# The Distribution of the Least-Squares Variance Estimator under Serial Correlation

*By* Ki-Jun Jeong*

Contents



## Introduction

The purpose of this paper is to derive the distribution of the sum of squares of least-squares (LS) residuals, $e'e$ in a linear model, and the distribution

* Assistant professor of economics, Seoul National University. This paper is based on a part of his Ph.D. dissertation submitted to Claremont Graduate School. He is indebted to Professor Donald H. Ebbeler for his comments and suggestions.

of the LS variance estimator, $s^2$, under serial correlation. Goldberger [1964, p. 242], Johnston [1972, pp. 248-49], Malinvaud [1970, Sec. 13.4], and Theil [1971, pp. 156-57] have discussed the bias of the variance estimator, $s^2$, under serial correlation. But no one seems to have discussed the higher moments of $s^2$ under serial correlation. In this paper we shall show that the increase in the variance of the variance estimator is very important under serial correlation. Based on the new findings we shall show that the chi-square test for the variance of disturbances may not be robust under serial correlation.

For this purpose we need some new tools which will be devised in Section I. The tools include the minor diagonal operators, the unit minor diagonal matrices and the minor traces. For this purpose we also need some new properties of the matrix $M=I-X(X'X)^{-1}X'$, where $X$ is the observation matrix of explanatory variables. The central role of the matrix $M$ in regression analysis is well known. But it seems that the properties of $M$ have not yet thoroughly been analysed. In Section II, we shall explain some new properties of matrix $M$ in terms of its minor traces.

Alternative models of serial correlation and properties of a quadratic form in normal variables will be discussed in Appendices.

## I. Useful Analytical Tools

In this section we shall derive some tools to be used in subsequent sections.

### 1.1. Minor Diagonal Operators

The $s$th minor diagonal operator $\Delta_s$ is defined by

(1.1) $\Delta_s A=B, \quad s=1,2,\cdots,n-1,$

where $A$ and $B$ are $n\times n$ matrices and $B$ is a matrix whose diagonals, $s$ elements above and $s$ elements below the main diagonal, are the same as those of $A$, and the other elements are all zero. The zeroth minor diagonal operator is defined as

(1.2) $\Delta_0=\Delta,$

where $\Delta$ is the diagonal operator which transforms an $n\times n$ matrix $A$ into an $n\times n$ diagonal matrix $\boldsymbol{B}$ whose diagonal is the same as that of $\boldsymbol{A}$:

(1.3) $\quad \Delta A=B.$

The complement of $\Delta$,

(1.4) $\quad \bar{\Delta}=1-\Delta$

will be called the off-diagonal operator, Some properties of the minor diagonal operators and the off-diagonal operators can be stated as follows:

(1.5) $\quad \Delta_s(aA+bB)=a\Delta_sA+b\Delta_sB,\ 0\leqq s<n,$

(1.6) $\quad \Delta_s(\Delta_tA)=\delta_{st}\Delta_sA,\ 0\leqq s,\ t<n,$

(1.6)′ $\quad \Delta\bar{\Delta}A=\bar{\Delta}\Delta A=0,$

(1.7) $\quad \Delta A\Delta B=\Delta(\Delta A)B=(\Delta A)(\Delta B),$

where $a$ and $b$ are scalars, and $\delta_{st}$ is the Kronecker delta function. For further properties and applications of the diagonal operator, see Jeong [1975].

**1.2. Unit Minor Diagonal Matricces**

The $s$th unit minor diagonal matrix, $\boldsymbol{D}_s$, is defined by

(1.8) $\quad D_s=\Delta_s(\iota\iota'),\ s=0,1,2,\cdots\cdots\cdots n-1,$

where $\iota$ is an $n\times 1$ sum vector, $[1\,1\,1\cdots\cdots1\,1]'$. The zeroth unit minor diagonal matrix, $\boldsymbol{D}_0$, is $\boldsymbol{I}_n$, the $n$th order identity matrix, by the definition of $\Delta_0$.

The unit minor diagonal matrices have the following properties:

(1.9) $\quad D_s=D_s',\ s=0,1,2,\cdots\cdots,n-1.$

(1.10) $\quad \begin{matrix} D_sD_t=D_{t-s}+D_{t+s}-B_{st}, \\ D_tD_s=D_{t-s}+D_{t+s}-B'_{st} \end{matrix}\ 0<s<t\leqq n/2.$

(1.11) $\quad D_s^2=2I_n+D_{2s}-C_s,\ 0<s<n/2,$

where

(1.12) $\quad B_{st}=\begin{bmatrix} O_{s,(t-s)} & I_s & O & O & O & \\ O & & O & O & O & O \\ O & & O & O & I_s & O_{s,(t-s)} \end{bmatrix},\qquad 0<s<t\leqq n/2,$

(1.13) $\quad C_s=\begin{bmatrix} I_s & O & O \\ O & O & O \\ O & O & I_s \end{bmatrix},\ 0<s<n/2.$

The orders of null matrices in $\boldsymbol{B}_{st}$ and $\boldsymbol{C}_s$ are determined conformably. If $n>>s,\ t$, then $\boldsymbol{B}_{st}$ and $\boldsymbol{C}_s$ are negligible. Therefore we have

(1.14) $\quad D_sD_t\doteqdot D_{t-s}+D_{t+s}\doteqdot D_tD_s,\ 0<s<t\leqslant n/2,$

so that $\boldsymbol{D}_s$ and $\boldsymbol{D}_t$ almost commute, and

(1.15) $D_s^2 \doteq 2I_n + D_{2s}, \quad 0 < s \ll n/2.$

Remark: From the properties above we are tempted to define $D_0 = 2I_n$, since the properties of the unit minor diagonal matrices can then be stated simply. But the present definition is convenient in other context as we shall see later.

Let $E_s$ be a lower triangular matrix of order $n$ defined by

(1.16) $E_s = \begin{bmatrix} O & O \\ I_{n-s} & O \end{bmatrix}, \quad 0 < s < n,$

where the orders of null submatrices are determined conformably. For convenience we shall denote the transpose of $E_s$ as $E_{-s}$. Further we shall define

(1.17) $E_0 = I_n.$

The $E_s$ matrices have the following properties:

(1.18) $E_s E_t = E_{s+t}, \quad -n < s,\ t,\ s+t < n.$

(1.19) $E_s + E_{-s} = D_s, \quad 0 < s < n.$

### 1.3. Minor Traces

The $s$th minor trace of an $n \times n$ square matrix $A$ is defined by

(1.20) $tr_s A = tr D_s A, \quad s = 0,\ 1,\ 2, \cdots\cdots n-1.$

By the definition of $D_0$, the zeroth minor trace of $A$ is $trA$. The above definition of an $s$th minor trace of a matrix amounts to the sum of elements of diagonals $s$ elements above and $s$ elements below the main diagonal of the matrix. In this context it is interesting to recall a property of the trace of a matrix,

(1.21) $trAB = \iota'(A * B')\iota$

given in Rao and Mitra [1971, p. 12], where * denotes the Hadamard product such that

(1.22) $A * B = [a_{ij}] * [b_{ij}] = [a_{ij} b_{ij}].$

Using this property we have

(1.23) $tr_s A = tr D_s A = \iota'(D_s * A')\iota = \iota'(D_s * A)\iota.$

The minor trace operator has the following properties:

(1.24) $tr_s(aA + bB) = a\,tr_s A + b\,tr_s B, \quad 0 \leqq s < n.$

(1.25) $tr_s AB = tr D_s AB = tr BD_s A \neq tr AD_s B = tr D_s BA = tr_s BA, \quad s \neq 0.$

(1.26) $tr_s D_t = 2(n-s)\delta_{st},\ s,t>0.$

(1.27) $tr_s E_t = (n-s)\delta_{st},\ s,t>0.$

where $a$ and $b$ are scalars, and $\delta_{st}$ is the Kronecker delta function. As (1.24) shows the minor trace operator is linear, but as (1.25) shows it has no commutative property in general except for $s=0$.

### 1.4. Autocorrelations and Cross Serial Correlations

Let $\boldsymbol{x}_i$ be an $n\times 1$ vector,

(1.28) $\boldsymbol{x}_i = [x_{1i}\ x_{2i}\cdots\cdots x_{ni}]',\ i=1,2,\cdots\cdots k.$

The normalized vector of $\boldsymbol{x}_i$ is defined by

(1.29) $\boldsymbol{x}_i^* = n^{-1/2} s_i^{-1} \boldsymbol{A}\boldsymbol{x}_i,$

where $\boldsymbol{A} = \boldsymbol{I}_n - n^{-1}\iota\iota'$, and $s_i^2 = n^{-1}\boldsymbol{x}_i\boldsymbol{A}\boldsymbol{x}_i$. As we shall see in Section II, if $\boldsymbol{x}_i$ is an $n\times 1$ sum vector, $\iota$, then it is convenient to define the normalization by

(1.30) $\boldsymbol{x}_i^* = n^{-1/2}\iota.$

The sample analogue of the $s$th autocorrelation coefficient of $\boldsymbol{x}_i$ is defined by

(1.31) $r_{ii,s} = \frac{1}{2}\boldsymbol{x}_i^{*\prime}\boldsymbol{D}_s\boldsymbol{x}_i^*,\ i=1,2,\cdots\cdots k;\ 0<s<n.$

(1.32) $r_{ii,o} = \boldsymbol{x}_i^{*\prime}\boldsymbol{x}_i^* = 1,\ i=1,2,\cdots\cdots k.$

The sampling analogue of the $s$th cross serial correlation coefficient between $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$ is defined by

(1.33) $r_{ij,s} = \frac{1}{2}\boldsymbol{x}^{*}_i{}'\boldsymbol{D}_s\boldsymbol{x}_j^*,\ i,j=1,2,\cdots\cdots k;\ 0<s<n.$

(1.34) $r_{ij,o} = r_{ij} = \boldsymbol{x}_i^{*\prime}\boldsymbol{x}_j^*,\ i,j=1,2,\cdots\cdots k.$

In practice, $s$ should be small compared to $n$, say, less than $\frac{1}{4}n$.

Remark: The usual definition of the sample analogues of autocorrelation and cross serial correlation coefficients is

(1.35) $r_{ij,s}^* = \sum_{t=1}^{n-s} x_{ti}^* x_{t+s,j}^* = \boldsymbol{x}_i^{*\prime}\boldsymbol{E}_{-s}\boldsymbol{x}_j^*,$

and conventionally it is assumed that

(1.36) $r_{ij,s}^* = r_{ji,-s}^* = r_{ji,s}^* = r_{ij,-s}^*,$

for example, see Malinvaud [1970, p.516], But this assumption is not necessarily consistent with the definition (1.35). From (1.35) we have

(1.37) $r_{ij,s}^* = \boldsymbol{x}_i^{*\prime}\boldsymbol{E}_{-s}\boldsymbol{x}_j^* = (\boldsymbol{x}_i^{*\prime}\boldsymbol{E}_{-s}\boldsymbol{x}_j^*)' = \boldsymbol{x}_j^{*\prime}\boldsymbol{E}'_{-s}\boldsymbol{x}_i^* = \boldsymbol{x}_j^{*\prime}\boldsymbol{E}_s\boldsymbol{x}_i^* = r_{ji,-s}^*.$

Therefore the first and the third equalities in (1.36) hold. But the second equality holds only when $i=j$.

From the property (1.19) of Section 1.2, we have

(1.38) $$r_{ij,s}=\frac{1}{2}x_i^{*\prime}D_s x_j^*=\frac{1}{2}(x_j^{*\prime}E_{-s}x_j^*+x_i^{*\prime}E_s x_j^*)=\frac{1}{2}(r_{ij,s}^*+r_{ij,-s}^*).$$

Therefore for $i=j$, $r_{ii,s}=r_{ii,s}^*$, and the statement,

(1.39) $$r_{ij,s}=r_{ji,-s}=r_{ji,s}=r_{ij,-s}$$

always holds.

### 1.5. Minor Diagonal Decomposition of a Correlation Matrix

Let $P$ be an $n\times n$ correlation matrix such that

(1.40) $$p_{ij}=p_{kl}=p_s \ \textit{if}\ |i-j|=|k-l|=s.$$

Then we can decompose it in terms of minor diagonals as follows:

(1.41) $$P=\sum_{s=0}^{n-1}\Delta_s P=\sum_{s=0}^{n-1}p_s D_s,\ p_o=1,$$

i.e., $P$ can be expressed as a linear combination of $D_s$'s. As we shall see later, we sometimes encounter the cases where we should evaluate the trace of a matrix product $PM$, where $M$ is an $n\times n$ matrix. Using the decomposition (1.41), we have

(1.42) $$trPM=tr(\sum_{s=0}^{n-1}p_s D_s)M=\sum_{s=0}^{n-1}p_s tr_s M,$$

i.e., the trace can be evaluated from a linear combination of the minor traces of $M$.

## II. Analysis of the Matrices $X$ and $M$

### 2.1. Introduction

Let $X$ be an $n\times k$ observation matrix of explanatory variables with full column rank such that

(2.1) $$X=[x_1\ x_2\cdots\cdots x_k]=[\iota Z],$$

where $x_i$, $i=1,2,\cdots\cdots k$, are $n\times 1$ vectors and $x_1=\iota$. Let us define the matrices,

(2.2) $$L=X(X'X)^{-1}X',\quad M=I-L.$$

These matrices, especially $M$, are very important in regression analysis. A well-known property of $L$ and $M$ is that each is an idempotent matrix:

$$(2.3)\quad \begin{aligned} &L^2=X(X'X)^{-1}X'X(X'X)^{-1}X'=X(X'X)^{-1}X'=L,\\ &M^2=(I-L)^2=I-2L+L^2=I-L=M. \end{aligned}$$

Since $L$ and $M$ are idempotent, their corresponding ranks are the same as their traces:

$$(2.4)\quad \begin{aligned} &rank(L)=trX(X'X)^{-1}X'=tr(X'X)^{-1}X'X=trI_k=k,\\ &rank(M)=tr(I_n-L)=trI_n-trL=n-k. \end{aligned}$$

The properties of $L$ and $M$ given above almost exhaust the properties which can be found in the literature. Contrary to the importance of $M$, inquiry into the properties of $M$ seems not yet satisfactory. In this section we shall investigate the properties of the matrix $X$ and its derived matrices, such as $L$ or $M$ rather thoroughly. Specifically we shall relate the properties of $L$ or $M$ with the properties of the columns of $X$. The investigation starts with the following theorem.

### 2.2. A Theorem

Theorem: If $A^*$ is a $k\times k$ nonsingular matrix and if we define

$$(2.5)\quad X^*=XA^*$$

$$(2.6)\quad L^*=X^*(X^{*\prime}X^*)^{-1}X^{*\prime}$$

$$(2.7)\quad M^*=I-L^*,$$

then

$$(2.8)\quad L^*=L$$

$$(2.9)\quad M^*=M.$$

Proof:

$$\begin{aligned} L^*&=X^*(X^{*\prime}X^*)^{-1}X^{*\prime}=XA^*(A^{*\prime}X'XA^*)^{-1}A^{*\prime}X'\\ &=XA^*A^{*-1}(X'X)^{-1}A^{*\prime-1}A^{*\prime}X'=X(X'X)^{-1}X'=L,\\ M^*&=I-L^*=I-L=M. \end{aligned}$$

Q.E.D.

This theorem says that any linear combination of the columns of $X$, so long as the rank of $X$ is preserved, can replace a column of $X$ without altering the resultant matrix $L$ or $M$. For example, we can multiply a column or add a multiple of a column to another without altering $L$ or $M$. Since some important quantities, such as the sum of squared residuals, derived from the LS regression depend on $X$ only through $L$ or $M$, the implication of this theorem is significant.

### 2.3. A Normalizing Transformation

The above theorem is valid for the nonsingular $A^*$ defined by

$$A^*=n^{-1/2}\begin{bmatrix}1 & -z'S_z^{-1}\\ 0 & S_z^{-1}\end{bmatrix} \tag{2.10}$$

where

$$z'=[\bar{x}_2\ \bar{x}_3\cdots\cdots\bar{x}_k]=n^{-1}\iota'Z \tag{2.11}$$

$$S_z=\begin{bmatrix}s_2 & & & \\ & s_3 & & \bigcirc \\ & & \cdot & \\ \bigcirc & & & \cdot \\ & & & s_k\end{bmatrix}=(n^{-1}\Delta Z'AZ)^{1/2} \tag{2.12}$$

$$A=I_n-n^{-1}\iota\iota'. \tag{2.13}$$

If we postmultiply $X$ by $A^*$ defined by (2.10), then we get

$$\begin{aligned}X^*=XA^*&=n^{-1/2}[\iota\ Z]\begin{bmatrix}1 & -z'S_z^{-1}\\ 0 & S_z^{-1}\end{bmatrix}\\ &=n^{-1/2}[\iota\ ZS_z^{-1}-\iota z'S_z^{-1}]\\ &=n^{-1/2}[\iota\ AZS_z^{-1}]\end{aligned} \tag{2.14}$$

$$\begin{aligned}X^{*\prime}X^*&=n^{-1}\begin{bmatrix}\iota'\\ S_z^{-1}Z'A\end{bmatrix}[\iota\ AZS_z^{-1}]\\ &=n^{-1}\begin{bmatrix}\iota'\iota & \iota'AZS_z^{-1}\\ S_z^{-1}Z'A\iota & S_z^{-1}Z'AZS_z^{-1}\end{bmatrix}\\ &=\begin{bmatrix}1 & O'\\ O & n^{-1}S_z^{-1}Z'AZS_z^{-1}\end{bmatrix}\end{aligned} \tag{2.15}$$

since

$$A\iota=O, \tag{2.16}$$

and

$$\iota'\iota=n. \tag{2.17}$$

From (2.14) and (2.15) we get

$$\iota'X^*=n^{-1/2}[\iota'\iota\ \iota'AZS_z^{-1}]=[n^{1/2}\ O'] \tag{2.18}$$

$$\begin{aligned}\Delta X^{*\prime}X^*&=\begin{bmatrix}1 & O'\\ O & n^{-1}S_z^{-1}(\Delta Z'AZ)S_z^{-1}\end{bmatrix}\\ &=\begin{bmatrix}1 & O'\\ O & S_z^{-1}S_z^2S_z^{-1}\end{bmatrix}=I_k.\end{aligned} \tag{2.19}$$

From (2.18) and (2.19) we conclude that the columns of $X^*$, $x_i^*$'s, are normalized vectors in the sense defined in Section 1.4. Therefore the matrix $X^{*\prime}X^*$ is the correlation matrix $R$ of the columns of $X$:

$$R=X^{*\prime}X^*=\begin{bmatrix}1 & O'\\ O & R_z\end{bmatrix}=\begin{bmatrix}1 & 0 & 0 & 0\cdots 0\\ 0 & 1 & r_{23} & r_{24}\cdots r_{2k}\\ 0 & r_{23} & 1 & r_{34}\cdots r_{3k}\\ \cdots & \cdots & \cdots & \cdots\\ 0 & r_{2k} & r_{3k} & r_{4k}\cdots 1\end{bmatrix} \tag{2.20}$$

We also observe that $R=I_k$ if and only if the columns of $X$ are uncorrelated.

### 2.4. Evaluation of the Matrix $X'X$ and Its Inverse

From (2.10) we get

$$A^{*-1}=n^{\frac{1}{2}}\begin{bmatrix}1 & z' \\ 0 & S_z\end{bmatrix} \tag{2.21}$$

Therefore we have

$$X'X=A^{*\prime-1}X^{*\prime}X^{*}A^{*-1}=n\begin{bmatrix}1 & O' \\ z & S_z\end{bmatrix}\begin{bmatrix}1 & O' \\ O & R_z\end{bmatrix}\begin{bmatrix}1 & z' \\ O & S_z\end{bmatrix}=n\begin{bmatrix}1 & z' \\ z & S_zR_zS_z+zz'\end{bmatrix} \tag{2.22}$$

$$\begin{aligned}(X'X)^{-1}=A^{*}(X^{*\prime}X^{*})^{-1}A^{*\prime}&=n^{-1}\begin{bmatrix}1 & -z'S_z^{-1} \\ O & S_z^{-1}\end{bmatrix}\begin{bmatrix}1 & O' \\ O & R_z^{-1}\end{bmatrix}\begin{bmatrix}1 & O' \\ -S_z^{-1}z & S_z^{-1}\end{bmatrix}\\ &=n^{-1}\begin{bmatrix}1+z'S_z^{-1}R_z^{-1}S_z^{-1}z & -z'S_z^{-1}R_z^{-1}S_z^{-1} \\ -S_z^{-1}R_z^{-1}S_z^{-1}z & S_z^{-1}R_z^{-1}S_z^{-1}\end{bmatrix}.\end{aligned} \tag{2.23}$$

Specifically if $x_i$'s are centralized so that $z=0$ and/or if $x_i$'s are uncorrelated so that $R_z=I_{k-1}$, then the expressions (2.22) and (2.23) become simplified.

### 2.5. Lagged Serial Correlation Matrices of Explanatory Variables

As we have defined in Section 1.4, the sampling analogue of cross serial correlation coefficeint of $x_i$ and $x_j$ at lag $s$ is given by

$$r_{ij,s}=\tfrac{1}{2}x_i^{*\prime}D_sx_j^{*},\ i,\ j=1,\ 2,\cdots\cdots k;\ 0<s\leqq n. \tag{2.24}$$

If $i=j$ then it is autocorrelation at lag $s$. Therefore we may call the matrix,

$$R_s=\tfrac{1}{2}X^{*\prime}D_sX^{*},\ 0<s\leqq n, \tag{2.25}$$

the serial correlation matrix of explanatory variables at lag $s$.

Malinvaud [1970] defined a matrix which is essentially the same as

$$R_s^{**}=\tfrac{1}{2}(X'X)^{-1}X'D_sX, \tag{2.26}$$

and he said (in our notation):

"The $R_s^{**}$ appear so to speak as autocorrelation matrix relating to the $n$ vector of the exogenous variables $X$." (p.516)

But this expression seems to be unsatisfactory. If we define a matrix $R_s^{*}$ analogous to (2.26) as

$$R_s^{*}=\tfrac{1}{2}(X^{*\prime}X^{*})^{-1}X^{*\prime}D_sX^{*},\ 0<s\leqq n, \tag{2.27}$$

then

$$R_s^{*}=\tfrac{1}{2}A^{*-1}(X'X)^{-1}X'D_sXA^{*}=A^{*-1}R_s^{**}A^{*}. \tag{2.28}$$

Therefore

(2.29) $$trR_s^*=trR_s^{**}.$$

But in general

(2.30) $$R_s^*\neq R_s^{**}$$

except when $k=2$.

If the columns of $X$ are uncorrelated then $R_s^*=R_s$. If the columns are moderately correlated, then $R_s^*$ is approximately the the serial correlation matrix of explanatory variables at lag $s$. But $R_s^{**}$ is not.

### 2.6. Evaluation of the Minor Traces of $L$ and $M$

In this section we shall evaluate the minor traces of $L$ and $M$ in terms of correlation coefficients of the columns of $X$.

*When* $k=2$. In this case the columns of $X$ are always uncorrelated so that

(2.31) $$X^{*\prime}X^*=I_2.$$

Therefore

(2.32) $$L=X^*X^{*\prime},$$

(2.33) $$trL=trX^*X^{*\prime}=tr(x_1^*x_1^{*\prime}+x_2^*x_2^{*\prime})=x_1^{*\prime}x_1^*+x_2^{*\prime}x_2^*=2,$$

(2.34) $$trM=n-trL=n-2.$$

The $s$th minor trace of $L$ is

(2.35) $$tr_sL=tr_sX^*X^{*\prime}=trD_s(x_1^*x_1^{*\prime}+x^*_2x^*_2{}^{\prime})$$
$$=x_1^{*\prime}D_sx_1^*+x_2^{*\prime}D_sx_2^*=2(r_{11,s}+r_{22,s}),\quad 0<s\leqslant n,$$

and the $s$th minor trace of $M$ is

(2.36) $$tr_sM=-tr_sL,\quad 0<s\leqslant n.$$

*When Columns of* $X$ *Are Uncorrelated.* This is a simple extension of the previous case. In this case

(2.37) $$X^{*\prime}X^*=I_k.$$

Therefore

(2.38) $$L=X^*X^{*\prime},$$

(2.39) $$trL=trX^{*\prime}X^*=trI_k=k,$$

(2.40) $$trM=n-trL=n-k.$$

The $s$th minor trace of $L$ is

(2.41) $$tr_s L = tr D_s X^* X^{*\prime} = tr X^{*\prime} D_s X^* = \sum_{i=1}^{k} x_i^{*\prime} D_s x_i^* = 2\sum_{i=1}^{k} r_{ii,s},\quad 0 < s << n,$$

and the $s$th minor trace of $M$ is

(2.42) $$tr_s M = -tr_s L = -2\sum_{i=1}^{k} r_{ii,s},\quad 0 < s << n.$$

*When* $k=3$ *and* $r_{23} \neq 0$. In this case

(2.43) $$(X^{*\prime} X^*)^{-1} = \begin{bmatrix} 1 & O' \\ O & R_c^{-1} \end{bmatrix}$$

where

(2.44) $$R_c^{-1} = \begin{bmatrix} 1 & r \\ r & 1 \end{bmatrix}^{-1} = (1-r^2)^{-1} \begin{bmatrix} 1 & -r \\ -r & 1 \end{bmatrix} = I_2 + r/(1-r^2) \begin{bmatrix} r & -1 \\ -1 & r \end{bmatrix}$$

(2.45) $$r = r_{23}.$$

Therefore

(2.46) $$(X^{*\prime} X^*)^{-1} = I_3 + r/(1-r^2) \begin{bmatrix} 0 & 0 & 0 \\ 0 & r & -1 \\ 0 & -1 & r \end{bmatrix}$$

(2.47) $$L = X^* (X^{*\prime} X^*)^{-1} X^{*\prime} = X^* X^{*\prime} + L_c$$

where

(2.48) $$L_c = r/(1-r^2) [x_1^* \; x_2^* \; x_3^*] \begin{bmatrix} 0 & 0 & 0 \\ 0 & r & -1 \\ 0 & -1 & r \end{bmatrix} \begin{bmatrix} x_1^{*\prime} \\ x_2^{*\prime} \\ x_3^{*\prime} \end{bmatrix}$$
$$= r(1-r^2)^{-1} \{ r(x_2^* x_2^{*\prime} + x_3^* x_3^{*\prime}) - (x_2^* x_3^{*\prime} + x_3^* x_2^{*\prime}) \}.$$

Since we have already evaluated the minor traces of $X^* X^{*\prime}$, to evaluate the minor traces of $L$ we only need to evaluate the minor traces of $L_c$.

(2.49) $$tr L_c = r(1-r^2)^{-1}(2r - 2r) = 0,$$

(2.50) $$tr_s L_c = r(1-r^2)^{-1} \{ r(x_2^{*\prime} D_s x_2^* + x_3^{*\prime} D_s x_3^*) - 2 x_2^{*\prime} D_s x_3^* \}$$
$$= 2r(1-r^2)^{-1}(r r_{22,s} + r r_{33,s} - 2 r_{23,s}).$$

Therefore we have

(2.51) $$tr L = tr X^* X^{*\prime} + tr L_c = 3,$$

(2.52) $$tr_s L = tr_s X^* X^{*\prime} + tr_s L_c = 2(r_{11,s} + r_{22,s} + r_{33,s})$$
$$+ 2r(1-r^2)^{-1}(r r_{22,s} + r r_{33,s} - 2 r_{23,s}).$$

In the case when the terms in (2.50) tend to cancel out each other, $tr_s L$ tends to be equal to $tr_s X^* X^{*\prime}$ even when $x_2$ and $x_3$ are correlated. We may expect this case frequently in economic data, see Ames and Reiter [1961].

As before we also have

$$trM = n-3, \tag{2.53}$$

$$tr_s M = -tr_s L. \tag{2.54}$$

*When Columns of **X** Are Uncorrelated Except the Last Two.* In this case

$$(X^{*\prime}X^*)^{-1} = I_k + r/(1-r^2)\begin{bmatrix} 0_{k-2} & 0 & 0 \\ 0 & r & -1 \\ 0 & -1 & r \end{bmatrix} \tag{2.55}$$

where

$$r = r_{k-1,k}. \tag{2.56}$$

Therefore

$$L = X^*(X^{*\prime}X^*)^{-1}X^{*\prime} = X^*X^{*\prime} + L_c \tag{2.57}$$

where

$$L_c = r(1-r^2)^{-1}\{r(x_{k-1}^* x_{k-1}^{*\prime} + x_k^* x_k^{*\prime}) - (x_{k-1}^* x_k^{*\prime} + x^* x_{k-1}^{*\prime})\}. \tag{2.58}$$

By the same manner as for (2.49) and (2.50) we have

$$trL_c = 0, \tag{2.59}$$

$$tr_s L_c = 2r(1-r^2)^{-1}\ (rr_{k-1,k-1,s} + rr_{kk,s} - 2r_{k-1,k,s}). \tag{2.60}$$

Therefore we have

$$trL = k, \tag{2.61}$$

$$tr_s L = tr_s X^*X^{*\prime} + tr_s L_c = 2\sum_{i=1}^{k} r_{ii,s} + 2r(1-r^2)^{-1}(rr_{k-1,k-1,s} + rr_{kk,s} - 2r_{k-1,k,s}), \tag{2.62}$$

$$trM = n-k, \tag{2.63}$$

$$tr_s M = -tr_s L. \tag{2.64}$$

When the terms in (2.60) tend to cancel out each other, $tr_s L$ tends to be equal to $tr_s X^*X^{*\prime}$ even when the last two columns of $X$ are correlated.

*When Columns of **X** Are Uncorrelated Except the Last Three.* Assume that the last three columns of $X$ are correlated with common correlation coefficient $r$. Then we get

$$(X^{*\prime}X^*)^{-1} = \begin{bmatrix} I_{k-3} & O' \\ O & R_c^{-1} \end{bmatrix} = I_k + \frac{r(1-r)}{1-3r^2+2r^3}\begin{bmatrix} 0_{k-3} & 0 & 0 & 0 \\ 0 & 2r & -1 & -1 \\ 0 & -1 & 2r & -1 \\ 0 & -1 & -1 & 2r \end{bmatrix} \tag{2.65}$$

with

(2.66) $$R_c=\begin{bmatrix}1 & r & r\\ r & 1 & r\\ r & r & 1\end{bmatrix}$$

Therefore

(2.67) $$L=X^*(X^{*\prime}X^*)^{-1}X^{*\prime}=X^*X^{*\prime}+L_c$$

where

(2.68) $$L_c=\frac{r(1-r)}{1-3r^2+2r^3}\{2r(x_{k-2}{}^*x_{k-2}{}^{*\prime}+x_{k-1}{}^{*\prime}x_{k-1}{}^{*\prime}+x_k{}^*x_k{}^{*\prime}) - \{(x_{k-1}{}^*+x_k{}^*)x_{k-2}{}^{*\prime}+(x_{k-2}{}^*+x_k{}^*)x_{k-1}{}^{*\prime}+(x_{k-2}{}^*+x_{k-1}{}^*)x_k{}^{*\prime})\},$$

The minor traces of $L_c$ are

(2.69) $$trL_c=\frac{r(1-r)}{1-3r^2+2r^3}(6r-6r)=0,$$

(2.70) $$tr_sL_c=\frac{r(1-r)}{1-3r^2+2r^3}\{2r(r_{k-2,k-2,s}+r_{k-1,k-1,s}+r_{kk,s})-6r_s\}$$

where we assume that $r_{ij,s}$ for $i,j=k-2, k-1, k$; $i\neq j$, are all equal to $r_s$. Now we can evaluate the minor traces of $L$ and $M$:

(2.71) $$trL=k,$$

(2.72) $$tr_sL=tr_sX^*X^{*\prime}+tr_sL_c=2\sum_{i=1}^{k}r_{ii,s} +\frac{r(1-r)}{1-3r^2+2r^3}\{2r(r_{k-2,k-2,s}+r_{k-1,k-1,s}+r_{kk,s})-6r_s\},$$

(2.73) $$trM=n-k,$$

(2.74) $$tr_sM=-tr_sL.$$

When the terms in (2.70) tend to cancel out each other, $tr_sL$ tends to be equal to $tr_sX^*X^{*\prime}$ even when the last three columns of $X$ are correlated. We may expect this case frequently in economic data, see Ames and Reiter [1961].

In general, even when the columns of $X$ are correlated the specific cases, (2.52), (2.62) and (2.72), and the specific examples given in Table 2.1 suggest that correlation may not have much effect on the minor traces of $L$ or $M$. Therefore we may reasonably assume that (2.41) and (2.42) are good approximations for the minor traces of $L$ and $M$ even when the columns of $X$ are moderately correlated. This assumption often leads to useful analytical results.

**Table 2.1. Minor traces of matrices $L$ and $X^*X^{*\prime}$**

| Data Sets | Trace | $tr_1$ | $tr_2$ | $tr_3$ | $tr_4$ | $tr_5$ | $tr_6$ | $tr_7$ |
|---|---|---|---|---|---|---|---|---|
| (A) Spirits | | $n=20$, $k=3$, max $r_{ij}=0.89$ | | | | | | |
| $L$ | 3.00 | 4.64 | 3.07 | 2.04 | 1.58 | 1.49 | 1.41 | 1.27 |
| $X^*X^{*\prime}$ | 3.00 | 5.21 | 4.34 | 3.55 | 2.80 | 2.17 | 1.71 | 1.35 |
| (B) Textile | | $n=17$, $k=3$, max $r_{ij}=0.22$ | | | | | | |
| $L$ | 3.00 | 5.04 | 3.85 | 2.46 | 1.43 | 0.66 | 0.18 | −0.09 |
| $X^*X^{*\prime}$ | 3.00 | 5.11 | 3.99 | 2.61 | 1.51 | 0.66 | 0.10 | −0.19 |
| (C) Pears | | $n=16$, $k=5$, max $r_{ij}=0.69$ | | | | | | |
| $L$ | 5.00 | 3.20 | 4.20 | 1.56 | 1.66 | 0.04 | −0.34 | 0.70 |
| $X^*X^{*\prime}$ | 5.00 | 5.86 | 5.31 | 2.28 | 1.33 | −0.26 | −0.83 | −0.57 |
| (D) Consumption | | $n=21$, $k=3$, max $r_{ij}=0.63$ | | | | | | |
| $L$ | 3.00 | 4.75 | 3.58 | 2.65 | 1.95 | 1.30 | 0.86 | 0.45 |
| $X^*X^{*\prime}$ | 3.00 | 4.53 | 3.19 | 2.09 | 1.40 | 0.73 | 0.22 | −0.08 |
| (E) Trend | | $n=21$, $k=2$, max $r_{ij}=0.00$ | | | | | | |
| $L$ | 2.00 | 3.62 | 3.24 | 2.87 | 2.50 | 2.15 | 1.81 | 1.48 |
| $X^*X^{*\prime}$ | 2.00 | 3.62 | 3.24 | 2.87 | 2.50 | 2.15 | 1.81 | 1.48 |
| (F) Artificial(1) | | $n=30$, $k=2$, max $r_{ij}=0.00$ | | | | | | |
| $L$ | 2.00 | 0.00 | 3.73 | 0.00 | 3.47 | 0.00 | 3.20 | 0.00 |
| $X^*X^{*\prime}$ | 2.00 | 0.00 | 3.73 | 0.00 | 3.47 | 0.00 | 3.20 | 0.00 |
| (G) Artificial(2) | | $n=30$, $k=3$, max $r_{ij}=0.06$ | | | | | | |
| $L$ | 3.00 | 1.80 | 5.33 | 1.40 | 4.67 | 1.02 | 4.03 | 0.65 |
| $X^*X^{*\prime}$ | 3.00 | 1.80 | 5.33 | 1.40 | 4.68 | 1.02 | 4.03 | 0.65 |

For the description of data, see Appendix III.

### 2.7. Commutability of $D_s$ and $M$

Consider the case when the columns of $X$ are uncorrelated. In this case we get from (2.38),

$$L = X^*X^{*\prime} = \sum_{i=1}^{k} x_i^* x_i^{*\prime}. \tag{2.75}$$

Consider the matrices, $x_i^*x_i^{*\prime}$, $i=1, 2, \cdots\cdots k$. The trace is

$$tr\, x_i^* x_i^{*\prime} = x_i^{*\prime} x_i^* = 1. \tag{2.76}$$

Therefore the average magnitude of the diagonal elements of $x_i^*x_i^{*\prime}$ is $n^{-1}$. From (2.24), the minor traces are

$$tr_s x_i^* x_i^{*\prime} = x_i^{*\prime} D_s x_i^* = 2r_{ii,s}. \tag{2.77}$$

Therefore the average magnitude of the elements of the $s$th minor diagonal of $\boldsymbol{x}_i^* \boldsymbol{x}_i^{*\prime}$ is

$$(2.78) \qquad (n-s)^{-1} r_{ii,s} = n^{-1}\{n r_{ii,s}/(n-s)\}$$

which is less than $n^{-1}$ in absolute value.

From the above reasoning, if the columns of $\boldsymbol{X}$ are uncorrelated, it follows that the average magnitude of the diagonal elements of $\boldsymbol{L}$ is $k/n$, and the average magnitude of the off-diagonal elements of $\boldsymbol{L}$ is less than $k/n$ in absolute value. Since $\boldsymbol{M}=\boldsymbol{I}-\boldsymbol{L}$, the average magnitude of the diagonal elements of $\boldsymbol{M}$ is $(n-k)/n$, and the average magnitude of the off-diagonal elements of $\boldsymbol{M}$ is less than $k/n$ in absolute value. Therefore, for moderately large $n$ and small $k$, the matrix $\boldsymbol{M}$ resembles the identity matrix $\boldsymbol{I}_n$ under certain regularity conditions, so that

$$(2.79) \qquad \boldsymbol{D}_s \boldsymbol{M} \doteq \boldsymbol{M} \boldsymbol{D}_s,$$

i.e., $\boldsymbol{D}_s$ and $\boldsymbol{M}$ commute approximately.

The statement (2.79) can be strengthened further if we can say that $\boldsymbol{D}_s$ and $\boldsymbol{L}$ commute at least approximately, or,

$$(2.80) \qquad \boldsymbol{D}_s \boldsymbol{L} \doteq \boldsymbol{L} \boldsymbol{D}_s.$$

Two important cases may be considered. If $\boldsymbol{x}_i$ is stationary in the sense that $\boldsymbol{x}_i$ is evenly scattered around the average, $\bar{x}_i$, without trend, then the matrix $\boldsymbol{x}_i^* \boldsymbol{x}_i^{*\prime}$ is approximately a linear combination of $\boldsymbol{D}_s$'s. If this holds for all $\boldsymbol{x}_i$, $i=1, 2, k$, then by (1.14) in Section 1.2, the property (2.80) holds.

Another important case when the property (2.80) holds is when the matrix $\boldsymbol{X}$ consists only of a constant term $\boldsymbol{x}_1$ and a trend or an arithmetic progression $\boldsymbol{x}_2$:

$$(2.81) \qquad \boldsymbol{X} = [\boldsymbol{x}_1 \; \boldsymbol{x}_2].$$

Then we have

$$(2.82) \qquad \boldsymbol{L} = \boldsymbol{X}^* \boldsymbol{X}^{*\prime} = \boldsymbol{x}_1^* \boldsymbol{x}_1^{*\prime} + \boldsymbol{x}_2^* \boldsymbol{x}_2^{*\prime}.$$

Let us denote the vector $\boldsymbol{x}_1^*$ or $\boldsymbol{x}_2^*$ simply by $\boldsymbol{x}$ such that

$$(2.83) \qquad \boldsymbol{x} = [x_1 \; x_2 \; \cdots \; x_n]'.$$

Let us define

$$(2.84) \qquad \boldsymbol{x}\boldsymbol{x}' = \boldsymbol{Y} = [y_{ij}]$$

(2.85) $\quad D_s Y = Y_s = [y_{sij}]$.

If we can show that $Y_s$ is approximately symmetric, then $D_s$ and $Y$ commute approximately and, in turn, $D_s$ and $L$ commute approximately. Therefore we shall prove that $Y_s$ is approximately symmetric.

The $(i,j)$th element of $Y_s$ is

(2.86) $\quad y_{sij} = y_{i+s,j} + y_{i-s,j}, \ s < i, \ j \leqq n-s.$

But the $(i,j)$th element of $Y$ is

(2.87) $\quad y_{ij} = x_i x_j.$

Therefore we get

(2.88) $\quad y_{sij} = x_{i+s}x_j + x_{i-s}x_j = (x_{i+s} + x_{i-s})x_j.$

For our specific $\boldsymbol{x}$,

(2.89) $x_i - x_{i-s} = x_{i+s} - x_i,$

or

(2.90) $\quad x_{i+s} + x_{i-s} = 2x_i, \ s < i \leqq n-s.$

Therefore (2.88) becomes

(2.91) $\quad y_{sij} = 2x_i x_j, \ s < i \leqq n-s.$

Likewise,

(2.92) $\quad y_{sji} = 2x_j x_i = y_{sij}, \ s < i \leqq n-s.$

Therefore, if $n$ is large and $s$ is small compared to $n$, then

(2.93) $\quad D_s Y \doteqdot Y D_s,$

and (2.80) holds.

In sum, if the explanatory variables are composed of a constant term, a linear trend, and stationary variables, and if they are uncorrelated, then (2.80) holds and (2.79) holds fairly accurately. In many cases we may expect economic data which approximately satisfy these conditions.

Even when the columns of $X$ are correlated, the specific cases, (2.47), (2.57), and (2.67), suggest that correlation may not have much effect if multicollinearity is not too extreme. Therefore, even when the columns of $X$ are moderately, correlated we may reasonably assume that $D_s$ and $M$ approximately commute. This assumption often leads to useful analytical

Table 2.2. Comparison of traces of $P^2M$ and $(PM)^2$

| Data | $(n, k)$ | $p$ | $trP^2M$ | $tr(PM)^2$ |
|---|---|---|---|---|
| | | 0 | 11.00 | 11.00 |
| Pears | | 0.3 | 9.50 | 9.30 |
| | (16.5) | 0.5 | 8.38 | 7.75 |
| | | 0.7 | 6.22 | 4.93 |
| | | 0 | 14.00 | 14.00 |
| Textile | | 0.3 | 11.85 | 11.75 |
| | (17.3) | 0.5 | 10.57 | 10.15 |
| | | 0.7 | 8.25 | 7.08 |
| | | 0 | 17.00 | 17.00 |
| Spirits | | 0.3 | 15.97 | 15.86 |
| | (20.3) | 0.5 | 16.56 | 16.08 |
| | | 0.7 | 16.01 | 14.41 |
| | | 0 | 19.00 | 19.00 |
| Trend | | 0.3 | 18.81 | 18.76 |
| | (21.2) | 0.5 | 19.99 | 19.72 |
| | | 0.9 | 19.27 | 18.18 |
| | | 0 | 18.00 | 18.00 |
| Consumption | | 0.3 | 16.87 | 16.68 |
| | (21.3) | 0.5 | 17.34 | 16.51 |
| | | 0.7 | 17.33 | 14.75 |
| | | 0 | 28.00 | 28.00 |
| Artificial(1) | | 0.3 | 32.10 | 32.08 |
| | (30.2) | 0.5 | 40.69 | 40.61 |
| | | 0.7 | 56.87 | 56.41 |
| | | 0 | 27.00 | 27.00 |
| Artificial(2) | | 0.3 | 29.06 | 29.01 |
| | (30.3) | 0.5 | 33.74 | 33.50 |
| | | 0.7 | 38.76 | 37.55 |

**P** is generated by the AR(1) process. For the description of data, see Appendix III. For alternative patterns of serial correlation see Appendix I.

results.

Commutability can be extended further: a matrix, which is a linear combination of $D_s$'s, and $M$ commute approximately. Especially, the autocorrelation matrix $P$ and $M$ commute approximately:

(2.94) $\quad PM \doteq MP,$

since $P$ is a linear combination of $D_s$'s as we have seen in Section 1.5. A direct consequence of it is the property,

(2.95) $\quad tr(PM)^2 = trPMPM \doteq trP^2M^2 = trP^2M.$

For the accuracy of this approximation, see Table 2.2.

## III. Distributions of $e'e$ and $s^2$ under Serial Correlation

### 3.1. Introduction

Consider a linear model,

(3.1) $\quad y = X\beta + u, \ u \sim N(O, \sigma^2 P)$

with $n$ observations and $k$ explanatory variables (including a constant term) and the autocorrelation matrix,

(3.2) $\quad P = \sum_{s=0}^{n-1} p_s D_s.$

If we denote the vector of LS residuals by $e$, then

(3.3) $\quad e = Mu, \ M = I - X(X'X)^{-1}X',$

and the sum of squared residuals is

(3.4) $\quad e'e = u'Mu.$

The LS variance estimator $s^2$ is given by

(3.5) $\quad s^2 = (n-k)^{-1}e'e.$

It is well-known that if disturbances are serially independent, then

(3.6) $\quad e'e \sim \sigma^2\chi^2(n-k),$

(3.7) $\quad s^2 \sim (n-k)^{-1}\sigma^2\chi^2(n-k),$

so that

(3.8) $\quad E(e'e) = (n-k)\sigma^2, \quad Var(e'e) = 2(n-k)\sigma^4,$

and

(3.9) $\quad E(s^2) = \sigma^2, \quad Var(s^2) = 2(n-k)^{-1}\sigma^4.$

But if the disturbances are serially correlated, then the properties, (3.6) to (3.9), are no longer true.

As (3.9) shows, $s^2$ is an unbiased estimator of $\sigma^2$ under serial independence. Goldberger [1964, p. 242], Johnston [1972, pp. 248-49], Malinvaud [1970, Sec. 13.4], and Theil [1971, pp. 256-57] have discussed the bias of the LS variance estimator, $s^2$, under serial correlation. But no one seems to have discussed the higher moments of $s^2$ under serial correlation. In this section we shall derive the distributions of $\boldsymbol{e}'\boldsymbol{e}$ and $s^2$ under serial correlation and consider the effect of serial correlation on the test for $\sigma^2$.

### 3.2. Evaluation of the Moments of $e'e$

Since $\boldsymbol{e}'\boldsymbol{e}$ is a quadratic form in normal variables, its moments can be evaluated in terms of traces of $(\boldsymbol{PM})^s$, see Appendix II. The first two moments are

$$(3.10)\qquad E(\boldsymbol{e}'\boldsymbol{e})=\sigma^2 tr\boldsymbol{PM},$$

$$(3.11)\qquad Var(\boldsymbol{e}'\boldsymbol{e})=2\sigma^4 tr(\boldsymbol{PM})^2 \doteq 2\sigma^4 tr\boldsymbol{P}^2\boldsymbol{M}.$$

The approximate equality in (3.11) follows from the argument given in Section 2.7.

According to the arguments in Section 1.5, we have

$$(3.12)\qquad tr\boldsymbol{PM}=\sum_{s=0}^{n-1} p_s tr_s \boldsymbol{M}=tr\boldsymbol{M}+\sum_{s=1}^{n-1} p_s tr_s \boldsymbol{M}.$$

Due to the properties of the unit minor diagonal matrices in Section 1.2, $\boldsymbol{P}^2$ can be approximated by a linear combination of $\boldsymbol{D}_s$'s:

$$(3.13)\qquad \boldsymbol{P}^2 \doteq \sum_{s=0}^{n-1} c_s \boldsymbol{D}_s.$$

Therefore the trace in (3.11) becomes

$$(3.14)\qquad tr\boldsymbol{P}^2\boldsymbol{M} \doteq tr\sum_{s=0}^{n-1} c_s \boldsymbol{D}_s \boldsymbol{M}=c_0 tr\boldsymbol{M}+\sum_{s=1}^{n-1} c_s tr_s \boldsymbol{M}.$$

Thus the evaluation of (3.10) and (3.11) is reduced to the evaluation of the minor traces of $\boldsymbol{M}$, which has already been done in Section 2.6.

A straightforward evaluation of $c_s$'s is as follows:

$$\begin{aligned}(3.15)\qquad \boldsymbol{P}^2 &=(\boldsymbol{I}+p_1\boldsymbol{D}_1+p_2\boldsymbol{D}_2+\cdots\cdots)^2\\ &=\boldsymbol{I}+(p_1\boldsymbol{D}_1)^2+(p_2\boldsymbol{D}_2)^2+\cdots\cdots+2(p_1\boldsymbol{D}_1+p_2\boldsymbol{D}_2+\cdots\cdots)\\ &\quad+2p_1\boldsymbol{D}_1(p_2\boldsymbol{D}_2+p_3\boldsymbol{D}_3+\cdots\cdots)+2p_2\boldsymbol{D}_2(p_3\boldsymbol{D}_3+p_4\boldsymbol{D}_4+\cdots\cdots)+\cdots\cdots\\ &\doteq \boldsymbol{I}+p_1^2(2\boldsymbol{I}+\boldsymbol{D}_2)+p_2^2(2\boldsymbol{I}+\boldsymbol{D}_4)+\cdots\cdots+2(p_1\boldsymbol{D}_1+p_2\boldsymbol{D}_2+\cdots\cdots)\end{aligned}$$

$$
\begin{aligned}
&+2\{p_1p_2(\boldsymbol{D}_1+\boldsymbol{D}_3)+p_1p_3(\boldsymbol{D}_2+\boldsymbol{D}_4)+\cdots\cdots\}\\
&+2\{p_2p_3(\boldsymbol{D}_1+\boldsymbol{D}_5)+p_2p_4(\boldsymbol{D}_2+\boldsymbol{D}_6)+\cdots\cdots\}+\cdots\cdots\\
=&(1+2p_1^2+2p_2^2+\cdots\cdots)\boldsymbol{I}\\
&+(2p_1+2p_1p_2+2p_2p_3+\cdots\cdots)\boldsymbol{D}_1\\
&+(p_1^2+2p_2+2p_1p_3+2p_2p_4+2p_3p_5+\cdots\cdots)\boldsymbol{D}_2\\
&+(2p_3+2p_1p_2+2p_1p_4+2p_2p_5+2_3p_6+\cdots\cdots)\boldsymbol{D}_3\cdots\cdots\\
&+(p_2^2+2p_4+2p_1p_3+2p_1p_5+2p_2p_6+2p_3p_7+\cdots\cdots)\boldsymbol{D}_4+\cdots\cdots\\
=&\sum_{s=0}^{n-1}\{p_s+\sum_{j=1}^{n-1-s}p_j(p_{j-s}+p_{j+s})\}\boldsymbol{D}_s
\end{aligned}
$$

where $p_o=1$, and $p_{-s}=p_s$. Equating the coefficients of (3.13) and (3.15), we get

$$
(3.16)\qquad c_s=p_s+\sum_{j=1}^{n-1-s}p_j(p_{j-s}+p_{j+s}),\qquad s=0,1,2,\cdots,n-1
$$

If the disturbances are generated according to the AR(1) process so that

$$
(3.17)\qquad p_s=p^s,\ s=0,1,2,\cdots\cdots,
$$

then, from (3.16), $c_s$ becomes

$$
\begin{aligned}
(3.18)\qquad c_s&=p_s+\sum_{j=1}^{n-1-s}p_j(p_{j-s}+p_{j+s})\\
&=p_s+\sum_{j=1}^{s}p_jp_{j-s}+\sum_{j=s+1}^{n-1-s}p_jp_{j-s}+\sum_{j=1}^{n-1-s}p_jp_{j+s}\\
&=p^s+\sum_{j=1}^{s}p^jp^{s-j}+\sum_{j=1}^{n-1-2s}p^{s+j}p^j+\sum_{j=1}^{n-1-s}p^jp^{j+s}\\
&=p^s+sp^s+2\sum_{j=1}^{\infty}p^sp^{2j}\\
&\doteqdot\{1+s+2p^2/(1-p^2)\}p^s=\{s+(1+p^2)/(1-p^2)\}p^s,\qquad s=0,1,2,\cdots\ll n.
\end{aligned}
$$

Specifically,

$$
(3.19)\qquad c_o=(1+p^2)/(1-p^2).
$$

If the disturbances are generated according to the MA(1) process so that

$$
(3.20)\qquad \begin{aligned}&p_o=1,\\ &p_1=p,\\ &p_s=0,\ s>1,\end{aligned}
$$

then, from (3.16), $c_s$'s become

$$
\begin{aligned}
&c_o=1+2p^2,\\
&c_1=2p,
\end{aligned}
$$

(3.21) $$c_2=p^2,$$
$$c_s=0,\quad s>2.$$

For alternative patterns of serial correlation, see Appendix I.

Specific results need specific assumptions. From now on, unless otherwise specified, we shall assume that

(3.22) $$\begin{aligned}&p_s=p^s,\quad s=0,1,2,\cdots\cdots\\&r_{ii,s}=r_{ii}^s,\quad i=1,2,\cdots\cdots k;\ s=0,1,2,\cdots\cdots\\&r_{ij,s}=0,\quad i\neq j,\ s=0,1,2,\cdots\cdots\end{aligned}$$

Then, from (3.10), (3.12), and (2.42), we get

(3.23) $$\begin{aligned}E(e'e)&=\sigma^2 tr\boldsymbol{M}+\sigma^2\sum_{s=1}^{n-1}p_s tr_s\boldsymbol{M}\\&=\sigma^2(n-k)-2\sigma^2\sum_{s=1}^{n-1}p^s\sum_{i=1}^{k}r_{ii}{}^s\\&\doteqdot\sigma^2(n-k)-2\sigma^2\sum_{i=1}^{k}\sum_{s=1}^{\infty}(pr_{ii})^s\\&=\sigma^2(n-k)(1-d_1),\end{aligned}$$

where

(3.24) $$d_1=2(n-k)^{-1}\sum_{i=1}^{k}pr_{ii}/(1-pr_{ii}).$$

From (3.11), (3.14), (3.18), (3.19), (3.22) and (2.42), we get

(3.25) $$\begin{aligned}Var(e'e)&\doteqdot 2\sigma^4\{c_o tr\boldsymbol{M}+\sum_{s=1}^{n-1}c_s tr_s\boldsymbol{M}\}\\&=2\sigma^4\{c_o(n-k)-2\sum_{s=1}^{n-1}(s+c_o)p^s\sum_{i=1}^{k}r_{ii}{}^s\}\\&\doteqdot 2\sigma^4\{c_o(n-k)-2c_o\sum_{i=1}^{k}\sum_{s=1}^{\infty}(c_o(pr_{ii})^s+s(pr_{ii})^s)\}\\&=2\sigma^4(n-k)(c_0-d_2),\end{aligned}$$

where

(3.26) $$d_2=2(n-k)^{-1}\{c_o\sum_{i=1}^{k}pr_{ii}/(1-pr_{ii})+\sum_{i=1}^{k}pr_{ii}/(1-pr_{ii})^2\}.$$

We may evaluate the moments of $e'e$ analogously when the columns of $\boldsymbol{X}$ are correlated. The moments are affected by the correlation only through the minor traces of $\boldsymbol{M}$. But as we have seen in Section 2.6, the minor traces are affected by the correlation only slightly when the correlation is moderate. Therefore we may consider the approximation of the moments of $e'e$ for uncorrelated $X$ also to be a good approximation for correlated $X$.

In (3. 24) and (3. 26), as $n\rightarrow\infty$, $d_1, d_2\rightarrow 0$, so that asymptotically we have

$$(3.27)\qquad E(e'e)\doteq\sigma^2(n-k),\quad Var(e'e)\doteq 2\sigma^4 c_0(n-k).$$

Note that the expressions in (3. 27) are independent of $X$. Asymptotically, the first moment is the same as that of the uncorrelated case, but the second moment is not.

### 3. 3. Distribution of e′e under Serial Correlation

Since $e'e$ is a positive semidefinite quadratic form in normal variables, we can approximate the distribution by Patnaik's two moment $\chi^2$ approximation, see Appendix II. By equating the first two moments of $e'e$ and $c\chi^2(f)$, we have

$$cf=\sigma^2(n-k)(1-d_1),\quad 2c^2f=2\sigma^4(n-k)(c_0-d_2),$$

or

$$(3.28)\qquad c=\sigma^2(c_0-d_2)/(1-d_1),\quad f=(n-k)(1-d_1)^2/(c_0-d_2).$$

Therefore we can say that the distribution of $e'e$ under serial correlation is approximately

$$(3.29)\qquad e'e\stackrel{\cdot}{\sim}\sigma^2(c_0-d_2)/(1-d_1)\chi^2\{(n-k)(1-d_1)^2/(c_0-d_2)\}.$$

Normally, $d_1$ and $d_2$ move in the same direction. Therefore the asymptotic approximation (3. 27) may be applicable to the case of moderately large $n$, in which case (3. 29) becomes

$$(3.30)\qquad e'e\stackrel{\cdot}{\sim}\sigma^2 c_0\chi^2\{(n-k)/c_0\}.$$

### 3. 4. Distribution of $s^2$ under Serial Correlation

The LS variance estimator $s^2$ defined by (3. 5) is an unbiased estimator of $\sigma^2$ under serial independence. But under serial correlation it may be biased. Under serial correlation generated by an AR(1) process, we get, from (3. 5) and (3. 23),

$$(3.31)\qquad E(s^2)=(n-k)^{-1}E(e'e)\doteq\sigma^2(1-d_1),$$

where $d_1$ is as defined by (3. 24):

$$d_1=2(n-k)^{-1}\sum_{i=1}^{k}pr_{ii}/(1-pr_{ii}).$$

$d_1$ represents the relative bias of the estimator. If all explanatory variables are serially uncorrelated then the bias is zero. But it is impossible when there is a constant term. The bias tends to zero as $n$ increases. Therefore $s^2$ is asymptotically unbiased under serial correlation.

The variance of $s^2$ can be obtained from the variance of $e'e$. From (3.25),

$$(3.32)\qquad Var(s^2)=(n-k)^{-2}Var(e'e)\doteq 2\sigma^4(n-k)^{-1}(c_0-d_2),$$

where $c_0$ and $d_2$ are as defined by (3.19) and (3.26). If $c_0-d_2=1$, then the variance is the same as the variance under serial independence. As $n$ goes to infinity $d_2$ goes to zero. But $c_0$ does not go to 1.

Patnaik's approximation of the distribution of $s^2$ leads to

$$(3.33)\qquad s^2\dot{\sim}\sigma^2(n-k)^{-1}(c_0-d_2)/(1-d_1)\chi^2(f)=\sigma^2(1-d_1)F(f,\infty)$$

where

$$(3.34)\qquad F(f,\infty)=f^{-1}\chi^2(f)$$

is the "chi-square over degrees of freedom" distribution, and

$$(3.35)\qquad f=(n-k)(1-d_1)^2/(c_0-d_2).$$

Analogous to (3.30), the asymptotic approximation of the distribution of $s^2$ is

$$(3.36)\qquad s^2\dot{\sim}\sigma^2(n-k)^{-1}c_0\chi^2((n-k)/c_0)=\sigma^2F((n-k)/c_0,\infty).$$

Table 3.1 gives some idea of the magnitudes of $d_1$, $d_2$, and the effective degrees of freedom, $f$. From this table we see that the primary determinant of $f$ is $p$, and $f$ is very close to $(n-k)/c_0$ even for $n-k=30$. The last column of this table is explained in Section 3.5.

### 3.5. Hypothesis Testing on $\sigma^2$ under Serial Correlation

As we have seen in Section 3.1, under serial independence $s^2$ is distributed as

$$(3.37)\qquad s^2\sim\sigma^2F\{(n-k),\infty\}.$$

Therefore if we test a null hypothesis,

$$(3.38)\qquad H_0:\sigma^2=\sigma_0^2,$$

against the alternative hypothesis,

**Table 3.1. Effective degrees of freedom and true significance level**

| $p$ | $r_{22}=r_{33}$ | $d_1$ | $d_2$ | $c_0^{-1}$ | $f/(n-k)$ | $f$ | $\alpha_J^*$ |
|---|---|---|---|---|---|---|---|
| | | $n=33,\ k=3,\ \alpha=0.05$ | | | | | |
| 0 | 0 | 0 | 0 | 1 | 1 | 30.00 | 0.050 |
| 0.3 | 0 | 0.0286 | 0.0750 | 0.8349 | 0.8405 | 25.21 | 0.072 |
| 0.5 | 0 | 0.0667 | 0.2444 | 0.6000 | 0.6125 | 18.38 | 0.129 |
| 0.7 | 0 | 0.1556 | 0.9730 | 0.3423 | 0.3660 | 10.98 | 0.266 |
| 0.3 | 0.3 | 0.0418 | 0.1053 | 0.8349 | 0.8405 | 25.22 | 0.073 |
| 0.5 | 0.3 | 0.0902 | 0.3113 | 0.6000 | 0.6107 | 18.32 | 0.135 |
| 0.7 | 0.3 | 0.1910 | 1.1214 | 0.3423 | 0.3636 | 10.91 | 0.285 |
| 0.3 | 0.5 | 0.0521 | 0.1309 | 0.8349 | 0.8422 | 25.27 | 0.073 |
| 0.5 | 0.5 | 0.1111 | 0.3778 | 0.6000 | 0.6130 | 18.39 | 0.141 |
| 0.7 | 0.5 | 0.2274 | 1.2932 | 0.3423 | 0.3666 | 11.00 | 0.307 |
| 0.3 | 0.7 | 0.0640 | 0.1624 | 0.8349 | 9.8461 | 25.38 | 0.074 |
| 0.5 | 0.7 | 0.1385 | 0.4746 | 0.6000 | 0.6226 | 18.68 | 0.149 |
| 0.7 | 0.7 | 0.2837 | 1.5984 | 0.3423 | 0.3878 | 11.63 | 0.345 |
| | | $n=63,\ k=3,\ \alpha=0.05$ | | | | | |
| 0 | 0 | 0 | 0 | 1 | 1 | 60.00 | 0.050 |
| 0.5 | 0 | 0.0333 | 0.1222 | 0.6000 | 0.6050 | 36.30 | 0.130 |
| 0.7 | 0 | 0.0778 | 0.4865 | 0.3423 | 0.3493 | 20.96 | 0.261 |
| 0.8 | 0 | 0.1333 | 1.2741 | 0.2195 | 0.2289 | 13.73 | 0.382 |
| 0.5 | 0.3 | 0.0451 | 0.1557 | 0.6000 | 0.6035 | 36.21 | 0.133 |
| 0.7 | 0.3 | 0.0955 | 0.5607 | 0.3423 | 0.3465 | 20.79 | 0.271 |
| 0.8 | 0.3 | 0.1544 | 1.3977 | 0.2195 | 0.2264 | 13.59 | 0.395 |
| 0.5 | 0.5 | 0.0556 | 0.1889 | 0.6000 | 0.6036 | 36.22 | 0.136 |
| 0.7 | 0.5 | 0.1137 | 0.6466 | 0.3423 | 0.3453 | 20.72 | 0.281 |
| 0.8 | 0.5 | 0.1778 | 1.5506 | 0.2195 | 0.2250 | 13.50 | 0.412 |
| 0.5 | 0.7 | 0.0692 | 0.2373 | 0.6000 | 0.6061 | 36.37 | 0.141 |
| 0.7 | 0.7 | 0.1418 | 0.7992 | 0.3423 | 0.3470 | 20.82 | 0.300 |
| 0.8 | 0.7 | 0.2182 | 1.8534 | 0.2195 | 0.2262 | 13.57 | 0.443 |

(3.39) $H_a: \sigma^2 \neq \sigma_0^2,$

then the conventional critical region of size $\alpha$ is defined by

(3.40) $\{s^2 | s^2 < \sigma_0^2 F_{\frac{1}{2}\alpha}(n-k, \infty)\} \cup \{s^2 | s^2 > \sigma_0^2 F_{1-\frac{1}{2}\alpha}(n-k, \infty)\}.$

But under serial correlation, $s^2$ is no longer distributed as (3.37) but as (3.33). Therefore, the true significance level of the test, $\alpha^*$, under serial correlation is approximately given by

(3.41) $\alpha_J^* = Pr\{F(f, \infty) < (1-d_1)^{-1} F_{\frac{1}{2}\alpha}(n-k, \infty)\}$

$$+Pr\{F(f,\infty)>(1-d_1)^{-1}F_{1-\frac{1}{2}\alpha}(n-k,\infty)\}.$$

The last column of Table 3.1 shows some examples of the values of $\alpha_J{}^*$ for different values of parameters, given $\alpha=0.05$.[1] From this table we can derive the following:

(1) Conventional $\chi^2$-test on $\sigma^2$ may not be robust for moderately correlated models.

(2) The dominant determinant of the true significance level is $p$, given the nominal level $\alpha$.

(3) For models with $n-k\geqq 30$, the effect of $n$ and $X$ on the true significance level is small.

## Appendix I. Alternative Models of Serial Correlation*

### A1.1. Linear Stochastic Models

Alternative serial correlation models can be defined by corresponding linear stochastic processes. A general linear stochastic process $u_t$, $t=0, \pm 1, \pm 2, \cdots\cdots$, can be represented by a weighted sum of present and past values of a "white noise" process, $v_t$:

$$\text{(A1.1)} \qquad u_t=v_t+\sum_{s=1}^{\infty}\phi_s v_{t-s}, \quad t=0, \ \pm 1, \ \pm 2, \cdots\cdots$$

The white noise process may be regarded as a series of shocks which drive the system. It consists of a sequence of uncorrelated normal random variables with mean zero and constant variance, i.e., for all $t$;

$$\text{(A1.2)} \qquad E(v_t)=0$$

$$E(v_t v_{t+s})=\gamma_s=\begin{cases}\sigma^2{}_v, & s=0.\\ 0, & s\neq 0,\end{cases}$$

---

[1] To check the accuracy of $\alpha_J{}^*$ we calculated

$$\alpha^*=\Pr(s^2<\sigma_0{}^2F_{\frac{1}{2}\alpha})+\Pr(s^2>\sigma_0{}^2F_{1-\frac{1}{2}\alpha})$$

for Consumption data explained in Appendix III using the Imhof [1961] procedure for evaluation of the distribution of a quadratic form in normal variables, see Appendix II. The result is as follows:

| | | | | |
|---|---|---|---|---|
| $p$ | 0.0 | 0.3 | 0.5 | 0.7 |
| $\alpha^*$ | 0.050 | 0.068 | 0.134 | 0.345 |

For this data set $n=21$, $k=3$, $r_{22}=0.681$, and $r_{33}=0.631$.

* This depends heavily on the presentation of Box and Jenkins[1970].

If we define the autocorrelation coefficient at lag $s$ by

(A1.3) $p_s=\gamma_s/\gamma_0,$

then, for the white noise process, $p_0=1$, and $p_s=0$, $s>0$.

Alternatively, the process $u_t$ can be written as a weighted sum of past values of $u_t$'s plus an added shock, $v_t$:

(A1.4) $u_t=\sum_{s=1}^{\infty}\pi_s u_{t-s}+v_t,\ t=0,\ \pm1,\ \pm2,\cdots\cdots$

If we introduce the lag operator $L$ such that

(A1.5) $Lu_t=u_{t-1},\ L^s u_t=u_{t-s},$

we can rewrite (A1.1) and (A1.4) as

(A1.6) $u_t=\Psi(L)v_t$

(A1.7) $\Pi(L)u_t=v_t,$

respectively, where

(A1.8) $\Psi(L)=1+\sum_{s=1}^{\infty}\psi_s L^s$

(A1.9) $\Pi(L)=1-\sum_{s=1}^{\infty}\pi_s L^s.$

**A1.2. Stationarity and Invertibility Conditions for a Linear Stochastic Process**

For the linear stochastic process represented by (A1.1) or (A1.6), the process is stationary if the sequence of $\psi$ weights of the process converges. The stationarity condition is that $\Psi(L)$ converges on or within the unit circle.

For the linear stochastic process represented by (A1.4) or (A1.7), the process is invertible if the sequence of $\pi$ weights of the process converges. The invertibility condition is that $\Pi(L)$ converges on or within the unit circle. The invertibility condition is independent of the stationarity condition.

**A1.3. Autoregressive, Moving Average, and Mixed Autoregressive Moving Average Processes**

The representations (A1.1) and (A1.4) of the general linear stochastic process would not be very useful in practice if they contained an infinite number of parameters, $\psi_s$ and $\pi_s$. We can introduce parsimony and yet retain models which are representationally useful by use of autoregressive, moving average, and mixed autoregressive moving average processes.

The special case of (A1.4), in which only the first $p$ of the weights are nonzero, may be written

(A1.10) $$u_t=\sum_{s=1}^{p}\phi_s u_{t-s}+v_t,\ t=0, \pm 1,\ \pm 2, \cdots\cdots,$$

where we now use symbols $\phi_s$ for the *finite* set of weight parameters. The process defined by (A1.10) is called an autoregressive process of order $p$, or more succinctly, an AR($p$) process. This process is always invertible.

The special case of (A1.1), in which only the first $q$ of the weights are nonzero, may be written

(A1.11) $$u_t=v_t-\sum_{s=1}^{q}\theta_s v_{t-s},\ t=0,\ \pm 1,\ \pm 2, \cdots\cdots,$$

where we now use symbols $-\theta_s$ for the *finite* set of weight parameters. The process defined by (A1.11) is called a moving average process of order $q$, or an MA($q$) process. The process is always stationary.

We may achieve parsimony by including both autoregressive and moving average terms in the presentation of a process. Thus we define the mixed autoregressive moving average process as

(A1.12) $$u_t=\sum_{s=1}^{p}\phi_s u_{t-s}+v_t-\sum_{s=1}^{q}\theta_s v_{t-s},\ t=0,\ \pm 1,\ \pm 2, \cdots\cdots,$$

which will be referred to as an ARMA($p$,$q$) process. This process is stationary if the autoregressive part is stationary and it is invertible if the moving average part is invertible.

In the following we shall discuss some important characteristics of the AR(1), AR(2), MA(1), MA(2), and ARMA(1,1) processes. The AR(1) process is the most widely assumed process in economics. The AR(2) process is sometimes assumed as a generalized alternative to the AR(1) process. The MA(1) process appears naturally in the Koyck[1954, p.32] transformation of a distributed lag model. The MA(2) process is a generalized alternative to the MA(1) process. The ARMA(1,1) process is a generalized alternative to both the AR(1) and MA(1) processes.

### A1.4. The First-Order and the Second-Order Autoregressive Processes

The AR(1) or the first order Markov process is defined by

(A1.13) $u_t=\phi_1 u_{t-1}+v_t,\ t=0,\ \pm1,\ \pm2,\cdots\cdots,$

For stationarity we require

(A1.14) $|\phi|<1.$

The serial correlation coefficients, $p_s$, satisfy the first order difference equation,

(A1.15) $p_s=\phi_1 p_{s-1},\ s>0,$

which, with the initial condition $p_0=1$, has the solution,

(A1.16) $p_s=\phi_1{}^s,\ s\geqq 0.$

Specifically $p_1=\phi_1$. Therefore the autocorrelation matrix $\boldsymbol{P}$ for $n$ successive observations is

(A1.17) $$\boldsymbol{P}=\sum_{s=0}^{n-1} p_1{}^s \boldsymbol{D}_s,$$

where the $\boldsymbol{D}_s$'s are the unit minor diagonal matrices defined in Section 1.2.

The AR(2) process is defined by

(A1.18) $u_t=\phi_1 u_{t-1}+\phi_2 u_{t-2}+v_t,\ t=0,\ \pm1,\ \pm2,\cdots\cdots$

For stationarity the parameters $\phi_1$ and $\phi_2$ must lie in the triangular region defined by

(A1.19) $$\begin{aligned}&\phi_2+\phi_1<1\\&\phi_2-\phi_1<1\\&-1<\phi_2<1\end{aligned}$$

The serial correlation coefficients, $p_s$, satisfy the second order difference equation,

(A1.20) $p_s=\phi_1 p_{s-1}+\phi_2 p_{s-2},\ s>1$

with the initial conditions,

(A1.21) $$\begin{aligned}&p_0=1\\&p_1=\phi_1/(1-\phi_2).\end{aligned}$$

Therefore the autocorrelation matrix for $n$ successive observations can be defined.

If $p_1$ and $p_2$ are specified, then $\phi_1$ and $\phi_2$ can be calculated by

(A1.22) $$\begin{aligned}&\phi_1=p_1(1-p_2)/(1-p_1{}^2)\\&\phi_2=(p_2-p_1{}^2)/(1-p_1{}^2)\end{aligned}$$

Here we note that if $p_1$ is specified beforehand, then the stationarity conditions, (A1.19), restrict the range of the permissible $p_2$ as

$$\text{(A1.23)} \qquad \begin{aligned} &\sup p_2 = 1 \\ &\inf p_2 = 2p_1^2 - 1. \end{aligned}$$

We shall examine the behavior of $p_s$ when $p_1 > 0$. From (A1.22) we see that it implies $\phi_1 > 0$. The characteristic roots of equation (A1.20) are

$$\text{(A1.24)} \qquad a_1, a_2 = \tfrac{1}{2}(\phi_1 \pm \sqrt{\phi^2_1 + 4\phi_2})$$

If $\phi_2 > 0$, then the dominant root $a_1$ lies between $\phi_1$ and 1 and $a_2$ is negative under stationarity so that as $s$ increases $p_s$ decreases with short oscillations. If $\phi_2 = 0$ then $a_1 = \phi_1$ and $a_2 = 0$ so that it corresponds to the AR(1) case. If $\phi_2 < 0$ but $\phi_1^2 + 4\phi_2 \geqq 0$, then $a_1$ and $a_2$ are both positive but less than one so that as $s$ increases $p_s$ decreases monotonically. If $\phi_1^2 + 4\phi_2 < 0$, then the roots are complex and as $s$ increases $p_s$ exhibits damping oscillations with damping factor $\sqrt{-\phi_2}$ and period of oscillation,

$$\text{(A1.25)} \qquad T = 2\pi / (\cos^{-1}(\phi_1 / 2\sqrt{-\phi_2})).$$

Since (A1.25) can be written as

$$\text{(A1.26)} \qquad \phi_1 = 2\sqrt{-\phi_2}\cos(2\pi/T),$$

for fixed $T$, the locus of (A1.26) in the $\phi_1 - \phi_2$ plane is a half parabola, and the parametric change of $T$ generates an isoperiod map (Jeong[1975b]). Since $p_1 > 0$ implies $\phi_1 > 0$, and it, in turn, implies $\cos(2\pi/T) > 0$, for $p_1 > 0$, the period of oscillation $T$ is always greater than 4.

### A1.5. The First-Order and the Second-Order Moving Average Processes

The MA(1) process is defined by

$$\text{(A1.27)} \qquad u_t = v_t - \theta_1 v_{t-1}, \quad t = 0, \pm 1, \pm 2, \cdots\cdots$$

This process is stationary for all values of $\theta_1$. But for invertibility we require

$$\text{(A1.28)} \qquad |\theta_1| < 1.$$

The serial correlation coefficients, $p_s$, are given by

$$\text{(A1.29)} \qquad \begin{aligned} &p_o = 1 \\ &p_1 = -\theta_1 / (1 + \theta_1^2) \\ &p_s = 0, \quad s > 1. \end{aligned}$$

Therefore the autocorrelation matrix $\boldsymbol{P}$ for $n$ successive observations is

(A1.30) $\boldsymbol{P}=\boldsymbol{I}_n+p_1\boldsymbol{D}_1.$

From (2.29) we see that

(A1.31) $|p_1|<\frac{1}{2}.$

If $p_1$ is specified then the $\theta_1$ which satisfies the invertibility condition, (A1.28), is given by

(A1.32) $\theta_1=(\sqrt{1-4p_1^2}-1)/2p_1$

uniquely.

The MA(2) process is defined by

(A1.33) $u_t=v_t-\theta_1 v_{t-1}-\theta_2 v_{t-2},\ t=0,\pm 1,\ \pm 2,\cdots\cdots$

The process is always stationary. But for invertibility we require

(A1.34)
$$\begin{aligned}&\theta_2+\theta_1<1\\&\theta_2-\theta_1<1\\&-2<\theta_2<1\end{aligned}$$

The serial correlation coefficients are given by

(A1.35)
$$\begin{aligned}&p_0=1\\&p_1=-\theta_1(1-\theta_2)/(1+\theta_1^2+\theta_2^2)\\&p_2=-\theta_2/(1+\theta_1^2+\theta_2^2)\\&p_s=0,\ s>2.\end{aligned}$$

Therefore the autocorrelation matrix $\boldsymbol{P}$ is of the form

(A1.26) $\boldsymbol{P}=\boldsymbol{I}_n+p_1\boldsymbol{D}_1+p_2\boldsymbol{D}_2.$

If $p_1$ is specified in advance, then the invertibility condition (A1.34) restricts the range of permissible $p_2$ as follows:

(A1.37)
$$\sup p_2=\tfrac{1}{2}p_1^2/(1-\sqrt{1-2p_1^2})$$
$$\inf p_2=\begin{cases}\tfrac{1}{2}p_1^2/(1+\sqrt{1-2p_1^2}), & \text{for } 2/3\leqq p_1<\sqrt{\tfrac{1}{2}}\\ p_1-\tfrac{1}{2}, & \text{for } 0<p_1<2/3.\end{cases}$$

### A1.6. The Mixed First-Order Autoregressive First-Order Moving Average Process

The ARMA(1,1) process is defined by

(A1.38) $u_t=\phi_1 u_{t-1}+v_t-\theta_1 v_{t-1},\ t=0,\ \pm 1,\ \pm 2,\cdots\cdots$

The stationarity and invertibility conditions for the process are

(A1.39) $|\phi_1|<1$

$|\theta_1|<1.$

The serial correlation coefficients are given by

(A1. 40) $p=1$
$p_1=(1-\phi_1\theta_1)(\phi_1-\theta_1)/(1+\theta_1^2-2\phi_1\theta_1)$
$p_s=\phi_1 p_{s-1},\ s>1.$

So for the ARMA(1, 1) process, $p_s$ decays exponentially from the starting value $p_1$ in contrast with the starting value $p_0=1$ for the AR(1) process.

If $p_1$ is specified in advance, the stationarity and invertibility conditions (A1. 39) restrict the range of permissible $p_2$ as

(A1. 41) $\sup p_2=|p_1|$
$\inf p_2=|p_1|(2|p_1|-1).$

## Appendix II. Quadratic Forms in Normal Variables

### A2. 1. Introduction

Suppose that $\boldsymbol{x}$ is an $n\times 1$ vector of central normal variables such that

(A2. 1) $\boldsymbol{x}\sim N(\boldsymbol{O},\boldsymbol{V}),$

where $\boldsymbol{V}$ is an $n\times n$ positive definite matrix. A quadratic form of $\boldsymbol{x}$, $Q(\boldsymbol{x})$, associated with a symmetric matrix $\boldsymbol{A}$ is defined

(A2. 2) $Q(\boldsymbol{x})=\boldsymbol{x}'\boldsymbol{A}\boldsymbol{x}.$

Many test statistics to be encountered in econometrics are intimately related to such quadratic forms. We shall discuss appropriate properties of the quadratic forms and the distribution of them. General references are Johnson and Kotz [1970b, Ch.29] and Rao [1971, Sec. 3b.4].

### A2. 2. Distribution of the Quadratic Forms

Finding the distribution of $Q(\boldsymbol{x})$ is equivalent to finding

(A2. 3) $Pr(Q(\boldsymbol{x})\leqq q)=Pr(\boldsymbol{x}'\boldsymbol{A}\boldsymbol{x}\leqq q),\ -\infty<q<\infty.$

Since $\boldsymbol{V}$ is a positive definite matrix, we can decompose it as

(A2. 4) $\boldsymbol{V}=\boldsymbol{T}\boldsymbol{T}',$

where $\boldsymbol{T}$ is a lower triangular matrix with positive diagonal elements. The matrix $\boldsymbol{T}'\boldsymbol{A}\boldsymbol{T}$ is symmetric, so if $\boldsymbol{\Lambda}$ is the diagonal matrix of eigenvalues of $\boldsymbol{T'AT}$ and $\boldsymbol{R}$ is the associated orthogonal matrix of eigenvectors of it, then we have

(A2.5) $\boldsymbol{R'T'ATR}=\boldsymbol{\Lambda}.$

Hence, if the linear transformation,

(A2.6) $\boldsymbol{w}=\boldsymbol{R'T}^{-1}\boldsymbol{x}$ or $\boldsymbol{x}=\boldsymbol{TRw}$,

is applied, then

(A2.7) $Var(\boldsymbol{w})=E(\boldsymbol{ww'})=\boldsymbol{R'T}^{-1}E(\boldsymbol{xx'})\boldsymbol{T}'^{-1}\boldsymbol{R}=\boldsymbol{R'R}=\boldsymbol{I}.$

From (A2.2) to (A2.6),

(A2.8) $Q(\boldsymbol{x})=\boldsymbol{x'Ax}=\boldsymbol{w'R'T'ATRw}=\boldsymbol{w'\Lambda w}=\sum_{j=1}^{n}\lambda_j w_j^2,$

where the numbers, $\lambda_1 \geqq \lambda_2 \geqq \cdots \geqq \lambda_n$, are the eigenvalues of $\boldsymbol{T'AT}$ or equivalently of $\boldsymbol{VA}$, and $w_j^2$'s are independent $\chi^2(1)$ variables due to (A2.7). If we denote the distribution function of $Q$ by $F(q)$, it now becomes

(A2.9) $F(q)=Pr(\sum_{j=1}^{n}\lambda_j w_j^2<q).$

If $\lambda_j$'s are bounded, then, by the central limit theorem, (standardized) $F(q)$ approaches a normal distribution as $n$ tends to infinity.

### A2.3. Characteristic Function and Moments

Since $w_j^2$'s are mutually independent $\chi^2(1)$ variables, the characteristic function of $Q$ is easily derivable from that of $\chi^2(1)$ as

(A2.10) $\phi(t)=E(e^{itQ})=\prod_{j=1}^{n}(1-2it\lambda_j)^{-\frac{1}{2}}.$

The cumulants are obtained from the cumulant generating function,

(A2.11)
$$K(t)=\ln E(e^{tQ})=-\tfrac{1}{2}\sum_{j=1}^{n}\ln(1-2\lambda_j t)$$
$$=\sum_{j=1}^{n}\sum_{s=1}^{\infty}2^{s-1}(\lambda_j t)^s/s$$
$$=\sum_{1=s}^{\infty}(t^s/s!)(2^{s-1}(s-1)!\sum_{j=1}^{n}\lambda_j^s),$$

see Lancaster(1953). Denoting the $s$th cumulant of $Q$ by $K_s(Q)$, we have in general

(A2.12) $K(t)=\sum_{s=1}^{\infty}(t^s/s!)K_s(Q).$

Comparing (A2.11) and (A2.12), we have

(A2.13)　$K_s(Q)=2^{s-1}(s-1)!\sum_{j=1}^{n}\lambda_j{}^s=2^{s-1}(s-1)!tr(\boldsymbol{VA})^s.$

If necessary, we can determine the mean $\mu$ and the $s$th central moments $\mu_s$ by

$$\mu=K_1(Q)=tr\boldsymbol{VA}$$
$$\mu_2=K_2(Q)=2tr(\boldsymbol{VA})^2$$
$$\mu_3=K_3(Q)=8tr(\boldsymbol{VA})^3$$
$$\mu_4-3\mu^2{}_2=K_4(Q)=48tr(\boldsymbol{VA})^4,$$

etc., see Rao[1973, p. 101].

### A2.4. The Imhof Procedure of Exact Computation of the Distribution of Quadratic Forms

There are several procedures available for exact computation of the distribution of quadratic forms, for example, Durbin and Watson [1971, Appendix 1]. But it seems that the Imhof procedure is the most popular among economists. Imhof [1961] proposed a computational procedure for exact calculation of the distribution of $Q(\boldsymbol{x})$ where $\boldsymbol{x}$ may be a noncentral normal random vector. But since we need only the central case in the present paper we shall confine the discussion to the central case only.

The Imhof procedure is a straightforward numerical integration of an inversion formula by Gil-Pelaez [1951], namely,

(A2.15)　$F(q)=\frac{1}{2}-\pi^{-1}\int_0^{\infty}t^{-1}Im(y)dt,$

where $Im(y)$ is the imaginary part of the complex number,

(A2.16)　$y=e^{-itq}\phi(t).$

If we denote the complex number by

(A2.17)　$y=re^{i\theta},$

then the imaginary part becomes

(A2.18)　$Im(y)=r\sin\theta.$

Therefore the main problem reduces to how to explain $r$ and $\theta$ in known quantities.

If we denote

(A2.19) $$1-2it\lambda_j=r_j\exp(i\theta_j),$$

then

(A2.20) $$r_j=(1+4\lambda_j^2t^2)^{\frac{1}{2}}$$
$$\theta_j=-\tan^{-1}(2\lambda_jt),$$

so that we have

(A2.21) $$y=e^{-itq}\prod_{j=1}^{n}(1-2i\lambda_jt)^{-\frac{1}{2}}$$
$$=\prod_{j=1}^{n}r_j^{-\frac{1}{2}}\exp\{i(-tq-\tfrac{1}{2}\sum_{k=1}^{n}\theta_k)\}$$
$$=\prod_{j=1}^{n}(1+4\lambda_j^2t^2)^{-\frac{1}{4}}\exp\{i(-tq+\tfrac{1}{2}\sum_{k=1}^{n}\tan^{-1}(2\lambda_kt))\}.$$

Comparing (A2.17) and (A2.21) we get

(A2.22) $$r=\prod_{j=1}^{n}(1+4\lambda_j^2t^2)^{-1/4}$$
$$\theta=-tq+\tfrac{1}{2}\sum_{j=1}^{n}\tan^{-1}(2\lambda_jt).$$

From (A2.15) and (A2.18), we get

(A2.23) $$F(q)=\tfrac{1}{2}-\pi^{-1}\int_0^{\infty}t^{-1}r\sin\theta dt,$$

$r$ and $\theta$ being defined by (A2.22). If we substitute

(A2.24) $$u=2t,$$

then (5.23) becomes

(A2.25) $$F(q)=\tfrac{1}{2}-\pi^{-1}\int_0^{\infty}\frac{\sin\theta(u)}{\gamma(u)}du$$

where

(A2.26) $$\theta(u)=\tfrac{1}{2}\{\sum_{j=1}^{n}\tan^{-1}(\lambda_ju)-qu\}$$

(A2.27) $$\gamma(u)=u\prod_{j=1}^{n}(1+\lambda^2{}_ju^2)^{1/4}.$$

Imhof [1961] also proved the following necessary result:

(A2.28) $$\lim_{u\to0}\frac{\sin\theta(u)}{\gamma(u)}=\tfrac{1}{2}(\sum_{j=1}^{n}\lambda_j-q).$$

This follows from the fact that as $u\to0$, (A2.29) $\sin\theta(u) \to \frac{1}{2}(\sum_{j=1}^{n} \lambda_j u - qu)$

$\gamma(u) \to u.$

The function $\gamma(u)$ is nonnegative and increases monotonically toward infinity. Therefore, in numerical work, the integration (A2.25) will be carried over a finite range, $0 \leqq u \leqq U$, only. The degree of approximation will depend, apart from rounding error, on two sorts of errors: first, the error arising from using an approximation rule to compute the integral over the range, $0 \leqq u \leqq U$, and secondly, a truncation error,

(A2.30) $$t_U = \pi^{-1} \int_U^\infty \frac{\sin\theta(u)}{\gamma(u)} du.$$

This truncation error can be bounded above as follows:

(A2.31) $$|t_U| \leqq \pi^{-1} \int_U^\infty \frac{|\sin\theta(u)|}{\gamma(u)} du \leqq \pi^{-1} \int_U^\infty \gamma^{-1}(u) du,$$

(A2.32) $$\gamma(u) = u \prod_{j=1}^{n} (1+\lambda_j^2 u^2)^{1/4} > u \prod_{j=1}^{n} |\lambda_j|^{1/2} u^{1/2} = u^{(1/2)n+1} \prod_{j=1}^{n} |\lambda_j|^{1/2}.$$

Therefore we have

(A2.33) $$|t_U| < (\pi \prod_{j=1}^{n} |\lambda_j|^{1/2})^{-1} \int_U^\infty u^{-(1/2)n-1} du = ((1/2)n\pi U^{(1/2)n} \prod_{j=1}^{n} |\lambda_j|^{(1/2)})^{-1}.$$

The errors can be controlled. Hence, $F(q)$ can be determined to any desired degree of accuracy. A computer program for this procedure was provided by Koerts and Abrahamse [1969, Chapter 9].

### A2.5. Some Theoretical Results

*Independence of Two Quadratic Forms.* Two quadratic forms $\boldsymbol{x'Ax}$ and $\boldsymbol{x'Bx}$ are independent if

(A2.34) $\boldsymbol{AVB} = 0,$

see, for example, Theorem 4.21 of Graybill [1961, p.88].

*Independence of a Linear Form and a Quadratic Form.* A linear form $\boldsymbol{a'x}$ and a quadratic form $\boldsymbol{x'Ax}$ are independent if

(A2.35) $\boldsymbol{a'VA} = 0,$

see, for example, Rao [1973, p.188].

*A Condition for a Quadratic Form to Be Distributed as $\chi^2$.* A necessary and sufficient condition for a quadratic form $\boldsymbol{x'Ax}$ to follow a $\chi^2$ distribution with $r$

degrees of freedom is that $VA$ has $r$ unit characteristic roots, with the rest being zero. See Johnson and Kotz [1970b, p. 178].

**A2.6. Approximations to the Distribution of a Quadratic Form**

Even though some exact computational procedures of the distribution of a quadratic form are available it is useful to have some simple accurate approximations to the distribution. There are a number of approximations to this distribution, see Johnson and Kotz [1970b, Sec. 29.5]. But for some analytical purposes the following approximations are useful.

Patnaik's[1949] two-moment $\chi^2$ approximation specifies approximating the distribution of a positive semidefinite quadratic form, $Q=\boldsymbol{x}'\boldsymbol{A}\boldsymbol{x}$, by

$$Q \stackrel{\cdot}{\sim} c\chi^2(f) \tag{A2.36}$$

with $c$ and $f$ chosen to make the first two moments agree, *i.e.*,

$$\begin{aligned} E(Q)&=tr\,\boldsymbol{VA}=cf \\ Var(Q)&=2tr(\boldsymbol{VA})^2=2c^2f, \end{aligned} \tag{A2.37}$$

which gives

$$\begin{aligned} c&=tr(\boldsymbol{VA})^2/tr\,\boldsymbol{VA} \\ f&=(tr\,\boldsymbol{VA})^2/tr(\boldsymbol{VA})^2. \end{aligned} \tag{A2.38}$$

This approximation gives fairly good results, see, for example, Imhof [1961, Table 1].

Pearson's[1959] three-moment $\chi^2$ approximation specifies approximating the distribution of $Q$ by

$$Q \stackrel{\cdot}{\sim} b+c\chi^2(f) \tag{A2.39}$$

with $b$, $c$ and $f$ chosen so that the first three moments of $Q$ agree, *i.e.*,

$$\begin{aligned} E(Q)&=tr\,\boldsymbol{VA}=b+cf \\ Var(Q)&=2tr(\boldsymbol{VA})^2=2c^2f \\ \mu_3(Q)&=8tr(\boldsymbol{VA})^3=8c^3f, \end{aligned} \tag{A2.40}$$

which gives

$$\begin{aligned} c&=tr(\boldsymbol{VA})^3/tr(\boldsymbol{VA})^2 \\ f&=(tr(\boldsymbol{VA})^2)^3/(tr(\boldsymbol{VA})^3)^2 \\ b&=tr\,\boldsymbol{VA}-(tr(\boldsymbol{VA})^2)^2/tr(\boldsymbol{VA})^3=tr\,\boldsymbol{VA}-cf. \end{aligned} \tag{A2.41}$$

This approximation is much better than the Patnaik's. But the existence of a constant $b$ is an inconvenience for some analytical purposes. The approximation can be applied to an indefinite form of $Q$, but one must assume that $tr(VA)^3$ is positive. Otherwise one approximates the distribution of $-Q$. For the accuracy of this approximation, see Imhof[1961, Sec. 4 and Table 1].

In both Patnaik's and Pearson's approximations, $f$ is usually fractional, so that an interpolation is needed if standard chi-square tables are used. Otherwise we can evaluate the probability using the exact formula,

$$(A2.42)\qquad Pr(\chi^2(f)<x_o)=e^{-\frac{1}{2}x_o}\sum_{j=0}^{\infty}\{(\tfrac{1}{2}x_o)^{\frac{1}{2}f+j}/(\tfrac{1}{2}f+j+1)\},$$

or the Wilson-Hiferty approximation,

$$(A2.43)\qquad Pr(\chi^2(f)<x_o)\doteqdot\Phi(((x_o/f)^{1/3}-1+2/(9f))(9f/2)^{1/2}),$$

where $\Phi(\ )$ is the distribution function of the standard normal variable. This approximation gives very good results, see Johnson and Kotz[1970a, Sec. 17.5].

### A2.7. Approximation to the Distribution of the Ratio of Independent Positive Semidefinite Quadratic Forms

Since the Patnaik approximation is fairly satisfactory we may attempt to approximate the distribution of the ratio of two independent positive semidefinite quadratic forms, $\boldsymbol{x}'\boldsymbol{A}\boldsymbol{x}$ and $\boldsymbol{x}'\boldsymbol{B}\boldsymbol{x}$, by an $F$-distribution constructed by fitting $\chi^2$ distributions in both the numerator and the denominator in the manner of Box [1954], *i.e.*,

$$(A2.44)\qquad (\boldsymbol{x}'\boldsymbol{A}\boldsymbol{x})/(\boldsymbol{x}'\boldsymbol{B}\boldsymbol{x})\backsim dF(f_1,f_2),$$
$$d=tr\,\boldsymbol{VA}/tr\,\boldsymbol{VB}$$

$$(A2.45)\qquad f_1=(tr\,\boldsymbol{VA})^2/tr(\boldsymbol{VA})^2$$
$$f_2=(tr\,\boldsymbol{VB})^2/tr(\boldsymbol{VB})^2.$$

This approximation also gives fairly good results, see, for example, Box [1954].

Usually $f_1$ and $f_2$ are fractional. We can evaluate the probability using the exact formula,

$$(A2.46)\qquad Pr(F(f_1,f_2)<F_o)=I_y((1/2)f_1,(1/2)f_2),$$

where $y=f_1F_0/(f_2+f_1F_0)$ and $I_y(a,b)$ is the incomplete beta function ratio. To save computing time we can make use of the relation, $I_y(a,b)=1-I_{1-y}(b,a)$. We can also evaluate the probability using Paulson's approximation, (A2.47) $$Pr(F(f_1,f_2)<F_0) \doteq \Phi\left\{\frac{(1-g_2)F_0^{1/3}-(1-g_1)}{(F_0^{2/3}g_2+g_1)^{1/2}}\right\}$$

where $\Phi(\ )$ is the distribution function of the standard normal variable, and $g_i=2/(9f_i)$, $i=1,2$. This approximation gives results to three significant figures in most cases, see Johnson and Kotz[1970b, p.83] and Ashby[1968].

### A2.8. Approximation to the Distribution of the Ratio of Independent Linear Form and Square Root of a Quadratic Form

The distribution of the ratio of a linear form $\boldsymbol{a}'\boldsymbol{x}$ to the square root of a positve semidefinite and independent quadratic form $\boldsymbol{x}'A\boldsymbol{x}$ can be approximated by the Student's $t$:

(A2.48) $$\boldsymbol{a}'\boldsymbol{x}/(\boldsymbol{x}'A\boldsymbol{x})^{\frac{1}{2}} \sim dt(f)$$

where

(A2.49) $$f=(trVA)^2/tr(VA)^2$$
$$d=(\boldsymbol{a}'V\boldsymbol{a})/(trVA)^{1/2}.$$

Probably it is Welch[1937] who used the approximation first.

In this case also, the degrees of freedom, $f$, is usually fractional. To evaluate the probability we can use the formula,

(A2.50) $$Pr(|t(f)|<t_0)=I_y(\tfrac{1}{2}, \tfrac{1}{2}f),$$

where $y=t_0^2/(f+t_0^2)$. Naturally, (A2.50) is the special case of (A2.46) when $f_1=1$, $f_2=f$, and $F_0=t_0^2$.

## Appendix III. Explanation of Data

The data sets used for illustrative purposes in the text include four sets

| Name of Data | Source |
|---|---|
| (A) Spirits (truncated) | Durbin and Watson [1951, Table 1] |
| (B) Textile | Theil and Nagar [1961, Table 4] |
| (C) Pears | Henshaw [1966, Table 1] |
| (D) Consumption | Klein [1950, p. 135, pp. 74-76] |
| (E) Trend | (−10 −9……9 10)′ |
| (F) Artificial (1) | (1 −1 1 −1……1 −1)′ |
| (G) Artificial (2) | Combination of trend and (F) |

All data sets contain constant terms.

of published data that have been used in the literature, e.g., Durbin and Watson[1971]. In addition we use one set of "trend" data and two sets of "artificial" data. The various data sets are described above.

## References


Abrahamse, A. P. J. and J. Koerts(1969), "A Comparison between the Power of the Durbin-Watson Test and the Power of the BLUS Test," *Journal of American Statistical Association*, 64, 938—48.

Ames, E. and S. Reiter(1961), "Distributions of Correlation Coefficients in Economic Time Series," *Journal of American Statistical Association*, 56, 637—56.

Anderson, T. W. (1971), *The Statistical Analysis of Time Series*, New York, Wiley.

Ashby, T. (1968), "A Modification to Paulson's Approximation to the Variance Ratio Distribution," *The Computer Journal*, 11, 109—10.

Box, G. E. P. (1954), "Some Theorems on Quardratic Forms Applied in the Study of Analysis of Variance Problems, I. Effect of Inequality of Variance in the One-Way Classification," *Annals of Mathematical Statistics*, 25, 290—303.

Box, G. P. and G. M. Jenkins(1970), *Time Series Analysis, Forecasting and Control*, San Francisco, Holden-Day.

Durbin, J. and G. S. Watson (1950), "Testing for Serial Correlation in Least Squares Regression," *Biometrika*, 37, 409—28.

Durbin, J. and G. S. Watson (1951), "Testing for Serial Correlation in Least Squares Regression," *Biometrika*, 38, 159—78.

Durbin, J. and G. S. Watson (1971), "Testing for Serial Correlation in Least Squares Regression," *Biometrika*, 58, 1—19.

Gil-Pelaez, J. (1951), "Note on the Inversion Theorem," *Biometrika*, 38, 481—82.

Goldberger, A. S. (1964), *Econometric Theory*, New York, Wiley.

Graybill, F. A. (1961), "*An Introduction to Linear Statistical Models*, Vol. I, New York, McGraw-Hill.

Henshaw, Jr., R. C. (1966), "Testing Single-Equation Least Squares Regression Models for Autocorrelated Disturbances," *Econometrica*, 34, 646—60.

Imhof, P. J. (1961), "Computing the Distribution of Quaratic Forms in Normal Variables," *Biometrika*, 48, 419—26.

Jeong, Ki-Jun (1975a), "Diagonal Operator with Its Economic Applications," Claremont Economic Papers, No. 128.

Jeong, Ki-Jun(1975b), "A Graphical Exposition of the Period of Oscillation in a Second Order Difference Equation," Claremont Economic Papers, No. 137.

Johnson, N. L. and S. Kotz(1970a), *Distributions in Statistics Continuous Univariate Distributions*-1, Boston, Houghton Mifflin Co. and New York, Wiley.

Johnson, N. L. and S. Kotz (1970b), *Distributions in Statistics : Continuous Univariate Distributions*-2, Boston, Houghtion Mifflin Co. and New York, Wiley.

Johnston, J. (1972), *Econometric Methods*, 2nd ed., New York, McGraw-Hill.

Klein, L. R. (1950), *Economic Fluctuations in the United States*, 1921—1941, New York, Wiley.

Koerts, J. (1967), "Some Further Notes on Disturbance Estimates in Regression Analysis," *Journal of American Statistical Association*, 62, 169—83.

Koerts, J. and A. P. J. Abrahamse(1968), "On the Power of the BLUS Procedure," *Journal of American Statistical Association*, 63, 1227—36.

Koerts, J. and A. P. J. Abrahamse (1969), *On the Theory and Application of the General Linear Model*, Rotterdam University Press.

Koyck, L. M. (1954), *Distributed Lags and Investment Analysis*, Amsterdam, North-Holland.

Lancaster, H. O. (1954), "Traces and Cumulants of Quadratic Forms in Normal Variables," *Journal of Royal Statistical Society*, Series, B, 16, 247—53.

Lehmann, E. (1959), *Testing Statistical Hypothesis*, New York, Wiley.

Malinvaud, E. (1970), *Statistical Methods oj Econometrics*, 2nd ed., Amsterdam, North-Holland.

Patnaik, P. B. (1949), "The Noncentral Chi-Square and F Distributions and Their Applications," *Biometrika*, 36, 202—32.

Pearson, E. S. (1959), "Note on an Approximation to the Distribution of Non-Central Chi-Square," *Biometrika*, 46, 364.

Rao, C. R. (1973), *Linear Statistical Inference and Its Applications*, 2nd ed., New York, Wiley.

Rao, C. R. and S. K. Mitra(1971), *Generalized Inverse of Matrices and Its Applications*, New York, Wiley.

Theil, H. (1971), *Principles of Econometrics*, New York, Wiley.

Theil, H. and A. L. Nagar(1961), "Testing the Independence of Regression Disturbances," *Journal Of American Statistical Association*, 56, 793—806.

Welch, B. L. (1937), "The Significance of the Difference Between Two Means When the Population Variances Are Unequal," *Biometrika*, 29, 350—62.