:[2]

$$\begin{aligned} M(\hat{\rho}) &= \frac{1-\rho^2}{T} + \frac{1}{T^2}\left[23\rho^2 + 10\rho - \frac{1+\rho}{1-\rho}\left(\frac{\alpha-(1-\rho)y_0}{\sigma}\right)^2 \right. \\ &\quad \left. - \frac{4\gamma_1^2\rho\left(1-\rho^2\right)^2}{1-\rho^3} - \gamma_2(1-\rho^2)\right] + o\left(T^{-2}\right). \end{aligned} \tag{11}$$

Now only the skewness and kurtosis coefficients matter for the approximate MSE, up to $O(T^{-2})$. The $O(T^{-1})$ MSE, $(1-\rho^2)/T$ is nothing but the asymptotic variance of $\hat{\rho}$ for the intercept model.

Figures 1-2 plots the true (solid line) and (feasible) approximate MSE, accommodating (short dashed line) and ignoring (dotted and dashed line) the presence of nonnormality, of $\hat{\rho}$ over 10,000 simulations when the error term ε_t follows a standardized asymmetric power distribution (APD) of Komunjer (2007).[3] It has a closed-form density function as shown in Komunjer (2007) with two parameters, $\alpha \in (0,1)$, which controls skewness, and $\lambda > 0$, which controls the tail properties. To prevent the signal-to-noise ratio going up as we increase ρ, we set $\beta = 1-\rho$. We experiment with $\sigma^2 = 0.5, 1$, $\alpha = 0.01, 0.05$, $\lambda = 0.5, 1$, $T = 50$.[4] To calculate the feasible approximate MSE under nonnormality, we put $\hat{\rho}, \hat{\sigma}^2, \hat{\gamma}_1$, and $\hat{\gamma}_2$ into (10). When ignoring nonnormality, we put $\hat{\rho}, \hat{\sigma}^2$, and $\gamma_1 = \gamma_2 = 0$ into (11) to calculate the approximate MSE. We use Fisher's (1928) k statistics to estimate γ_1 and γ_2, see Dressel (1940) and Stuart and Ord (1987) for the expressions of the k statistics in terms of sample moments.

As is clear from the two figures, in the presence of nonnormality, (11) approximates the true MSE remarkably well for different degrees of nonnormality and magnitudes of the error variance. Ignoring the effects of nonnormality produces overestimated MSE; accommodating nonnormality, (11) produces very accurate estimate of the true MSE. Lastly, when the degree of nonnormality goes down (λ goes from 0.5 to 1), the gap between the approximate results accommodating and ignoring nonnormality gets closer, as we can expect.

4.2 MA(1) Model

For the invertible MA(1) model,

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

where $|\phi| < 1$, $\varepsilon_t \sim$i.i.d.$(0,\sigma^2)$ and the moments of ε_t/σ follow (6), the parameter vector $\theta = (\phi, \sigma^2)'$ is usually estimated by the conditional QMLE (conditional on $\varepsilon_0 = 0$) that maximizes the likelihood function under a normal density of ε_t for a sample of size T. Let C be a lower triangular $T \times T$ matrix with unit diagonal

[2] Note that Bao (2007) also derived (11) by working with a direct Nagar (1959) expansion of a ratio of quadratic forms for the OLS estimator in the intercept model. In contrast, here we arrive at (11) from simplifying the general result (10).

[3] In the experiment, we set $y_0 = 1$. However, the results are not sensitive to the choice of y_0. For other possible fixed or random start-up values, we get similar patterns for the two figures.

[4] When $\alpha = 0.01$, as λ goes from 0.5 to 1, γ_1 goes from 4.3124 to 1.9997, γ_2 from 34.5737 to 5.9988; when $\alpha = 0.05$, as λ goes from 0.5 to 1, γ_1 goes from 4.3477 to 1.9914, γ_2 from 35.1736 to 5.9669.

and its tt'-th lower off-diagonal element being $-\phi_0$ if $t - t' = 1$ and zero otherwise, $B = -\partial C/\partial\phi$, $N_1 = C^{-1}B$, $N_2 = 2(C^{-1}B)^2 + (C^{-1}B)'(C^{-1}B)$, $N_3 = 6(C^{-1}B)^3 + 6(C^{-1}B)'(C^{-1}B)^2$, and $N_4 = 24(C^{-1}B)^4 + 24(C^{-1}B)'(C^{-1}B)^3 + 12[(C^{-1}B)^2]'(C^{-1}B)^2$. Bao and Ullah (2007b) showed that the second-order MSE (up to $O(T^{-2})$) of the QMLE $\hat{\phi}$ is given by

$$\begin{aligned} M(\hat{\phi}) &= \frac{6\delta_{200}}{[\mathrm{tr}(N_2)]^2} - \frac{8\delta_{210}}{[\mathrm{tr}(N_2)]^3} + \frac{3\delta_{220} + \delta_{301} + 4\mathrm{tr}(N_3)\delta_{300}}{[\mathrm{tr}(N_2)]^4} \\ &\quad - \frac{4\mathrm{tr}(N_3)\delta_{310} + \mathrm{tr}(N_4)\delta_{400}/3}{[\mathrm{tr}(N_2)]^5} + \frac{5[\mathrm{tr}(N_3)]^2\delta_{400}}{4[\mathrm{tr}(N_2)]^6}, \end{aligned} \tag{13}$$

where $\delta_{ijk} = E[(\varepsilon' N_1 \varepsilon)^i \cdot (\varepsilon' N_2 \varepsilon)^j \cdot (\varepsilon' N_3 \varepsilon)^k / \sigma^{2(i+j+k)}]$, $\varepsilon = (\varepsilon_1, \cdots, \varepsilon_T)'$. Note that δ_{200}, δ_{210}, and δ_{300} could be evaluated straightforwardly by using the results collected in Ullah (2004). But δ_{220}, δ_{301}, δ_{310}, δ_{400} are expectations of quadratic forms of order 4 in the nonnormal vector ε, which could not be evaluated without the analytical result (7).[5] Now thanks to (7), together with help of Mathematica for symbolic calculation, we simplify all the terms in (13) explicitly in terms of model parameters as in Appendix B.

Note that

$$\begin{aligned} [\mathrm{tr}(N_2)]^{-2} &= \left[\frac{T}{1-\phi^2} - \frac{1}{(1-\phi^2)^2} + o(T^{-2})\right]^{-2} \\ &= \left[\frac{T}{1-\phi^2} - \frac{1}{(1-\phi^2)^2}\right]^{-2} \left[1 + \frac{o(T^{-2})}{\frac{T}{1-\phi^2} - \frac{1}{(1-\phi^2)^2}}\right]^{-2} \\ &= \left[\frac{T}{1-\phi^2} - \frac{1}{(1-\phi^2)^2}\right]^{-2} + o(T^{-3}) \\ &= \left(\frac{T}{1-\phi^2}\right)^{-2} \left[1 - \frac{\frac{1}{(1-\phi^2)^2}}{\frac{T}{1-\phi^2}}\right]^{-2} + o(T^{-3}) \\ &= \frac{(1-\phi^2)^2}{T^2} - \frac{2(1-\phi^2)}{T^3} + o(T^{-3}) \end{aligned}$$

by using the expansion $(1+x)^{-2} = 1 - 2x + 3x^2 + \cdots$. In general, we can write

$$[\mathrm{tr}(N_2)]^{-i} = \frac{(1-\phi^2)^i}{T^i} - \frac{i(1-\phi^2)^{i-1}}{T^{i+1}} + o(T^{-i-1}),\ i = 2, 3, 4, 5, 6. \tag{14}$$

Substituting all the δ's from Appendix B and (14) into (13) yields the following analytical second-order MSE of the QMLE $\hat{\phi}$,

$$M(\hat{\phi}) = \frac{1-\phi^2}{T} + \frac{1}{T^2}\left[9 - \phi^2 - \left(1-\phi^2\right)\gamma_2 + \frac{2\phi(1+\phi)^2\left(1 - 2\phi + \phi^3\right)\gamma_1^2}{\left(1+\phi+\phi^2\right)^2}\right] + o(T^{-2}). \tag{15}$$

Omitting the higher-order terms, (15) immediately gives the asymptotic / first-order variance of $\hat{\phi}$, $(1-\phi^2)/T$, which is in fact robust to the nonnormal behavior of the error term. The effects of nonnormality come into play only through the order $O(T^{-2})$ terms. Numerically, if one simply plugs (7) into (13) to

[5] Given this, Bao and Ullah (2007b) had to assume normailty to evaluate numerically the second-order MSE.

calculate the second-order MSE of $\hat{\phi}$, one might think that all of $\gamma_1, \gamma_2, \gamma_3, \gamma_4, \gamma_6$ will matter. But in fact the analytical result (15) shows that only γ_1 and γ_2, i.e., the skewness and excess kurtosis of ε, contribute to the order $O(T^{-2})$ terms, similar to the AR(1) model. Under normality, the asymptotic variance is always smaller than the second-order MSE. Moreover, as the absolute value of ϕ increases, the gap between them goes up.

5 Conclusions

We have derived a recursive formula for evaluating the expectation of the product of an arbitrary number of quadratic forms in normal variables and the expectations of quadratic form of order 4 in nonnormal variables. The recursive feature of the result under normality makes it straightforward to program and in terms of computer time, the recursive formula may have advantage over that based on the traditional moment generating function approach when the matrices have high dimension and the order of quadratic forms is large. For the nonnormal quadratic forms, we express the results explicitly as functions of the cumulants of the underlying nonnormal distribution. Setting all the nonnormality parameters equal to zero gives the results under normality as a special case. We apply the nonnormal results to study the finite sample MSE of the OLS estimator in an AR(1) model with exogenous regressors and the QMLE in a simple MA(1) model when the errors are nonnormally distributed.

Appendix A: Proofs

We first introduce fours lemmas. They are needed when we later try to expand $d'A_1 d \cdot E\left(\prod_{i=2}^{n} Q_i\right)$. In the following derivations, for multiple summations, the indices are not equal to each other, e.g., in Lemma 1, $j_1 \neq j_2 \neq \cdots \neq j_i$ in $\sum_{j_1=1}^{n} \cdots \sum_{j_i=1}^{n}$. We also write $y = \mu + \varepsilon$.

Lemma 1: The gradient of $E\left(\prod_{i=1}^{n} Q_i\right)$ with respect to μ is

$$\frac{\partial E\left(\prod_{i=1}^{n} Q_i\right)}{\partial \mu} = \sum_{i=1}^{n} \sum_{j_1=1}^{n} \cdots \sum_{j_i=1}^{n} 2^i A_{j_1} \cdots A_{j_i} \mu E\left(\frac{Q_1 \cdots Q_n}{Q_{j_1} \cdots Q_{j_i}}\right). \tag{A.1}$$

Proof: Changing the order of derivative and expectation signs,

$$\begin{aligned}
\frac{\partial E\left(\prod_{i=1}^{n} Q_i\right)}{\partial \mu} &= E\left[\frac{\partial \prod_{i=1}^{n} Q_i}{\partial \mu}\right] \\
&= E\left[\frac{\prod_{i=1}^{n} (\varepsilon+\mu)' A_i (\varepsilon+\mu)}{\partial \mu}\right] \\
&= 2\sum_{j=1}^{n} A_j E\left[y\left(\frac{Q_1 \cdots Q_n}{Q_j}\right)\right] \\
&= 2\sum_{j=1}^{n} A_j \mu E\left(\frac{Q_1 \cdots Q_n}{Q_j}\right) + 2\sum_{j=1}^{n} A_j \frac{\partial}{\partial \mu} E\left(\frac{Q_1 \cdots Q_n}{Q_j}\right),
\end{aligned}$$

where the last equality follows from (1) by replacing y with the operator d in $E\left[y\left(\frac{Q_1 \cdots Q_n}{Q_j}\right)\right]$. Upon successive substitutions, the result follows. ■

It should be noted that in (A.1), the products involved in $A_{j_1} \cdots A_{j_{i-1}}$ and $\frac{Q_1 \cdots Q_n}{Q_{j_1} \cdots Q_{j_i}}$ are assumed to be one when $i < 2$ for the former or when $i > n-1$ for the latter, following the standard convention that an empty product is to be interpreted as one. For example, $\partial E(Q_1)/\partial \mu = 2A_1\mu$, $\partial E(Q_1Q_2)/\partial\mu = 2[A_1\mu E(Q_2) + A_2\mu E(Q_1)] + 4(A_1A_2\mu + A_2A_1\mu)$.

Lemma 2: The Hessian of $E\left(\prod_{i=1}^{n} Q_i\right)$ with respect to μ is

$$\frac{\partial^2 E\left(\prod_{i=1}^{n} Q_i\right)}{\partial \mu' \partial \mu} = 2\sum_{j=1}^{n} A_j E\left(\frac{Q_1 \cdots Q_n}{Q_j}\right) + 4\sum_{j=1}^{n}\sum_{k=1}^{n} E\left(A_j y y' A_k \frac{Q_1 \cdots Q_n}{Q_j Q_k}\right). \tag{A.2}$$

Proof: From the proof of Lemma 1, $\partial E\left(\prod_{i=1}^n Q_i\right)/\partial\mu' = 2\sum_{j=1}^n y'A_jE\left(\frac{Q_1\cdots Q_n}{Q_j}\right)$, so

$$\begin{aligned}\frac{\partial^2 E\left(\prod_{i=1}^n Q_i\right)}{\partial\mu'\partial\mu} &= 2E\left(\sum_{j=1}^n \frac{\partial y'A_j\frac{Q_1\cdots Q_n}{Q_j}}{\partial\mu}\right) \\ &= 2E\left(\sum_{j=1}^n A_j\frac{Q_1\cdots Q_n}{Q_j}\right) + 2E\left(\sum_{j=1}^n \frac{\partial\frac{Q_1\cdots Q_n}{Q_j}}{\partial\mu}y'A_j\right) \\ &= 2\sum_{j=1}^n A_jE\left(\frac{Q_1\cdots Q_n}{Q_j}\right) + 4\sum_{j=1}^n\sum_{k=1}^n E\left(A_jyy'A_k\frac{Q_1\cdots Q_n}{Q_jQ_k}\right). \quad \blacksquare\end{aligned}$$

Lemma 3: Suppose $b = b(\mu)$ is a scalar, A is symmetric, $m\times m$, and does not depend on μ, then $\partial'_\mu A\mu b = b\mathrm{tr}(A) + \mu'A\frac{\partial b}{\partial\mu}$.

Proof:

$$\begin{aligned}\partial'_\mu A\mu b &= (\partial/\partial\mu_1,\cdots,\partial/\partial\mu_m)\left[\begin{pmatrix}\sum_{j=1}^m A_{1,j}\mu_j \\ \vdots \\ \sum_{j=1}^m A_{m,j}\mu_j\end{pmatrix} b\right] \\ &= \sum_{j=1}^m A_{j,j}b + \left[(\partial/\partial\mu_1,\cdots,\partial/\partial\mu_m)\,b\right]\begin{pmatrix}\sum_{j=1}^m A_{1,j}\mu_j \\ \vdots \\ \sum_{j=1}^m A_{m,j}\mu_j\end{pmatrix} \\ &= b\mathrm{tr}(A) + \left(\sum_{j=1}^m A_{1,j}\mu_j,\cdots,\sum_{j=1}^m A_{m,j}\mu_j\right)\begin{pmatrix}\partial b/\partial\mu_1 \\ \vdots \\ \partial b/\partial\mu_m\end{pmatrix} \\ &= b\mathrm{tr}(A) + \mu'A\frac{\partial b}{\partial\mu}. \quad \blacksquare\end{aligned}$$

Lemma 4: Suppose $b = b(\mu)$ is a scalar, A is symmetric, $m\times m$, and does not depend on μ, then $\partial'_\mu A\partial_\mu b = \mathrm{tr}\left(A\frac{\partial^2 b}{\partial\mu'\partial\mu}\right)$.

Proof:

$$\begin{aligned}\partial'_\mu A\partial_\mu b &= (\partial/\partial\mu_1,\cdots,\partial/\partial\mu_m)\left[\begin{pmatrix}\sum_{j=1}^m A_{1,j}\frac{\partial b}{\partial\mu_j} \\ \vdots \\ \sum_{j=1}^m A_{m,j}\frac{\partial b}{\partial\mu_j}\end{pmatrix}\right] \\ &= \sum_{i=1}^m\sum_{j=1}^m A_{i,j}\frac{\partial^2 b}{\partial\mu_i\partial\mu_j} \\ &= \mathrm{tr}\left(A\frac{\partial^2 b}{\partial\mu'\partial\mu}\right). \quad \blacksquare\end{aligned}$$

Proof of Theorem 1: Using (1) and replacing $y'A_1y$ with $d'A_1d$,

$$\begin{aligned}
E\left(\prod_{i=1}^{n} Q_i\right) &= d'A_1d \cdot E\left(\prod_{i=2}^{n} Q_i\right) \\
&= d'A_1\left(\mu+\partial_\mu\right)E\left(\prod_{i=2}^{n} Q_i\right) \\
&= d'A_1\mu E\left(\prod_{i=2}^{n} Q_i\right) + d'A_1\frac{\partial E\left(\prod_{i=2}^{n} Q_i\right)}{\partial\mu} \\
&= \left(\mu'+\partial_\mu'\right)A_1\mu E\left(\prod_{i=2}^{n} Q_i\right) + \left(\mu'+\partial_\mu'\right)A_1\frac{\partial E\left(\prod_{i=2}^{n} Q_i\right)}{\partial\mu} \\
&= \mu'A_1\mu \cdot E\left(\prod_{i=2}^{n} Q_i\right) + \partial_\mu' A_1\mu E\left(\prod_{i=2}^{n} Q_i\right) + \mu'A_1\frac{\partial E\left(\prod_{i=2}^{n} Q_i\right)}{\partial\mu} + \partial_\mu' A_1\frac{\partial E\left(\prod_{i=2}^{n} Q_i\right)}{\partial\mu}.
\end{aligned}$$

From Lemma 3,

$$\partial_\mu' A_1\mu E\left(\prod_{i=2}^{n} Q_i\right) = \operatorname{tr}(A_1)E\left(\prod_{i=2}^{n} Q_i\right) + \mu'A_1\frac{\partial E\left(\prod_{i=2}^{n} Q_i\right)}{\partial\mu}.$$

From Lemma 4,

$$\partial_\mu' A_1\frac{\partial E\left(\prod_{i=2}^{n} Q_i\right)}{\partial\mu} = \operatorname{tr}\left[A_1\frac{\partial^2 E\left(\prod_{i=2}^{n} Q_i\right)}{\partial\mu'\partial\mu}\right].$$

Thus

$$\begin{aligned}
E\left(\prod_{i=1}^{n} Q_i\right) &= \mu'A_1\mu \cdot E\left(\prod_{i=2}^{n} Q_i\right) + \operatorname{tr}(A_1)E\left(\prod_{i=2}^{n} Q_i\right) + 2\mu'A_1\frac{\partial E\left(\prod_{i=2}^{n} Q_i\right)}{\partial\mu} + \operatorname{tr}\left[A_1\frac{\partial^2 E\left(\prod_{i=2}^{n} Q_i\right)}{\partial\mu'\partial\mu}\right] \\
&= E(Q_1)E\left(\prod_{i=2}^{n} Q_i\right) + \sum_{i=1}^{n-1} 2^{i+1}\sum_{j_1=2}^{n}\cdots\sum_{j_i=2}^{n}\mu'A_1A_{j_1}\cdots A_{j_i}\mu E\left(\frac{Q_2\cdots Q_n}{Q_{j_1}\cdots Q_{j_i}}\right) \\
&\quad +2\sum_{j=2}^{n}\operatorname{tr}(A_1A_j)E\left(\frac{Q_2\cdots Q_n}{Q_j}\right) + 4\sum_{j=2}^{n}\sum_{k=2}^{n}E\left(y'A_jA_1A_ky\frac{Q_2\cdots Q_n}{Q_jQ_k}\right) \\
&= E(Q_1)E\left(\prod_{i=2}^{n} Q_i\right) + \sum_{i=1}^{n-1} 2^{i+1}\sum_{j_1=2}^{n}\cdots\sum_{j_i=2}^{n}\mu'A_1A_{j_1}\cdots A_{j_i}\mu E\left(\frac{Q_2\cdots Q_n}{Q_{j_1}\cdots Q_{j_i}}\right) \\
&\quad +2\sum_{j=2}^{n}\operatorname{tr}(A_1\Omega A_j)E\left(\frac{Q_2\cdots Q_n}{Q_j}\right) + 4\sum_{\substack{j=2 \\ j<k}}^{n}\sum_{k=2}^{n}E\left(y'\left(A_jA_1A_k+A_kA_1A_j\right)y\frac{Q_2\cdots Q_n}{Q_jQ_k}\right),
\end{aligned}$$

where the second last equality follows by substituting the results for $\partial E\left(\prod_{i=2}^{n} Q_i\right)/\partial\mu$ and $\partial^2 E\left(\prod_{i=2}^{n} Q_i\right)/\partial\mu'\partial\mu$ from Lemmas 1 and 2 and noting that $E(Q_1) = \mu'A_1\mu + \operatorname{tr}(A_1)$, and the last equality follows by noting that $A_jA_1A_k$ is not symmetric, and we replace it with $\left(A_jA_1A_k + A_kA_1A_j\right)/2$, and that the two indices j and k have symmetric roles. Note that in the last equality there is a term $E\left[y'\left(A_jA_1A_k+A_kA_1A_j\right)y\frac{Q_2\cdots Q_n}{Q_jQ_k}\right]$, which is a quadratic form of order $n-2$. Upon successive substitution, result (2) follows.■

Proof of Corollary 2: Comparing the proof of (2) and (4), we need to show that, when $\mu = 0$,

$$W_n \equiv \sum_{j=2}^{n} E\left(y'A_jA_1y \cdot \frac{Q_2\cdots Q_n}{Q_j}\right)$$

is in fact equal to

$$U_n \equiv \sum_{j=2}^{n} \mathrm{tr}(A_1 A_j) E\left(\frac{Q_2 \cdots Q_n}{Q_j}\right) + 2\sum_{j=2}^{n}\sum_{\substack{i=2\\ i\neq j}}^{n} E\left[y' A_i A_1 A_j y \cdot \left(\frac{Q_2\cdots Q_n}{Q_i Q_j}\right)\right].$$

It is enough to show that the jth summands of W_n and U_n, denoted by $w_{n,j}$ and $u_{n,j}$, respectively, are equal. We proceed with our proof by induction. When $n=3$, $w_{3,2} = E\left(y' A_2 A_1 y \cdot y' A_3 y\right)$, $u_{3,2} = \mathrm{tr} A_1 A_2 \cdot E(Q_3) + E\left[y'\left(A_2 A_1 A_3 + A_3 A_1 A_2\right) y\right]$. Obviously, $w_{3,2} = u_{3,2}$ when $\mu = 0$. Similarly, $w_{3,3} = u_{3,3}$. Now suppose $w_{k,j} = u_{k,j}$, $j = 2, \cdots, k$, i.e.

$$\begin{aligned} E\left(y' A_j A_1 y \cdot \frac{Q_2\cdots Q_k}{Q_j}\right) &= \mathrm{tr} A_1 A_j \cdot E\left(\frac{Q_2\cdots Q_k}{Q_j}\right) \\ &\quad + \sum_{\substack{i=2\\ i\neq j}}^{k} E\left[y'\left(A_i A_1 A_j + A_j A_1 A_i\right) y \cdot \left(\frac{Q_2\cdots Q_k}{Q_i Q_j}\right)\right]. \end{aligned} \tag{A.3}$$

When $n = k+1$, using (2),

$$\begin{aligned} w_{k+1,j} &= E\left(y' A_j A_1 y \cdot \frac{Q_2\cdots Q_{k+1}}{Q_j}\right) \\ &= \mathrm{tr} A_1 A_j \cdot E\left(\frac{Q_2\cdots Q_{k+1}}{Q_j}\right) + 2\sum_{\substack{i=2\\ i\neq j}}^{k+1} \mathrm{tr} A_j A_1 A_i \cdot E\left(\frac{Q_2\cdots Q_{k+1}}{Q_i Q_j}\right) \\ &\quad +2\sum_{\substack{i=2\\ i\neq l\neq j}}^{k+1}\sum_{l=2}^{k+1} E\left[y'\left(A_i A_j A_1 A_l + A_l A_1 A_j A_l\right) y \cdot \left(\frac{Q_2\cdots Q_{k+1}}{Q_i Q_j Q_l}\right)\right]. \end{aligned}$$

To establish the equivalence of $w_{k+1,j}$ and

$$\begin{aligned} u_{k+1,j} &= \mathrm{tr} A_1 A_j \cdot E\left(\frac{Q_2\cdots Q_{k+1}}{Q_j}\right) + \sum_{\substack{i=2\\ i\neq j}}^{k+1} E\left[y'\left(A_i A_1 A_j + A_j A_1 A_i\right) y \cdot \left(\frac{Q_2\cdots Q_{k+1}}{Q_i Q_j}\right)\right] \\ &= \mathrm{tr} A_1 A_j \cdot E\left(\frac{Q_2\cdots Q_{k+1}}{Q_j}\right) + 2\sum_{\substack{i=2\\ i\neq j}}^{k+1} E\left[y'\left(A_i A_1 A_j\right) y \cdot \left(\frac{Q_2\cdots Q_{k+1}}{Q_i Q_j}\right)\right], \end{aligned}$$

it is enough to show the sum of the last two terms of $w_{k+1,j}$ is equal to the last term of $u_{k+1,j}$, which is apparently true following (A.3). ■

Proof of Theorem 2: For $E\left(\prod_{i=1}^{4} Q_i\right) = \mathrm{tr}\{E[(y^{\otimes})(y^{\otimes})']A^{\otimes}\}$, $(y^{\otimes})(y^{\otimes})$ has elements $\prod_{i=1}^{m} y_i^{\alpha(i)} = y_1^{\alpha(1)} \cdots y_m^{\alpha(m)}$ with $\alpha(1) + \cdots + \alpha(m) = 2n = 8$. We put $\prod_{i=1}^{m} y_i^{\alpha(i)} = y_{i_1} \cdots y_{i_8}$, which has nonzero expectation only in the following seven situations:

1. All the eight indices $i_1, \cdots, i_8$ are equal.
2. The eight indices consist of two different groups, with two equal indices in the first group and six equal indices in the second group, e.g., $i_1 = i_2, i_3 = i_4 = \cdots = i_8, i_1 \neq i_3$.

3. The eight indices consist of two different groups, with three equal indices in the first group and five equal indices in the second group, e.g., $i_1 = i_2 = i_3, i_4 = i_5 = \cdots = i_8, i_1 \neq i_4$.

4. The eight indices consist of two different groups, with four equal indices in each group, e.g., $i_1 = i_2 = i_3 = i_4, i_5 = j_6 = i_7 = i_8, i_1 \neq i_5$.

5. The eight indices consist of three different groups, with two equal indices in the first group, two equal indices in the second group, and four equal indices in the third group, e.g., $i_1 = i_2, i_3 = i_4, i_5 = i_6 = i_7 = i_8, i_1 \neq i_3 \neq i_5$.

6. The eight indices consist of three different groups, with two equal indices in the first group, three equal indices in the second group, and three equal indices in the third group, e.g., $i_1 = i_2, i_3 = i_4 = i_5, i_6 = i_7 = i_8, i_1 \neq i_3 \neq i_6$.

7. The eight indices consist of four different groups, with two equal indices in each group, e.g., $i_1 = i_2, i_3 = i_4, i_5 = i_6, i_7 = i_8, i_1 \neq i_3 \neq i_5 \neq i_7$.

By some tedious algebra, we can write down the result for $n = 4$ as in (7) and the f_γ's are as follows (where $\odot$ denotes the Hadamard product symbol):

$$
\begin{aligned}
f_{\gamma_2} =& \mathrm{tr}(A_1)\mathrm{tr}(A_2)\mathrm{tr}(A_3 \odot A_4) + \mathrm{tr}(A_1)\mathrm{tr}(A_3)\mathrm{tr}(A_2 \odot A_4) + \mathrm{tr}(A_1)\mathrm{tr}(A_4)\mathrm{tr}(A_2 \odot A_3) + \mathrm{tr}(A_2)\mathrm{tr}(A_3)\mathrm{tr}(A_1 \odot A_4) \\
&+ \mathrm{tr}(A_2)\mathrm{tr}(A_4)\mathrm{tr}(A_1 \odot A_3) + \mathrm{tr}(A_3)\mathrm{tr}(A_4)\mathrm{tr}(A_1 \odot A_2) + 2[\iota'(A_1 \odot A_2)\iota\mathrm{tr}(A_3 \odot A_4) + \iota'(A_1 \odot A_3)\iota\mathrm{tr}(A_2 \odot A_4) \\
&+ \iota'(A_1 \odot A_4)\iota\mathrm{tr}(A_2 \odot A_3) + \iota'(A_2 \odot A_3)\iota\mathrm{tr}(A_1 \odot A_4) + \iota'(A_2 \odot A_4)\iota\mathrm{tr}(A_1 \odot A_3) + \iota'(A_3 \odot A_4)\iota\mathrm{tr}(A_1 \odot A_2)] \\
&+ 4[\mathrm{tr}(A_1)\mathrm{tr}(A_2 \odot (A_3A_4)) + \mathrm{tr}(A_1)\mathrm{tr}(A_3 \odot (A_2A_4)) + \mathrm{tr}(A_1)\mathrm{tr}(A_4 \odot (A_2A_3)) + \mathrm{tr}(A_2)\mathrm{tr}(A_1 \odot (A_3A_4)) \\
&+ \mathrm{tr}(A_2)\mathrm{tr}(A_3 \odot (A_1A_4)) + \mathrm{tr}(A_2)\mathrm{tr}(A_4 \odot (A_1A_3)) + \mathrm{tr}(A_3)\mathrm{tr}(A_1 \odot (A_2A_4)) + \mathrm{tr}(A_3)\mathrm{tr}(A_2 \odot (A_1A_4)) \\
&+ \mathrm{tr}(A_3)\mathrm{tr}(A_4 \odot (A_1A_2)) + \mathrm{tr}(A_4)\mathrm{tr}(A_1 \odot (A_2A_3)) + \mathrm{tr}(A_4)\mathrm{tr}(A_2 \odot (A_1A_3)) + \mathrm{tr}(A_4)\mathrm{tr}(A_3 \odot (A_1A_2))] \\
&+ 8[\mathrm{tr}((I \odot A_1)A_2A_3A_4) + \mathrm{tr}((I \odot A_1)A_2A_4A_3) + \mathrm{tr}((I \odot A_1)A_3A_2A_4) + \mathrm{tr}((I \odot A_2)A_1A_3A_4) + \mathrm{tr}((I \odot A_2)A_1A_4A_3) \\
&+ \mathrm{tr}((I \odot A_2)A_3A_1A_4) + \mathrm{tr}((I \odot A_3)A_1A_2A_4) + \mathrm{tr}((I \odot A_3)A_1A_4A_2) + \mathrm{tr}((I \odot A_3)A_2A_1A_4) + \mathrm{tr}((I \odot A_4)A_1A_2A_3) \\
&+ \mathrm{tr}((I \odot A_4)A_1A_3A_2) + \mathrm{tr}((I \odot A_4)A_2A_1A_3)] + 16[\iota'(I \odot (A_1A_2))(I \odot (A_3A_4))\iota + \iota'(I \odot (A_1A_3))(I \odot (A_2A_4))\iota \\
&+ \iota'(I \odot (A_1A_4))(I \odot (A_2A_3))\iota], \\
f_{\gamma_4} =& \mathrm{tr}(A_1)\mathrm{tr}(A_2 \odot A_3 \odot A_4) + \mathrm{tr}(A_2)\mathrm{tr}(A_1 \odot A_3 \odot A_4) + \mathrm{tr}(A_3)\mathrm{tr}(A_1 \odot A_2 \odot A_4) + \mathrm{tr}(A_4)\mathrm{tr}(A_1 \odot A_2 \odot A_3) \\
&+ 4[\mathrm{tr}(A_1 \odot A_2 \odot (A_3A_4)) + \mathrm{tr}(A_1 \odot A_3 \odot (A_2A_4)) + \mathrm{tr}(A_1 \odot A_4 \odot (A_2A_3)) + \mathrm{tr}(A_2 \odot A_3 \odot (A_1A_4)) \\
&+ \mathrm{tr}(A_2 \odot A_4 \odot (A_1A_3)) + \mathrm{tr}(A_3 \odot A_4 \odot (A_1A_2))], \\
f_{\gamma_6} =& \mathrm{tr}(A_1 \odot A_2 \odot A_3 \odot A_4), \\
f_{\gamma_1^2} =& 2[\iota'(I \odot A_2)A_3(I \odot A_4)\iota\mathrm{tr}(A_1) + \iota'(I \odot A_2)A_4(I \odot A_3)\iota\mathrm{tr}(A_1) + \iota'(I \odot A_3)A_2(I \odot A_4)\iota\mathrm{tr}(A_1) \\
&+ \iota'(I \odot A_1)A_3(I \odot A_4)\iota\mathrm{tr}(A_2) + \iota'(I \odot A_1)A_4(I \odot A_3)\iota\mathrm{tr}(A_2) + \iota'(I \odot A_3)A_1(I \odot A_4)\iota\mathrm{tr}(A_2) \\
&+ \iota'(I \odot A_1)A_2(I \odot A_4)\iota\mathrm{tr}(A_3) + \iota'(I \odot A_1)A_4(I \odot A_2)\iota\mathrm{tr}(A_3) + \iota'(I \odot A_2)A_1(I \odot A_4)\iota\mathrm{tr}(A_3) \\
&+ \iota'(I \odot A_1)A_2(I \odot A_3)\iota\mathrm{tr}(A_4) + \iota'(I \odot A_1)A_3(I \odot A_2)\iota\mathrm{tr}(A_4) + \iota'(I \odot A_2)A_1(I \odot A_3)\iota\mathrm{tr}(A_4)] \\
&+ 4[\iota'(I \odot A_1)A_2A_3(I \odot A_4)\iota + \iota'(I \odot A_1)A_2A_4(I \odot A_3)\iota + \iota'(I \odot A_1)A_3A_4(I \odot A_2)\iota
\end{aligned}
$$

$$
\begin{aligned}
&+\iota'(I\odot A_2)A_1A_3(I\odot A_4)\iota+\iota'(I\odot A_2)A_1A_4(I\odot A_3)\iota+\iota'(I\odot A_2)A_3A_4(I\odot A_1)\iota\\
&+\iota'(I\odot A_3)A_1A_2(I\odot A_4)\iota+\iota'(I\odot A_3)A_1A_4(I\odot A_2)\iota+\iota'(I\odot A_3)A_2A_4(I\odot A_1)\iota\\
&+\iota'(I\odot A_4)A_1A_2(I\odot A_3)\iota+\iota'(I\odot A_4)A_1A_3(I\odot A_2)\iota+\iota'(I\odot A_4)A_2A_3(I\odot A_1)\iota]\\
&+4[\iota'(A_2\odot A_3\odot A_4)\iota\mathrm{tr}(A_1)+\iota'(A_1\odot A_3\odot A_4)\iota\mathrm{tr}(A_2)+\iota'(A_1\odot A_2\odot A_4)\iota\mathrm{tr}(A_3)\\
&+\iota'(A_1\odot A_2\odot A_3)\iota\mathrm{tr}(A_4)]+8[\iota'(A_1\odot A_2)A_3(I\odot A_4)\iota+\iota'(A_1\odot A_2)A_4(I\odot A_3)\iota\\
&+\iota'(A_1\odot A_3)A_2(I\odot A_4)\iota+\iota'(A_1\odot A_3)A_4(I\odot A_2)\iota+\iota'(A_1\odot A_4)A_2(I\odot A_3)\iota\\
&+\iota'(A_1\odot A_4)A_3(I\odot A_2)\iota+\iota'(A_2\odot A_3)A_1(I\odot A_4)\iota+\iota'(A_2\odot A_3)A_4(I\odot A_1)\iota\\
&+\iota'(A_2\odot A_4)A_1(I\odot A_3)\iota+\iota'(A_2\odot A_4)A_3(I\odot A_1)\iota+\iota'(A_3\odot A_4)A_1(I\odot A_2)\iota\\
&+\iota'(A_3\odot A_4)A_2(I\odot A_1)\iota]+16[\mathrm{tr}(A_1(A_2\odot A_3)A_4)+\mathrm{tr}(A_1(A_2\odot A_4)A_3)+\mathrm{tr}(A_1(A_3\odot A_4)A_2)\\
&+\mathrm{tr}(A_2(A_1\odot A_3)A_4)+\mathrm{tr}(A_2(A_1\odot A_4)A_3)+\mathrm{tr}(A_3(A_1\odot A_2)A_4)],\\
f_{\gamma_2^2}=&\mathrm{tr}(A_1\odot A_2)\mathrm{tr}(A_3\odot A_4)+\mathrm{tr}(A_1\odot A_3)\mathrm{tr}(A_2\odot A_4)+\mathrm{tr}(A_1\odot A_4)\mathrm{tr}(A_2\odot A_3)\\
&+4[\iota'(I\odot A_1)(A_2\odot A_3)(I\odot A_4)\iota+\iota'(I\odot A_1)(A_2\odot A_4)(I\odot A_3)\iota+\iota'(I\odot A_1)(A_3\odot A_4)(I\odot A_2)\iota\\
&+\iota'(I\odot A_2)(A_1\odot A_3)(I\odot A_4)\iota+\iota'(I\odot A_2)(A_1\odot A_4)(I\odot A_3)\iota+\iota'(I\odot A_3)(A_1\odot A_2)(I\odot A_4)\iota]\\
&+8\iota'(A_1\odot A_2\odot A_3\odot A_4)\iota,\\
f_{\gamma_1\gamma_3}=&\,2[\iota'(I\odot A_1)A_2(I\odot A_3\odot A_4)\iota+\iota'(I\odot A_1)A_3(I\odot A_2\odot A_4)\iota\\
&+\iota'(I\odot A_1)A_4(I\odot A_2\odot A_3)\iota+\iota'(I\odot A_2)A_1(I\odot A_3\odot A_4)\iota+\iota'(I\odot A_2)A_3(I\odot A_1\odot A_4)\iota\\
&+\iota'(I\odot A_2)A_4(I\odot A_1\odot A_3)\iota+\iota'(I\odot A_3)A_1(I\odot A_2\odot A_4)\iota+\iota'(I\odot A_3)A_2(I\odot A_1\odot A_4)\iota\\
&+\iota'(I\odot A_3)A_4(I\odot A_1\odot A_2)\iota+\iota'(I\odot A_4)A_1(I\odot A_2\odot A_3)\iota+\iota'(I\odot A_4)A_2(I\odot A_1\odot A_3)\iota\\
&+\iota'(I\odot A_4)A_3(I\odot A_1\odot A_2)\iota]+8[\iota'(I\odot A_1)(A_2\odot A_3\odot A_4)\iota+\iota'(I\odot A_2)(A_1\odot A_3\odot A_4)\iota\\
&+\iota'(I\odot A_3)(A_1\odot A_2\odot A_4)\iota+\iota'(I\odot A_4)(A_1\odot A_2\odot A_3)\iota].
\end{aligned}
$$

Similarly, $E\left(y\prod_{i=1}^3 y'A_iy\right)$ have a representative element $E\left(y_j\prod_{i=1}^3 y'A_iy\right)$, $j=1,\cdots,m$. Then following (5), $E\left(y_j\prod_{i=1}^3 y'A_iy\right)$ =tr$\{E[y_j(y^{\otimes})(y^{\otimes})']A^{\otimes}\}$, where $y_j(y^{\otimes})(y^{\otimes})'$ has elements $y_j\prod_{i=1}^m y_i^{\alpha(i)}=y_jy_1^{\alpha(1)}\cdots y_m^{\alpha(m)}$ with $\alpha(1)+\cdots+\alpha(m)=2(n-1)=6$. We put $y_j\prod_{i=1}^m y_i^{\alpha(i)}=y_jy_{i_1}\cdots y_{i_6}$, which has nonzero expectation only in the following four situations:

1. All the seven indices $j,i_1,\cdots,i_6$ are equal.

2. The seven indices consist of two different groups, with two equal indices in the first group and five equal indices in the second group, e.g., $j=i_1$, $i_2=i_3=\cdots=i_6$, $j\neq i_2$, or $i_1=i_2$, $j=i_3=i_4=\cdots=i_6$, $i_1\neq j$.

3. The seven indices consist of two different groups, with three equal indices in the first group and four equal indices in the second group, e.g., $j=i_1=i_2$, $i_3=i_4=\cdots=i_6$, $j\neq i_3$, or $i_1=i_2=i_3$, $j=i_4=i_5=i_6$, $i_1\neq j$.

4. The seven indices consist of three different groups, with two equal indices in the first group, two equal indices in the second group, and three equal indices in the third group, e.g., $j=i_1$, $i_2=i_3$, $i_4=i_5=i_6$, $j\neq i_2\neq i_4$, or $i_1=i_2$, $i_3=i_4$, $j=i_5=i_6$, $i_1\neq i_3\neq j$.

By some tedious algebra, we can write down the result as in (8) and the h_γ's are

$h_{\gamma_5} = (I \odot A_1 \odot A_2 \odot A_3)\,\iota,$

$h_{\gamma_3} = 4[I \odot A_1 \odot (A_2A_3)]\iota + 4[I \odot A_2 \odot (A_1A_3)]\iota + 4[I \odot A_3 \odot (A_1A_2)]\iota + 2A_1(I \odot A_2 \odot A_3)\iota$
$\quad +2A_2(I \odot A_1 \odot A_3)\iota + 2A_3(I \odot A_1 \odot A_2)\iota + \mathrm{tr}(A_1)(I \odot A_2 \odot A_3)\iota + \mathrm{tr}(A_2)(I \odot A_1 \odot A_3)\iota$
$\quad +\mathrm{tr}(A_3)(I \odot A_1 \odot A_2)\iota,$

$h_{\gamma_1} = 4\mathrm{tr}(A_1)[I \odot (A_2A_3)]\iota + 4\mathrm{tr}(A_2)[I \odot (A_1A_3)]\iota + 4\mathrm{tr}(A_3)[I \odot (A_1A_2)]\iota + 8A_1(A_2 \odot A_3)\iota$
$\quad +8A_2(A_1 \odot A_3)\iota + 8A_3(A_1 \odot A_2)\iota + \mathrm{tr}(A_1)\mathrm{tr}(A_2)\,(I \odot A_3)\iota + \mathrm{tr}(A_1)\mathrm{tr}(A_3)\,(I \odot A_2)\iota$
$\quad +\mathrm{tr}(A_2)\mathrm{tr}(A_3)\,(I \odot A_1)\iota + 2\iota'(A_1 \odot A_2)\iota(I \odot A_3)\iota + 2\iota'(A_1 \odot A_3)\iota(I \odot A_2)\iota + 2\iota'(A_2 \odot A_3)\iota(I \odot A_1)\iota$
$\quad +8[I \odot (A_1A_2A_3)]\iota + 8[I \odot (A_1A_3A_2)]\iota + 8[I \odot (A_2A_1A_3)]\iota + 4A_1A_2(I \odot A_3)\iota + 4A_1A_3(I \odot A_2)\iota$
$\quad +4A_2A_1(I \odot A_3)\iota + 4A_2A_3(I \odot A_1)\iota + 4A_3A_1(I \odot A_2)\iota + 4A_3A_2(I \odot A_1)\iota + 2\mathrm{tr}(A_1)A_2(I \odot A_3)\iota$
$\quad +2\mathrm{tr}(A_1)A_3(I \odot A_2)\iota + 2\mathrm{tr}(A_2)A_1(I \odot A_3)\iota + 2\mathrm{tr}(A_2)A_3(I \odot A_1)\iota + 2\mathrm{tr}(A_3)A_1(I \odot A_2)\iota$
$\quad +2\mathrm{tr}(A_3)A_2(I \odot A_1)\iota,$

$h_{\gamma_1\gamma_2} = 8(A_1 \odot A_2 \odot A_3)\iota + 4(A_1 \odot A_2)(I \odot A_3)\iota + 4(A_1 \odot A_3)(I \odot A_2)\iota + 4(A_2 \odot A_3)(I \odot A_1)\iota$
$\quad +\mathrm{tr}(A_1 \odot A_2)(I \odot A_3)\iota + \mathrm{tr}(A_1 \odot A_3)(I \odot A_2)\iota + \mathrm{tr}(A_2 \odot A_3)(I \odot A_1)\iota + 2(I \odot A_1)A_2(I \odot A_3)\iota$
$\quad +2(I \odot A_1)A_3(I \odot A_2)\iota + 2(I \odot A_2)A_1(I \odot A_3)\iota + 2(I \odot A_2)A_3(I \odot A_1)\iota + 2(I \odot A_3)A_1(I \odot A_2)\iota$
$\quad +2(I \odot A_3)A_2(I \odot A_1)\iota.$ ■

Appendix B: Terms for $M(\hat{\phi})$ in MA(1)

The following terms that are needed to evaluate the MSE result (13) of the QMLE $\hat{\phi}$ can be derived using the nonnormal order-4 quadratic form result (7) as well as those for nonnormal quadratic forms of lower orders. A Mathematica code used to help simply the expressions is available upon request from the authors.

$\mathrm{tr}(N_2) = \frac{T}{1-\phi^2} - \frac{1}{(1-\phi^2)^2} + o(T^{-2}),$

$\mathrm{tr}(N_3) = \frac{6T\phi}{(1-\phi^2)^2} - \frac{12\phi}{(1-\phi^2)^3} + o(T^{-2}),$

$\mathrm{tr}(N_4) = \frac{12T(1+3\phi^2)}{(1-\phi^2)^3} - \frac{24(1+5\phi^2)}{(1-\phi^2)^4} + o(T^{-2}),$

$\delta_{200} = \mathrm{tr}(N_2) = \frac{T}{1-\phi^2} - \frac{1}{(1-\phi^2)^2} + o(T^{-2}),$

$\delta_{210} = \frac{T^2}{(1-\phi^2)^2} + T\left[\frac{6+20\phi^2}{(1-\phi^2)^3} + \frac{2\phi\left(1+2\phi+2\phi^2\right)\gamma_1^2}{(1+\phi)(1-\phi^3)^2} + \frac{2\gamma_2}{(1-\phi^2)^2}\right] - \frac{13+68\phi^2}{(1-\phi^2)^4} - \frac{2\phi\left(2+6\phi+9\phi^2+5\phi^3+\phi^4\right)\gamma_1^2}{(1+\phi)^2(1-\phi^3)^3}$
$\quad -\frac{\left(3+4\phi^2\right)\gamma_2}{(1-\phi^2)^3(1+\phi^2)} + o(T^{-2}),$

$\delta_{300} = T\left[\frac{6\phi}{(1-\phi^2)^2} + \frac{\gamma_1^2}{1-\phi^3}\right] - \frac{12\phi}{(1-\phi^2)^3} - \frac{\gamma_1^2}{(1-\phi^3)^2} + o(T^{-2}),$

$\delta_{220} = \frac{T^3}{(1-\phi^2)^3} + T^2\left[\frac{\left(19+86\phi^2\right)}{(1-\phi^2)^4} + \frac{4\phi\left(1+2\phi+2\phi^2\right)\gamma_1^2}{(1-\phi)^3(1+\phi)^2(1+\phi+\phi^2)^2} + \frac{5\gamma_2}{(1-\phi^2)^3}\right]$
$\quad +T\left[\frac{3\left(21+230\phi^2+136\phi^4\right)}{(1-\phi^2)^5} + \frac{\left(60+361\phi^2+304\phi^4\right)\gamma_2}{(1-\phi^2)^4(1+\phi^2)} + \frac{2\,\gamma_4}{(1-\phi^2)^3}\right.$
$\qquad + \frac{4(12+97\phi+322\phi^2+728\phi^3+1122\phi^4+1312\phi^5+1172\phi^6+873\phi^7+458\phi^8+142\phi^9)\gamma_1^2}{(1+\phi)^3(1-\phi^3)^4}$
$\qquad \left.+\frac{2\,\left(1+7\phi^2+11\phi^4+9\phi^6\right)\gamma_2^2}{(1-\phi^4)^3} + \frac{8\,\phi\left(1+2\phi+2\phi^2\right)\gamma_1\gamma_3}{(1-\phi)^3(1+\phi)^2(1+\phi+\phi^2)^2}\right]$
$\quad -\frac{9\left(27+392\phi^2+344\phi^4\right)}{(1-\phi^2)^6} - \frac{\left(141+1456\phi^2+3633\phi^4+3454\phi^6+1288\phi^8\right)\gamma_2}{(1-\phi^2)^5(1+\phi^2)^3} - \frac{\left(3+9\phi^2+8\phi^4+6\phi^6\right)\gamma_4}{(1-\phi^2)^4(1+2\phi^2+2\phi^4+\phi^6)}$

$$
\begin{aligned}
&-\frac{2(53+556\phi+2487\phi^2+7911\phi^3+18534\phi^4+35375\phi^5+55735\phi^6+75214\phi^7+86169\phi^8+85487\phi^9+72074\phi^{10}+51669\phi^{11})\gamma_1^2}{(1+\phi)^4(1+\phi^2)^2(1-\phi^3)^5}\\
&-\frac{2(30280\phi^{12}+14366\phi^{13}+4804\phi^{14}+1038\phi^{15}+64\phi^{16})\gamma_1^2}{(1+\phi)^4(1+\phi^2)^2(1-\phi^3)^5}-\frac{2\left(2+16\phi^2+37\phi^4+46\phi^6+11\phi^8\right)\gamma_2^2}{(1-\phi^4)^4}\\
&-\frac{4\phi(4+21\phi+63\phi^2+132\phi^3+228\phi^4+333\phi^5+421\phi^6+463\phi^7+451\phi^8+381\phi^9+278\phi^{10}+171\phi^{11}+86\phi^{12}+30\phi^{13}+6\phi^{14})\gamma_1\gamma_3}{(1-\phi)^4(1+\phi)^3(1+\phi^2)(1+\phi+\phi^2)^3(1+\phi+\phi^2+\phi^3+\phi^4)^2}+o(T^{-2}),
\end{aligned}
$$

$$
\begin{aligned}
\delta_{301}&=T^2\Big[\frac{18\left(1+5\phi^2\right)}{(1-\phi^2)^4}+\frac{6\phi\gamma_1^2}{(1-\phi)^3(1+\phi)^2(1+\phi+\phi^2)}\Big]+T\Big[\frac{18\left(5+53\phi^2+30\phi^4\right)}{(1-\phi^2)^5}\\
&+\frac{6\phi\left(57+207\phi+499\phi^2+761\phi^3+909\phi^4+821\phi^5+634\phi^6+336\phi^7+108\phi^8\right)\gamma_1^2}{(1+\phi)^3(1-\phi^3)^4}\\
&+\frac{36\left(2+9\phi^2\right)\gamma_2}{(1-\phi^2)^4}+\frac{6\left(1+3\phi^2+4\phi^4+3\phi^6\right)\gamma_2^2}{(1-\phi^4)^3}+\frac{12\phi\gamma_1\gamma_3}{(1-\phi)^3(1+\phi)^2(1+\phi+\phi^2)}\Big]\\
&-\frac{36\left(9+129\phi^2+112\phi^4\right)}{(1-\phi^2)^6}-\frac{18\left(9+91\phi^2+223\phi^4+208\phi^6+76\phi^8\right)\gamma_2}{(1-\phi^2)^5(1+\phi^2)^3}-\frac{6\phi(160+751\phi+2561\phi^2+6012\phi^3+11708\phi^4+18480\phi^5)\gamma_1^2}{(1+\phi)^4(1+\phi^2)^2(1-\phi^3)^5}\\
&-\frac{6\phi(25218\phi^6+28931\phi^7+28917\phi^8+24350\phi^9+17530\phi^{10}+10204\phi^{11}+4842\phi^{12}+1578\phi^{13}+336\phi^{14}+12\phi^{15})\gamma_1^2}{(1+\phi)^4(1+\phi^2)^2(1-\phi^3)^5}\\
&-\frac{6\left(2+9\phi^2+18\phi^4+18\phi^6+4\phi^8\right)\gamma_2^2}{(1-\phi^4)^4}-\frac{6\phi\left(5+15\phi+29\phi^2+43\phi^3+57\phi^4+61\phi^5+56\phi^6+44\phi^7+30\phi^8+14\phi^9+4\phi^{10}\right)\gamma_1\gamma_3}{(1-\phi)^4(1+\phi)^3(1+\phi+\phi^2)^2(1+\phi+\phi^2+\phi^3+\phi^4)^2}+o(T^{-2}),
\end{aligned}
$$

$$
\begin{aligned}
\delta_{310}&=T^2\left[\frac{18\ \phi}{(1-\phi^2)^3}+\frac{\gamma_1^2}{1-\phi^2-\phi^3+\phi^5}\right]+T\left[\frac{6\ \phi\left(19+22\phi^2\right)}{(1-\phi^2)^4}+\frac{66\ \phi\gamma_2}{(1-\phi^2)^3}+\frac{2\ \phi\left(1+2\phi^2\right)\gamma_2^2}{(1-\phi^4)^2}\right.\\
&\left.+\frac{\left(34+92\phi+211\phi^2+297\phi^3+383\phi^4+300\phi^5+144\phi^6\right)\gamma_1^2}{(1+\phi)^2(1-\phi^3)^3}+\frac{2\ \gamma_1\gamma_3}{1-\phi^2-\phi^3+\phi^5}\right]-\frac{12\phi\left(33+58\phi^2\right)}{(1-\phi^2)^5}-\frac{6\phi\left(26+55\phi^2+33\phi^4\right)\gamma_2}{(1-\phi^2)^4(1+\phi^2)^2}-\\
&\frac{(65+255\phi+816\phi^2+1750\phi^3+3184\phi^4+4320\phi^5+4719\phi^6+3887\phi^7+2550\phi^8+1056\phi^9+264\phi^{10}+12\phi^{11})\gamma_1^2}{(1+\phi)^3(1+\phi^2)(1-\phi^3)^4}\\
&-\frac{2\phi\left(2+5\phi^2+\phi^4\right)\gamma_2^2}{(1-\phi^4)^3}-\frac{\left(3+6\phi+8\phi^2+9\phi^3+10\phi^4+6\phi^5+2\phi^6\right)\gamma_1\gamma_3}{(1-\phi)^3(1+\phi)^2(1+\phi+\phi^2)^2(1+\phi+\phi^2+\phi^3+\phi^4)}+o(T^{-2}),
\end{aligned}
$$

$$
\begin{aligned}
\delta_{400}&=\frac{3T^2}{(1-\phi^2)^2}+T\left[\frac{6\ \left(2+7\phi^2\right)}{(1-\phi^2)^3}+\frac{12\ \gamma_2}{(1-\phi^2)^2}+\frac{12\ \phi\left(2+3\phi+3\phi^2\right)\gamma_1^2}{(1+\phi)(1-\phi^3)^2}+\frac{\gamma_2^2}{1-\phi^4}\right]\\
&-\frac{9\left(3+16\phi^2\right)}{(1-\phi^2)^4}-\frac{6\left(3+4\phi^2\right)\gamma_2}{(1-\phi^2)^3(1+\phi^2)}-\frac{24\phi\left(2+5\phi+7\phi^2+4\phi^3+\phi^4\right)\gamma_1^2}{(1+\phi)^2(1-\phi^3)^3}-\frac{\gamma_2^2}{(1-\phi^4)^2}+o(T^{-2}).
\end{aligned}
$$

References

Bao, Y. (2007) The approximate moments of the least squares estimator for the stationary autoregressive model under a general error distribution. *Econometric Theory* 23, 1013-1021.

Bao, Y. & A. Ullah (2007a) Finite sample properties of maximum likelihood estimator in spatial models. *Journal of Econometrics* 137, 396-413.

Bao, Y. & A. Ullah (2007b) The second-order bias and mean squared error of estimators in time series models. *Journal of Econometrics*, 140,650-669.

Chandra, R. (1983) Estimation of Econometric Models when Disturbances are not Necessarily Normal. Ph.D. Dissertation, Department of Statistics, Lucknow University, India.

Don, F.J.H. (1979) The expectations of products of quadratic forms in normal variables. *Statistica Neerlandica* 33, 73-79.

Dressel, P.L. (1940) Statistical semi-invariants and their estimates with particular emphasis on their relation to algebraic invariants. *Annals of Mathematical Statistics* 11, 33-57.

Dufour, J.-M. (1984) Unbiasedness of predictions from estimated autoregressions when the true order is unknown. *Econometrica* 52, 209-215

Fisher, R.A. (1928) Moments and product moments of sampling distributions. *Proceedings of the London Mathematical Society* 30, 199-238.

Ghazal, G.A. (1996) Recurrence formula for expectations of products of quadratic forms. *Statistics and Probability Letters* 27, 101-109.

Hoque, A., J.R. Magnus & B. Pesaran (1988) The exact multi-period mean-square forecast error for the first-order autoregressive model. *Journal of Econometrics* 39, 327-346.

Kadane, J.B. (1971) Comparison of k-class estimators when the disturbances are small. *Econometrica* 39, 723–737.

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

Komunjer, I. (2007) Asymmetric power distribution: theory and applications to risk measurement. *Journal of Applied Econometrics* 22, 891-921.

Kumar, A. (1973) Expectations of product of quadratic forms. *Sanhkya (Series B)* 35, 359-362.

Lieberman, O. (1994) Saddlepoint approximations for the least squares estimator in first-order autoregression. *Biometrika* 81, 807-811.

Lieberman, O. (1997) The effect of nonnormality. *Econometric Theory* 13, 52-78.

Magee, L. (1985) Efficiency of iterative estimators in the regression model with AR(1) disturbances. *Journal of Econometrics* 29, 275-287

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.

Magnus, J.R. & H. Neudecker (1979) The commutation matrix: some properties and applications. *Annals of Statistics* 7, 381-394

Mathai, A.M. & S.B. Provost (1992) *Quadratic Forms in Random Variables*. New York: Marcel Dekker.

Mishra, P.N (1972) Recurrence formula for mathematical expectation of products of matrices of structural disturbances. *Sanhkya (Series B)* 34, 379-384.

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.

Pesaran, M.H. & T. Yamagata (2005) Testing slope homogeneity in large panels. Working paper, Cambridge University.

Smith, M.D. (1993) Expectations of ratios of quadratic forms in normal variables: evaluating some top order invariant polynomials. *Australian Journal of Statistics* 35, 271-282.

Srivastava, V.K. & K. Maekawa (1995) Efficiency properties of feasible generalized least squares estimators in SURE models under nonnormal disturbances. *Journal of Econometrics* 66, 99-121.

Srivastava, V.K. & R. Tiwari (1976) Evaluation of expectations of products of stochastic matrices. *Scandinavian Journal of Statistics* 3, 135-138.

Stuart, A. & J.K. Ord (1987) *Kendall's Advanced Theory of Statistics* (Vol. 1, 5th ed.). New York: Oxford University Press.

Theil, H. & A.L. Nagar (1961) Testing the independence of regression disturbances. *Journal of the American Statistical Association* 56, 793-806.

Ullah, A. (1990) Finite sample econometrics: a unified approach. In Carter, R.A.L., J. Dutta and A. Ullah (eds.), *Contributions to Econometric Theory and Application, Volume in Honor of A.L. Nagar*. New York: Springer-Verlag, pp. 242-292.

Ullah, A. (2004) *Finite Sample Econometrics*. New York: Oxford University Press.

Ullah, A., V.K. Srivastava & R. Chandra (1983) Properties of shrinkage estimators when the disturbances are not normal. *Journal of Econometrics* 21, 389-402.

White, J.S. (1957) Approximate moments for the serial correlation coefficients. *Annals of Mathematical Statistics* 28, 798-803

Zivot, E, R. Startz & C.R. Nelson (1998) Valid confidence intervals and inference in the presence of weak instruments. *International Economic Review* 39, 1119-1144.

Figure 1: MSE for the Intercept Model, T = 50, σ^2 = 0.5
(True: Solid; Nonnormal: Short Dashes; Normal: Dots and Dashes)

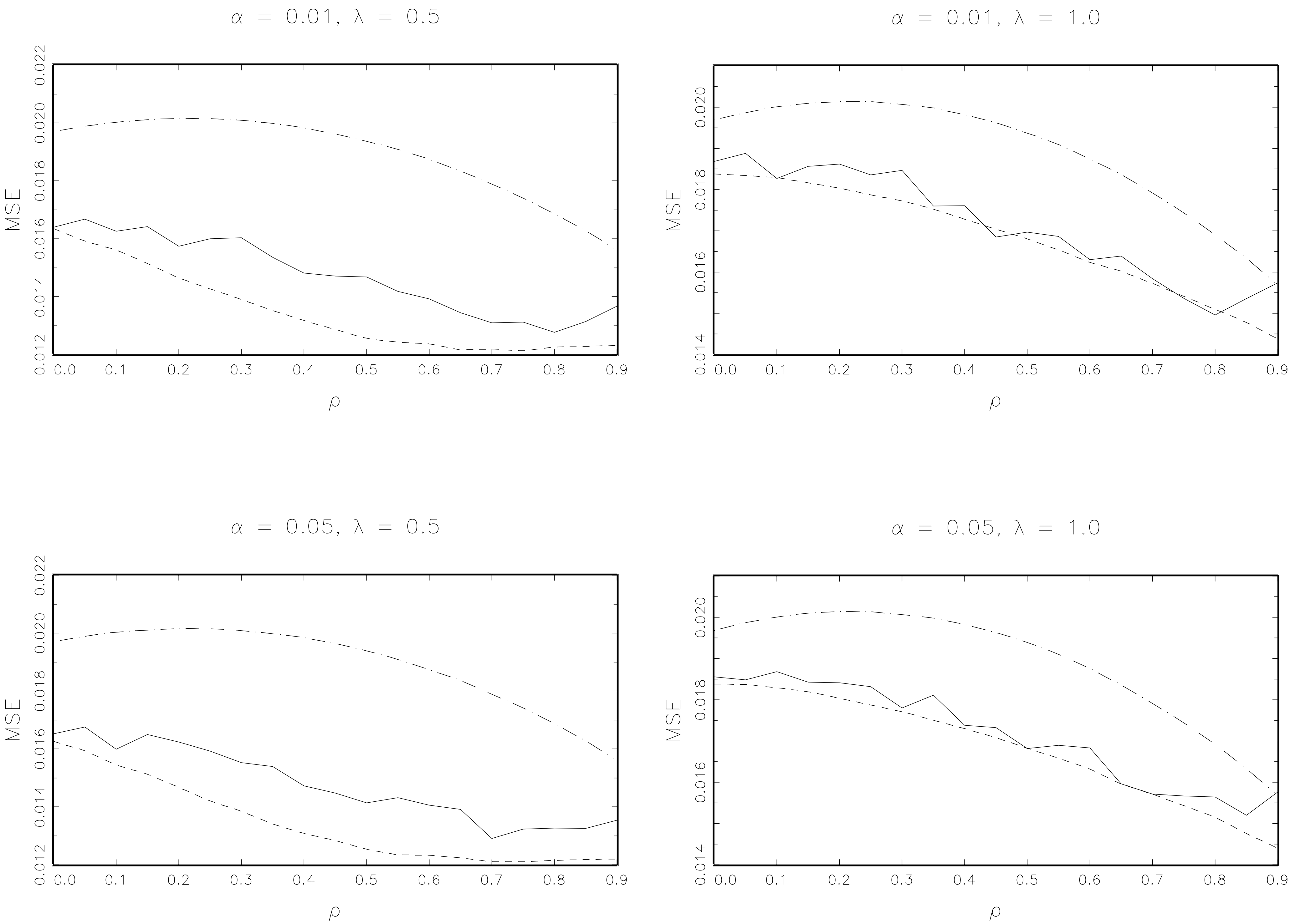

Figure 2: MSE for the Intercept Model, T = 50, σ^2 = 1.0
(True: Solid; Nonnormal: Short Dashes; Normal: Dots and Dashes)