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

Yong Bao[†]
Department of Economics
Temple University

Aman Ullah[‡]
Department of Economics
University of California, Riverside

August 20, 2007

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 normal variables with nonzero mean follows a recurrence formula. The formula includes the existing result for normal variables with zero mean as a special case. For nonnormal variables, while the existing results are available only for quadratic forms of order up to 3, we derive analytical results for quadratic form of order 4 and half quadratic form of order 3. The results involve the cumulants of the nonnormal distribution up to the eighth order for order 4 quadratic from, and up to the seventh order for order 3 half quadratic from. Setting all the nonnormality parameters equal to zero gives the results for normal vectors. Some econometric applications using our normal and nonnormal results are presented.

Keywords: Expectation; Quadratic form; Nonnormality

JEL Classification: C10; C19

*We are grateful to Peter Phillips (the editor) and three anonymous referees for their constructive and detailed feedback that greatly improved the paper. We are also thankful to Jason Abrevaya, Fathali Firoozi, and seminar participants at Purdue University for their comments and suggestions. All remaining errors are our own.

[†]Department of Economics, Temple University, Philadelphia, PA 19122, (215) 204-8165, Fax: (215) 204-8173, e-mail: yong.bao@temple.edu.

[‡]Corresponding Author: Department of Economics, University of California, Riverside, CA, 92521, (951) 827-1591, Fax: (951) 787-5685, 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 (1994a), 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 $E(\prod_{i=1}^{n} Q_i)$, 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 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 4.2, the expectation of "half" quadratic form $E(y\prod_{i=1}^{n-1} y'A_iy)$, 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 $E(\prod_{i=1}^{n} y'A_iy)$ or $E(y\prod_{i=1}^{n-1} y'A_iy)$, though for some special nonnormal distributions, including mixtures of normal, we may invoke 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 normally distributed. We are going to utilize a nonstochastic operator proposed by Ullah (2004) to facilitate the derivation. Moreover, we are going to relate this approach to the moment generating function approach and we discuss possible advantage of our recursive approach in terms of computer time. When y has zero mean, the recursive result degenerates to the result given in Ghazal (1996). Secondly, we try to derive the expectations of products of quadratic form of order 4 and half quadratic form of order 3 in a general nonnormal random vector y. 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 gives econometric applications of the main results. Section 5 concludes.

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 first the less known approach of Ullah (2004) and the more popular moment generating function approach.

2.1 A Nonstochastic Operator

In his monograph, Ullah (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 $x \sim N(\mu, \Omega)$,

$$Eh\left(x\right) = h\left(d\right) \cdot 1 = h\left(d\right) \text{ and } Eh\left(x\right) g\left(x\right) = h\left(d\right) Eg\left(x\right) \tag{1}$$

for the real-valued analytic functions $h\left(\cdot\right)$ and $g\left(\cdot\right)$, where $d = \mu + \Omega \frac{\partial}{\partial \mu}$ is a nonstochastic operator. 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 $Ad \cdot 1 = A\mu$ given the recursive nature of d (alternatively, $d'A_1 d \cdot 1 = \text{tr}(Add' \cdot 1) = \mu'A\mu + \text{tr}(A\Omega)$ by applying d to $d' \cdot 1 = \mu'$). 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 not correct.

Given (1), a recursive procedure $E(\prod_{i=1}^{n} Q_i) = d'A_1 d \cdot E(\prod_{i=2}^{n} Q_i)$ immediately follows for $Q_i = y'A_i y$, $y \sim N(\mu, I)$. Apparently, we need to expand the operation $d'A_1 d$ 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 $d'A_1 d \cdot E(\prod_{i=2}^{n} Q_i)$ explicitly. As a consequence, a recursive formula is derived for $E(\prod_{i=1}^{n} Q_i)$ for any n.

2.2 The Moment Generating Function

Alternatively, it is possible to write down the joint moment generating function of Q_i, $i = 1, \cdots, n$,

$$M_Q(t) = E[\exp(\sum_{i=1}^{n} t_i Q_i)] = E[\exp(y'By)], \tag{2}$$

where $t = (t_1, \cdots, t_n)'$ and $B = \sum_{i=1}^{n} t_i A_i$. Since y is multivariate normal, it can be easily seen

$$M_Q(t) = \exp\left\{ \frac{1}{2}\mu' \left[(I - 2B)^{-1} - I \right] \mu - \frac{1}{2} \ln |I - 2B| \right\}.$$

For t sufficiently close to zero, by expanding $(I - 2B)^{-1}$ and $\ln|I - 2B| = \text{tr}[\ln(I - 2B)]$, one can write

$$M_Q(t) = \exp\left\{ \frac{1}{2} \sum_{i=1}^{\infty} \left[2^i \mu' B^i \mu + \frac{2^i}{i} \text{tr}(B^i) \right] \right\}. \tag{3}$$

Then using Wilf's (1994) notation, $E(\prod_{i=1}^n Q_i) = [t_1 t_2 \cdots t_n] M_Q(t)$.[2] Equivalently

$$E(\prod_{i=1}^{n} Q_i) = \frac{\partial^n}{\partial t_1 \partial t_2 \cdots \partial t_n} \exp\left\{ \frac{1}{2} \sum_{i=1}^{\infty} \left[2^i \mu' B^i \mu + \frac{2^i}{i} \mathrm{tr}(B^i) \right] \right\} \Bigg|_{t=0}, \tag{4}$$

which is accessible via Faà di Bruno's formula (see, e.g., Barndoff-Nielsen and Cox, 1989).

Modern symbolic algebra packages such as Maple and Mathematica can readily produce coefficients of this type. Here we do not claim that using the nonstochastic operator has any conceptual advantage over the moment generating function. However, we should point out that the computerized output may not necessarily reveal the recursive nature of the algorithm in calculating $E(\prod_{i=1}^n Q_i)$. In terms of numerical programming and applications, this has strong implications. Based on $[t_1 t_2 \cdots t_n] M_Q(t)$, the computer may quickly produce algebraic expressions in (4), whose number of terms can on the other hand explode also very quickly as n goes up. Then reevaluating these algebraic terms numerically given μ and A_i may be far time consuming, especially when m and n are both large. But if one programs the recursive formula for $E(\prod_{i=1}^n Q_i)$ (see (7) in the next subsection) derived using the nonstochastic operator, then computational time can be substantially saved since the in-between results $E(\prod_{i=1}^p Q_i)$, $p < n$ can be saved numerically. This will be demonstrated more clearly in the next subsection.

Ghazal (1996) found a recursive relationship for the case of zero mean normal random vector by following the expansion (4). In this next subsection, we derive results for the more general case of $y \sim N(\mu, I)$ by using Ullah's (2004) nonstochastic operator.

2.3 A Recursive Procedure

We first introduce two lemmas regarding the gradient vector and the Hessian matrix of $E(\prod_{i=1}^n Q_i)$ with respect to the mean vector μ. They are needed when we later try to expand $d' A_1 d \cdot E(\prod_{i=2}^n Q_i)$. For the starting case when $n = 1$, $E(Q_1) = \mu' A_1 \mu + \mathrm{tr}(A_1)$, as shown before. 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 $E(\prod_{i=1}^n Q_i)$ with respect to μ is

$$\frac{\partial E(\prod_{i=1}^n Q_i)}{\partial \mu} = \sum_{i=1}^{n} \sum_{j_1=1}^{n} \cdots \sum_{j_i=1}^{n} 2^i A_{j_1} \cdots A_{j_i} \mu E\left(\frac{Q_1 \cdots Q_n}{Q_{j_1} \cdots Q_{j_i}} \right) \tag{5}$$

Proof: Change the order of derivative and expectation signs, and use successively the operator (1), then the result follows. ■

It should be noted that in (5), 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, $\partial E(Q_1)/\partial \mu = 2A_1 \mu$, $\partial E(Q_1 Q_2)/\partial \mu = 2[A_1 \mu E(Q_2) + A_2 \mu E(Q_1)] + 4(A_1 \Omega A_2 \mu + A_2 \Omega A_1 \mu)$.

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

$$\frac{\partial^2 E(\prod_{i=1}^n Q_i)}{\partial\mu'\partial\mu} = 2\sum_{j=1}^n A_j E\left(\frac{Q_1\cdots Q_n}{Q_j}\right) + 4\sum_{j=1}^n\sum_{k=1}^n E\left(A_j yy' A_k \frac{Q_1\cdots Q_n}{Q_j Q_k}\right). \tag{6}$$

Proof: Using Lemma 1, direct derivative rules give the result ■

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

$$E(\prod_{i=1}^n Q_i) = \sum_{i=0}^{n-1} 2^i \sum_{j_1=2}^n \cdots \sum_{j_i=2}^n \left[g_{j_1\cdots j_i} E\left(\frac{Q_2\cdots Q_n}{Q_{j_1}\cdots Q_{j_i}}\right)\right], \tag{7}$$

where $g_{j_1\cdots j_i} = \mu'(A_1A_{j_1}\cdots A_{j_i} + A_{j_1}A_1A_{j_2}\cdots A_{j_i} + \cdots + A_{j_1}\cdots A_{j_i}A_1)\mu + \mathrm{tr}(A_1A_{j_1}\cdots A_{j_i})$.

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

$$\begin{aligned}
E(\prod_{i=1}^n Q_i) &= d'A_1d\cdot E(\prod_{i=2}^n Q_i)\\
&= \mathrm{tr}\left[A_1dd'E(\prod_{i=2}^n Q_i)\right]\\
&= \mathrm{tr}\left[A_1d\left(\mu'E(\prod_{i=2}^n Q_i)\right) + A_1d\left(\frac{\partial E(\prod_{i=2}^n Q_i)}{\partial\mu'}\right)\right]\\
&= \mu'A_1\mu\cdot E(\prod_{i=2}^n Q_i) + \mathrm{tr}\left[A_1\frac{\partial\mu' E(\prod_{i=2}^n Q_i)}{\partial\mu}\right] + \mathrm{tr}\left[A_1\mu\frac{\partial E(\prod_{i=2}^n Q_i)}{\partial\mu'}\right] + \mathrm{tr}\left[A_1\frac{\partial^2 E(\prod_{i=2}^n Q_i)}{\partial\mu'\partial\mu}\right]\\
&= \mu'A_1\mu\cdot E(\prod_{i=2}^n Q_i) + \mathrm{tr}\left[A_1\left(E(\prod_{i=2}^n Q_i) + \frac{\partial E(\prod_{i=2}^n Q_i)}{\partial\mu}\mu'\right)\right]\\
&\quad + \mathrm{tr}\left[A_1\mu\frac{\partial E(\prod_{i=2}^n Q_i)}{\partial\mu'}\right] + \mathrm{tr}\left[A_1\frac{\partial^2 E(\prod_{i=2}^n Q_i)}{\partial\mu'\partial\mu}\right]\\
&= \mu'A_1\mu\cdot E(\prod_{i=2}^n Q_i) + \mathrm{tr}(A_1)E(\prod_{i=2}^n Q_i) + 2\mu'A_1\frac{\partial E(\prod_{i=2}^n Q_i)}{\partial\mu} + \mathrm{tr}(A_1)\frac{\partial^2 E(\prod_{i=2}^n Q_i)}{\partial\mu'\partial\mu}\\
&= E(Q_1)E(\prod_{i=2}^n Q_i) + \sum_{i=1}^{n-1} 2^{i+1}\sum_{j_1=2}^n\cdots\sum_{j_i=2}^n \mu'A_1A_{j_1}\cdots A_{j_i}\mu E\left(\frac{Q_2\cdots Q_n}{Q_{j_1}\cdots Q_{j_i}}\right)\\
&\quad + 2\sum_{j=2}^n \mathrm{tr}(A_1A_j)E\left(\frac{Q_2\cdots Q_n}{Q_j}\right) + 4\sum_{j=2}^n\sum_{k=2}^n E\left(y'A_jA_1A_k y\frac{Q_2\cdots Q_n}{Q_jQ_k}\right)\\
&= E(Q_1)E(\prod_{i=2}^n Q_i) + \sum_{i=1}^{n-1} 2^{i+1}\sum_{j_1=2}^n\cdots\sum_{j_i=2}^n \mu'A_1A_{j_1}\cdots A_{j_i}\mu E\left(\frac{Q_2\cdots Q_n}{Q_{j_1}\cdots Q_{j_i}}\right)\\
&\quad + 2\sum_{j=2}^n \mathrm{tr}(A_1\Omega A_j)E\left(\frac{Q_2\cdots Q_n}{Q_j}\right) + 4\sum_{\substack{j=2\\ j<k}}^n\sum_{k=2}^n E\left(y'\left(A_jA_1A_k + A_kA_1A_j\right)y\frac{Q_2\cdots Q_n}{Q_jQ_k}\right),
\end{aligned}$$

where the second last equality follows by substituting the results from Lemmas 1 and 2 and noting that $E(Q_1) = \mu' A_1 \mu + \mathrm{tr}(A_1)$, and the last equality follows by noting that $A_j A_1 A_k$ is not symmetric, and we replace it with $(A_j A_1 A_k + A_k A_1 A_j)/2$, and that the two indices j and k have symmetric roles. Note that in the last equality there is a term $E[y'(A_j A_1 A_k + A_k A_1 A_j)\, y \frac{Q_2 \cdots Q_n}{Q_j Q_k}]$, which is a quadratic form of order $n-2$. Upon successive substitution, result (7) follows.■

Note that in (7), as before, 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. In particular, when $i = 0$, $g_{j_1 \cdots j_i} = \mu' A_1 \mu + \mathrm{tr}(A_1) = E(Q_1)$, $\frac{Q_2 \cdots Q_n}{Q_{j_1} \cdots Q_{j_i}} = Q_2 \cdots Q_n$. Now applying the theorem, we have the following results for n up to 4,

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

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

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

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

Upon substitution, we have the results for expectations of quadratic forms up to order 4 as given in Ullah (2004). Note that the above results can alternatively follow from (4). For example, consider $n = 3$ so that $B = t_1 A_1 + t_2 A_2 + t_3 A_3$. Using Faà di Bruno's formula,

$$\begin{aligned}
E(\prod_{i=1}^{3} Q_i) &= \left.\frac{\partial^3}{\partial t_1 \partial t_2 \partial t_3} \exp(u)\right|_{t=0} \\
&= \underbrace{\left.\exp(u)\frac{\partial^3 u}{\partial t_1 \partial t_2 \partial t_3}\right|_{t=0}}_{\text{Part I}} + \underbrace{\left.\exp(u)\left(\frac{\partial u}{\partial t_1}\frac{\partial^2 u}{\partial t_2 \partial t_3} + \frac{\partial u}{\partial t_2}\frac{\partial^2 u}{\partial t_1 \partial t_3} + \frac{\partial u}{\partial t_3}\frac{\partial^2 u}{\partial t_1 \partial t_2}\right)\right|_{t=0}}_{\text{Part II}} \\
&\quad \underbrace{+\left.\exp(u)\frac{\partial u}{\partial t_1}\frac{\partial u}{\partial t_2}\frac{\partial u}{\partial t_3}\right|_{t=0}}_{\text{Part III}},
\end{aligned}$$

where $u = \frac{1}{2}\sum_{i=1}^{\infty}\left[2^i \mu' B^i \mu + \frac{2^i}{i}\mathrm{tr}(B^i)\right]$. Obviously,

$$\begin{aligned}
\text{Part I} &= 8\mu' A_1 A_2 A_3 \mu + 8\mu' A_1 A_3 A_2 \mu + 8\mu' A_2 A_1 A_3 \mu + 8\mathrm{tr}(A_1 A_2 A_3), \\
\text{Part II} &= [4\mu' A_2 A_3 \mu + 2\mathrm{tr}(A_2 A_3)]E(Q_1) + [4\mu' A_1 A_3 \mu + 2\mathrm{tr}(A_1 A_3)]E(Q_2) \\
&\quad + [4\mu' A_1 A_2 \mu + 2\mathrm{tr}(A_1 A_2)]E(Q_3), \\
\text{Part III} &= E(Q_1)E(Q_2)E(Q_3),
\end{aligned}$$

so Part I + Part II + Part III gives the same expression for $E(\prod_{i=1}^{3} Q_i)$, as derived previously using the recursive formula (7). In terms of numerical calculation, we notice that the main difference is that Faà di

Bruno's formula breaks $E(\prod_{i=1}^3 Q_i)$ into functions of single quadratic forms $E(Q_1)$, $E(Q_2)$, and $E(Q_3)$, but the recursive formula (7) breaks $E(\prod_{i=1}^3 Q_i)$ into functions of quadratic forms of lower orders (2 and 1). (Interested readers may try with any n and can verify that Faà di Bruno's formula always breaks the result into single quadratic forms, whereas the recursive formula (7) breaks the result into quadratic forms of lower orders.) In terms of computer time, evaluating $E(Q_1)E(Q_2)E(Q_3) + [\mu' A_2 A_3 \mu + 2\mathrm{tr}(A_2 A_3)]E(Q_1)$ (needed using Faà di Bruno's formula) may be more time consuming than evaluating $E(Q_1)E(Q_2 Q_3)$ (needed using (7)). The reason is as follows. To evaluate $E(Q_1)E(Q_2Q_3)$, $E(Q_2Q_3)$ can be calculated from a subroutine and once it is calculated, only a scalar return value is needed and all matrices (A_2 and A_3) involved can be cleared out from computer memory, but in evaluating $E(Q_1)E(Q_2)E(Q_3) + [\mu' A_2 A_3 \mu + 2\mathrm{tr}(A_2 A_3)]E(Q_1)$, all the matrices are not cleared out of computer memory until the calculations are finished. This effect will be most apparent when both m and n are large.

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 (7).

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

$$E[(y'Ay)^n] = \sum_{i=0}^{n-1} g_i E\left(y'Ay\right)^{n-i-1}, \tag{8}$$

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

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

$$E(\prod_{i=1}^n Q_i) = E(Q_1) \cdot E(\prod_{i=2}^n Q_i) + 2\sum_{j=2}^n E(y'A_jA_1y \cdot y'A_2y \cdots y'A_{j-1}y \cdot \prod_{k=j+1}^n Q_k). \tag{9}$$

Proof: Comparing the proof of (7) and (9), 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

$$U_n \equiv \sum_{j=2}^n \mathrm{tr}(A_1A_j) E\left(\frac{Q_2 \cdots Q_n}{Q_j}\right) + 2\sum_{j=2}^n \sum_{\substack{i=2 \\ i \neq j}}^n E\left[y'A_iA_1A_jy \cdot \left(\frac{Q_2 \cdots Q_n}{Q_iQ_j}\right)\right].$$

It is enough to show that the jth summands of W_n and U_n are equal. This can be proved straightforwardly by induction. ■

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

$$\begin{aligned} E(\prod_{i=1}^{n} Q_i) &= E(\otimes_{i=1}^{n} Q_i) \\ &= \mathrm{tr}\{E[\otimes_{i=1}^{n}(yy'A_i)]\} \\ &= \mathrm{tr}\{E[\otimes_{i=1}^{n}(yy')](\otimes_{i=1}^{n}A_i)\} \\ &= \mathrm{tr}\{E[(y^{\otimes})(y^{\otimes})']A^{\otimes}\}, \end{aligned} \tag{10}$$

where $\otimes$ denotes and 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_is 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, an example of which is provided in the next section.

To facilitate our derivation, suppose now that the mean vector $\mu = 0$ and y_i is IID. As it turns out, as n goes up, the work is more demanding. In the literature, results are only available for n up to three, 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:

$$\begin{aligned} m_1 &= 0, \quad m_2 = 1, \quad m_3 = \gamma_1, \quad m_4 = \gamma_2 + 3, \\ m_5 &= \gamma_3 + 10\gamma_1, \quad m_6 = \gamma_4 + 15\gamma_2 + 10\gamma_1^2 + 15, \\ m_7 &= \gamma_5 + 21\gamma_3 + 35\gamma_2\gamma_1 + 105\gamma_1, \\ m_8 &= \gamma_6 + 28\gamma_4 + 56\gamma_3\gamma_1 + 35\gamma_2^2 + 210\gamma_2 + 280\gamma_1^2 + 105, \end{aligned} \tag{11}$$

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.

For $\prod_{i=1}^{m} y_i^{\alpha(i)} = y_1^{\alpha(1)} \cdots y_m^{\alpha(m)}$ 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 follows, where $\odot$ denotes the Hadamard product symbol:

$$
\begin{aligned}
&E(\textstyle\prod_{i=1}^4 Q_i) = \gamma_2\{\sum_{i<j;k<l}[\mathrm{tr}(A_i)\mathrm{tr}(A_j)\mathrm{tr}(A_k \odot A_l) + 2\iota'(A_i \odot A_j)\iota\mathrm{tr}(A_k \odot A_l)] \\
&\quad +4\textstyle\sum_{k<l}\mathrm{tr}(A_i)\mathrm{tr}(A_j \odot (A_kA_l)) + 8\sum_{j<l}\mathrm{tr}((I \odot A_i)A_jA_kA_l) + 16\sum_{k<l}\iota'(I \odot (A_1A_j))(I \odot (A_kA_l))\iota\} \\
&\quad +\gamma_4\{\textstyle\sum_{j<k<l}\mathrm{tr}(A_i)\mathrm{tr}(A_j \odot A_k \odot A_l) + 4\sum_{i<j;k<l}\mathrm{tr}[A_i \odot A_j \odot (A_kA_l)]\} + \gamma_6\mathrm{tr}(A_1 \odot A_2 \odot A_3 \odot A_4) \\
&\quad +2\gamma_1^2\{\textstyle\sum_{j<l}\mathrm{tr}(A_i)\iota'(I \odot A_j)A_k(I \odot A_l)\iota + 2\sum_{j<k}\iota'(I \odot A_i)A_jA_k(I \odot A_l)\iota + 2\sum_{j<k<l}\mathrm{tr}(A_i)\iota'(A_j \odot A_k \odot A_l)\iota \\
&\quad +4\textstyle\sum_{i<j}\iota'(A_i \odot A_j)A_k(I \odot A_l)\iota + 8\sum_{i<l;j<k}\mathrm{tr}[A_i(A_j \odot A_k)A_l]\} + \gamma_2^2[\sum_{k<l}\mathrm{tr}(A_1 \odot A_j)\mathrm{tr}(A_k \odot A_l) \\
&\quad +4\textstyle\sum_{i<l;j<k}\iota'(I \odot A_i)(A_j \odot A_k)(I \odot A_l)\iota + 8\iota'(A_1 \odot A_2 \odot A_3 \odot A_4)\iota] + 2\gamma_1\gamma_3[\sum_{k<l}\iota'(I \odot A_i)A_j(I \odot A_k \odot A_l)\iota \\
&\quad +4\textstyle\sum_{j<k<l}\iota'(I \odot A_i)(A_j \odot A_k \odot A_l)\iota] + \prod_{i=1}^4\mathrm{tr}(A_i) + 2\sum_{i<j;k<l}\mathrm{tr}(A_i)\mathrm{tr}(A_j)\mathrm{tr}(A_kA_l) \\
&\quad +4\textstyle\sum_{k<l}\mathrm{tr}(A_1A_j)\mathrm{tr}(A_kA_l) + 8\sum_{j<k<l}\mathrm{tr}(A_i)\mathrm{tr}(A_jA_kA_l) + 16\sum_{j<l}\mathrm{tr}(A_1A_jA_kA_l)\,,
\end{aligned}
$$

where all the summations are over the indices under the summation signs running from 1 to 4, unequal to each other, and sometimes with inequality restrictions. For example, $\sum_{i<j;k<l}$ indicates summing over the four indices i, j, k, l from 1 to 4, $i \neq j \neq k \neq l$, $i < j$, and $k < l$.

Note that sometimes we may also be interested in the expectations of a "half" quadratic form $y\prod_{i=1}^n y'A_iy$, which is the product of linear function and quadratic forms. Let $E(y\prod_{i=1}^n y'A_iy)$ have a representative element $E(y_j\prod_{i=1}^n y'A_iy)$, $j = 1, \cdots, m$. Then following (10),

$$E(y_j \prod_{i=1}^n y'A_iy) = \mathrm{tr}\{E[y_j(y^{\otimes})(y^{\otimes})']A^{\otimes}\}.$$

Analytical results are available for n up to 2 in the literature. Now we focus on the case when $n = 3$ for the half quadratic form. For $y_j\prod_{i=1}^m y_i^{\alpha(i)} = y_jy_1^{\alpha(1)}\cdots y_m^{\alpha(m)}$ with $\alpha(1) + \cdots + \alpha(m) = 2n = 6$, 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 for the half quadratic form when $n = 3$:

$$E(y \prod_{i=1}^{3} y'A_i y) = \gamma_5 \left(I \odot A_1 \odot A_2 \odot A_3\right) \iota + \gamma_3 \sum_{j<k} \{4[I \odot A_i \odot (A_j A_k)]\iota + 2A_i(I \odot A_j \odot A_k)\iota$$
$$+\mathrm{tr}(A_i)(I \odot A_j \odot A_k)\iota\} + \gamma_1 \{4 \sum_{j<k} [\mathrm{tr}(A_i)(I \odot (A_j A_k))\iota + 2A_i(A_j \odot A_k)\iota] + 8 \sum_{i<k} [I \odot (A_i A_j A_k)]\iota$$
$$+ \sum_{i<j} [\mathrm{tr}(A_i)\mathrm{tr}(A_j)\left(I \odot A_k\right)\iota + 2\iota'(A_i \odot A_j)\iota(I \odot A_k)\iota] + 2 \sum [2A_i A_j (I \odot A_k)\iota + \mathrm{tr}(A_i) A_j (I \odot A_k)\iota]\}$$
$$+\gamma_1\gamma_2 \{8(A_1 \odot A_2 \odot A_3)\iota + \sum_{i<j} [4(A_i \odot A_j)(I \odot A_k)\iota + \mathrm{tr}(A_i \odot A_j)(I \odot A_k)\iota] + 2 \sum (I \odot A_i) A_j (I \odot A_k)\iota\},$$

where all the summations are over the indices under the summation signs running from 1 to 3, unequal to each other, and sometimes with inequality restrictions. For example, $\sum_{i<j}$ stands for summation over $i, j, k = 1, 2, 3$, $i \neq j \neq k$, and $i < j$.

Note that setting $\gamma_1, \cdots, \gamma_6$ all equal to zero, then the result for quadratic form of order 4 derived in this section degenerates into the result under normality as given in Ullah (2004), and trivially, the result for half quadratic form of order 3 degenerates to zero.

4 Econometric Applications

Now we give some applications of the results for normal and nonnormal variables from the previous sections. We consider the ratio of normal quadratic forms whose expectation is equal to the ratio of expectations. We also study the effects of nonnormality on the approximate mean squared error (MSE) of the estimator in a simple time series model.

4.1 Moments of Ratios of Quadratic Forms

For an m-variate normal vector $y \sim N(0, I)$, consider the ratio of quadratic form $R = y'By/y'Ay$.[3] In general, the moments of this ratio is not equal to the ratio of moments. However, with some restrictions put on the symmetric matrices A and B, we can have a separation result

$$E\left[\left(\frac{y'By}{y'Ay}\right)^k\right] = \frac{E[(y'By)^k]}{E[(y'Ay)^k]}. \tag{12}$$

The conditions are that A and B are of rank $r \leq m$, commutative, $AB \neq 0$, and A is idempotent (see Ghazal, 1994). The numerator and denominator in (12) can be easily evaluated using the normal result (9). For A

being idempotent, we can verify $E[(y'Ay)^k] = 2^k\Gamma\left(r/2+k\right)/\Gamma\left(r/2\right) = r(r+2)\cdots(r+2(k-1))$. In fact, we can generalize the separation result (12) to the case

$$E\left[\frac{\prod_{i=1}^{q}\left(y'B_iy\right)^{k_i}}{\left(y'Ay\right)^k}\right] = \frac{E\left[\prod_{i=1}^{q}\left(y'B_iy\right)^{k_i}\right]}{E\left(y'Ay\right)^k}, \tag{13}$$

where $k = \sum_{i=1}^{q} k_i$, and each pair $(A,\, B_i)$ satisfies the same conditions as $(A,\, B)$ in (12). This can be proved by following Sawa (1972)'s result,

$$E\left[\frac{\prod_{i=1}^{q}\left(y'B_iy\right)^{k_i}}{\left(y'Ay\right)^k}\right] = \frac{1}{\Gamma\left(k\right)}\int_{-\infty}^{0}\left(-t\right)^{k-1}\left[\frac{\partial^{k_1+\cdots k_q}|S|^{-1/2}}{\partial t_1^{k_1}\cdots\partial t_q^{k_q}}\right]_{t_1=\cdots=t_q=0} dt,$$

where $S = I-2tA-2t_1B_1-\cdots-2t_qB_q$. Since $E[(y'Ay)^k] = 2^k\Gamma\left(r/2+k\right)/\Gamma\left(r/2\right)$ and $E\left[\prod_{i=1}^{q}\left(y'B_iy\right)^{k_i}\right] = \left[\partial^{k_1+\cdots k_q}|\Delta|^{-1/2}/\partial t_1^{k_1}\cdots\partial t_q^{k_q}\right]_{t_1=\cdots=t_q=0}$ for $\Delta = I-2t_1B_1-\cdots-2t_qB_q$, and also noting that

$$\int_{-\infty}^{0}\left(-t_1\right)^{k-1}\left(1-2t_1\right)^{-(r/2+k)} = \frac{\Gamma\left(k\right)\Gamma\left(r/2\right)}{2^k\Gamma\left(r/2+k\right)},$$

to show (13), it is sufficient to prove

$$\left.\frac{\partial^{k_1+\cdots k_q}|S|^{-1/2}}{\partial t_1^{k_1}\cdots\partial t_q^{k_q}}\right|_{t_1=\cdots=t_q=0} = (1-2t)^{-(r/2+k)}\left.\frac{\partial^{k_1+\cdots k_q}|\Delta|^{-1/2}}{\partial t_1^{k_1}\cdots\partial t_q^{k_q}}\right|_{t_1=\cdots=t_q=0},$$

which can be shown to be true by the method of induction.

The separation result (13) can be used to study the finite sample properties of the maximum likelihood estimator in the spatial autoregressive model

$$y = \rho Wy + \varepsilon, \tag{14}$$

where y is an $T\times 1$ vector of observations on the dependent spatial variable, Wy is the corresponding spatially lagged dependent variable for weights matrix W, which is assumed to be known *a priori*, $\rho \in (-1,1)$ is the spatial autoregressive parameter, and ε is the vector of IID $N(0,\sigma^2)$ error terms. To estimate ρ, the concentrated likelihood function (by concentrating out σ^2) is $L(\rho) = T^{-1}\sum_{i=1}^{T}[\ln|A| - \ln(2\pi y'Dy/n)/2 - n/2]$, where $A = I-\rho W$, $D = I-\rho(W+W')+\rho^2W'W$. Since $y = A^{-1}\varepsilon$, then immediately we can write down the score function as well as its higher order derivatives in terms of $B_i = \partial^i \ln|A|/\partial\rho^i$ and $(\varepsilon'M_1\varepsilon)^i(\varepsilon'M_2\varepsilon)^j/(\varepsilon'\varepsilon)^{i+j}$, where $M_i = A^{-1\prime}(\partial^iD/\partial\rho^i)A^{-1}$. For examples, $\partial L(\rho)/\partial\rho = n^{-1}B_1 - 2^{-1}(\varepsilon'M_1\varepsilon)/(\varepsilon'\varepsilon)$, $\partial^2L(\rho)/\partial\rho^2 = n^{-1}B_2 - 2^{-1}(\varepsilon'M_2\varepsilon)/(\varepsilon'\varepsilon) + 2^{-1}(\varepsilon'M_1\varepsilon)^2/(\varepsilon'\varepsilon)^2$, $\partial^3L(\rho)/\partial\rho^3 = n^{-1}B_3 + 1.5(\varepsilon'M_1\varepsilon)(\varepsilon'M_2\varepsilon)/(\varepsilon'\varepsilon)^2 - (\varepsilon'M_1\varepsilon)^3/(\varepsilon'\varepsilon)^3$, where $B_1 = -\mathrm{tr}(A^{-1}W)$, $B_2 = -\mathrm{tr}[(A^{-1}W)^2]$, $B_3 = -2\mathrm{tr}[(A^{-1}W)^3]$. As such, Bao and Ullah (2007a) showed that the approximate bias and MSE of the maximum likelihood estimator $\hat{\rho}$ can be expressed in terms of $E[(\varepsilon'M_1\varepsilon)^i(\varepsilon'M_2\varepsilon)^j/(\varepsilon'\varepsilon)^{i+j}]$. They followed the top-order invariant polynomial approach of Smith (1989) to work out the expectations. It is obvious that the separation result (13) could have been used by setting $A = I$ and $B = M_1, M_2$. By using (7), we can easily check that the two methods yield the same results as given in Bao and Ullah (2007a).

4.2 Effects of Nonnormality on the MSE in AR Model

Given the results under nonnormality from Section 3, we can study the effects of nonnormality on the approximate bias and MSE of the OLS estimator of the autoregressive coefficient in the following AR(1) model with exogenous regressors

$$y_t = \rho y_{t-1} + x_t'\beta + \sigma\varepsilon_t,\ t = 1, \cdots, T \tag{15}$$

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 IID and follows some nonnormal distribution with moments (11). 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 $\rho^{t-t'-1}$, $A = MC$, $r = M(y_0 F + CX\beta)$, where we assume that the first observation y_0 has been observed.

Bao and Ullah (2007b) found that nonnormality affects the 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 given in Section 3. Now we are ready to study the effects of nonnormality on the MSE. From Bao and Ullah (2007b), we can express the approximate MSE of the OLS estimator $\hat{\rho}$ of the autoregressive coefficient as

$$\begin{aligned} M(\hat{\rho}) \quad = \quad & \frac{6(\lambda_{10000} + \lambda_{02000} + 2r'\omega_{01000})}{(r'r + \lambda_{00100})^2} \\ & - \frac{8[r'r(\lambda_{10000}+\lambda_{02000}+4\lambda_{01010})+\lambda_{10100}+\lambda_{02100}+2r'(rr'\omega_{01000}+A\omega_{10000}+A\omega_{02000}+\omega_{01100})]}{(r'r + \lambda_{00100})^3} \\ & + \frac{3[(r'r)^2(\lambda_{10000} + \lambda_{02000}) + 2r'r(\lambda_{10100} + 4\lambda_{01010} + \lambda_{02100}) + 4\lambda_{10001} + \lambda_{10200} + 4\lambda_{02001}]}{(r'r + \lambda_{00100})^4} \\ & + \frac{3[8\lambda_{01110} + \lambda_{02200} + 4r'rr'(A\omega_{10000} + A\omega_{02000} + \omega_{01100}) + 2(r'r)^2 r'\omega_{01000}]}{(r'r + \lambda_{00100})^4} \\ & + \frac{6r'(\omega_{01200} + 4\omega_{01001} + 2A\omega_{10100} + 2A\omega_{02100})}{(r'r + \lambda_{00100})^4} + o(T^{-2}), \end{aligned} \tag{16}$$

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 $\omega_{ijklm} = \sigma^{2(i+j+k+l+m)+1} E[\varepsilon \cdot (\varepsilon' rr'\varepsilon)^i \cdot (\varepsilon' A\varepsilon)^j \cdot (\varepsilon' A'A\varepsilon)^k \cdot (\varepsilon' A'rr'\varepsilon)^l \cdot (\varepsilon' A'rr'A\varepsilon)^m]$. Obviously, expectations of quadratic forms up to order 4 (λ_{ijklm}) and half quadratic form up to 3 (ω_{ijklm}) are needed to evaluate (16). So numerically, it is possible for us to investigate explicitly the effects of nonnormality on $M(\hat{\rho})$ by using Section 3 (with A and $A'rr'$ in the quadratic forms symmetrized) and the results in Ullah (2004).

For the special case when x_t is a vector of ones, i.e., when we have the so-called intercept model, Bao (2007) showed that one can simplify the MSE result (16) analytically to express the result explicitly in terms of model parameters and the nonnormality parameters:

$$M(\hat{\rho}) = \frac{1-\rho^2}{T} + \frac{1}{T^2}\left[23\rho^2 + 10\rho - \frac{1+\rho}{1-\rho}\left(\frac{\alpha-(1-\rho)y_0}{\sigma}\right)^2 \right.$$
$$\left. -\frac{4\gamma_1^2\rho\left(1-\rho^2\right)^2}{1-\rho^3} - \gamma_2(1-\rho^2)\right] + o\left(T^{-2}\right). \quad (17)$$

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).[4] 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 $\sigma^2 = 0.5, 1$, $\alpha = 0.01, 0.05$, $\lambda = 0.5, 1$, $T = 50$.[5] To calculate the feasible approximate MSE under nonnormality, we put $\hat{\rho}, \hat{\sigma}^2, \hat{\gamma}_1$, and $\hat{\gamma}_2$ into (16). When ignoring nonnormality, we put $\hat{\rho}, \hat{\sigma}^2$, and $\gamma_1 = \gamma_2 = 0$ into (17) 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, (17) 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, (17) 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.

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 and half quadratic form of order 3 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. The nonnormal results involve the cumulants of the nonnormal distribution up to the eighth order for order 4 quadratic from, and up to the seventh order for order 3 half quadratic from. Setting all the nonnormality parameters equal to zero gives the results under normality as a special case. We apply our normal and nonnormal results to two econometric problems. For the spatial autoregressive model, since the score function and its higher order derivatives can be rewritten as ratio of quadratic forms, whose expectation is equal to the ratio of expectations, the normal recursive formula derived in this paper can be applied directly to obtain the finite sample bias and MSE of the maximum likelihood estimator.

For the time series AR(1) model with an intercept, our numerical simulations demonstrate that in the presence of nonnormality, the analytical MSE formula approximates the true MSE of the OLS estimator of the autoregressive coefficient remarkably well.

Notes

[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 (11) 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 IID, as is usually the case considered for the error vector in a standard regression model

[2]We thank a referee for bringing the notation of Wilf (1994) to our attention. Here if $\nu = (\nu_1, \nu_2, \cdots, \nu_n)$ is a vector of nonnegative integers with $|\nu| = \sum_{i=1}^n \nu_i$, the symbol $[t_1^{\nu_1} t_2^{\nu_2} \cdots t_n^{\nu_n}]$ is interpreted as "the coefficient of $t_1^{\nu_1} t_2^{\nu_2} \cdots t_n^{\nu_n}$ in the expansion of the following function."

[3]We note that when y is nonnormally distributed, various attempts have been made to study the *approximate* moments and distribution of the ratio, see Ali and Sharma (1993), Lieberman (1994b, 1997), and Ullah and Srivastava (1994), although the exact properties of the ratio under a general nonnormal density are quite difficult to derive.

[4]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.

[5]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.

References

Ali, M.M. & S.C. Sharma (1993) Robustness of nonnormality of the Durbin-Watson test for autocorrelation. *Journal of Econometrics* 57, 117-136.

Barndoff-Nielsen, O.E. & D.R. Cox (1989). *Asymptotic Techniques for Use in Statistics.* London: Chapman & Hall.

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. (1994) Moments of the ratio of two dependent quadratic forms. *Statistics and Probability Letters* 20, 313-319.

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*, forthcoming.

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

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

Lieberman, O. (1994b) A Laplace approximation for the moments of a ratio of quadratic forms. *Biometrika* 81, 681-690.

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.

Sawa, T. (1972) Finite sample properties of the k-class estimators. *Econometrica* 40, 653-680.

Smith, M.D. (1989) On the expectation of a ratio of quadratic forms in normal variables. *Journal of Multivariate Analysis* 31, 244-257.

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. (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

Wilf, H.S. (1994) *Generatingfunctionology.* San Diego: Academic Press.

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.