Appendices

A

Vectors and Matrices

The following summary of matrix and vector algebra is not meant to be an introduction to the subject but is just a brief review of terms and rules used in the text. Most of them can be found in books such as Graybill (1969), Searle (1982), Anderson (1984, Appendix), Magnus & Neudecker (1988), Magnus (1988) or Lütkepohl (1996a). Therefore proofs or further references are only provided in exceptional cases.

A.1 Basic Definitions

A *matrix* is a rectangular array of numbers. For instance,

$$\begin{bmatrix} 3 & -5 \\ .3 & 0 \end{bmatrix}, \quad (0,1,0), \quad \begin{bmatrix} 3 & 5 & .3 & .3 \\ 2 & 2 & 2 & 2 \end{bmatrix}$$

are matrices. More generally,

$$A = (a_{ij}) = \begin{bmatrix} a_{11} & \dots & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & \dots & a_{mn} \end{bmatrix} \tag{A.1.1}$$

is a matrix with m rows and n columns. Such a matrix is briefly called $(m \times n)$ matrix, m being the *row dimension* and n being the *column dimension*. The numbers a_{ij} are the *elements* or *components* of A. In the following, it is assumed that the elements of all matrices considered are real numbers unless otherwise stated. In other words, we will be concerned with real rather than complex matrices. If the dimensions m and n are clear from the context or if they are of no importance, the notation $A = (a_{ij})$ means that a_{ij} is a *typical element* of A, that is, A consists of elements a_{ij}, $i = 1, \dots, m, j = 1, \dots, n$.

A $(1 \times n)$ matrix is a *row vector* and an $(m \times 1)$ matrix is a *column vector* which is often denoted by a lower case letter in the text. If not otherwise

noted, all vectors will be column vectors in the following. Instead of $(m \times 1)$ matrix we sometimes say $(m \times 1)$ vector or simply *m-vector* or *m-dimensional* vector.

An $(m \times m)$ matrix with the number of rows equal to the number of columns is a *square matrix*. An $(m \times m)$ square matrix

$$\begin{bmatrix} a_{11} & 0 & \dots & 0 \\ 0 & a_{22} & & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \dots & a_{mn} \end{bmatrix}$$

with zeros off the main diagonal is a *diagonal matrix*. If all the diagonal elements of a diagonal matrix are one, it is an *identity* or *unit matrix*. An $(m \times m)$ identity matrix is denoted by I_m or simply by I if the dimension is unimportant or obvious from the context. A square matrix with all elements below (above) the main diagonal being zero is called *upper* (*lower*) *triangular* or simply *triangular matrix*. A matrix consisting of zeros only is a *null matrix* or *zero matrix*. Usually, in this text, such a matrix is simply denoted by 0 and its dimensions have to be figured out from the context.

The *transpose* of the $(m \times n)$ matrix A given in (A.1.1) is the $(n \times m)$ matrix

$$A' = \begin{bmatrix} a_{11} & \dots & a_{m1} \\ \vdots & & \vdots \\ a_{1n} & \dots & a_{mn} \end{bmatrix},$$

the n rows of A' being the n columns of A. The matrix A is *symmetric* if $A' = A$. For instance,

$$\begin{bmatrix} 3 & 3 & 1 \\ 0 & 1 & 0 \end{bmatrix} \quad \text{is the transpose of} \quad \begin{bmatrix} 3 & 0 \\ 3 & 1 \\ 1 & 0 \end{bmatrix}$$

and

$$\begin{bmatrix} 2 & -1 \\ -1 & 0 \end{bmatrix}$$

is a symmetric matrix.

A.2 Basic Matrix Operations

Let $A = (a_{ij})$ and $B = (b_{ij})$ be $(m \times n)$ matrices. The two matrices are *equal*, $A = B$, if $a_{ij} = b_{ij}$ for all i, j. The following matrix operations are basic:

$$A + B := (a_{ij} + b_{ij}). \qquad (addition)$$
$$A - B := (a_{ij} - b_{ij}). \qquad (subtraction)$$

For a real constant c,

$$cA = Ac := (ca_{ij}). \qquad (multiplication\ by\ a\ scalar)$$

Let $C = (c_{ij})$ be an $(n \times r)$ matrix, then the product

$$AC := \left(\sum_{j=1}^{n} a_{ij} c_{jk} \right) \qquad (multiplication)$$

is an $(m \times r)$ matrix. For instance,

$$\begin{bmatrix} 3 & 3 \\ 2 & 1 \end{bmatrix} \begin{bmatrix} 2 & 3 & 0 \\ 2 & 4 & -1 \end{bmatrix} = \begin{bmatrix} 3 \cdot 2 + 3 \cdot 2 & 3 \cdot 3 + 3 \cdot 4 & 3 \cdot 0 - 3 \cdot 1 \\ 2 \cdot 2 + 1 \cdot 2 & 2 \cdot 3 + 1 \cdot 4 & 2 \cdot 0 - 1 \cdot 1 \end{bmatrix}$$
$$= \begin{bmatrix} 12 & 21 & -3 \\ 6 & 10 & -1 \end{bmatrix}.$$

If the column dimension of A is the same as the row dimension of C so that A and C can be multiplied, the two matrices are *conformable*. In the product AC the matrix C is *premultiplied* by A and A is *postmultiplied* by C.

Rules: Suppose A, B, and C are matrices with suitable dimensions so that the following operations are defined and c is a scalar.

(1) $A + B = B + A$.
(2) $(A + B) + C = A + (B + C)$.
(3) $A(B + C) = AB + AC$.
(4) $c(A + B) = cA + cB$.
(5) $AB \neq BA$ in general.
(6) $(AB)C = A(BC)$.
(7) $(AB)' = B'A'$.
(8) $AI = IA = A$.
(9) AA' and $A'A$ are symmetric matrices.

A.3 The Determinant

The *determinant* of an $(m \times m)$ square matrix $A = (a_{ij})$ is the sum of all products

$$(-1)^p a_{1i_1} a_{2i_2} \cdots a_{mi_m}$$

consisting of precisely one element from each row and each column multiplied by -1 or 1, depending on the permutation $i_1, \ldots, i_m$ of the subscripts. The -1

is used if the number of inversions of $i_1, \ldots, i_m$ to obtain the order $1, 2, \ldots, m$ is odd and 1 is used otherwise. The sum is taken over all $m!$ permutations of the column subscripts.

For a (1×1) matrix the determinant equals the value of the single element and for $m > 1$ the determinant may be defined recursively as follows. Suppose

$$A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$$

is a (2×2) matrix. Then the determinant is

$$\det(A) = |A| = a_{11}a_{22} - a_{12}a_{21}. \tag{A.3.1}$$

For instance,

$$\det \begin{bmatrix} 3 & 1 \\ 2 & 2 \end{bmatrix} = 4.$$

To specify the determinant of a general $(m \times m)$ matrix $A = (a_{ij})$ we define the *minor* of the ij-th element a_{ij} as the determinant of the $((m-1) \times (m-1))$ matrix that is obtained by deleting the i-th row and j-th column from A. The *cofactor* of a_{ij}, denoted by A_{ij}, is the minor multiplied by $(-1)^{i+j}$. Now

$$\det(A) = |A| = a_{i1}A_{i1} + \cdots + a_{im}A_{im} = a_{1j}A_{1j} + \cdots + a_{mj}A_{mj} \tag{A.3.2}$$

for any i or $j \in \{1, \ldots, m\}$. It does not matter which row or column is chosen in (A.3.2) because the determinant of a matrix is a unique number.

For example, for the (3×3) matrix

$$A = \begin{bmatrix} 2 & 1 & 3 \\ 0 & 2 & 1 \\ 1 & -1 & 4 \end{bmatrix} \tag{A.3.3}$$

the minor of the upper right-hand corner element is

$$\det \begin{bmatrix} 0 & 2 \\ 1 & -1 \end{bmatrix} = -2.$$

The cofactor is also -2 because $(-1)^{1+3} = 1$. Developing by the first row gives

$$|A| = 2 \cdot \det \begin{bmatrix} 2 & 1 \\ -1 & 4 \end{bmatrix} - 1 \cdot \det \begin{bmatrix} 0 & 1 \\ 1 & 4 \end{bmatrix} + 3 \cdot \det \begin{bmatrix} 0 & 2 \\ 1 & -1 \end{bmatrix} = 13.$$

The same result is obtained by developing by any other row or column, e.g., using the first column gives

$$|A| = 2 \cdot \det \begin{bmatrix} 2 & 1 \\ -1 & 4 \end{bmatrix} - 0 \cdot \det \begin{bmatrix} 1 & 3 \\ -1 & 4 \end{bmatrix} + 1 \cdot \det \begin{bmatrix} 1 & 3 \\ 2 & 1 \end{bmatrix} = 13.$$

Rules: In the following rules, $A = (a_{ij})$ and $B = (b_{ij})$ are $(m \times m)$ matrices and c is a scalar.

(1) $\det(I_m) = 1$.
(2) If A is a diagonal matrix, $\det(A) = a_{11} \cdot a_{22} \cdots a_{mm}$.
(3) If A is a lower or upper triangular matrix, $|A| = a_{11} \cdots a_{mm}$.
(4) If A contains a row or column of zeros, $|A| = 0$.
(5) If B is obtained from A by adding to one row (column) a scalar multiple of another row (column), then $|A| = |B|$.
(6) If A has two identical rows or columns, then $|A| = 0$.
(7) $\det(cA) = c^m \det(A)$.
(8) $|AB| = |A||B|$.
(9) If C is an $(m \times n)$ matrix, $\det(I_m + CC') = \det(I_n + C'C)$.

A.4 The Inverse, the Adjoint, and Generalized Inverses

A.4.1 Inverse and Adjoint of a Square Matrix

An $(m \times m)$ square matrix A is *nonsingular* or *regular* or *invertible* if there exists a unique $(m \times m)$ matrix B such that $AB = I_m$. The matrix B is denoted by A^{-1}. It is the *inverse* of A,

$$AA^{-1} = A^{-1}A = I_m.$$

For $m > 1$, the $(m \times m)$ matrix of cofactors,

$$A^{adj} = \begin{bmatrix} A_{11} & \dots & A_{1m} \\ \vdots & \ddots & \vdots \\ A_{m1} & \dots & A_{mm} \end{bmatrix}'$$

is the *adjoint* of A. For a (1×1) matrix A, we define the adjoint to be 1, that is, $A^{adj} = 1$. To compute the inverse of the $(m \times m)$ matrix A, the relation

$$A^{-1} = |A|^{-1} A^{adj} \tag{A.4.1}$$

is sometimes useful. For this expression to be meaningful, $|A|$ has to be nonzero. Indeed, A is nonsingular if and only if $\det(A) \neq 0$.

As an example consider the matrix given in (A.3.3). Its adjoint is

$$A^{adj} = \begin{bmatrix} \begin{vmatrix} 2 & 1 \\ -1 & 4 \end{vmatrix} & -\begin{vmatrix} 0 & 1 \\ 1 & 4 \end{vmatrix} & \begin{vmatrix} 0 & 2 \\ 1 & -1 \end{vmatrix} \\ -\begin{vmatrix} 1 & 3 \\ -1 & 4 \end{vmatrix} & \begin{vmatrix} 2 & 3 \\ 1 & 4 \end{vmatrix} & -\begin{vmatrix} 2 & 1 \\ 1 & -1 \end{vmatrix} \\ \begin{vmatrix} 1 & 3 \\ 2 & 1 \end{vmatrix} & -\begin{vmatrix} 2 & 3 \\ 0 & 1 \end{vmatrix} & \begin{vmatrix} 2 & 1 \\ 0 & 2 \end{vmatrix} \end{bmatrix}' = \begin{bmatrix} 9 & -7 & -5 \\ 1 & 5 & -2 \\ -2 & 3 & 4 \end{bmatrix}.$$

Consequently,

$$A^{-1} = \frac{1}{13} \begin{bmatrix} 9 & -7 & -5 \\ 1 & 5 & -2 \\ -2 & 3 & 4 \end{bmatrix}.$$

Multiplying this matrix by A is easily seen to result in the (3×3) identity matrix.

Rules:

(1) For an $(m \times m)$ square matrix A, $AA^{adj} = A^{adj}A = |A|I_m$.
(2) An $(m \times m)$ matrix A is nonsingular if and only if $\det(A) \neq 0$.

In the following, $A = (a_{ij})$ and B are nonsingular $(m \times m)$ matrices and $c \neq 0$ is a scalar constant.

(3) $A^{-1} = A^{adj}/|A|$.
(4) $(A')^{-1} = (A^{-1})'$.
(5) $(AB)^{-1} = B^{-1}A^{-1}$.
(6) $(cA)^{-1} = \frac{1}{c}A^{-1}$.
(7) $I_m^{-1} = I_m$.
(8) If A is a diagonal matrix, then A^{-1} is also diagonal with diagonal elements $1/a_{ii}$.
(9) For an $(m \times n)$ matrix C, $(I_m + CC')^{-1} = I_m - C(I_n + C'C)^{-1}C'$.

A.4.2 Generalized Inverses

Let A be an $(m \times n)$ matrix. Any matrix B satisfying $ABA = A$ is a *generalized inverse* of A. For example, if

$$A = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix},$$

the following matrices are generalized inverses of A:

$$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \quad \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \quad \begin{bmatrix} 1 & 0 \\ 0 & \frac{1}{2} \end{bmatrix}.$$

Obviously, a generalized inverse is not unique in general. An $(n \times m)$ matrix B is called *Moore-Penrose* (*generalized*) *inverse* of A if it satisfies the following four conditions:

$$\begin{aligned} &ABA = A, \\ &BAB = B, \\ &(AB)' = AB, \\ &(BA)' = BA. \end{aligned} \tag{A.4.2}$$

The Moore-Penrose inverse of A is denoted by A^+, it exists for any $(m \times n)$ matrix and is unique.

Rules: (See Magnus & Neudecker (1988, p. 33, Theorem 5).)

(1) $A^+ = A^{-1}$ if A is nonsingular.
(2) $(A^+)^+ = A$.
(3) $(A')^+ = (A^+)'$.
(4) $A'AA^+ = A^+AA' = A'$.
(5) $A'A^{+\prime}A^+ = A^+A^{+\prime}A' = A^+$.
(6) $(A'A)^+ = A^+A^{+\prime}, (AA')^+ = A^{+\prime}A^+$.
(7) $A^+ = (A'A)^+A' = A'(AA')^+$.

A.5 The Rank

Let $x_1, \ldots, x_n$ be $(m \times 1)$ vectors. They are *linearly independent* if, for the constants $c_1, \ldots, c_n$,

$$c_1x_1 + \cdots + c_nx_n = 0$$

implies $c_1 = \cdots = c_n = 0$. Equivalently, defining the $(n \times 1)$ vector $c = (c_1, \ldots, c_n)'$ and the $(m \times n)$ matrix $X = (x_1, \ldots, x_n)$, the columns of X are linearly independent if $Xc = 0$ implies $c = 0$. The columns of X are *linearly dependent* if $c_1x_1 + \cdots + c_nx_n = 0$ holds with at least one $c_i \neq 0$. In that case,

$$x_i = d_1x_1 + \cdots + d_{i-1}x_{i-1} + d_{i+1}x_{i+1} + \cdots + d_nx_n,$$

where $d_j = -c_j/c_i$. In other words, $x_1, \ldots, x_n$ are linearly dependent if at least one of the vectors is a linear combination of the other vectors.

If $n > m$, the columns of X are linearly dependent. Consequently, if $x_1, \ldots, x_n$ are linearly independent, then $n \leq m$.

Let $a_1, \ldots, a_n$ be the columns of the $(m \times n)$ matrix $A = (a_1, \ldots, a_n)$. That is, the a_i are $(m \times 1)$ vectors. The *rank* of A, briefly $\mathrm{rk}(A)$, is the maximum number of linearly independent columns of A. Thus, if $n \leq m$ and the $a_1, \ldots, a_n$ are linearly independent, $\mathrm{rk}(A) = n$. The maximum number of linearly independent columns of A equals the maximum number of linearly independent rows. Hence, the rank may be defined equivalently as the maximum number of linearly independent rows. If $m \geq n$ $(m \leq n)$ then we say that A has *full rank* if $\mathrm{rk}(A) = n$ $(\mathrm{rk}(A) = m)$.

Rules: Let A be an $(m \times n)$ matrix.

(1) $\mathrm{rk}(A) \leq \min(m, n)$.
(2) $\mathrm{rk}(A) = \mathrm{rk}(A')$.
(3) $\mathrm{rk}(AA') = \mathrm{rk}(A'A) = \mathrm{rk}(A)$.
(4) If B is a nonsingular $(n \times n)$ matrix, then $\mathrm{rk}(AB) = \mathrm{rk}(A)$.

(5) If $\mathrm{rk}(A) = m$, then $A^+ = A'(AA')^{-1}$.
(6) If $\mathrm{rk}(A) = n$, then $A^+ = (A'A)^{-1}A'$.
(7) If B is an $(n \times r)$ matrix, $\mathrm{rk}(AB) \leq \min\{\mathrm{rk}(A), \mathrm{rk}(B)\}$.
(8) If A is $(m \times m)$, then $\mathrm{rk}(A) = m$ if and only if $|A| \neq 0$.

A.6 Eigenvalues and -vectors – Characteristic Values and Vectors

The *eigenvalues* or *characteristic values* or *characteristic roots* of an $(m \times m)$ square matrix A are the roots of the polynomial in λ given by $\det(A - \lambda I_m)$ or $\det(\lambda I_m - A)$. The determinant is sometimes called the *characteristic determinant* and the polynomial is called the *characteristic polynomial* of A. Because the roots of a polynomial are complex numbers, the eigenvalues are also complex in general. A number λ_i is an eigenvalue of A, if the columns of $(A - \lambda_i I_m)$ are linearly dependent. Consequently, there exists an $(m \times 1)$ vector $v_i \neq 0$ such that

$$(A - \lambda_i I_m)v_i = 0 \quad \text{or} \quad Av_i = \lambda_i v_i.$$

A vector with this property is an *eigenvector* or *characteristic vector* of A associated with the eigenvalue λ_i. Of course, any nonzero scalar multiple of v_i is also an eigenvector of A associated with λ_i.

As an example consider the matrix

$$A = \begin{bmatrix} 1 & 0 \\ 1 & 3 \end{bmatrix}.$$

Its eigenvalues are the roots of

$$|A - \lambda I_2| = \det \begin{bmatrix} 1-\lambda & 0 \\ 1 & 3-\lambda \end{bmatrix} = (1-\lambda)(3-\lambda).$$

Hence, $\lambda_1 = 1$ and $\lambda_2 = 3$ are the eigenvalues of A. Associated eigenvectors are obtained by solving

$$\begin{bmatrix} 1 & 0 \\ 1 & 3 \end{bmatrix} \begin{bmatrix} v_{11} \\ v_{21} \end{bmatrix} = \begin{bmatrix} v_{11} \\ v_{21} \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} 1 & 0 \\ 1 & 3 \end{bmatrix} \begin{bmatrix} v_{12} \\ v_{22} \end{bmatrix} = 3 \begin{bmatrix} v_{12} \\ v_{22} \end{bmatrix}.$$

Thus,

$$\begin{bmatrix} v_{11} \\ v_{21} \end{bmatrix} = \begin{bmatrix} 1 \\ -\frac{1}{2} \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} v_{12} \\ v_{22} \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

are eigenvectors of A associated with λ_1 and λ_2, respectively.

In the following rules, the *modulus* of a complex number $z = z_1 + iz_2$ is used. Here z_1 and z_2 are the real and imaginary parts of z, respectively, and $i = \sqrt{-1}$. The modulus $|z|$ of z is defined as

$$|z| := \sqrt{z_1^2 + z_2^2}.$$

If $z_2 = 0$ so that z is a real number, the modulus is just the absolute value of z, which justifies the notation.

Rules:

(1) If A is symmetric, then all its eigenvalues are real numbers.
(2) The eigenvalues of a diagonal matrix are its diagonal elements.
(3) The eigenvalues of a triangular matrix are its diagonal elements.
(4) An $(m \times m)$ matrix has at most m eigenvalues.
(5) Let $\lambda_1, \ldots, \lambda_m$ be the eigenvalues of the $(m \times m)$ matrix A, then $|A| = \lambda_1 \cdots \lambda_m$, that is, the determinant is the product of the eigenvalues.
(6) Let λ_i and λ_j be *distinct* eigenvalues of A with associated eigenvectors v_i and v_j. Then v_i and v_j are linearly independent.
(7) All eigenvalues of the $(m \times m)$ matrix A have modulus less than 1 if and only if $\det(I_m - Az) \neq 0$ for $|z| \leq 1$, that is, the polynomial $\det(I_m - Az)$ has no roots in and on the complex unit circle.

A.7 The Trace

The *trace* of an $(m \times m)$ square matrix $A = (a_{ij})$ is the sum of its diagonal elements,

$$\operatorname{tr} A = \operatorname{tr}(A) := a_{11} + \cdots + a_{mm}.$$

For example,

$$\operatorname{tr}\begin{bmatrix} 1 & 0 \\ 1 & 3 \end{bmatrix} = 4.$$

Rules: A and B are $(m \times m)$ matrices and $\lambda_1, \ldots, \lambda_m$ are the eigenvalues of A.

(1) $\operatorname{tr}(A + B) = \operatorname{tr}(A) + \operatorname{tr}(B)$.
(2) $\operatorname{tr} A = \operatorname{tr} A'$.
(3) If C is $(m \times n)$ and D is $(n \times m)$, $\operatorname{tr}(CD) = \operatorname{tr}(DC)$.
(4) $\operatorname{tr} A = \lambda_1 + \cdots + \lambda_m$.

A.8 Some Special Matrices and Vectors

A.8.1 Idempotent and Nilpotent Matrices

An $(m \times m)$ matrix A is *idempotent* if $AA = A^2 = A$. Examples of idempotent matrices are $A = I_m$, $A = 0$, and

$$A = \begin{bmatrix} \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \end{bmatrix}.$$

An $(m \times m)$ matrix A is *nilpotent* if there exists a positive integer i such that $A^i = 0$. For instance, the (2×2) matrices

$$A = \begin{bmatrix} 0 & 3 \\ 0 & 0 \end{bmatrix} \quad \text{and} \quad B = \begin{bmatrix} 1 & -1 \\ 1 & -1 \end{bmatrix}$$

are nilpotent because $A^2 = B^2 = 0$.

Rules: In the following rules, A is an $(m \times m)$ matrix.

(1) If A is a diagonal matrix, it is idempotent if and only if all the diagonal elements are either zero or one.
(2) If A is symmetric and idempotent, $\mathrm{rk}(A) = \mathrm{tr}(A)$.
(3) If A is idempotent and $\mathrm{rk}(A) = m$, then $A = I_m$.
(4) If A is idempotent, then $I_m - A$ is idempotent.
(5) If A is symmetric and idempotent, then $A^+ = A$.
(6) If B is an $(m \times n)$ matrix, then BB^+ and B^+B are idempotent.
(7) If A is idempotent, then all its eigenvalues are zero or one.
(8) If A is nilpotent, then all its eigenvalues are zero.

A.8.2 Orthogonal Matrices and Vectors and Orthogonal Complements

Two $(m \times 1)$ *vectors* x and y are *orthogonal* if $x'y = 0$. They are *orthonormal* if they are orthogonal and have unit length, where the length of a vector x is $||x|| := \sqrt{x'x}$.

An $(m \times k)$ matrix B is *orthogonal* to the $(m \times n)$ matrix A if $A'B = 0$. If A is an $(m \times n)$ matrix of full column rank, an *orthogonal complement* of A, denoted by $A_\perp$, is an $(m \times (m - n))$ matrix of full column rank such that $A'A_\perp = 0$. The orthogonal complement of a nonsingular square matrix is zero and the orthogonal complement of a zero matrix is an identity matrix of suitable dimension.

An $(m \times m)$ square *matrix* A is *orthogonal* if its transpose is its inverse, $A'A = AA' = I_m$. In other words, A is orthogonal if its rows and columns are orthonormal vectors.

Examples of orthogonal vectors are

$$x = \begin{bmatrix} 5 \\ 0 \\ 0 \end{bmatrix} \quad \text{and} \quad y = \begin{bmatrix} 0 \\ 2 \\ 0 \end{bmatrix}.$$

The following four matrices are orthogonal matrices:

$$\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \quad \begin{bmatrix} \cos\varphi & \sin\varphi \\ -\sin\varphi & \cos\varphi \end{bmatrix},$$

$$\begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad \begin{bmatrix} 1/\sqrt{3} & 1/\sqrt{3} & 1/\sqrt{3} \\ 1/\sqrt{2} & -1/\sqrt{2} & 0 \\ 1/\sqrt{6} & 1/\sqrt{6} & -2/\sqrt{6} \end{bmatrix}.$$

Suppose

$$A = \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 0 & 2 \end{bmatrix},$$

then

$$\begin{bmatrix} 1 \\ -1 \\ \frac{1}{2} \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} 2 \\ -2 \\ 1 \end{bmatrix}$$

are orthogonal complements of A.

Rules:

(1) I_m is an orthogonal matrix.
(2) If A is an orthogonal matrix, then $\det(A) = 1$ or -1.
(3) If A and B are orthogonal and conformable matrices, then AB is orthogonal.
(4) If λ_i, λ_j are distinct eigenvalues of a *symmetric* matrix A, then the corresponding eigenvectors v_i and v_j are orthogonal.
(5) For an $(m \times n)$ matrix A of full column rank and $n < m$, the matrix $[A : A_\perp]$ is invertible.

A.8.3 Definite Matrices and Quadratic Forms

Let A be a *symmetric* $(m \times m)$ matrix and x an $(m \times 1)$ vector. The function $x'Ax$ is called a *quadratic form* in x. The symmetric matrix A or the corresponding quadratic form is

(i) *positive definite* if $x'Ax > 0$ for all m-vectors $x \neq 0$;
(ii) *positive semidefinite* if $x'Ax \geq 0$ for all m-vectors x;
(iii) *negative definite* if $x'Ax < 0$ for all m-vectors $x \neq 0$;
(iv) *negative semidefinite* if $x'Ax \leq 0$ for all m-vectors x;
(v) *indefinite* if $x'Ax > 0$ for some x and $x'Ax < 0$ for another x.

Rules: In the following rules, A is a symmetric $(m \times m)$ matrix.

(1) $A = (a_{ij})$ is positive definite if and only if all its principle minors are positive, where

$$\det \begin{bmatrix} a_{11} & \dots & a_{1i} \\ \vdots & \ddots & \vdots \\ a_{i1} & \dots & a_{ii} \end{bmatrix}$$

is the i-th principle minor of A.
(2) A is negative definite (semidefinite) if and only if $-A$ is positive definite (semidefinite).
(3) If A is positive or negative definite, it is nonsingular.
(4) All eigenvalues of a positive (negative) definite matrix are greater (smaller) than zero.
(5) A diagonal matrix is positive (negative) definite if and only if all its diagonal elements are positive (negative).
(6) If A is positive definite and B an $(m \times n)$ matrix, then $B'AB$ is positive semidefinite.
(7) If A is positive definite and B an $(m \times n)$ matrix with $\mathrm{rk}(B) = n$, then $B'AB$ is positive definite.
(8) If A is positive definite, then A^{-1} is positive definite.
(9) If A is idempotent, then it is positive semidefinite.

With these rules it is easy to check that

$$\begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} 3 & 1 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 4 \end{bmatrix}$$

are positive definite matrices and

$$\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$$

are positive semidefinite matrices.

A.9 Decomposition and Diagonalization of Matrices

A.9.1 The Jordan Canonical Form

Let A be an $(m \times m)$ matrix with eigenvalues $\lambda_1, \dots, \lambda_n$. Then there exists a nonsingular matrix P such that

$$P^{-1}AP = \begin{bmatrix} \Lambda_1 & & 0 \\ & \ddots & \\ 0 & & \Lambda_n \end{bmatrix} =: \Lambda \quad \text{or} \quad A = P\Lambda P^{-1}, \tag{A.9.1}$$

where

$$\Lambda_i = \begin{bmatrix} \lambda_i & 1 & 0 & \dots & 0 \\ 0 & \lambda_i & 1 & & 0 \\ \vdots & & \ddots & \ddots & \vdots \\ 0 & 0 & & \ddots & 1 \\ 0 & 0 & \dots & \dots & \lambda_i \end{bmatrix}.$$

This decomposition of A is the *Jordan canonical form*. Because the eigenvalues of A may be complex numbers, Λ and P may be complex matrices. If multiple roots of the characteristic polynomial exist, they may have to appear more than once in the list $\lambda_1, \dots, \lambda_n$.

The Jordan canonical form has some important implications. For instance, it implies that

$$A^j = (P\Lambda P^{-1})^j = P\Lambda^j P^{-1}$$

and it can be shown that

$$\Lambda_i^j = \begin{bmatrix} \lambda_i^j & \binom{j}{1}\lambda_i^{j-1} & \dots & \binom{j}{r_i-1}\lambda_i^{j-r_i+1} \\ 0 & \lambda_i^j & \dots & \binom{j}{r_i-2}\lambda_i^{j-r_i+2} \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \dots & \lambda_i^j \end{bmatrix},$$

where

$$\binom{p}{q} = \frac{p!}{(p-q)!q!}$$

denotes a binomial coefficient. We have the following rules.

Rules: Suppose A is a real $(m \times m)$ matrix with eigenvalues $\lambda_1, \dots, \lambda_n$ which have all modulus less than 1, that is, $|\lambda_i| < 1$ for $i = 1, \dots, n$. Furthermore, let Λ and P be the matrices given in (A.9.1).

(1) $A^j = P\Lambda^j P^{-1} \underset{j\to\infty}{\longrightarrow} 0$.

(2) $\sum_{j=0}^{\infty} A^j = (I_m - A)^{-1}$ exists.

(3) The sequence A^j, $j = 0, 1, 2, \dots$, is absolutely summable, that is,

$$\sum_{j=0}^{\infty} |\alpha_{kl,j}|$$

is finite for all $k, l = 1, \dots, m$, where $\alpha_{kl,j}$ is a typical element of A^j. (See Section C.3 regarding the concept of absolute summability.)

A.9.2 Decomposition of Symmetric Matrices

If A is a symmetric $(m \times m)$ matrix, then there exists an orthogonal matrix P such that

$$P'AP = \Lambda = \begin{bmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_m \end{bmatrix} \quad \text{or} \quad A = P\Lambda P', \tag{A.9.2}$$

where the λ_i's are the eigenvalues of A and the columns of P are the corresponding eigenvectors. Here all matrices are real again because the eigenvalues of a symmetric matrix are real numbers. Denoting the i-th column of P by p_i and using that $p_i'p_j = 0$ for $i \neq j$, we get

$$A = P\Lambda P' = \sum_{i=1}^{m} \lambda_i p_i p_i'. \tag{A.9.3}$$

Moreover,

$$A^2 = P\Lambda P' P\Lambda P' = P\Lambda^2 P'$$

and, more generally,

$$A^k = P\Lambda^k P'.$$

If A is a positive definite symmetric $(m \times m)$ matrix, then all eigenvalues are positive so that the notation

$$\Lambda^{1/2} := \begin{bmatrix} \sqrt{\lambda_1} & & 0 \\ & \ddots & \\ 0 & & \sqrt{\lambda_m} \end{bmatrix}$$

makes sense. Defining $Q = P\Lambda^{1/2}P'$, we get $QQ = A$. In generalization of the terminology for positive real numbers, Q may be called a *square root* of A and may be denoted by $A^{1/2}$.

A.9.3 The Choleski Decomposition of a Positive Definite Matrix

If A is a positive definite $(m \times m)$ matrix, then there exists a lower (upper) triangular matrix P with positive main diagonal such that

$$P^{-1}AP'^{-1} = I_m \quad \text{or} \quad A = PP'. \tag{A.9.4}$$

Similarly, if A is positive semidefinite with $\mathrm{rk}(A) = n < m$, then there exists a nonsingular matrix P such that

$$P^{-1}AP'^{-1} = \begin{bmatrix} I_n & 0 \\ 0 & 0 \end{bmatrix}. \tag{A.9.5}$$

Alternatively, $A = QQ'$, where

$$Q = P\begin{bmatrix} I_n & 0 \\ 0 & 0 \end{bmatrix}.$$

For instance,

$$\begin{bmatrix} 26 & 3 & 0 \\ 3 & 9 & 0 \\ 0 & 0 & 81 \end{bmatrix} = \begin{bmatrix} 5 & 1 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 9 \end{bmatrix}\begin{bmatrix} 5 & 0 & 0 \\ 1 & 3 & 0 \\ 0 & 0 & 9 \end{bmatrix}$$
$$= \begin{bmatrix} \sqrt{26} & 0 & 0 \\ 3/\sqrt{26} & 15/\sqrt{26} & 0 \\ 0 & 0 & 9 \end{bmatrix}\begin{bmatrix} \sqrt{26} & 3/\sqrt{26} & 0 \\ 0 & 15/\sqrt{26} & 0 \\ 0 & 0 & 9 \end{bmatrix}.$$

The decomposition $A = PP'$, where P is lower triangular with positive main diagonal, is sometimes called *Choleski decomposition*. Computer programs are available to determine the matrix P for a given positive definite matrix A. If a lower triangular matrix P is supplied by the program, an upper triangular matrix Q can be obtained as follows: Define an $(m \times m)$ matrix

$$G = \begin{bmatrix} 0 & \dots & 0 & 1 \\ 0 & \dots & 1 & 0 \\ \vdots & \ddots & & \vdots \\ 1 & & 0 & 0 \end{bmatrix}$$

with ones on the diagonal from the upper right-hand corner to the lower left-hand corner and zeros elsewhere. Note that $G' = G$ and $G^{-1} = G$. Suppose a decomposition of the $(m \times m)$ matrix A is desired. Then decompose $B = GAG$ as $B = PP'$, where P is lower triangular. Hence,

$$A = GBG = GPGGP'G = QQ',$$

where $Q = GPG$ is upper triangular.

A.10 Partitioned Matrices

Let the $(m \times n)$ matrix A be *partitioned* into *submatrices* $A_{11}, A_{12}, A_{21}, A_{22}$ with dimensions $(p \times q), (p \times (n-q)), ((m-p) \times q)$, and $((m-p) \times (n-q))$, respectively, so that

$$A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}. \tag{A.10.1}$$

For such a partitioned matrix, a number of useful results hold.

Rules:

(1) $A' = \begin{bmatrix} A'_{11} & A'_{21} \\ A'_{12} & A'_{22} \end{bmatrix}$.

(2) If $n = m$ and $q = p$ and A, A_{11}, and A_{22} are nonsingular, then

$$\begin{aligned} A^{-1} &= \begin{bmatrix} D & -DA_{12}A_{22}^{-1} \\ -A_{22}^{-1}A_{21}D & A_{22}^{-1} + A_{22}^{-1}A_{21}DA_{12}A_{22}^{-1} \end{bmatrix} \\ &= \begin{bmatrix} A_{11}^{-1} + A_{11}^{-1}A_{12}GA_{21}A_{11}^{-1} & -A_{11}^{-1}A_{12}G \\ -GA_{21}A_{11}^{-1} & G \end{bmatrix}, \end{aligned}$$

where $D := (A_{11} - A_{12}A_{22}^{-1}A_{21})^{-1}$ and $G := (A_{22} - A_{21}A_{11}^{-1}A_{12})^{-1}$.

(3) Under the conditions of (2),

$$(A_{11} - A_{12}A_{22}^{-1}A_{21})^{-1} = A_{11}^{-1} + A_{11}^{-1}A_{12}(A_{22} - A_{21}A_{11}^{-1}A_{12})^{-1}A_{21}A_{11}^{-1}.$$

(4) Under the conditions of (2), if A_{12} and A_{21} are null matrices,

$$A^{-1} = \begin{bmatrix} A_{11}^{-1} & 0 \\ 0 & A_{22}^{-1} \end{bmatrix}.$$

(5) If A is a square matrix ($n = m$) and A_{11} is square and nonsingular, then $|A| = |A_{11}| \cdot |A_{22} - A_{21}A_{11}^{-1}A_{12}|$.

(6) If A is a square matrix and A_{22} is square and nonsingular, then $|A| = |A_{22}| \cdot |A_{11} - A_{12}A_{22}^{-1}A_{21}|$.

A.11 The Kronecker Product

Let $A = (a_{ij})$ and $B = (b_{ij})$ be $(m \times n)$ and $(p \times q)$ matrices, respectively. The $(mp \times nq)$ matrix

$$A \otimes B := \begin{bmatrix} a_{11}B & \dots & a_{1n}B \\ \vdots & & \vdots \\ a_{m1}B & \dots & a_{mn}B \end{bmatrix} \tag{A.11.1}$$

is the *Kronecker product* or *direct product* of A and B. For example, the Kronecker product of

$$A = \begin{bmatrix} 3 & 4 & -1 \\ 2 & 0 & 0 \end{bmatrix} \quad \text{and} \quad B = \begin{bmatrix} 5 & -1 \\ 3 & 3 \end{bmatrix} \tag{A.11.2}$$

is

$$A \otimes B = \begin{bmatrix} 15 & -3 & 20 & -4 & -5 & 1 \\ 9 & 9 & 12 & 12 & -3 & -3 \\ 10 & -2 & 0 & 0 & 0 & 0 \\ 6 & 6 & 0 & 0 & 0 & 0 \end{bmatrix}$$

and

$$B \otimes A = \begin{bmatrix} 15 & 20 & -5 & -3 & -4 & 1 \\ 10 & 0 & 0 & -2 & 0 & 0 \\ 9 & 12 & -3 & 9 & 12 & -3 \\ 6 & 0 & 0 & 6 & 0 & 0 \end{bmatrix}.$$

Rules: In the following rules, suitable dimensions are assumed.

(1) $A \otimes B \neq B \otimes A$ in general.
(2) $(A \otimes B)' = A' \otimes B'$.
(3) $A \otimes (B + C) = A \otimes B + A \otimes C$.
(4) $(A \otimes B)(C \otimes D) = AC \otimes BD$.
(5) If A and B are invertible, then $(A \otimes B)^{-1} = A^{-1} \otimes B^{-1}$.
(6) If A and B are square matrices with eigenvalues λ_A, λ_B, respectively, and associated eigenvectors v_A, v_B, then $\lambda_A \lambda_B$ is an eigenvalue of $A \otimes B$ with eigenvector $v_A \otimes v_B$.
(7) If A and B are $(m \times m)$ and $(n \times n)$ square matrices, respectively, then $|A \otimes B| = |A|^n |B|^m$.
(8) If A and B are square matrices,

$$\operatorname{tr}(A \otimes B) = \operatorname{tr}(A)\operatorname{tr}(B).$$

(9) $(A \otimes B)^+ = A^+ \otimes B^+$.

A.12 The vec and vech Operators and Related Matrices

A.12.1 The Operators

Let $A = (a_1, \ldots, a_n)$ be an $(m \times n)$ matrix with $(m \times 1)$ columns a_i. The *vec operator* transforms A into an $(mn \times 1)$ vector by stacking the columns, that is,

$$\operatorname{vec}(A) = \begin{bmatrix} a_1 \\ \vdots \\ a_n \end{bmatrix}.$$

For instance, if A and B are as in (A.11.2), then

$$\operatorname{vec}(A) = \begin{bmatrix} 3 \\ 2 \\ 4 \\ 0 \\ -1 \\ 0 \end{bmatrix} \quad \text{and} \quad \operatorname{vec}(B) = \begin{bmatrix} 5 \\ 3 \\ -1 \\ 3 \end{bmatrix}.$$

Rules: Let A, B, C be matrices with appropriate dimensions.

(1) $\mathrm{vec}(A+B) = \mathrm{vec}(A) + \mathrm{vec}(B)$.
(2) $\mathrm{vec}(ABC) = (C' \otimes A)\mathrm{vec}(B)$.
(3) $\mathrm{vec}(AB) = (I \otimes A)\mathrm{vec}(B) = (B' \otimes I)\mathrm{vec}(A)$.
(4) $\mathrm{vec}(ABC) = (I \otimes AB)\mathrm{vec}(C) = (C'B' \otimes I)\mathrm{vec}(A)$.
(5) $\mathrm{vec}(B')'\mathrm{vec}(A) = \mathrm{tr}(BA) = \mathrm{tr}(AB) = \mathrm{vec}(A')'\mathrm{vec}(B)$.
(6) $\begin{aligned}\mathrm{tr}(ABC) &= \mathrm{vec}(A')'(C' \otimes I)\mathrm{vec}(B) \\ &= \mathrm{vec}(A')'(I \otimes B)\mathrm{vec}(C) \\ &= \mathrm{vec}(B')'(A' \otimes I)\mathrm{vec}(C) \\ &= \mathrm{vec}(B')'(I \otimes C)\mathrm{vec}(A) \\ &= \mathrm{vec}(C')'(B' \otimes I)\mathrm{vec}(A) \\ &= \mathrm{vec}(C')'(I \otimes A)\mathrm{vec}(B).\end{aligned}$

The vech operator is closely related to vec. It only stacks the elements on and below the main diagonal of a square matrix. For instance,

$$\mathrm{vech}\begin{bmatrix} \alpha_{11} & \alpha_{12} & \alpha_{13} \\ \alpha_{21} & \alpha_{22} & \alpha_{23} \\ \alpha_{31} & \alpha_{32} & \alpha_{33} \end{bmatrix} = \begin{bmatrix} \alpha_{11} \\ \alpha_{21} \\ \alpha_{31} \\ \alpha_{22} \\ \alpha_{32} \\ \alpha_{33} \end{bmatrix}.$$

In general, if A is an $(m \times m)$ matrix, $\mathrm{vech}(A)$ is an $m(m+1)/2$-dimensional vector. The vech operator is usually applied to symmetric matrices to collect the separate elements only.

A.12.2 Elimination, Duplication, and Commutation Matrices

The vec and vech operators are related by the *elimination matrix*, $\mathbf{L}_m$, and the *duplication matrix*, $\mathbf{D}_m$. The former is an $(\frac{1}{2}m(m+1) \times m^2)$ matrix such that, for an $(m \times m)$ square matrix A,

$$\mathrm{vech}(A) = \mathbf{L}_m \mathrm{vec}(A). \tag{A.12.1}$$

Thus, e.g., for $m = 3$,

$$\mathbf{L}_3 = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}.$$

The duplication matrix $\mathbf{D}_m$ is $(m^2 \times \frac{1}{2}m(m+1))$ and is defined so that, for any symmetric $(m \times m)$ matrix A,

$$\mathrm{vec}(A) = \mathbf{D}_m \mathrm{vech}(A). \tag{A.12.2}$$

For instance, for $m = 3$,

$$\mathbf{D}_3 = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}.$$

Because the rank of $\mathbf{D}_m$ is easily seen to be $m(m+1)/2$, the matrix $\mathbf{D}'_m\mathbf{D}_m$ is invertible. Thus, left-multiplication of (A.12.2) by $(\mathbf{D}'_m\mathbf{D}_m)^{-1}\mathbf{D}'_m$ gives

$$(\mathbf{D}'_m\mathbf{D}_m)^{-1}\mathbf{D}'_m\text{vec}(A) = \text{vech}(A). \tag{A.12.3}$$

Note, however, that $(\mathbf{D}'_m\mathbf{D}_m)^{-1}\mathbf{D}'_m \neq \mathbf{L}_m$ in general because (A.12.3) holds for symmetric matrices A only while (A.12.1) holds for arbitrary square matrices A.

The *commutation matrix*, $\mathbf{K}_{mn}$, is another matrix that is occasionally useful in dealing with the vec operator. $\mathbf{K}_{mn}$ is an $(mn \times mn)$ matrix defined such that, for any $(m \times n)$ matrix A,

$$\text{vec}(A') = \mathbf{K}_{mn}\text{vec}(A)$$

or, equivalently,

$$\text{vec}(A) = \mathbf{K}_{nm}\text{vec}(A').$$

For example,

$$\mathbf{K}_{32} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

because for

$$A = \begin{bmatrix} \alpha_{11} & \alpha_{12} \\ \alpha_{21} & \alpha_{22} \\ \alpha_{31} & \alpha_{32} \end{bmatrix},$$

$$\text{vec}(A') = \begin{bmatrix} \alpha_{11} \\ \alpha_{12} \\ \alpha_{21} \\ \alpha_{22} \\ \alpha_{31} \\ \alpha_{32} \end{bmatrix} = \mathbf{K}_{32}\begin{bmatrix} \alpha_{11} \\ \alpha_{21} \\ \alpha_{31} \\ \alpha_{12} \\ \alpha_{22} \\ \alpha_{32} \end{bmatrix} = \mathbf{K}_{32}\text{vec}(A).$$

Rules:

(7) $\mathbf{L}_m\mathbf{D}_m = I_{m(m+1)/2}$.
(8) $\mathbf{K}_{mm}\mathbf{D}_m = \mathbf{D}_m$.
(9) $\mathbf{K}_{m1} = \mathbf{K}_{1m} = I_m$.
(10) $\mathbf{K}'_{mn} = \mathbf{K}^{-1}_{mn} = \mathbf{K}_{nm}$.
(11) $\text{tr}\ \mathbf{K}_{mm} = m$.
(12) $\det(\mathbf{K}_{mn}) = (-1)^{mn(m-1)(n-1)/4}$.
(13) $\text{tr}(\mathbf{D}'_m\mathbf{D}_m) = m^2$, $\text{tr}(\mathbf{D}'_m\mathbf{D}_m)^{-1} = m(m+3)/4$.
(14) $\det(\mathbf{D}'_m\mathbf{D}_m) = 2^{m(m-1)/2}$.
(15) $\text{tr}(\mathbf{D}_m\mathbf{D}'_m) = m^2$.
(16) $|\mathbf{D}'_m(A \otimes A)\mathbf{D}_m| = 2^{m(m-1)/2}|A|^{m+1}$, where A is an $(m \times m)$ matrix.
(17) $(\mathbf{D}'_m(A \otimes A)\mathbf{D}_m)^{-1} = (\mathbf{D}'_m\mathbf{D}_m)^{-1}\mathbf{D}'_m(A^{-1} \otimes A^{-1})\mathbf{D}_m(\mathbf{D}'_m\mathbf{D}_m)^{-1}$, if A is a nonsingular $(m \times m)$ matrix.
(18) $\mathbf{L}_m\mathbf{L}'_m = I_{m(m+1)/2}$.
(19) $\mathbf{L}_m\mathbf{L}'_m$ and $\mathbf{L}_m\mathbf{K}_{mm}\mathbf{L}'_m$ are idempotent.

Let A and B be lower triangular $(m \times m)$ matrices. Then we have the following rules:

(20) $\mathbf{L}_m(A \otimes B)\mathbf{L}'_m$ is lower triangular.
(21) $\mathbf{L}'_m\mathbf{L}_m(A' \otimes B)\mathbf{L}'_m = (A' \otimes B)\mathbf{L}'_m$.
(22) $[\mathbf{L}_m(A' \otimes B)\mathbf{L}'_m]^s = \mathbf{L}_m((A')^s \otimes B^s)\mathbf{L}'_m$ for $s = 0, 1, \ldots$ and for $s = \ldots, -2, -1$, if A^{-1} and B^{-1} exist.

Let G be $(m \times n)$, F $(p \times q)$, and $\mathbf{b}$ $(p \times 1)$. Then the following results hold:

(23) $\mathbf{K}_{pm}(G \otimes F) = (F \otimes G)\mathbf{K}_{qn}$.
(24) $\mathbf{K}_{pm}(G \otimes F)\mathbf{K}_{nq} = F \otimes G$.
(25) $\mathbf{K}_{pm}(G \otimes \mathbf{b}) = \mathbf{b} \otimes G$.
(26) $\mathbf{K}_{pm}(\mathbf{b} \otimes G) = G \otimes \mathbf{b}$.
(27) $\text{vec}(G \otimes F) = (I_n \otimes \mathbf{K}_{qm} \otimes I_p)(\text{vec}(G) \otimes \text{vec}(F))$.
(28) $(\mathbf{D}'_m\mathbf{D}_m)^{-1}\mathbf{D}'_m\mathbf{K}_{mm} = (\mathbf{D}'_m\mathbf{D}_m)^{-1}\mathbf{D}'_m$.

A.13 Vector and Matrix Differentiation

In the following, it will be assumed that all derivatives exist and are continuous. Let $f(\beta)$ be a scalar function that depends on the $(n \times 1)$ vector $\beta = (\beta_1, \ldots, \beta_n)'$.

$$\frac{\partial f}{\partial \beta} := \begin{bmatrix} \dfrac{\partial f}{\partial \beta_1} \\ \vdots \\ \dfrac{\partial f}{\partial \beta_n} \end{bmatrix}, \quad \frac{\partial f}{\partial \beta'} := \left[\frac{\partial f}{\partial \beta_1}, \ldots, \frac{\partial f}{\partial \beta_n}\right]$$

are $(n \times 1)$ and $(1 \times n)$ vectors of first order partial derivatives, respectively, and

$$\frac{\partial^2 f}{\partial \beta \partial \beta'} := \left[\frac{\partial^2 f}{\partial \beta_i \partial \beta_j}\right] = \begin{bmatrix} \frac{\partial^2 f}{\partial \beta_1 \partial \beta_1} & \cdots & \frac{\partial^2 f}{\partial \beta_1 \partial \beta_n} \\ \vdots & & \vdots \\ \frac{\partial^2 f}{\partial \beta_n \partial \beta_1} & \cdots & \frac{\partial^2 f}{\partial \beta_n \partial \beta_n} \end{bmatrix}$$

is the $(n \times n)$ *Hessian matrix* of second order partial derivatives. If $f(A)$ is a scalar function of an $(m \times n)$ matrix $A = (a_{ij})$, then

$$\frac{\partial f}{\partial A} := \left[\frac{\partial f}{\partial a_{ij}}\right]$$

is an $(m \times n)$ matrix of partial derivatives. If the $(m \times n)$ matrix $A = (a_{ij})$ depends on the scalar β, then

$$\frac{\partial A}{\partial \beta} := \left[\frac{\partial a_{ij}}{\partial \beta}\right]$$

is an $(m \times n)$ matrix. If $y(\beta) = (y_1(\beta), \ldots, y_m(\beta))'$ is an $(m \times 1)$ vector that depends on the $(n \times 1)$ vector β, then

$$\frac{\partial y}{\partial \beta'} := \begin{bmatrix} \frac{\partial y_1}{\partial \beta_1} & \cdots & \frac{\partial y_1}{\partial \beta_n} \\ \vdots & & \vdots \\ \frac{\partial y_m}{\partial \beta_1} & \cdots & \frac{\partial y_m}{\partial \beta_n} \end{bmatrix}$$

is an $(m \times n)$ matrix and

$$\frac{\partial y'}{\partial \beta} := \left(\frac{\partial y}{\partial \beta'}\right)'.$$

For example, if $\beta = (\beta_1, \beta_2)'$ and $f(\beta) = \beta_1^2 - 2\beta_1\beta_2$, then

$$\frac{\partial f}{\partial \beta} = \begin{bmatrix} \frac{\partial f}{\partial \beta_1} \\ \\ \frac{\partial f}{\partial \beta_2} \end{bmatrix} = \begin{bmatrix} 2\beta_1 - 2\beta_2 \\ -2\beta_1 \end{bmatrix}.$$

If

$$y(\beta) = \begin{bmatrix} \beta_1^3 + \beta_2 \\ e^{\beta_1} \end{bmatrix}, \quad \text{then} \quad \frac{\partial y}{\partial \beta'} = \begin{bmatrix} 3\beta_1^2 & 1 \\ e^{\beta_1} & 0 \end{bmatrix}.$$

The following two propositions are useful for deriving rules for vector and matrix differentiation.

Proposition A.1 (*Chain Rule for Vector Differentiation*)
Let α and β be $(m \times 1)$ and $(n \times 1)$ vectors, respectively, and suppose $h(\alpha)$ is $(p \times 1)$ and $g(\beta)$ is $(m \times 1)$. Then, with $\alpha = g(\beta)$,

$$\frac{\partial h(g(\beta))}{\partial \beta'} = \frac{\partial h(\alpha)}{\partial \alpha'} \frac{\partial g(\beta)}{\partial \beta'} \qquad (p \times n).$$

■

Proposition A.2 (*Product Rules for Vector Differentiation*)
(a) Suppose β is $(m \times 1)$, $a(\beta) = (a_1(\beta), \ldots, a_n(\beta))'$ is $(n \times 1)$, $c(\beta) = (c_1(\beta), \ldots, c_p(\beta))'$ is $(p \times 1)$ and $A = (a_{ij})$ is $(n \times p)$ and does not depend on β. Then

$$\frac{\partial [a(\beta)' A c(\beta)]}{\partial \beta'} = c(\beta)' A' \frac{\partial a(\beta)}{\partial \beta'} + a(\beta)' A \frac{\partial c(\beta)}{\partial \beta'}.$$

(b) If β is a (1×1) scalar, $A(\beta)$ is $(m \times n)$ and $B(\beta)$ is $(n \times p)$, then

$$\frac{\partial AB}{\partial \beta} = \frac{\partial A}{\partial \beta} B + A \frac{\partial B}{\partial \beta}.$$

(c) If β is an $(m \times 1)$ vector, $A(\beta)$ is $(n \times p)$ and $B(\beta)$ is $(p \times q)$, then

$$\frac{\partial \operatorname{vec}(AB)}{\partial \beta'} = (I_q \otimes A) \frac{\partial \operatorname{vec}(B)}{\partial \beta'} + (B' \otimes I_n) \frac{\partial \operatorname{vec}(A)}{\partial \beta'}.$$

■

Proof:

(a)

$$\begin{aligned} \frac{\partial (a' A c)}{\partial \beta'} &= \frac{\partial \left(\sum_{i,j} a_i a_{ij} c_j \right)}{\partial \beta'} \\ &= \sum_{i,j} \left[\frac{\partial a_i}{\partial \beta'} a_{ij} c_j + a_i a_{ij} \frac{\partial c_j}{\partial \beta'} \right] \\ &= c' A' \frac{\partial a}{\partial \beta'} + a' A \frac{\partial c}{\partial \beta'}. \end{aligned}$$

(b)

$$AB = \left[\sum_j a_{ij} b_{jk} \right] \text{ and}$$

$$\frac{\partial \left(\sum_j a_{ij} b_{jk}\right)}{\partial \beta} = \sum_j \left[\frac{\partial a_{ij}}{\partial \beta} b_{jk} + a_{ij} \frac{\partial b_{jk}}{\partial \beta}\right].$$

(c) Follows from (b) by stacking the columns of AB and writing the resulting columns $\partial \operatorname{vec}(AB)/\partial \beta_i$ for $i = 1, \ldots, m$ in one matrix. ■

The following rules are now easy to verify.

Rules:

(1) Let A be an $(m \times n)$ matrix and β be an $(n \times 1)$ vector. Then

$$\frac{\partial A\beta}{\partial \beta'} = A \quad \text{and} \quad \frac{\partial \beta' A'}{\partial \beta} = A'.$$

Proof: This result is a special case of Proposition A.2(a). ■

(2) Let A be $(m \times m)$ and β be $(m \times 1)$. Then

$$\frac{\partial \beta' A\beta}{\partial \beta} = (A + A')\beta \quad \text{and} \quad \frac{\partial \beta' A\beta}{\partial \beta'} = \beta'(A' + A).$$

Proof: See Proposition A.2(a). ■

(3) If A is $(m \times m)$ and β is $(m \times 1)$, then

$$\frac{\partial^2 \beta' A\beta}{\partial \beta \partial \beta'} = A + A'.$$

Proof: Follows from (1) and (2). ■

(4) If A is a symmetric $(m \times m)$ matrix and β an $(m \times 1)$ vector then

$$\frac{\partial^2 \beta' A\beta}{\partial \beta \partial \beta'} = 2A.$$

Proof: See (3). ■

(5) Let Ω be a symmetric $(n \times n)$ matrix and $c(\beta)$ an $(n \times 1)$ vector that depends on the $(m \times 1)$ vector β. Then

$$\frac{\partial c(\beta)' \Omega c(\beta)}{\partial \beta'} = 2c(\beta)' \Omega \frac{\partial c(\beta)}{\partial \beta'}$$

and

$$\frac{\partial^2 c(\beta)' \Omega c(\beta)}{\partial \beta \partial \beta'} = 2\left[\frac{\partial c(\beta)'}{\partial \beta} \Omega \frac{\partial c(\beta)}{\partial \beta'} + [c(\beta)' \Omega \otimes I_m] \frac{\partial \operatorname{vec}(\partial c(\beta)'/\partial \beta)}{\partial \beta'}\right].$$

In particular, if y is an $(n \times 1)$ vector and X an $(n \times m)$ matrix,

$$\frac{\partial (y - X\beta)' \Omega (y - X\beta)}{\partial \beta'} = -2(y - X\beta)' \Omega X$$

and

$$\frac{\partial^2 (y - X\beta)' \Omega (y - X\beta)}{\partial \beta \partial \beta'} = 2X' \Omega X.$$

Proof: Follows from Proposition A.2(a). ■

(6) Suppose β is $(m \times 1)$, $B(\beta)$ is $(n \times p)$, A is $(k \times n)$, and C is $(p \times q)$ and the latter two matrices do not depend on β. Then

$$\frac{\partial \operatorname{vec}(ABC)}{\partial \beta'} = (C' \otimes A) \frac{\partial \operatorname{vec}(B)}{\partial \beta'}.$$

Proof: Follows from Rule (2), Section A.12, and Proposition A.1. ■

(7) Suppose β is $(m \times 1)$, $A(\beta)$ is $(n \times p)$, $D(\beta)$ is $(q \times r)$, and C is $(p \times q)$ and does not depend on β. Then

$$\frac{\partial \operatorname{vec}(ACD)}{\partial \beta'} = (I_r \otimes AC) \frac{\partial \operatorname{vec}(D)}{\partial \beta'} + (D'C' \otimes I_n) \frac{\partial \operatorname{vec}(A)}{\partial \beta'}.$$

Proof: Follows from Proposition A.2(c) by setting $B = CD$ and noting that $\partial \operatorname{vec}(CD)/\partial \beta' = (I_r \otimes C) \partial \operatorname{vec}(D)/\partial \beta'$. ■

(8) If β is $(m \times 1)$ and $A(\beta)$ is $(n \times n)$, then, for any positive integer h,

$$\frac{\partial \operatorname{vec}(A^h)}{\partial \beta'} = \left[\sum_{i=0}^{h-1} (A')^{h-1-i} \otimes A^i \right] \frac{\partial \operatorname{vec}(A)}{\partial \beta'}.$$

Proof: Follows inductively from Proposition A.2(c). The result is evident for $h = 1$. Assuming it holds for $h - 1$ gives

$$\begin{aligned} \frac{\partial \operatorname{vec}(AA^{h-1})}{\partial \beta'} &= (I_n \otimes A) \left[\sum_{i=0}^{h-2} (A')^{h-2-i} \otimes A^i \right] \frac{\partial \operatorname{vec}(A)}{\partial \beta'} \\ &\quad + ((A')^{h-1} \otimes I_n) \frac{\partial \operatorname{vec}(A)}{\partial \beta'}. \end{aligned}$$

■

(9) If A is a nonsingular $(m \times m)$ matrix, then

$$\frac{\partial \operatorname{vec}(A^{-1})}{\partial \operatorname{vec}(A)'} = -(A^{-1})' \otimes A^{-1}.$$

Proof: Using Proposition A.2(c),

$$0 = \frac{\partial \operatorname{vec}(I_m)}{\partial \operatorname{vec}(A)'} = \frac{\partial \operatorname{vec}(A^{-1}A)}{\partial \operatorname{vec}(A)'} = (I_m \otimes A^{-1})\frac{\partial \operatorname{vec}(A)}{\partial \operatorname{vec}(A)'} + (A' \otimes I_m)\frac{\partial \operatorname{vec}(A^{-1})}{\partial \operatorname{vec}(A)'}.$$

■

(10) Let A be a symmetric positive definite $(m \times m)$ matrix and let P be a lower triangular $(m \times m)$ matrix with positive elements on the main diagonal such that $A = PP'$. Moreover, let $\mathbf{L}_m$ be an $(\frac{1}{2}m(m+1) \times m^2)$ elimination matrix such that $\mathbf{L}_m \operatorname{vec}(A) = \operatorname{vech}(A)$ consists of the elements on and below the main diagonal of A only. Then

$$\begin{aligned}\frac{\partial \operatorname{vech}(P)}{\partial \operatorname{vech}(A)'} &= \{\mathbf{L}_m[(I_m \otimes P)\mathbf{K}_{mm} + (P \otimes I_m)]\mathbf{L}_m'\}^{-1} \\ &= \{\mathbf{L}_m(I_{m^2} + \mathbf{K}_{mm})(P \otimes I_m)\mathbf{L}_m'\}^{-1},\end{aligned}$$

where $\mathbf{K}_{mm}$ is an $(m^2 \times m^2)$ commutation matrix such that $\mathbf{K}_{mm}\operatorname{vec}(P) = \operatorname{vec}(P')$.

Proof: See Lütkepohl (1989a). ■

(11) If $A = (a_{ij})$ is an $(m \times m)$ matrix, then

$$\frac{\partial \operatorname{tr}(A)}{\partial A} = I_m.$$

Proof: $\operatorname{tr}(A) = a_{11} + \cdots + a_{mm}$. Hence,

$$\frac{\partial \operatorname{tr}(A)}{\partial a_{ij}} = \begin{cases} 0 & \text{if } i \neq j, \\ 1 & \text{if } i = j. \end{cases}$$

■

(12) If $A = (a_{ij})$ is $(m \times n)$ and $B = (b_{ij})$ is $(n \times m)$, then

$$\frac{\partial \operatorname{tr}(AB)}{\partial A} = B'.$$

Proof: Follows because $\operatorname{tr}(AB) = \sum_{j=1}^{n} a_{1j}b_{j1} + \cdots + \sum_{j=1}^{n} a_{mj}b_{jm}$. ■

(13) Suppose A is an $(m \times n)$ matrix and B, C are $(m \times m)$ and $(n \times m)$, respectively. Then

$$\frac{\partial \operatorname{tr}(BAC)}{\partial A} = B'C'.$$

Proof: Follows from Rule (12) because $\operatorname{tr}(BAC) = \operatorname{tr}(ACB)$. ■

(14) Let A, B, C, D be $(m \times n)$, $(n \times n)$, $(m \times n)$, and $(n \times m)$ matrices, respectively. Then

$$\frac{\partial \operatorname{tr}(DABA'C)}{\partial A} = CDAB + D'C'AB'.$$

Proof: See Murata (1982, Appendix, Theorem 6a). ■

(15) Let A, B, and C be $(m \times m)$ matrices and suppose A is nonsingular. Then

$$\frac{\partial \operatorname{tr}(BA^{-1}C)}{\partial A} = -(A^{-1}CBA^{-1})'.$$

Proof: By Rule (6) of Section A.12,

$$\operatorname{tr}(BA^{-1}C) = \operatorname{vec}(B')'(C' \otimes I_m) \operatorname{vec}(A^{-1}).$$

Hence, using (9),

$$\begin{aligned} \frac{\partial \operatorname{tr}(BA^{-1}C)}{\partial \operatorname{vec}(A)'} &= -\operatorname{vec}(B')'(C' \otimes I_m)((A^{-1})' \otimes A^{-1}) \\ &= -\left[(A^{-1}C \otimes A^{-1\prime}) \operatorname{vec}(B')\right]' \\ &= -\left[\operatorname{vec}(A^{-1\prime}B'C'A^{-1\prime})\right]' \end{aligned}$$

by Rule (2) of Section A.12. ■

(16) Let $A = (a_{ij})$ be an $(m \times m)$ matrix. Then

$$\frac{\partial |A|}{\partial A} = (A^{adj})',$$

where A^{adj} is the adjoint of A.

Proof: Developing by the i-th row of A gives

$$|A| = a_{i1}A_{i1} + \cdots + a_{im}A_{im},$$

where A_{ij} is the cofactor of a_{ij}. Hence,

$$\frac{\partial |A|}{\partial a_{ij}} = A_{ij}$$

because A_{ij} does not contain a_{ij}. ■

(17) If A is a nonsingular $(m \times m)$ matrix with $|A| > 0$, then

$$\frac{\partial \ln |A|}{\partial A} = (A')^{-1}.$$

Proof: Using Proposition A.1 (chain rule),

$$\frac{\partial \ln |A|}{\partial A} = \frac{\partial \ln |A|}{\partial |A|} \cdot \frac{\partial |A|}{\partial A} = \frac{1}{|A|}(A^{adj})' = (A')^{-1}.$$

■

Proposition A.3 (*Taylor's Theorem*)
Let $f(\beta)$ be a scalar valued function of the $(m \times 1)$ vector β. Suppose $f(\beta)$ is at least twice continuously differentiable on an open set $\mathbb{S}$ that contains β_0, β, and the entire line segment between β_0 and β. Then there exists a point $\overline{\beta}$ on the line segment such that

$$f(\beta) = f(\beta_0) + \frac{\partial f(\beta_0)}{\partial \beta'}(\beta - \beta_0) + \frac{1}{2}(\beta - \beta_0)' \frac{\partial^2 f(\overline{\beta})}{\partial \beta \partial \beta'}(\beta - \beta_0), \qquad \text{(A.13.1)}$$

where $\partial f(\beta_0)/\partial \beta' := (\partial f/\partial \beta'|_{\beta_0})$. ■

The expansion of f given in (A.13.1) is a *second order Taylor expansion* at or around β_0.

A.14 Optimization of Vector Functions

Suppose $f(\beta)$ is a real valued (scalar) differentiable function of the $(m \times 1)$ vector β. A necessary condition for a local optimum (minimum or maximum) at $\widetilde{\beta}$ is that

$$\frac{\partial f}{\partial \beta} = 0 \qquad \text{for } \beta = \widetilde{\beta}, \quad \text{that is,} \quad \frac{\partial f(\widetilde{\beta})}{\partial \beta} := \left[\frac{\partial f}{\partial \beta}\Big|_{\widetilde{\beta}} \right] = 0.$$

In other words, $f(\cdot)$ has a *stationary point* at $\widetilde{\beta}$. If this condition is satisfied and the *Hessian matrix* of second order partial derivatives

$$\frac{\partial^2 f}{\partial \beta \partial \beta'}$$

is negative (positive) definite for $\beta = \widetilde{\beta}$, then $\widetilde{\beta}$ is a local maximum (minimum).

If a set of constraints is given in the form

$$\varphi(\beta) = (\varphi_1(\beta), \ldots, \varphi_n(\beta))' = 0,$$

that is, $\varphi(\beta)$ is an $(n \times 1)$ vector, then a local optimum, subject to these constraints, is obtained at a stationary point of the *Lagrange function*

$$\mathcal{L}(\beta, \lambda) = f(\beta) - \lambda' \varphi(\beta),$$

where λ is an $(n \times 1)$ vector of *Lagrange multipliers*. In other words, a necessary condition for a constrained local optimum is that

$$\frac{\partial \mathcal{L}}{\partial \beta} = 0 \quad \text{and} \quad \frac{\partial \mathcal{L}}{\partial \lambda} = 0$$

hold simultaneously.

The following results are useful in some optimization problems.

Proposition A.4 (*Maximum of* $\operatorname{tr}(B'\Omega B)$)
Let Ω be a positive semidefinite symmetric $(K \times K)$ matrix with eigenvalues $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_K$ and corresponding orthonormal $(K \times 1)$ eigenvectors $v_1, v_2, \ldots, v_K$. Moreover, let B be a $(K \times r)$ matrix with $B'B = I_r$. Then the maximum of $\operatorname{tr}(B'\Omega B)$ with respect to B is obtained for

$$B = \widehat{B} = [v_1, \ldots, v_r]$$

and

$$\max_B \operatorname{tr}(B'\Omega B) = \lambda_1 + \cdots + \lambda_r.$$

■

Proof: The proposition follows from Theorem 6, p. 205, of Magnus & Neudecker (1988) by induction. For $r = 1$, our result is just a special case of that theorem. For $r > 1$, assuming that the proposition holds for $r-1$ and denoting the columns of B by $b_1, \ldots, b_r$,

$$\begin{aligned}
\operatorname{tr}(B'\Omega B) &= \operatorname{tr}\begin{bmatrix} b_1' \\ \vdots \\ b_r' \end{bmatrix} \Omega [b_1, \ldots, b_r] = \operatorname{tr}\begin{bmatrix} b_1'\Omega b_1 & & * \\ & \ddots & \\ * & & b_r'\Omega b_r \end{bmatrix} \\
&= b_1'\Omega b_1 + \cdots + b_{r-1}'\Omega b_{r-1} + b_r'\Omega b_r \\
&= \lambda_1 + \cdots + \lambda_{r-1} + b_r'\Omega b_r
\end{aligned}$$

and $\max b_r'\Omega b_r = v'\Omega v_r = \lambda_r$, under the conditions of the proposition, by the aforementioned theorem from Magnus & Neudecker (1988). ■

The next proposition may be regarded as a corollary of Proposition A.4.

Proposition A.5 (*Minimum of* $\operatorname{tr}(Y - BCX)'\Sigma_u^{-1}(Y - BCX)$)
Let Y, X, Σ_u, B, and C be matrices of dimensions $(K \times T), (Kp \times T), (K \times K), (K \times r)$, and $(r \times Kp)$, respectively, with Σ_u positive definite, $\operatorname{rk}(B) = \operatorname{rk}(C) = r$, $\operatorname{rk}(X) = Kp$, and $\operatorname{rk}(Y) = K$. Then a minimum of

$$\operatorname{tr}[(Y - BCX)'\Sigma_u^{-1}(Y - BCX)] \tag{A.14.1}$$

with respect to B and C is obtained for

$$B = \widehat{B} = \Sigma_u^{1/2}\widehat{V} \quad \text{and} \quad C = \widehat{C} = \widehat{V}'\Sigma_u^{-1/2}YX'(XX')^{-1}, \tag{A.14.2}$$

where $\widehat{V} = [\widehat{v}_1, \ldots, \widehat{v}_r]$ is the $(K \times r)$ matrix of the orthonormal eigenvectors corresponding to the r largest eigenvalues of

$$\frac{1}{T}\Sigma_u^{-1/2}YX'(XX')^{-1}XY'\Sigma_u^{-1/2}$$

in nonincreasing order. ■

Proof: We first assume $\Sigma_u = I_K$.

$$\begin{aligned}
&\mathrm{tr}[(Y - BCX)'(Y - BCX)] \\
&\quad = \mathrm{tr}[(Y - BCX)(Y - BCX)'] \\
&\quad = [\mathrm{vec}(Y) - \mathrm{vec}(BCX)]'[\mathrm{vec}(Y) - \mathrm{vec}(BCX)] \\
&\quad = [\mathrm{vec}(Y) - (X' \otimes B)\mathrm{vec}(C)]' [\mathrm{vec}(Y) - (X' \otimes B)\mathrm{vec}(C)] .
\end{aligned} \tag{A.14.3}$$

A derivation similar to that in Section 3.2.1 shows that this sum of squares is minimized with respect to $\mathrm{vec}(C)$ when this vector is chosen to be

$$\begin{aligned}
\mathrm{vec}(\widehat{C}) &= [(X \otimes B')(X' \otimes B)]^{-1} (X \otimes B')\mathrm{vec}(Y) \\
&= (XX' \otimes B'B)^{-1} \mathrm{vec}(B'YX') \\
&= \mathrm{vec}[(B'B)^{-1}B'YX'(XX')^{-1}].
\end{aligned}$$

Because we may normalize the columns of B, we choose $B'B = I_r$ without loss of generality. Hence,

$$\widehat{C} = B'YX'(XX')^{-1}. \tag{A.14.4}$$

Substituting for C in (A.14.3) gives

$$\begin{aligned}
&\mathrm{tr}[(Y - BB'YX'(XX')^{-1}X)(Y - BB'YX'(XX')^{-1}X)'] \\
&\quad = \mathrm{tr}(YY') - \mathrm{tr}(BB'YX'(XX')^{-1}XY') - \mathrm{tr}(YX'(XX')^{-1}XY'BB') \\
&\qquad + \mathrm{tr}(BB'YX'(XX')^{-1}XX'(XX')^{-1}XY'BB') \\
&\quad = \mathrm{tr}(YY') - \mathrm{tr}(B'YX'(XX')^{-1}XY'B),
\end{aligned}$$

where again $B'B = I_r$ has been used. This expression is minimized with respect to B, where

$$\frac{1}{T}\mathrm{tr}B'YX'(XX')^{-1}XY'B$$

assumes its maximum. By Proposition A.4, the maximum is attained if B consists of the eigenvectors corresponding to the r largest eigenvalues of

$$\frac{1}{T}YX'(XX')^{-1}XY'$$

which proves the proposition for $\Sigma_u = I_K$.

If $\Sigma_u \neq I_K$,

$$\mathrm{tr}[(Y - BCX)'\Sigma_u^{-1}(Y - BCX)] = \mathrm{tr}[(Y^{\#} - B^{\#}CX)'(Y^{\#} - B^{\#}CX)]$$

has to be minimized with respect to $B^{\#}$ and C. Here $Y^{\#} = \Sigma_u^{-1/2}Y$ and $B^{\#} = \Sigma_u^{-1/2}B$. From the above derivation the solution is $\widehat{B}^{\#} = \widehat{V}$ and

$$\widehat{C} = \widehat{B}^{\#\prime}Y^{\#}X'(XX')^{-1} = \widehat{V}'\Sigma_u^{-1/2}YX'(XX')^{-1},$$

where the columns of $\widehat{V}$ are the eigenvectors corresponding to the r largest eigenvalues of

$$\frac{1}{T}Y^{\#}X'(XX')^{-1}XY^{\#\prime} = \frac{1}{T}\Sigma_u^{-1/2}YX'(XX')^{-1}XY'\Sigma_u^{-1/2}.$$

Hence, $\widehat{B} = \Sigma_u^{1/2}\widehat{B}^{\#} = \Sigma_u^{1/2}\widehat{V}$. ■

A result similar to that in Proposition A.4 also holds for the maximum and minimum of a determinant. The following proposition is a slight modification of Theorem 15 of Magnus & Neudecker (1988, Chapter 11).

Proposition A.6 (*Maximum and Minimum of* $|C\Omega C'|$)
Let Ω be a positive definite symmetric $(K \times K)$ matrix with eigenvalues $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_K$ and corresponding orthonormal $(K \times 1)$ eigenvectors $v_1, \ldots, v_K$. Furthermore, let C be an $(r \times K)$ matrix with $CC' = I_r$. Then

$$\max_C |C\Omega C'| = \lambda_1 \cdots \lambda_r$$

and the maximum is attained for

$$C = \widehat{C} = [v_1, \ldots, v_r]'.$$

Moreover,

$$\min_C |C\Omega C'| = \lambda_K \lambda_{K-1} \cdots \lambda_{K-r+1}$$

and the minimum is attained for

$$C = \widehat{C} = [v_K, \ldots, v_{K-r+1}]'.$$

■

An important implication of this proposition is used in Chapter 7 and is stated next.

Proposition A.7 (*Minimum of* $|T^{-1}(Y - BCX)(Y - BCX)'|$)
Let Y and X be $(K \times T)$ matrices of rank K and let B and C be of rank r and dimensions $(K \times r)$ and $(r \times K)$, respectively. Furthermore, let $\lambda_1 \geq \cdots \geq \lambda_K$ be the eigenvalues of

$$(XX')^{-1/2}XY'(YY')YX'(XX')^{-1/2\prime}$$

and the corresponding orthonormal eigenvectors are $v_1, \ldots, v_K$. Here

$$(XX')^{-1/2}$$

is some matrix satisfying

$$(XX')^{-1/2}(XX')(XX')^{-1/2\prime} = I_K.$$

Then

$$\min_{B,C} |T^{-1}(Y - BCX)(Y - BCX)'| = |T^{-1}YY'|(1 - \lambda_1) \cdots (1 - \lambda_r)$$

and the minimum is attained for

$$C = \widehat{C} = [v_1, \ldots, v_r]'(XX')^{-1/2}$$

and

$$B = \widehat{B} = YX'\widehat{C}'(\widehat{C}XX'\widehat{C}')^{-1}.$$

■

A proof of this proposition can be found in Tso (1981). It should be noted that the minimizing matrices $\widehat{B}$ and $\widehat{C}$ are not unique. Any nonsingular $(r \times r)$ matrix F leads to another set of minimizing matrices $F\widehat{C}, \widehat{B}F^{-1}$.

A.15 Problems

The following problems refer to the matrices

$$A = \begin{bmatrix} 5 & 2 \\ -1 & 1 \end{bmatrix}, \quad B = \begin{bmatrix} 6 & 0 & 0 \\ -6 & 1 & 0 \end{bmatrix}, \quad C = \begin{bmatrix} 1 & 4 & 0 \\ 2 & 2 & 2 \\ 1 & 2 & 0 \end{bmatrix},$$

$$D = \begin{bmatrix} 5 & 2 \\ 2 & 1 \end{bmatrix}, \quad H(\beta) = \begin{bmatrix} 4\beta_1 & 2\beta_1 + \beta_2 \\ 1 + \beta_2 & 3 \end{bmatrix}.$$

Problem A.1
Determine $A + D$, $A - 2D$, A', AB, BC, $B'A$, $B'A'$, $A \otimes D$, $B \otimes D$, $D \otimes B$, $B' \otimes D'$, $B + BC$, tr A, tr D, det A, $|D|$, $|C|$, vec(B), vec(B'), vech(C), $\mathbf{K}_{33}$, A^{-1}, D^{-1}, $(A \otimes D)^{-1}$, rk(C), rk(B), det$(A \otimes D)$, tr$(A \otimes D)$, C^{-1} (use the rules for the partitioned inverse).

Problem A.2
Determine the eigenvalues of A, D, and $A \otimes D$.

Problem A.3
Find an upper triangular matrix Q such that $D = QQ'$ and find an orthogonal matrix P such that $D = P\Lambda P'$, where Λ is a diagonal matrix with the eigenvalues of D on the main diagonal. Compute D^5.

Problem A.4
Is $F = I_2 - BB'$ idempotent? Is BB' positive definite?

Problem A.5
Determine the following derivatives.

$$\frac{\partial \det(H)}{\partial \beta}, \qquad \frac{\partial^2 \det(H)}{\partial \beta \partial \beta'}, \qquad \frac{\partial \operatorname{tr}(H)}{\partial \beta},$$

$$\frac{\partial \operatorname{vec}(H)}{\partial \beta'}, \qquad \frac{\partial \operatorname{vec}(H^2)}{\partial \beta'}, \qquad \frac{\partial H(\beta)\beta}{\partial \beta'},$$

where $\beta = (\beta_1, \beta_2)'$.

Problem A.6
Determine the stationary points of $|H|$ with respect to β. Are they local extrema?

Problem A.7
Give a second order Taylor expansion of $\det(H)$ around $\beta = (0, 0)'$.

B

Multivariate Normal and Related Distributions

B.1 Multivariate Normal Distributions

A K-dimensional vector of continuous random variables $y = (y_1, \ldots, y_K)'$ has a *multivariate normal distribution* with mean vector $\mu = (\mu_1, \ldots, \mu_K)'$ and covariance matrix Σ, briefly

$$y \sim \mathcal{N}(\mu, \Sigma),$$

if its distribution has the probability density function (p.d.f.)

$$f(y) = \frac{1}{(2\pi)^{K/2}} |\Sigma|^{-1/2} \exp\left[-\frac{1}{2}(y-\mu)'\Sigma^{-1}(y-\mu)\right]. \tag{B.1.1}$$

Alternatively, $y \sim \mathcal{N}(\mu, \Sigma)$, if for any K-vector c for which $c'\Sigma c \neq 0$ the linear combination $c'y$ has a univariate normal distribution, that is, $c'y \sim \mathcal{N}(c'\mu, c'\Sigma c)$ (see Rao (1973, Chapter 8)). This definition of a multivariate normal distribution is useful because it carries over to the case where Σ is positive semidefinite and singular, while the multivariate density in (B.1.1) is only meaningful, if Σ is positive definite and, hence, nonsingular. It must be emphasized, however, that the two definitions are equivalent, if Σ is positive definite rather than just positive semidefinite. Another possibility to define a multivariate normal distribution with singular covariance matrix may be found in Anderson (1984).

The following results regarding the multivariate normal and related distributions are useful. Many of them are stated in Judge et al. (1985, Appendix A). Proofs can be found in Rao (1973, Chapter 8) and Hogg & Craig (1978, Chapter 12).

Proposition B.1 (*Marginal and Conditional Distributions of a Multivariate Normal*)
Let y_1 and y_2 be two random vectors such that

$$\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}, \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix} \right),$$

where the partitioning of the mean vector and covariance matrix corresponds to that of the vector $(y_1', y_2')'$. Then,

$$y_1 \sim \mathcal{N}(\mu_1, \Sigma_{11})$$

and the conditional distribution of y_1 given $y_2 = c$ is also multivariate normal,

$$(y_1 | y_2 = c) \sim \mathcal{N}(\mu_1 + \Sigma_{12}\Sigma_{22}^{-1}(c - \mu_2), \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}).$$

If Σ_{22} is singular, the inverse can be replaced by a generalized inverse. Moreover, y_1 and y_2 are independent if and only if $\Sigma_{12} = \Sigma_{21}' = 0$. ■

Proposition B.2 (*Linear Transformation of a Multivariate Normal Random Vector*)
Suppose $y \sim \mathcal{N}(\mu, \Sigma)$ is $(K \times 1)$, A is an $(M \times K)$ matrix and c an $(M \times 1)$ vector. Then

$$x = Ay + c \sim \mathcal{N}(A\mu + c, A\Sigma A').$$

■

B.2 Related Distributions

Suppose $y \sim \mathcal{N}(0, I_K)$. The distribution of $z = y'y$ is a (central) *chi-square distribution* with K degrees of freedom,

$$z \sim \chi^2(K).$$

Proposition B.3 (*Distributions of Quadratic Forms*)

(1) Suppose $y \sim \mathcal{N}(0, I_K)$ and A is a symmetric idempotent $(K \times K)$ matrix with $\text{rk}(A) = n$. Then $y'Ay \sim \chi^2(n)$.
(2) If $y \sim \mathcal{N}(0, \Sigma)$, where Σ is a positive definite $(K \times K)$ matrix, then $y'\Sigma^{-1}y \sim \chi^2(K)$.
(3) Let $y \sim \mathcal{N}(0, QA)$, where Q is a symmetric idempotent $(K \times K)$ matrix with $\text{rk}(Q) = n$ and A is a positive definite $(K \times K)$ matrix. Then $y'A^{-1}y \sim \chi^2(n)$.
(4) Suppose $y \sim \mathcal{N}(0, \Sigma)$, where Σ is a nonsingular $(K \times K)$ covariance matrix. Furthermore, let A be a $(K \times K)$ matrix with $\text{rk}(A) = n$. Then

$$y'Ay \sim \chi^2(n) \quad \Rightarrow \quad A\Sigma A = A$$

and

$$A\Sigma A = A \quad \Rightarrow \quad y'Ay \sim \chi^2(n).$$

■

Proposition B.4 (*Independence of a Normal Vector and a Quadratic Form*) Suppose $y \sim \mathcal{N}(\mu, \sigma^2 I_K)$, A is a symmetric, idempotent $(K \times K)$ matrix, B is an $(M \times K)$ matrix and $BA = 0$. Then By is stochastically independent of the random variable $y'Ay$. ■

Proposition B.5 (*Independence of Quadratic Forms*)
Suppose $y \sim \mathcal{N}(\mu, \sigma^2 I_K)$ and A and B are symmetric, idempotent $(K \times K)$ matrices with $AB = 0$, then $y'Ay$ and $y'By$ are stochastically independent. ■

If $z \sim \mathcal{N}(0, 1)$ and $u \sim \chi^2(m)$ are stochastically independent, then

$$T = \frac{z}{\sqrt{u/m}}$$

has a *t-distribution* with m degrees of freedom, $T \sim t(m)$. If $u \sim \chi^2(m)$ and $v \sim \chi^2(n)$ are independent, then

$$\frac{u/m}{v/n} \sim F(m, n),$$

that is, the ratio of two independent χ^2 random variables, each divided by its degrees of freedom, has an *F-distribution* with m and n degrees of freedom. The numbers m and n indicate the numerator and denominator degrees of freedom, respectively.

Proposition B.6 (*Distributions of Ratios of Quadratic Forms*)

(1) Suppose $x \sim \mathcal{N}(0, I_m)$ and $y \sim \mathcal{N}(0, I_n)$ are independent. Then

$$\frac{x'x/m}{y'y/n} \sim F(m, n).$$

(2) If $y \sim \mathcal{N}(0, I_K)$ and A and B are symmetric, idempotent $(K \times K)$ matrices with $\mathrm{rk}(A) = m$, $\mathrm{rk}(B) = n$ and $AB = 0$, then

$$\frac{y'Ay/m}{y'By/n} \sim F(m, n).$$

(3) $z \sim F(m, n) \Rightarrow \dfrac{1}{z} \sim F(n, m).$

■

If $y \sim \mathcal{N}(\mu, I_K)$, then $y'y$ has a *noncentral* χ^2*-distribution* with K degrees of freedom and *noncentrality parameter* (or simply *noncentrality*) $\tau = \mu'\mu$. Briefly,

$$y'y \sim \chi^2(K; \tau).$$

The noncentrality parameter is sometimes defined differently in the literature. For instance, $\lambda = \frac{1}{2}\mu'\mu$ is sometimes called noncentrality parameter. Let $w \sim \chi^2(m;\tau)$ and $v \sim \chi^2(n)$ be independent random variables, then

$$\frac{w/m}{v/n} \sim F(m,n;\tau),$$

that is, the ratio has a *noncentral F-distribution* with m and n degrees of freedom and noncentrality parameter τ.

Proposition B.7 (*Quadratic Form with Noncentral χ^2-Distribution*)
If $y \sim \mathcal{N}(\mu, \Sigma)$ with positive definite $(K \times K)$ covariance matrix Σ, then $y'\Sigma^{-1}y \sim \chi^2(K;\mu'\Sigma^{-1}\mu)$. ■

C

Stochastic Convergence and Asymptotic Distributions

It is often difficult to derive the exact distributions of estimators and test statistics. In that case, their asymptotic or limiting properties, when the sample size gets large, are of interest. The limiting properties are then regarded as approximations to the properties for the sample size available. In order to study the limiting properties, some concepts of convergence of sequences of random variables and vectors are useful. They are discussed in Sections C.1 and C.2. Infinite sums of random variables are treated in Section C.3. Laws of large numbers and central limit theorems are given in Section C.4. Asymptotic properties of estimators are considered in Section C.5. Maximum likelihood estimators and their asymptotic properties are discussed in Section C.6 and some common testing principles are treated in Section C.7. Finally, asymptotic properties of nonstationary processes with unit roots are dealt with in Section C.8.

This appendix contains a brief summary of results used in the text. Many of these results can be found in Judge et al. (1985, Section 5.8). A more complete discussion and proofs are provided in Fuller (1976), Roussas (1973), Serfling (1980), Davidson (1994, 2000) and other more advanced books on statistics. Further references will be given in the following.

C.1 Concepts of Stochastic Convergence

Let $x_1, x_2, \ldots$ or $\{x_T\}$, $T = 1, 2, \ldots$, be a sequence of scalar random variables which are all defined on a common probability space $(\Omega, \mathcal{F}, \Pr)$. The sequence $\{x_T\}$ *converges in probability* to the random variable x (which is also defined on $(\Omega, \mathcal{F}, \Pr)$) if for every $\epsilon > 0$,

$$\lim_{T\to\infty} \Pr(|x_T - x| > \epsilon) = 0$$

or, equivalently,

$$\lim_{T\to\infty} \Pr(|x_T - x| < \epsilon) = 1.$$

This type of stochastic convergence is abbreviated as

$$\text{plim } x_T = x \qquad \text{or} \qquad x_T \xrightarrow{p} x.$$

The limit x may be a fixed, nonstochastic real number which is then regarded as a degenerate random variable that takes on one particular value with probability one.

The sequence $\{x_T\}$ *converges almost surely* (a.s.) or *with probability one* to the random variable x if for every $\epsilon > 0$,

$$\Pr\left(\lim_{T\to\infty} |x_T - x| < \epsilon\right) = 1.$$

This type of stochastic convergence is often written as $x_T \xrightarrow{a.s.} x$ and is sometimes called *strong convergence*.

The sequence $\{x_T\}$ *converges in quadratic mean* or *mean square error* to x, briefly $x_T \xrightarrow{q.m.} x$, if

$$\lim_{T\to\infty} E(x_T - x)^2 = 0.$$

This type of convergence requires that the mean and variance of the x_T's and x exist.

Finally, denoting the distribution functions of x_T and x by F_T and F, respectively, the sequence $\{x_T\}$ is said to *converge in distribution* or *weakly* or *in law* to x, if for all real numbers c for which F is continuous,

$$\lim_{T\to\infty} F_T(c) = F(c).$$

This type of convergence is abbreviated as $x_T \xrightarrow{d} x$. It must be emphasized that we do not require the convergence of the sequence of p.d.f.s of the x_T's to the p.d.f. of x. In fact, we do not even require that the distributions of the x_T's have p.d.f.s. Even if they do have p.d.f.s, convergence in distribution does not imply their convergence to the p.d.f. of x.

All these concepts of stochastic convergence can be extended to sequences of random vectors (multivariate random variables). Suppose $\{x_T = (x_{1T}, \ldots, x_{KT})'\}$, $T = 1, 2, \ldots$, is a sequence of K-dimensional random vectors and $x = (x_1, \ldots, x_K)'$ is a K-dimensional random vector. Then the following definitions are used:

$$\text{plim } x_T = x \quad \text{or} \quad x_T \xrightarrow{p} x \quad \text{if} \quad \text{plim } x_{kT} = x_k \text{ for } k = 1, \ldots, K.$$

$$x_T \xrightarrow{a.s.} x \quad \text{if} \quad x_{kT} \xrightarrow{a.s.} x_k \text{ for } k = 1, \ldots, K.$$

$$x_T \xrightarrow{q.m.} x \quad \text{if} \quad \lim E[(x_T - x)'(x_T - x)] = 0.$$

$$x_T \xrightarrow{d} x \quad \text{if} \quad \lim F_T(c) = F(c) \text{ for all continuity points of } F.$$

Here F_T and F are the joint distribution functions of x_T and x, respectively. Almost sure convergence and convergence in probability can be defined for matrices in the same way in terms of convergence of the individual elements. Convergence in quadratic mean and in distribution is easily extended to sequences of random matrices by vectorizing them. In the following proposition, the relationships between the different modes of convergence are given.

Proposition C.1 (*Convergence Properties of Sequences of Random Variables*)
Suppose $\{x_T\}$ is a sequence of K-dimensional random variables. Then the following relations hold:

(1) $x_T \xrightarrow{a.s.} x \Rightarrow x_T \xrightarrow{p} x \Rightarrow x_T \xrightarrow{d} x$.
(2) $x_T \xrightarrow{q.m.} x \Rightarrow x_T \xrightarrow{p} x \Rightarrow x_T \xrightarrow{d} x$.
(3) If x is a fixed, nonstochastic vector, then

$$x_T \xrightarrow{q.m.} x \quad \Leftrightarrow \quad [\lim E(x_T) = x \text{ and } \lim E\{(x_T - Ex_T)'(x_T - Ex_T)\} = 0].$$

(4) If x is a fixed, nonstochastic random vector, then

$$x_T \xrightarrow{p} x \quad \Leftrightarrow \quad x_T \xrightarrow{d} x.$$

(5) (Slutsky's Theorem) If $g : \mathbb{R}^K \to \mathbb{R}^m$ is a continuous function, then

$$x_T \xrightarrow{p} x \quad \Rightarrow \quad g(x_T) \xrightarrow{p} g(x) \quad [\text{plim } g(x_T) = g(\text{plim } x_T)],$$

$$x_T \xrightarrow{d} x \quad \Rightarrow \quad g(x_T) \xrightarrow{d} g(x),$$

and

$$x_T \xrightarrow{a.s.} x \quad \Rightarrow \quad g(x_T) \xrightarrow{a.s.} g(x).$$

■

Proposition C.2 (*Properties of Convergence in Probability and in Distribution*)
Suppose $\{x_T\}$ and $\{y_T\}$ are sequences of $(K \times 1)$ random vectors, $\{A_T\}$ is a sequence of $(K \times K)$ random matrices, x is a $(K \times 1)$ random vector, c is a fixed $(K \times 1)$ vector, and A is a fixed $(K \times K)$ matrix.

(1) If plim x_T, plim y_T, and plim A_T exist, then
  (a) $\text{plim}\,(x_T \pm y_T) = \text{plim } x_T \pm \text{plim } y_T$;
  (b) $\text{plim}\,(c'x_T) = c'(\text{plim } x_T)$;
  (c) $\text{plim } x_T'y_T = (\text{plim } x_T)'(\text{plim } y_T)$;
  (d) $\text{plim } A_Tx_T = \text{plim}\,(A_T)\text{plim}\,(x_T)$.
(2) If $x_T \xrightarrow{d} x$ and $\text{plim}\,(x_T - y_T) = 0$, then $y_T \xrightarrow{d} x$.
(3) If $x_T \xrightarrow{d} x$ and plim $y_T = c$, then

(a) $x_T \pm y_T \xrightarrow{d} x \pm c$;
(b) $y_T' x_T \xrightarrow{d} c'x$.

(4) If $x_T \xrightarrow{d} x$ and plim $A_T = A$, then $A_T x_T \xrightarrow{d} Ax$.
(5) If $x_T \xrightarrow{d} x$ and plim $A_T = 0$, then plim $A_T x_T = 0$.

■

Proposition C.3 (*Limits of Sequences of t and F Random Variables*)

(1) $t(T) \xrightarrow{d}_{T\to\infty} \mathcal{N}(0,1)$
(that is, a sequence of random variables with t-distributions with T degrees of freedom converges to a standard normal distribution as the degrees of freedom go to infinity).
(2) $JF(J,T) \xrightarrow{d}_{T\to\infty} \chi^2(J)$.

■

C.2 Order in Probability

Let $\{a_T\}$ be a sequence of real numbers and $\{b_T\}$ a sequence of positive real numbers. Then a_T is said to be *of smaller order than* b_T ($a_T = o(b_T)$) if $\lim_{T\to\infty} a_T/b_T = 0$ and a_T is said to be *at most of order* b_T ($a_T = O(b_T)$) if there exists a number c such that for all T, $|a_T|/b_T \leq c$.

Proposition C.4 (*Order of Convergence Results*)
For sequences of real numbers $\{a_T\}$, $\{b_T\}$ and sequences of positive real numbers $\{c_T\}$, $\{d_T\}$, the following results hold:

(1) $a_T = o(c_T), b_T = o(d_T) \Rightarrow a_T b_T = o(c_T d_T)$, $a_T + b_T = o(\max[c_T, d_T])$ and $|a_T|^s = o(c_T^s)$ for $s > 0$.
(2) $a_T = O(c_T), b_T = O(d_T) \Rightarrow a_T b_T = O(c_T d_T)$, $a_T + b_T = O(\max[c_T, d_T])$ and $|a_T|^s = O(c_T^s)$ for $s > 0$.
(3) $a_T = o(c_T), b_T = O(d_T) \Rightarrow a_T b_T = o(c_T d_T)$.

■

Let $\{A_T = (a_{ij,T})\}$ be a sequence of random $(m \times n)$ matrices and $\{b_T\}$ a sequence of positive real numbers. Then A_T is said to be *of smaller order in probability than* b_T ($A_T = o_p(b_T)$) if plim $_{T\to\infty} A_T/b_T = 0$ and A_T is said to be *at most of order in probability* b_T or *bounded in probability by* b_T ($A_T = O_p(b_T)$) if, for every $\epsilon > 0$, there exists a number c_ϵ such that for all T, $\Pr\{|a_{ij,T}| \geq c_\epsilon b_T\} \leq \epsilon$ for $i = 1, \ldots, m$, $j = 1, \ldots, n$. The following results hold for sequences of random matrices.

Proposition C.5 (*Order in Probability Results*)
For sequences of random matrices of suitable fixed dimensions $\{A_T\}$, $\{B_T\}$ and sequences of positive real numbers $\{c_T\}$, $\{d_T\}$ the following results hold:

(1) $A_T = o_p(c_T), B_T = o_p(d_T) \quad \Rightarrow \quad A_T B_T = o_p(c_T d_T)$ and $A_T + B_T = o_p(\max[c_T, d_T])$.
(2) $A_T = O_p(c_T), B_T = O_p(d_T) \quad \Rightarrow \quad A_T B_T = O_p(c_T d_T)$ and $A_T + B_T = O_p(\max[c_T, d_T])$.
(3) $A_T = o_p(c_T), B_T = O_p(d_T) \quad \Rightarrow \quad A_T B_T = o_p(c_T d_T)$.

■

For the next result see, e.g., Fuller (1976, p. 192).

Proposition C.6 (*Taylor's Theorem for Functions of Random Vectors*) Let $y_T = (y_{1T}, \ldots, y_{KT})' = a + O_p(r_T)$ be a K-dimensional random vector sequence, where $r_T = o(1)$, and let $g : \mathbb{R}^K \to \mathbb{R}$ be a function with continuous partial derivatives of order two at $a = (a_1, \ldots, a_K)'$. Then

$$g(y_T) = g(a) + \frac{\partial g(a)}{\partial y'}(y_T - a) + O_p(r_T^2).$$

If g has continuous partial derivatives of order three,

$$g(y_T) = g(a) + \frac{\partial g(a)}{\partial y'}(y_T - a) + \frac{1}{2}(y_T - a)' \frac{\partial^2 g(a)}{\partial y \partial y'}(y_T - a) + O_p(r_T^3).$$

■

C.3 Infinite Sums of Random Variables

The MA representation of a VAR process is often an infinite sum of random vectors. As in the study of infinite sums of real numbers, we must specify what we mean by such an infinite sum. The concept of absolute convergence is basic in the following. A doubly infinite sequence of real numbers $\{a_i\}$, $i = 0, \pm 1, \pm 2, \ldots$, is *absolutely summable* if

$$\lim_{n \to \infty} \sum_{i=-n}^{n} |a_i|$$

exists and is finite. The limit is usually denoted by

$$\sum_{i=-\infty}^{\infty} |a_i|.$$

The following theorem provides a justification for working with infinite sums of random variables. A proof may be found in Fuller (1976, pp. 29-31).

Proposition C.7 (*Existence of Infinite Sums of Random Variables*) Suppose $\{a_i\}$ is an absolutely summable sequence of real numbers and $\{z_t\}$, $t = 0, \pm 1, \pm 2, \ldots$, is a sequence of random variables satisfying

$$E(z_t^2) \leq c, \quad t = 0, \pm 1, \pm 2, \ldots,$$

for some finite constant c. Then there exists a sequence of random variables $\{y_t\}$, $t = 0, \pm 1, \pm 2, \ldots$, such that

$$\sum_{i=-n}^{n} a_i z_{t-i} \xrightarrow[n\to\infty]{q.m.} y_t$$

and, thus,

$$\operatorname*{plim}_{n\to\infty} \sum_{i=-n}^{n} a_i z_{t-i} = y_t.$$

The random variables y_t are uniquely determined except on a set of probability zero. If, in addition, the z_t are independent random variables, then

$$\sum_{i=-n}^{n} a_i z_{t-i} \xrightarrow{a.s.} y_t.$$

■

This theorem makes precise what we mean by a (univariate) infinite MA

$$y_t = \sum_{i=0}^{\infty} \Phi_i u_{t-i},$$

where u_t is univariate zero mean white noise with variance $\sigma_u^2 < \infty$. Defining $a_i = 0$ for $i < 0$ and $a_i = \Phi_i$ for $i \geq 0$ and assuming that $\{a_i\}$ is absolutely summable, the proposition guarantees that the process y_t is uniquely defined as a limit in mean square, except on a set of probability zero. The latter qualification may be ignored for practical purposes because we may always change a random variable on a set of probability zero without changing its probability characteristics. The requirement for the MA coefficients to be absolutely summable is satisfied if y_t is a stable AR process. For instance, if $y_t = \alpha y_{t-1} + u_t$ is an AR(1) process, $\Phi_i = \alpha^i$ which is an absolutely summable sequence for $|\alpha| < 1$. With respect to the moments of an infinite sum of random variables the following result holds:

Proposition C.8 (*Moments of Infinite Sums of Random Variables*) Suppose z_t satisfies the conditions of Proposition C.7, $\{a_i\}$ and $\{b_i\}$ are absolutely summable sequences of real numbers,

$$y_t = \sum_{i=-\infty}^{\infty} a_i z_{t-i}, \quad \text{and} \quad x_t = \sum_{i=-\infty}^{\infty} b_i z_{t-i}.$$

Then

$$E(y_t) = \lim_{n\to\infty} \sum_{i=-n}^{n} a_i E(z_{t-i})$$

and

$$E(y_t x_t) = \lim_{n\to\infty} \sum_{i=-n}^{n} \sum_{j=-n}^{n} a_i b_j E(z_{t-i} z_{t-j})$$

and, in particular,

$$E(y_t^2) = \lim_{n\to\infty} \sum_{i=-n}^{n} \sum_{j=-n}^{n} a_i a_j E(z_{t-i} z_{t-j}).$$

■

Proof: Fuller (1976, pp. 32-33). ■

All these concepts and results may be extended to vector processes. A sequence of $(K \times K)$ matrices $\{A_i = (a_{mn,i})\}$, $i = 0, \pm 1, \pm 2, \ldots$, is *absolutely summable* if each sequence $\{a_{mn,i}\}$, $m, n = 1, \ldots, K$; $i = 0, \pm 1, \pm 2, \ldots$, is absolutely summable. Equivalently, $\{A_i\}$ may be defined to be absolutely summable if the sequence $\{\|A_i\|\}$ is summable, where

$$\|A_i\| = [\mathrm{tr}(A_i A_i')]^{1/2} = \left(\sum_m \sum_n a_{mn,i}^2 \right)^{1/2}$$

is the Euclidean norm of A_i. To see the equivalence of the two definitions, note that

$$|a_{mn,i}| \le \|A_i\| \le \sum_m \sum_n |a_{mn,i}|.$$

Hence,

$$\sum_{i=-\infty}^{\infty} |a_{mn,i}| \tag{C.3.1}$$

exists and is finite if

$$\sum_{i=-\infty}^{\infty} \|A_i\| \tag{C.3.2}$$

is finite. In turn, if (C.3.1) is finite for all m, n, then, for all h,

$$\sum_{i=-h}^{h} \|A_i\| \leq \sum_{i=-h}^{h} \sum_{m} \sum_{n} |a_{mn,i}|$$

so that (C.3.2) is finite. Thus, the two definitions are indeed equivalent.

Proposition C.9 (*Existence of Infinite Sums of Random Vectors*) Suppose $\{A_i\}$ is an absolutely summable sequence of real $(K \times K)$ matrices and $\{z_t\}$ is a sequence of K-dimensional random variables satisfying

$$E(z_t' z_t) \leq c, \quad t = 0, \pm 1, \pm 2, \ldots,$$

for some finite constant c. Then there exists a sequence of K-dimensional random variables $\{y_t\}$ such that

$$\sum_{i=-n}^{n} A_i z_{t-i} \xrightarrow[n \to \infty]{q.m.} y_t.$$

The sequence is uniquely determined except on a set of probability zero. ■

Proof: Analogous to Fuller (1976, pp. 29-31); replace the absolute value by $\|\cdot\|$. ■

This proposition ensures that the infinite MA representations of the VAR processes considered in this text are well-defined because it can be shown that the MA coefficient matrices of a stable VAR process form an absolutely summable sequence. With respect to moments of infinite sums, we have the following result.

Proposition C.10 (*Moments of Infinite Sums of Random Vectors*) Suppose z_t satisfies the conditions of Proposition C.9, $\{A_i\}$ and $\{B_i\}$ are absolutely summable sequences of $(K \times K)$ matrices,

$$y_t = \sum_{i=-\infty}^{\infty} A_i z_{t-i} \quad \text{and} \quad x_t = \sum_{i=-\infty}^{\infty} B_i z_{t-i}.$$

Then

$$E(y_t) = \lim_{n \to \infty} \sum_{i=-n}^{n} A_i E(z_{t-i})$$

and

$$E(y_t x_t') = \lim_{n \to \infty} \sum_{i=-n}^{n} \sum_{j=-n}^{n} A_i E(z_{t-i} z_{t-j}') B_j',$$

where the limit of the sequence of matrices is the matrix of limits of the sequences of individual elements. ■

Proof: Along similar lines as the proof of Fuller (1976, Theorem 2.2.2, pp. 32-33). ■

While we have restricted the discussion to absolutely summable sequences of coefficients, it may be worth mentioning that infinite sums of random variables and vectors can be defined in more general terms.

C.4 Laws of Large Numbers and Central Limit Theorems

The derivation of asymptotic properties of estimators and test statistics is largely based on *laws of large numbers* (LLNs) and *central limit theorems* (CLTs) some examples of which are listed in the following. So-called *weak LLNs* specify conditions under which a sample mean converges in probability to the population mean and *strong LLNs* state the corresponding results for almost sure convergence.

In stating some of the results, martingale difference processes are useful tools. Suppose $\{x_t\}$ $(t = 1, 2, \dots)$ is a sequence of zero mean random variables and let Ω_t be an information set available at time t which includes at least $\{x_1, \dots, x_t\}$ and possibly other random variables. The sequence $\{x_t\}$ is said to be a *martingale difference sequence with respect to the sequence* Ω_t if $E(x_t|\Omega_{t-1}) = 0$ for all $t = 2, 3, \dots$. It is simply referred to as *martingale difference sequence* if $E(x_t) = 0$ for $t = 1, 2, \dots$, and $E(x_t|x_{t-1}, \dots, x_1) = 0$ for $t = 2, 3, \dots$. More generally, a sequence $\{x_t\}$ of K-dimensional vector random variables satisfying $E(x_t) = 0$ for all t and $E(x_t|x_{t-1}, \dots, x_1) = 0$ for $t = 2, 3, \dots$, is a *vector martingale difference sequence.*

It is sometimes useful to allow the x_t's to depend on the sample size. This way a different sequence for each sample size T is obtained. Denoting by $x_{T,t}$ the t-th element of the T-th sequence, not just a sequence but an *array* of random variables $\{x_{T,t}\}$ $(t = 1, 2, \dots, T; T = 1, 2, \dots)$ is obtained. Such an array is called a *martingale difference array* if $E(x_{T,t}) = 0$ for all t and T and $E(x_{T,t}|x_{T,t-1}, \dots, x_{T,1}) = 0$ for all t and $T > 1$. This definition also applies for vector arrays.

The following inequality is a useful device for deriving asymptotic results. It is therefore presented here (see, e.g., Fuller (1976, Theorem 5.1.1)).

Proposition C.11 (*Chebyshev's Inequality*)
Given $r \in \mathbb{N}$, $r > 0$, let x be a random variable such that $E(|x|^r)$ exists. Then, for any $c \in \mathbb{R}$ and $\epsilon > 0$,

$$\Pr\{|x - c| \geq \epsilon\} \leq \frac{E(|x - c|^r)}{\epsilon^r}.$$

■

The next proposition collects some weak LLNs (see, e.g., Davidson (1994, Part IV)).

Proposition C.12 (*Weak Laws of Large Numbers*)

(1) (Khinchine's Theorem) (Rao (1973, p. 112))
Let $\{x_t\}$ be a sequence of i.i.d. random variables with $E(x_t) = \mu < \infty$. Then

$$\overline{x}_T := \frac{1}{T}\sum_{t=1}^{T} x_t \xrightarrow{p} \mu.$$

(2) Let $\{x_t\}$ be a sequence of independent random variables with $E(x_t) = \mu < \infty$ and $E|x_t|^{1+\epsilon} \leq c < \infty$ $(t = 1, 2, \dots)$ for some $\epsilon > 0$ and a finite constant c. Then $\overline{x}_T \xrightarrow{p} \mu$.
(3) (Chebyshev's Theorem) (Rao (1973, p. 112))
Let $\{x_t\}$ be a sequence of uncorrelated random variables with $E(x_t) = \mu < \infty$ and $\lim_{T\to\infty} E(\overline{x}_T - \mu)^2 = 0$. Then $\overline{x}_T \xrightarrow{p} \mu$.
(4) (Corollary to Chebyshev's Theorem)
Let $\{x_t\}$ be a sequence of independent random variables with $E(x_t) = \mu < \infty$ and $\mathrm{Var}(x_t) \leq c < \infty$ $(t = 1, 2, \dots)$ for some finite constant c. Then $\overline{x}_T \xrightarrow{p} \mu$.
(5) (LLN for Martingale Differences)
Let $\{x_t\}$ be a strictly stationary martingale difference sequence with $E|x_t| < \infty$ $(t = 1, 2, \dots)$. Then $\overline{x}_T \xrightarrow{p} 0$.
(6) (LLN for Martingale Difference Arrays)
Let $\{x_{T,t}\}$ be a martingale difference array with $E|x_{T,t}|^{1+\epsilon} \leq c < \infty$ for all t and T for some $\epsilon > 0$ and a finite constant c. Then $\overline{x}_T := T^{-1}\sum_{t=1}^{T} x_{T,t} \xrightarrow{p} 0$.
(7) (Stationary Processes) (Hamilton (1994, Proposition 7.5))
Let $\{x_t\}$ be a stationary stochastic process with $E(x_t) = \mu < \infty$ and $E[(x_t - \mu)(x_{t-j} - \mu)] = \gamma_j$ $(t = 1, 2, \dots)$ such that $\sum_{j=0}^{\infty} |\gamma_j| < \infty$. Then $\overline{x}_T \xrightarrow{q.m.} \mu$ and, hence, $\overline{x}_T \xrightarrow{p} \mu$, and $\lim_{T\to\infty} TE(\overline{x}_T - \mu)^2 = \sum_{j=-\infty}^{\infty} \gamma_j$.

■

Notice that the i.i.d. assumption in Khinchine's theorem may be replaced by the requirement that moments exist of order larger than one. In fact, Chebyshev's theorem even requires the existence of second order moments. It is actually sufficient that the variances of the x_t are bounded. It may be worth noting that heterogenous variances are allowed for the weak LLN to hold, if the variances are bounded. The last result in the proposition shows that uncorrelated elements of the sequence under consideration are not required. Actually, a martingale difference sequence does not necessarily have independent elements so that for most of the above results independence of the sequence elements is not assumed.

Notice, that the proposition generalizes straightforwardly to sequences of random vectors because convergence in probability for a sequence of random vectors is defined in terms of convergence of the sequences of the individual

elements. The following CLTs are stated for vector sequences and, of course, hold for univariate sequences as special cases.

Proposition C.13 (*Central Limit Theorems*)

(1) (Lindeberg-Levy CLT)
Let $\{x_t\}$ be a sequence of K-dimensional i.i.d. random vectors with mean μ and covariance matrix Σ_x. Then

$$\sqrt{T}(\overline{x}_T - \mu) \xrightarrow{d} \mathcal{N}(0, \Sigma_x).$$

(2) (CLT for Martingale Difference Arrays) (see Hamilton (1994, Proposition 7.9))
Let $\{x_{T,t} = (x_{1T,t}, \ldots, x_{KT,t})'\}$ be a K-dimensional martingale difference array with covariance matrices $E(x_{T,t}x'_{T,t}) = \Sigma_{Tt}$ such that $T^{-1}\sum_{t=1}^{T} \Sigma_{Tt} \to \Sigma$, where Σ is positive definite. Moreover, suppose that $T^{-1}\sum_{t=1}^{T} x_{T,t}x'_{T,t} \xrightarrow{p} \Sigma$ and $E(x_{iT,t}x_{jT,t}x_{kT,t}x_{lT,t}) < \infty$ for all t and T and all $1 \le i, j, k, l \le K$. Then

$$\sqrt{T}\,\overline{x}_T \xrightarrow{d} \mathcal{N}(0, \Sigma).$$

(3) (CLT for Stationary Processes)
Let $x_t = \mu + \sum_{j=0}^{\infty} \Phi_j u_{t-j}$ be a K-dimensional stationary stochastic process with $E(x_t) = \mu < \infty$, $\sum_{j=0}^{\infty} \|\Phi_j\| < \infty$ and $u_t \sim (0, \Sigma_u)$ i.i.d. white noise. Then

$$\sqrt{T}(\overline{x}_T - \mu) \xrightarrow{d} \mathcal{N}\left(0, \sum_{j=-\infty}^{\infty} \Gamma_x(j)\right),$$

where $\Gamma_x(j) := E[(x_t - \mu)(x_{t-j} - \mu)']$.

■

The results in Proposition C.13 are just examples of useful CLTs. A variety of similar results exists for different sets of conditions. More discussion of CLTs and proofs can be found in Davidson (1994, Part V). For the CLT for stationary processes see Anderson (1971, Chapters 7 and 8).

To derive the asymptotic distribution of a vector sequence it is actually sufficient to consider univariate series. This is a consequence of the following result.

Proposition C.14 (*Cramér-Wold Device*) (Rao (1973, p. 123))
Let x_T be a K-dimensional sequence of random vectors and x a K-dimensional random vector. If $c'x_T \xrightarrow{d} c'x$ for all $c \in \mathbb{R}^K$, then $x_T \xrightarrow{d} x$. ■

Therefore, to show asymptotic normality of a sequence, $\sqrt{T}(\widehat{\beta}_T - \beta) \xrightarrow{d} \mathcal{N}(0, \Sigma)$, it suffices to show for all K-vectors c with $c'\Sigma c \neq 0$,

$$\frac{\sqrt{T}c'(\widehat{\beta}_T - \beta)}{(c'\Sigma c)^{1/2}} \xrightarrow{d} \mathcal{N}(0, 1).$$

Hence, CLTs for univariate series can in fact be used to show multivariate results.

C.5 Standard Asymptotic Properties of Estimators and Test Statistics

Suppose we have a sequence of $(m \times n)$ estimators $\{\widehat{B}_T\}$ for an $(m \times n)$ parameter matrix B, where T denotes the sample sizes (time series lengths) on which the estimators are based. For simplicity we will delete the subscript T in the following and we will mean the sequence of estimators when we use the term "estimator".

The estimator $\widehat{B}$ is *consistent* if plim $\widehat{B} = B$. In the related literature, this type of consistency is sometimes called *weak consistency*. However, in this text, we simply use the term consistency instead. The estimator is *strongly consistent* if $\widehat{B} \xrightarrow{a.s.} B$, and the estimator is *mean square consistent* if $\widehat{B} \xrightarrow{q.m.} B$. By Proposition C.1, both strong consistency and mean square consistency imply consistency.

Let $\widehat{\beta}$ be an estimator (a sequence of estimators) of a $(K \times 1)$ vector β. The estimator is said to have an *asymptotic normal distribution* if $\sqrt{T}(\widehat{\beta} - \beta)$ converges in distribution to a random vector with multivariate normal distribution $\mathcal{N}(0, \Sigma)$, that is,

$$\sqrt{T}(\widehat{\beta} - \beta) \xrightarrow{d} \mathcal{N}(0, \Sigma). \tag{C.5.1}$$

In that case, for large T, $\mathcal{N}(\beta, \Sigma/T)$ is usually used as an approximation to the distribution of $\widehat{\beta}$. Equivalently, by the Cramér-Wold device (Proposition C.14), (C.5.1) may be defined by requiring that

$$\frac{\sqrt{T}c'(\widehat{\beta} - \beta)}{(c'\Sigma c)^{1/2}} \xrightarrow{d} \mathcal{N}(0, 1),$$

for any $(K \times 1)$ vector c for which $c'\Sigma c \neq 0$. The following proposition provides some useful rules for determining the asymptotic distributions of estimators and test statistics.

Proposition C.15 (*Asymptotic Properties of Estimators*)
Suppose $\widehat{\beta}$ is an estimator of the $(K \times 1)$ vector β with $\sqrt{T}(\widehat{\beta} - \beta) \xrightarrow{d} \mathcal{N}(0, \Sigma)$. Then the following rules hold:

(1) If plim $\widehat{A} = A$, then $\sqrt{T}\widehat{A}(\widehat{\beta} - \beta) \xrightarrow{d} \mathcal{N}(0, A\Sigma A')$ (see Schmidt (1976, p. 251)).
(2) If $R \neq 0$ is an $(M \times K)$ matrix, then $\sqrt{T}(R\widehat{\beta} - R\beta) \xrightarrow{d} \mathcal{N}(0, R\Sigma R')$.
(3) (Delta method)
If $g(\beta) = (g_1(\beta), \ldots, g_m(\beta))'$ is a vector-valued continuously differentiable function with $\partial g/\partial\beta' \neq 0$ at β, then

$$\sqrt{T}[g(\widehat{\beta}) - g(\beta)] \xrightarrow{d} \mathcal{N}\left(0, \frac{\partial g(\beta)}{\partial \beta'} \Sigma \frac{\partial g(\beta)'}{\partial \beta}\right).$$

If $\partial g/\partial\beta' = 0$ at β, $\sqrt{T}[g(\widehat{\beta}) - g(\beta)] \xrightarrow{p} 0$. (See Serfling (1980, pp. 122-124)).
(4) If Σ is nonsingular, $T(\widehat{\beta} - \beta)'\Sigma^{-1}(\widehat{\beta} - \beta) \xrightarrow{d} \chi^2(K)$.
(5) If Σ is nonsingular and plim $\widehat{\Sigma} = \Sigma$, then $T(\widehat{\beta} - \beta)'\widehat{\Sigma}^{-1}(\widehat{\beta} - \beta) \xrightarrow{d} \chi^2(K)$.
(6) If $\Sigma = QA$, where Q is symmetric, idempotent of rank n and A is positive definite, then $T(\widehat{\beta} - \beta)'A^{-1}(\widehat{\beta} - \beta) \xrightarrow{d} \chi^2(n)$.

■

C.6 Maximum Likelihood Estimation

Suppose $y_1, y_2, \ldots$ is a sequence of K-dimensional random vectors, the first T of which have a joint probability density function $f_T(y_1, \ldots, y_T; \delta_0)$, where δ_0 is an unknown $(M \times 1)$ vector of parameters that does not depend on T. It is assumed to be from a subset $\mathbb{D}$ of the M-dimensional Euclidean space $\mathbb{R}^M$. Suppose further that $f_T(\cdot; \delta)$ has a known functional form and one wishes to estimate δ_0.

For a fixed realization $y_1, \ldots, y_T$, the function

$$l(\delta) = l(\delta | y_1, \ldots, y_T) = f_T(y_1, \ldots, y_T; \delta),$$

viewed as a function of δ, is the *likelihood function*. Its natural logarithm $\ln l(\delta|\cdot)$ is the *log-likelihood function*. A vector $\widetilde{\delta}$, maximizing the likelihood function or log-likelihood function, is called a *maximum likelihood* (ML) *estimate*, that is, if

$$l(\widetilde{\delta}) = \sup_{\delta \in \mathbb{D}} l(\delta),$$

then $\widetilde{\delta}$ is an ML estimate. Here sup denotes the supremum, that is, the least upper bound, which may exist even if the maximum does not. In general, $\widetilde{\delta}$ depends on $y_1, \ldots, y_T$, that is, $\widetilde{\delta} = \widetilde{\delta}(y_1, \ldots, y_T)$. Replacing the fixed values $y_1, \ldots, y_T$ by their corresponding random vectors, $\widetilde{\delta}$ is an *ML estimator* of δ_0 if the functional dependence on $y_1, \ldots, y_T$ is such that $\widetilde{\delta}$ is a random vector.

If $l(\delta)$ is a differentiable function of δ, the vector of first order partial derivatives of $\ln l(\delta)$, that is,

$$s(\delta) = \partial \ln l(\delta)/\partial\delta,$$

regarded as a random vector (a function of the random vectors $y_1, \ldots, y_T$), is the *score vector*. It vanishes at $\delta = \widetilde{\delta}$ if the maximum of $\ln l(\delta)$ is attained at an interior point of the parameter space $\mathbb{D}$. The *information matrix* for δ_0 is minus the expectation of the matrix of second order partial derivatives of $\ln l$, evaluated at the true parameter vector δ_0,

$$\mathcal{I}(\delta_0) = -E\left[\left.\frac{\partial^2 \ln l}{\partial\delta\partial\delta'}\right|_{\delta_0}\right].$$

The matrix

$$\mathcal{I}_a(\delta_0) = \lim_{T\to\infty} \mathcal{I}(\delta_0)/T,$$

if it exists, is the *asymptotic information matrix* for δ_0. If it is nonsingular, its inverse is a lower bound for the covariance matrix of the asymptotic distribution of any consistent estimator with asymptotic normal distribution. In other words, if $\widehat{\delta}$ is a consistent estimator of δ_0 with

$$\sqrt{T}(\widehat{\delta} - \delta_0) \xrightarrow{d} \mathcal{N}(0, \Sigma_{\widehat{\delta}}),$$

then $\mathcal{I}_a(\delta_0)^{-1} \leq \Sigma_{\widehat{\delta}}$, that is, $\Sigma_{\widehat{\delta}} - \mathcal{I}_a(\delta_0)^{-1}$ is positive semidefinite. Under quite general regularity conditions, an ML estimator $\widetilde{\delta}$ for δ_0 is consistent and

$$\sqrt{T}(\widetilde{\delta} - \delta_0) \xrightarrow{d} \mathcal{N}(0, \mathcal{I}_a(\delta_0)^{-1}).$$

Thus, in large samples, $\widetilde{\delta}$ is approximately distributed as $\mathcal{N}(\delta_0, \mathcal{I}_a(\delta_0)^{-1}/T)$.

C.7 Likelihood Ratio, Lagrange Multiplier, and Wald Tests

Three principles for constructing tests of statistical hypotheses are employed frequently in the text. We consider testing of

$$H_0 : \varphi(\delta_0) = 0 \quad \text{against} \quad H_1 : \varphi(\delta_0) \neq 0, \tag{C.7.1}$$

where δ_0 is the true $(M \times 1)$ parameter vector, as in the previous section, and $\varphi : \mathbb{R}^M \to \mathbb{R}^N$ is a continuously differentiable function so that $\varphi(\delta)$ is of dimension $(N \times 1)$. We assume that $[\partial\varphi/\partial\delta'|_{\delta_0}]$ has rank N. This condition implies that $N \leq M$ and the N restrictions for the parameter vector are

distinguishable in a neighborhood of δ_0. Often the hypotheses can be written alternatively as

$$H_0 : \delta_0 = g(\gamma_0) \quad \text{against} \quad H_1 : \delta_0 \neq g(\gamma_0), \tag{C.7.2}$$

where γ_0 is an $(M - N)$-dimensional vector and $g : \mathbb{R}^{M-N} \to \mathbb{R}^M$ is a continuously differentiable function in a neighborhood of γ_0 (see Gallant (1987, pp. 57-58)).

The *likelihood ratio* (LR) *test* of (C.7.1) or (C.7.2) is based on the statistic

$$\lambda_{LR} = 2[\ln l(\widetilde{\delta}) - \ln l(\widetilde{\delta}_r)],$$

where $\widetilde{\delta}$ denotes the unconstrained ML estimator and $\widetilde{\delta}_r$ is the restricted ML estimator of δ_0, subject to the restrictions specified under H_0, that is, $\widetilde{\delta}_r$ is obtained by maximizing $\ln l$ over the parameter space restricted by the conditions stated in H_0. Under suitable regularity conditions, we have

$$\lambda_{LR} \xrightarrow{d} \chi^2(N). \tag{C.7.3}$$

The *Lagrange multiplier* (LM) *statistic* for testing (C.7.1) or (C.7.2) is of the form

$$\lambda_{LM} = s(\widetilde{\delta}_r)' \mathcal{I}(\widetilde{\delta}_r)^{-1} s(\widetilde{\delta}_r), \tag{C.7.4}$$

where $s(\delta)$ denotes the score vector and $\mathcal{I}(\delta)$ the information matrix, as before. In the LM statistic, both functions are evaluated at the restricted estimator of δ_0. Under H_0, λ_{LM} has an asymptotic $\chi^2(N)$-distribution, if weak regularity conditions are satisfied. The name derives from the fact that it can be written as

$$\lambda_{LM} = \widetilde{\lambda}' \left[\frac{\partial \varphi}{\partial \delta'} \Big|_{\widetilde{\delta}_r} \right] \mathcal{I}(\widetilde{\delta}_r)^{-1} \left[\frac{\partial \varphi'}{\partial \delta} \Big|_{\widetilde{\delta}_r} \right] \widetilde{\lambda}, \tag{C.7.5}$$

where $\widetilde{\lambda}$ is the vector of Lagrange multipliers for which the Lagrange function has a stationary point corresponding to the constrained estimator (see Appendix A.14).

The equivalence of (C.7.4) and (C.7.5) can be seen by recalling that the constrained minimum of $-\ln l$ is attained at a stationary point of the Lagrange function

$$\mathcal{L}(\delta, \lambda) = -\ln l(\delta) + \lambda' \varphi(\delta).$$

In other words, $\widetilde{\delta}_r$ satisfies

$$0 = \left[\frac{\partial \mathcal{L}}{\partial \delta'} \Big|_{\widetilde{\delta}_r, \widetilde{\lambda}} \right] = - \left[\frac{\partial \ln l}{\partial \delta'} \Big|_{\widetilde{\delta}_r} \right] + \widetilde{\lambda}' \left[\frac{\partial \varphi}{\partial \delta'} \Big|_{\widetilde{\delta}_r} \right] = -s(\widetilde{\delta}_r)' + \widetilde{\lambda}' \left[\frac{\partial \varphi}{\partial \delta'} \Big|_{\widetilde{\delta}_r} \right].$$

The LM statistic is often computed via an auxiliary regression. To see how this can be done, consider a normal regression model of the form

$$\mathbf{y} = X\boldsymbol{\beta} + Z\boldsymbol{\gamma} + \mathbf{u},$$

where $\mathbf{y}$ and $\mathbf{u}$ are $(T \times 1)$ vectors, X and Z are $(T \times M)$ and $(T \times N)$ regressor matrices, respectively, $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ are $(M \times 1)$ and $(N \times 1)$ parameter vectors and $\mathbf{u} \sim \mathcal{N}(0, \sigma_u^2 I_T)$. Suppose we wish to test the pair of hypotheses

$$H_0 : \boldsymbol{\gamma} = 0 \quad \text{versus} \quad H_1 : \boldsymbol{\gamma} \neq 0.$$

In this case, the score vector is

$$s\begin{pmatrix} \boldsymbol{\beta} \\ \boldsymbol{\gamma} \end{pmatrix} = \frac{1}{\sigma_u^2}\begin{bmatrix} X' \\ Z' \end{bmatrix}(\mathbf{y} - X\boldsymbol{\beta} - Z\boldsymbol{\gamma}),$$

the inverse information matrix is

$$\sigma_u^2 \begin{bmatrix} X'X & X'Z \\ Z'X & Z'Z \end{bmatrix}^{-1},$$

and the restricted estimator is

$$\begin{bmatrix} \widehat{\boldsymbol{\beta}} \\ 0 \end{bmatrix} = \begin{bmatrix} (X'X)^{-1}X'\mathbf{y} \\ 0 \end{bmatrix}.$$

Notice that the first order conditions for computing this estimator imply

$$X'(y - X\widehat{\boldsymbol{\beta}}) = X'\widehat{\mathbf{u}} = 0.$$

Here $\widehat{\mathbf{u}} := y - X\widehat{\boldsymbol{\beta}}$ is the residual vector of the restricted estimation. Hence, the score vector evaluated at the restricted estimator is

$$s\begin{pmatrix} \widehat{\boldsymbol{\beta}} \\ 0 \end{pmatrix} = \frac{1}{\sigma_u^2}\begin{bmatrix} X'(y - X\widehat{\boldsymbol{\beta}}) \\ Z'(y - X\widehat{\boldsymbol{\beta}}) \end{bmatrix} = \frac{1}{\sigma_u^2}\begin{bmatrix} 0 \\ Z'\widehat{\mathbf{u}} \end{bmatrix}$$

and the LM statistic becomes

$$\begin{aligned} \lambda_{LM} &= [0 : \widehat{\mathbf{u}}'Z]\begin{bmatrix} X'X & X'Z \\ Z'X & Z'Z \end{bmatrix}^{-1}\begin{bmatrix} 0 \\ Z'\widehat{\mathbf{u}} \end{bmatrix}/\sigma_u^2 \\ &= \widehat{\mathbf{u}}'Z(Z'Z - Z'X(X'X)^{-1}X'Z)^{-1}Z'\widehat{\mathbf{u}}/\sigma_u^2, \end{aligned}$$

where the rules for the partitioned inverse have been used (see Appendix A.10).

The same statistic is obtained by using the usual χ^2-statistic for testing $\boldsymbol{\gamma} = 0$ in the auxiliary regression model

$$\widehat{\mathbf{u}} = X\boldsymbol{\beta} + Z\boldsymbol{\gamma} + \mathbf{e},$$

where $\mathbf{e}$ is an error vector. The LS estimator from this model is

$$\begin{bmatrix} \widetilde{\boldsymbol{\beta}} \\ \widetilde{\boldsymbol{\gamma}} \end{bmatrix} = \begin{bmatrix} X'X & X'Z \\ Z'X & Z'Z \end{bmatrix}^{-1} \begin{bmatrix} X'\widehat{\mathbf{u}} \\ Z'\widehat{\mathbf{u}} \end{bmatrix}.$$

Using $X'\widehat{\mathbf{u}} = 0$ and the rules for the partitioned inverse gives

$$\begin{aligned} \widetilde{\boldsymbol{\gamma}} &= (Z'Z - Z'X(X'X)^{-1}X'Z)^{-1}Z'\widehat{\mathbf{u}} \\ &\sim \mathcal{N}(\boldsymbol{\gamma}, \sigma_u^2(Z'Z - Z'X(X'X)^{-1}X'Z)^{-1}). \end{aligned}$$

Hence, the χ^2-statistic

$$\widetilde{\boldsymbol{\gamma}}'(Z'Z - Z'X(X'X)^{-1}X'Z)\widetilde{\boldsymbol{\gamma}}/\sigma_u^2$$

is easily seen to be identical to the previously obtained expression for λ_{LM}. Of course, algebraically the same result is obtained if σ_u^2 is replaced by an estimator. Using the usual modifications, the statistic has an F-distribution in this case. More precisely,

$$\frac{\widetilde{\boldsymbol{\gamma}}'(Z'Z - Z'X(X'X)^{-1}X'Z)\widetilde{\boldsymbol{\gamma}}}{N\widehat{\sigma}_u^2} \sim F(N, T - M - N).$$

Although we have used a normal regression model with nonstochastic regressors in this illustration, a similar reasoning often applies for more general situations and it implies an auxiliary regression model from which the LM statistic can be obtained. The reason is that much of the derivation rests on the algebraic properties of the quantities involved. Therefore, similar arguments can be used, for example, if the regressors are stochastic or a GLS estimation is used. In Chapters 4 and 5, the LM statistics for residual autocorrelation in VAR models are, for instance, derived in this way.

The *Wald statistic* is based on an unconstrained estimator which is asymptotically normal,

$$\sqrt{T}(\widetilde{\delta} - \delta_0) \xrightarrow{d} \mathcal{N}(0, \Sigma_{\widetilde{\delta}}).$$

By Proposition C.15(3), it follows that

$$\sqrt{T}[\varphi(\widetilde{\delta}) - \varphi(\delta_0)] \xrightarrow{d} \mathcal{N}\left(0, \left[\frac{\partial \varphi}{\partial \delta'}\bigg|_{\delta_0}\right] \Sigma_{\widetilde{\delta}} \left[\frac{\partial \varphi'}{\partial \delta}\bigg|_{\delta_0}\right]\right).$$

Thus, by Proposition C.15(5), if $H_0 : \varphi(\delta_0) = 0$ is true and the covariance matrix is invertible,

$$\lambda_W = T\varphi(\widetilde{\delta})' \left(\left[\frac{\partial \varphi}{\partial \delta'}\bigg|_{\widetilde{\delta}}\right] \widetilde{\Sigma}_{\widetilde{\delta}} \left[\frac{\partial \varphi'}{\partial \delta}\bigg|_{\widetilde{\delta}}\right]\right)^{-1} \varphi(\widetilde{\delta}) \xrightarrow{d} \chi^2(N), \tag{C.7.6}$$

where $\widetilde{\Sigma}_{\widetilde{\delta}}$ is a consistent estimator of $\Sigma_{\widetilde{\delta}}$. The statistic λ_W is the Wald statistic. For further discussion of the three test statistics and proofs of their asymptotic distributions see also Hayashi (2000, Chapter 7).

In summary, we have three test statistics with equivalent asymptotic distributions under the null hypothesis. The LR statistic involves both the restricted and the unrestricted ML estimators, the LM statistic is based on the restricted estimator only, and the Wald statistic requires just the unrestricted estimator. The choice among the three statistics is often based on computational convenience. Wald tests have the disadvantage that they are not invariant under transformations of the restrictions. In other words, if the restrictions can be written in two equivalent ways (e.g., $\delta_i = 0$ and $\delta_i^2 = 0$) the corresponding Wald tests may have different small sample properties. Their small sample power may be low (see Gregory & Veall (1985), Breusch & Schmidt (1988)).

C.8 Unit Root Asymptotics

C.8.1 Univariate Processes

In deriving asymptotic results for processes with unit roots, it is helpful to consider also continuous stochastic processes. An important example is a *standard Brownian motion* or a *standard Wiener process* $\mathbf{W}(\cdot)$ which is a function defined on the unit interval $[0, 1]$ and it assigns a random variable $\mathbf{W}(t)$ to each $t \in [0, 1]$ such that the following conditions hold:

(1) $\mathbf{W}(0) = 0$ with probability one.
(2) $\mathbf{W}(t)$ is continuous in t with probability one.
(3) For any partitioning of the unit interval, $0 \leq t_1 < t_2 < \cdots < t_k \leq 1$, the vector

$$\begin{bmatrix} \mathbf{W}(t_2) - \mathbf{W}(t_1) \\ \vdots \\ \mathbf{W}(t_k) - \mathbf{W}(t_{k-1}) \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix}, \begin{bmatrix} t_2 - t_1 & & 0 \\ & \ddots & \vdots \\ 0 & \dots & t_k - t_{k-1} \end{bmatrix} \right),$$

that is, the differences have a multivariate normal distribution with independent components, means of zero, and variances $t_i - t_{i-1}$.

Wiener processes play an important role in the asymptotic theory for unit root processes. Nonstandard versions of the type $Z(t) = \sigma \mathbf{W}(t)$ are often encountered. Their increments are still independent but $Z(t) - Z(s) \sim \mathcal{N}(0, \sigma^2(t-s))$ for $s < t$. Notice also that $Z(t) \sim \mathcal{N}(0, \sigma^2 t)$.

In developing unit root asymptotics, we are often interested in quantities of the form

$$X_T(r) = \frac{1}{T} \sum_{t=1}^{[Tr]} w_t,$$

where w_t is a stationary stochastic process, $r \in [0, 1]$ denotes a fraction and $[Tr]$ signifies the largest integer less than or equal to Tr. If the $w_t = u_t$ are i.i.d. $(0, \sigma_u^2)$, we know from a central limit theorem (see Proposition C.13) that

$$\sqrt{T}X_T(r) = \frac{\sqrt{[Tr]}}{\sqrt{T}} \frac{1}{\sqrt{[Tr]}} \sum_{t=1}^{[Tr]} u_t \xrightarrow{d} \mathcal{N}(0, r\sigma_u^2),$$

for every $r \in [0,1]$, because $\sqrt{[Tr]}/\sqrt{T} \to \sqrt{r}$. Moreover,

$$\sqrt{T}[X_T(r_2) - X_T(r_1)]/\sigma_u \xrightarrow{d} \mathcal{N}(0, r_2 - r_1),$$

for $r_1 < r_2$. For nonoverlapping partitions of the unit interval, the partial sums will be made up of independent terms and they are therefore independent. Hence, it is plausible to write

$$\sqrt{T}X_T(\cdot)/\sigma_u \xrightarrow{d} \mathbf{W}(\cdot). \tag{C.8.1}$$

This notation and result generalizes the previously defined concept of convergence in distribution because now convergence is stated for a sequence of continuous time stochastic processes. The result is often referred to as a *functional central limit theorem* (FCLT) or *invariance principle* or *Donsker's theorem.*

Giving a precise definition of the related concept of convergence in distribution is simplified by considering convergence of probability measures. A sequence of probability measures $\Pr_T$ is said to converge to the probability measure Pr or $\Pr_T$ *converges weakly* to Pr, if $\Pr_T(\mathbb{A}) \to \Pr(\mathbb{A})$ for all measurable sets $\mathbb{A}$, with the exception of sets for which the boundary points have nonzero probability mass. Instead of considering the distribution functions, we may define convergence in distribution via weak convergence of the corresponding sequence of probability measures. Thus, constructing a probability space on a suitable space of functions defined on the unit interval, the convergence in (C.8.1) can be defined rigorously. Although this type of convergence is more properly called weak convergence, we will still use the symbol $\xrightarrow{d}$ for signifying it. For more precise discussions see, for example, Davidson (1994, 2000) or Johansen (1995).

We may also generalize the concept of convergence in probability to the case of sequences of random functions. For a sequence $G_T(\cdot)$ and a random function $G(\cdot)$ we write $G_T \xrightarrow{p} G$ if

$$\sup_{t\in[0,1]} |G_T(t) - G(t)| \xrightarrow{p} 0.$$

Another useful tool in dealing with unit root process is the *continuous mapping theorem* which states that, given a sequence of stochastic functions $\{G_T(\cdot)\}$, a stochastic function $G(\cdot)$ and a continuous functional $g(\cdot)$ (a function defined on a space of functions), we have

$$G_T \xrightarrow{d} G \quad \Rightarrow \quad g(G_T) \xrightarrow{d} g(G).$$

Using the FCLT, this theorem implies, for instance, that

$$\int_0^1 \sqrt{T} X_T(r)dr \xrightarrow{d} \sigma_u \int_0^1 \mathbf{W}(r)dr$$

because the integral is a continuous functional.

These tools are useful in proving the following proposition from Hamilton (1994, Proposition 17.1) which summarizes a number of helpful results from the literature, many of which were derived, e.g., by Phillips (1987).

Proposition C.16 (*Properties of Random Walks and Related Quantities*) Suppose $x_t = x_{t-1} + u_t$ is a random walk with i.i.d. white noise, $u_t \sim (0, \sigma_u^2)$, and $x_0 = 0$. Then the following results hold:

(1) $T^{-1/2} \sum_{t=1}^T u_t \xrightarrow{d} \sigma_u \mathbf{W}(1) = \mathcal{N}(0, \sigma_u^2)$.
(2) $T^{-1} \sum_{t=1}^T x_{t-1} u_t \xrightarrow{d} \frac{1}{2}\sigma_u^2[\mathbf{W}(1)^2 - 1] = \frac{1}{2}\sigma_u^2[\chi^2(1) - 1]$.
(3) $T^{-3/2} \sum_{t=1}^T t u_t \xrightarrow{d} \sigma_u \mathbf{W}(1) - \sigma_u \int_0^1 \mathbf{W}(r)dr = \mathcal{N}(0, \sigma_u^2/3)$.
(4) $T^{-3/2} \sum_{t=1}^T x_{t-1} \xrightarrow{d} \sigma_u \int_0^1 \mathbf{W}(r)dr = \mathcal{N}(0, \sigma_u^2/3)$.
(5) $T^{-2} \sum_{t=1}^T x_{t-1}^2 \xrightarrow{d} \sigma_u^2 \int_0^1 \mathbf{W}(r)^2 dr$.
(6) $T^{-5/2} \sum_{t=1}^T t x_{t-1} \xrightarrow{d} \sigma_u \int_0^1 r\mathbf{W}(r)dr$.
(7) $T^{-3} \sum_{t=1}^T t x_{t-1}^2 \xrightarrow{d} \sigma_u^2 \int_0^1 r\mathbf{W}(r)^2 dr$.
(8) $T^{-(n+1)} \sum_{t=1}^T t^n \to 1/(n+1)$ for $n = 0, 1, \ldots$.

■

From these results, the following asymptotic distributions of Dickey-Fuller (DF) statistics for unit roots can be derived. For details see, e.g., Hamilton (1994, Section 17.4). It is assumed that estimation is based on a sample $y_1, \ldots, y_T$ and a presample value y_0 is also available.

Proposition C.17 (*Asymptotic Distributions of Dickey-Fuller Test Statistics*)

(1) Suppose $\widehat{\rho} = \sum_{t=1}^T y_{t-1} y_t / \sum_{t=1}^T y_{t-1}^2$ is the LS estimator of the coefficient ρ of the AR(1) process $y_t = \rho y_{t-1} + u_t$, $t = 1, 2, \ldots$, where $u_t \sim (0, \sigma_u^2)$ is i.i.d. white noise. Here y_0 is a fixed starting value or a stochastic variable with a given fixed distribution (which does not depend on the sample size). Then, if $\rho = 1$,

$$T(\widehat{\rho} - 1) \xrightarrow{d} \frac{\frac{1}{2}[\mathbf{W}(1)^2 - 1]}{\int_0^1 \mathbf{W}(r)^2 dr}$$

and the t-statistic

$$t_{\widehat{\rho}-1} = \frac{\widehat{\rho} - 1}{\widehat{\sigma}_{\widehat{\rho}}} \xrightarrow{d} \frac{\frac{1}{2}[\mathbf{W}(1)^2 - 1]}{\left[\int_0^1 \mathbf{W}(r)^2 dr\right]^{1/2}},$$

where $\widehat{\sigma}_{\widehat{\rho}}^2 = T^{-1} \sum_{t=1}^T (y_t - \widehat{\rho} y_{t-1})^2 / \sum_{t=1}^T y_{t-1}^2$ is the usual LS estimator of the variance of $\widehat{\rho}$.

(2) Suppose $y_t = \mu + x_t$, $t = 1, 2, \ldots$, with $x_t = \rho x_{t-1} + u_t$, where $u_t \sim (0, \sigma_u^2)$ is i.i.d. white noise and μ is a fixed mean term. Moreover, let $x_0 = 0$ and y_0 be a fixed starting value or a stochastic variable with a given fixed distribution. Furthermore, $\widehat{\rho}$ is the LS estimator of ρ from a regression $y_t = \nu + \rho y_{t-1} + u_t$. Then, if $\rho = 1$,

$$T(\widehat{\rho} - 1) \xrightarrow{d} \frac{\frac{1}{2}[\mathbf{W}(1)^2 - 1] - \mathbf{W}(1) \int_0^1 \mathbf{W}(r) dr}{\int_0^1 \mathbf{W}(r)^2 dr - \left[\int_0^1 \mathbf{W}(r) dr\right]^2}$$

and the t-statistic

$$t_{\widehat{\rho}-1} = \frac{\widehat{\rho} - 1}{\widehat{\sigma}_{\widehat{\rho}}} \xrightarrow{d} \frac{\frac{1}{2}[\mathbf{W}(1)^2 - 1] - \mathbf{W}(1) \int_0^1 \mathbf{W}(r) dr}{\left\{\int_0^1 \mathbf{W}(r)^2 dr - \left[\int_0^1 \mathbf{W}(r) dr\right]^2\right\}^{1/2}},$$

where $\widehat{\sigma}_{\widehat{\rho}}^2$ is the usual LS estimator of the variance of $\widehat{\rho}$.

(3) Suppose $y_t = \nu + y_{t-1} + u_t$, $t = 1, 2, \ldots$, where $u_t \sim (0, \sigma_u^2)$ is i.i.d. white noise and $\nu \neq 0$ is a constant term. Moreover, let y_0 be a fixed starting value or a stochastic variable with a given fixed distribution. Furthermore, $\widehat{\rho}$ is the LS estimator of ρ from a regression $y_t = \nu + \rho y_{t-1} + u_t$. Then, if $\rho = 1$,

$$T^{3/2}(\widehat{\rho} - 1) \xrightarrow{d} \mathcal{N}(0, 12\sigma_u^2/\nu^2)$$

and the t-statistic

$$t_{\widehat{\rho}-1} = \frac{\widehat{\rho} - 1}{\widehat{\sigma}_{\widehat{\rho}}} \xrightarrow{d} \mathcal{N}(0, 1),$$

where $\widehat{\sigma}_{\widehat{\rho}}^2$ is the usual LS estimator of the variance of $\widehat{\rho}$.

(4) Suppose $y_t = \mu_0 + \mu_1 t + x_t$, $t = 1, 2, \ldots$, with $x_t = \rho x_{t-1} + u_t$, where $u_t \sim (0, \sigma_u^2)$ is i.i.d. white noise and μ_0 and μ_1 are fixed intercept and trend slope terms. Moreover, let $x_0 = 0$ and y_0 be a fixed starting value or a stochastic variable with a given fixed distribution. Furthermore, $\widehat{\rho}$ is the LS estimator of ρ from a regression $y_t = \nu_0 + \nu_1 t + \rho y_{t-1} + u_t$. Then, if $\rho = 1$,

$$T(\widehat{\rho} - 1) \xrightarrow{d} a/b$$

and

$$t_{\widehat{\rho}-1} = \frac{\widehat{\rho} - 1}{\widehat{\sigma}_{\widehat{\rho}}} \xrightarrow{d} a/\sqrt{b},$$

where

$$
\begin{aligned}
a &= \int_0^1 \mathbf{W}(r)d\mathbf{W}(r) \\
&\quad +12\left[\int_0^1 r\mathbf{W}(r)dr - \frac{1}{2}\int_0^1 \mathbf{W}(r)dr\right]\left[\int_0^1 \mathbf{W}(r)dr - \frac{1}{2}\mathbf{W}(1)\right] \\
&\quad -\mathbf{W}(1)\int_0^1 \mathbf{W}(r)dr
\end{aligned}
$$

and

$$
\begin{aligned}
b &= \int_0^1 \mathbf{W}(r)^2 dr - 12\left(\int_0^1 r\mathbf{W}(r)dr\right)^2 \\
&\quad +12\int_0^1 \mathbf{W}(r)dr\int_0^1 r\mathbf{W}(r)dr - 4\left(\int_0^1 \mathbf{W}(r)dr\right)^2.
\end{aligned}
$$

Furthermore, $\widehat{\sigma}^2_{\widehat{\rho}}$ is the usual LS estimator of the variance of $\widehat{\rho}$.

■

Obviously, most of the asymptotic distributions obtained for $\widehat{\rho}$ are nonstandard if $\rho = 1$. In fact, even the convergence rate of the estimator is nonstandard. It converges at a much faster rate to its true value of 1 than usual estimators based on stationary processes. More precisely, $\widehat{\rho} - \rho = O_p(T^{-1})$ if $\rho = 1$ in Cases 1, 2, and 4 in the proposition, whereas in the stationary case of an AR(1) process $y_t = \rho y_{t-1} + u_t$, say, we have for the LS estimator of ρ, $\widehat{\rho} - \rho = O_p(T^{-1/2})$. The latter rate also holds if y_t is stationary and has a nonzero mean term. In Case 3 of Proposition C.17, the convergence rate of $\widehat{\rho}$ is even larger because in that case the estimator is dominated by the linear trend which is generated by the drift term.

It is important to note that the limiting distributions in Cases 1, 2, and 4 are free of unknown nuisance parameters. Therefore, it is easy to compute percentage points of the limiting distributions by simulation methods. To do that, it is strictly speaking not even necessary to know the exact form of the asymptotic distributions of the estimators. It is sufficient to know that well-defined asymptotic distributions are obtained which do not depend on unknown nuisance parameters. Of course, there are also situations when a more detailed knowledge of the asymptotic distributions and closed form expressions are helpful.

The results of Proposition C.16 can be generalized in different ways. First of all, the process x_t may have a more complicated dependence structure. In particular, the error process may be a stationary process. Consider, for instance, a process $x_t = x_{t-1} + w_t$, where $w_t = \sum_{j=0}^{\infty} \theta_j u_{t-j} = \theta(L)u_t$ is a stationary process with $\sum_{j=0}^{\infty} j|\theta_j| < \infty$ and $u_t \sim (0, \sigma_u^2)$ is white noise, then x_t can be rewritten as

$$
x_t = x_0 + w_1 + \cdots + w_t = x_0 + \theta(1)(u_1 + \cdots + u_t) + \sum_{j=0}^{\infty} \theta_j^* u_{t-j} - w_0^*,
$$

where $\theta(1) = \sum_{j=0}^{\infty} \theta_j$, $\theta_j^* = -\sum_{i=j+1}^{\infty} \theta_i$, $j = 0, 1, \ldots$, and $w_0^* = \sum_{j=0}^{\infty} \theta_j^* u_{-j}$ contains initial values. Thus, x_t is the sum of a random walk, a stationary process and initial values. Note that the condition $\sum_{j=0}^{\infty} j|\theta_j| < \infty$ ensures that $\sum_{j=0}^{\infty} |\theta_j^*| < \infty$, so that $\sum_{j=0}^{\infty} \theta_j^* u_{t-j}$ is indeed well-defined according to Proposition C.7. Although the condition for the θ_j is stronger than absolute summability, it is satisfied for many processes of practical interest. For example, the MA representation of a stable AR or ARMA process satisfies the condition. The decomposition of x_t in a random walk, a stationary component, and initial values is known as the *Beveridge-Nelson decomposition*. It is a convenient tool in generalizing the results in Propositions C.16 and C.17.

In fact, if y_t is a finite order AR process, $y_t = \alpha_1 y_{t-1} + \cdots + \alpha_p y_{t-p} + u_t$, where u_t is again white noise, y_t can be rewritten as

$$y_t = \rho y_{t-1} + \gamma_1 \Delta y_{t-1} + \cdots + \gamma_{p-1} \Delta y_{t-p+1} + u_t$$

or, subtracting y_{t-1} on both sides,

$$\Delta y_t = (\rho - 1) y_{t-1} + \gamma_1 \Delta y_{t-1} + \cdots + \gamma_{p-1} \Delta y_{t-p+1} + u_t.$$

Estimating ρ or $\rho - 1$ from these equations by LS, it turns out that the resulting estimators have the same asymptotic properties as in Proposition C.17 (see, e.g., Hamilton (1994)).

Another possible generalization of these results may be obtained by considering multivariate processes. We will tackle both generalizations at once in the following.

C.8.2 Multivariate Processes

For the present purposes, multivariate Brownian motions or Wiener processes are of central importance. The univariate definition can be generalized as follows. A K-dimensional *standard Brownian motion* or *standard Wiener process* $\mathbf{W}(\cdot)$ is a function defined on the unit interval $[0, 1]$ which assigns a K-dimensional random vector $\mathbf{W}(t)$ to each $t \in [0, 1]$ such that:

(1) $\mathbf{W}(0) = 0$ with probability one.
(2) A realization $\mathbf{W}(t)$ is a continuous function in t on the unit interval with probability one.
(3) For any partitioning of the unit interval, $0 \le t_1 < t_2 < \cdots < t_k \le 1$, the vector

$$\begin{bmatrix} \mathbf{W}(t_2) - \mathbf{W}(t_1) \\ \vdots \\ \mathbf{W}(t_k) - \mathbf{W}(t_{k-1}) \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix}, \begin{bmatrix} (t_2 - t_1) I_K & & 0 \\ & \ddots & \vdots \\ 0 & \ldots & (t_k - t_{k-1}) I_K \end{bmatrix} \right),$$

that is, the differences have multivariate normal distributions with independent components, means of zero, and variances of the form $t_i - t_{i-1}$, depending on their difference in time.

Again, for any nonsingular $(K \times K)$ matrix P, a nonstandard version of a Wiener process $Z(t) := P\mathbf{W}(t)$ is obtained for which the increments are still independent but $Z(t) - Z(s) \sim \mathcal{N}(0, (t-s)PP')$ for $s < t$. Moreover, $Z(t) \sim \mathcal{N}(0, tPP')$.

For a sequence $G_T(\cdot)$ of multivariate random functions, we define convergence in probability to a random function G, $G_T \xrightarrow{p} G$, to hold if

$$\sup_{t\in[0,1]} \|G_T(t) - G(t)\| \xrightarrow[T\to\infty]{p} 0.$$

Also, the continuous mapping theorem remains valid in the multivariate case.

As in the univariate case, it is of interest to consider quantities of the form

$$X_T(r) = \frac{1}{T}\sum_{t=1}^{[Tr]} w_t,$$

where w_t is a stationary stochastic process, $r \in [0,1]$ denotes a fraction and $[Tr]$ signifies the largest integer less than or equal to Tr. If $w_t = u_t \sim (0, \Sigma_u)$ is i.i.d. white noise, it follows from a multivariate version of a suitable CLT (see Proposition C.13) that

$$\sqrt{T}[X_T(r_2) - X_T(r_1)] \xrightarrow{d} \mathcal{N}(0, (r_2 - r_1)\Sigma_u)$$

for $r_1 < r_2$. Hence, using the same ideas as in the univariate case,

$$\sqrt{T}\Sigma_u^{-1/2} X_T(\cdot) \xrightarrow{d} \mathbf{W}(\cdot),$$

which is a multivariate version of the previously stated FCLT also referred to as *invariance principle* or *Donsker's theorem*.

If $x_t = x_{t-1} + w_t$, where

$$w_t = \boldsymbol{\Xi}(L)u_t = \sum_{j=0}^{\infty} \boldsymbol{\Xi}_j u_{t-j}, \quad \text{with} \quad \sum_{j=0}^{\infty} j\|\boldsymbol{\Xi}_j\| < \infty,$$

and $u_t \sim (0, \Sigma_u = (\sigma_{ij}))$ is white noise, then a *multivariate Beveridge-Nelson decomposition* is available,

$$x_t = x_0 + w_1 + \cdots + w_t = x_0 + \boldsymbol{\Xi}(1)\sum_{s=1}^{t} u_s + \sum_{j=0}^{\infty} \boldsymbol{\Xi}_j^* u_{t-j} - w_0^*,$$

where $\boldsymbol{\Xi}(1) = \sum_{j=0}^{\infty} \boldsymbol{\Xi}_j$, $\boldsymbol{\Xi}_j^* = -\sum_{i=j+1}^{\infty} \boldsymbol{\Xi}_i$, $j = 0, 1, \ldots$, and $w_0^* = \sum_{j=0}^{\infty} \boldsymbol{\Xi}_j^* u_{-j}$ contains initial values. Now x_t is a sum of a multivariate random walk, a stationary process, and initial values (see also Proposition 6.1). Using these concepts, the following generalized version of Proposition C.16 can be established. It also goes back to Phillips and others (see Phillips & Durlauf (1986), Park & Phillips (1988, 1989), Phillips & Solo (1992), Sims et al. (1990), Johansen (1995)) and may be found, e.g., in Hamilton (1994, Proposition 18.1).

Proposition C.18 (*Properties of Multivariate Unit Root Processes*) Suppose $x_t = x_{t-1} + w_t$, $t = 1, 2, \ldots$, is a K-dimensional generalized random walk with initial vector $x_0 = 0$ and stationary error term

$$w_t = \Xi(L)u_t = \sum_{j=0}^{\infty} \Xi_j u_{t-j}, \qquad t \in \mathbb{Z},$$

where

$$\sum_{j=0}^{\infty} j\|\Xi_j\| < \infty,$$

and $u_t \sim (0, \Sigma_u = (\sigma_{ij}))$, $t \in \mathbb{Z}$, is i.i.d. white noise with finite fourth moments. Let P be a lower triangular matrix such that $\Sigma_u = PP'$,

$$\Gamma_w(h) := E(w_t w'_{t-h}) = \sum_{j=0}^{\infty} \Xi_{j+h} \Sigma_u \Xi'_j, \qquad h = 0, 1, 2, \ldots,$$

for an arbitrary positive integer n, $W_t := (w'_{t-1}, \ldots, w'_{t-n})'$ is a Kn-dimensional vector with $\Sigma_W := E(W_t W'_t)$, and the $(K \times K)$ matrix $\Lambda := \Xi(1)P$. Then the following results hold:

(1) $T^{-1/2} \sum_{t=1}^{T} w_t \xrightarrow{d} \Lambda \mathbf{W}(1)$.
(2) $T^{-1/2} \sum_{t=1}^{T} W_t u_{it} \xrightarrow{d} \mathcal{N}(0, \sigma_{ii}\Sigma_W)$ for $i = 1, \ldots, K$.
(3) $T^{-1} \sum_{t=1}^{T} w_t w'_{t-h} \xrightarrow{p} \Gamma_w(h)$ for $h = 0, 1, 2, \ldots$.
(4) $T^{-1} \sum_{t=1}^{T} (x_{t-1} w'_{t-h} + w_{t-h} x'_{t-1})$

$$\xrightarrow{d} \begin{cases} \Lambda \mathbf{W}(1)\mathbf{W}(1)'\Lambda' - \Gamma_w(0) & \text{for } h = 0, \\ \Lambda \mathbf{W}(1)\mathbf{W}(1)'\Lambda' - \Gamma_w(0) + \sum_{j=-h+1}^{h-1} \Gamma_w(j) & \text{for } h = 1, 2, \ldots. \end{cases}$$

(5) $T^{-1} \sum_{t=1}^{T} x_{t-1} w'_t \xrightarrow{d} \Lambda \left\{ \int_0^1 \mathbf{W}(r) d\mathbf{W}(r)' \right\} \Lambda' + \sum_{j=1}^{\infty} \Gamma_w(j)$.
(6) $T^{-1} \sum_{t=1}^{T} x_{t-1} u'_t \xrightarrow{d} \Lambda \left\{ \int_0^1 \mathbf{W}(r) d\mathbf{W}(r)' \right\} P'$.
(7) $T^{-3/2} \sum_{t=1}^{T} x_{t-1} \xrightarrow{d} \Lambda \int_0^1 \mathbf{W}(r) dr$.
(8) $T^{-3/2} \sum_{t=1}^{T} t w_{t-h} \xrightarrow{d} \Lambda \left\{ \mathbf{W}(1) - \int_0^1 \mathbf{W}(r) dr \right\}$ for $h = 0, 1, 2, \ldots$.
(9) $T^{-2} \sum_{t=1}^{T} x_{t-1} x'_{t-1} \xrightarrow{d} \Lambda \left\{ \int_0^1 \mathbf{W}(r)\mathbf{W}(r)' dr \right\} \Lambda'$.
(10) $T^{-5/2} \sum_{t=1}^{T} t x_{t-1} \xrightarrow{d} \Lambda \int_0^1 r\mathbf{W}(r) dr$.
(11) $T^{-3} \sum_{t=1}^{T} t x_{t-1} x'_{t-1} \xrightarrow{d} \Lambda \left\{ \int_0^1 r\mathbf{W}(r)\mathbf{W}(r)' dr \right\} \Lambda'$.

■

These results are the basis for much of the asymptotic theory related to multivariate VAR processes with unit roots. Extensions exist for more general processes w_t and u_t.

D

Evaluating Properties of Estimators and Test Statistics by Simulation and Resampling Techniques

If asymptotic theory is difficult or only small samples are available, properties of estimators and test statistics are sometimes investigated by heavy use of the computer. The idea is to simulate the distribution (or some of its properties) of the random variables of interest by artificially sampling from some known distribution. Generally, if the random variable or vector of interest, say $q = q(z)$, is a function of a random vector z with a known distribution F_z, then samples $z_1, \ldots, z_n$ are drawn from F_z and the empirical distribution of q given by $q_n = q(z_n)$, $n = 1, \ldots, N$, is determined. The characteristics of the actual distribution of q are then inferred from the empirical distribution.

Often the statistics of interest in this book are functions of multiple time series generated by VAR(p) processes. Therefore, we will briefly describe in the next section how to simulate such time series. Afterwards, some more details are given on simulation and resampling techniques for evaluating estimators and test statistics.

D.1 Simulating a Multiple Time Series with VAR Generation Process

To simulate a multiple time series of dimension K and length T, we first generate a series of (often independent) disturbance vectors $u_{-s}, \ldots, u_0, u_1, \ldots, u_T$. If a series of Gaussian disturbances is desired, i.e., $u_t \sim \mathcal{N}(0, \Sigma_u)$, we may choose K independent univariate standard normal variates $v_1, \ldots, v_K$ and multiply by a $(K \times K)$ matrix P for which $PP' = \Sigma_u$, that is,

$$u_t = P \begin{bmatrix} v_1 \\ \vdots \\ v_K \end{bmatrix}.$$

This process is repeated $T + s + 1$ times until we have the desired series of disturbances. Programs for generating (pseudo) standard normal variates are

available on most computers. Also facilities for generating random numbers from other distributions are usually available and may be used in a similar manner to obtain disturbances from other distributions of interest.

For a given set of parameters $\nu, A_1, \ldots, A_p$, where ν is $(K \times 1)$ and the A_i are $(K \times K)$, and a given set of starting values $y_{-p+1}, \ldots, y_0$, the u_t may be used to simulate a time series $y_1, \ldots, y_T$ with VAR(p) generation process recursively as

$$y_t = \nu + A_1 y_{t-1} + \cdots + A_p y_{t-p} + u_t$$

starting with $t = 1, t = 2$, etc. until $t = T$. There are different ways to obtain the initial values. Assuming that the desired process is stable, they may be set to zero or to the process mean $\mu = (I_K - A_1 - \cdots - A_p)^{-1}\nu$. Because the choice of initial values has some impact on the generated time series, a number of presample values $y_t, t = -s, \ldots, 0$, is often generated and then discarded in the subsequent analysis.

A possible way to ensure the same correlation structure for the initial values and the rest of the time series is to determine the covariance matrix of p consecutive y_t vectors, say Σ_Y. Using the results of Chapter 2, Section 2.1, that matrix may be obtained from

$$\mathrm{vec}(\Sigma_Y) = (I_{(Kp)^2} - \mathbf{A} \otimes \mathbf{A})^{-1}\mathrm{vec}(\Sigma_U),$$

where

$$\mathbf{A} = \underset{(Kp \times Kp)}{\begin{bmatrix} A_1 & A_2 & \ldots & A_{p-1} & A_p \\ I_K & 0 & \ldots & 0 & 0 \\ 0 & I_K & & 0 & 0 \\ \vdots & & \ddots & \vdots & \vdots \\ 0 & 0 & \ldots & I_K & 0 \end{bmatrix}} \quad \text{and} \quad \Sigma_U = \underset{(Kp \times Kp)}{\begin{bmatrix} \Sigma_u & 0 & \ldots & 0 \\ 0 & 0 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & 0 \end{bmatrix}}.$$

Then a $(Kp \times Kp)$ matrix Q is chosen such that $QQ' = \Sigma_Y$ and p initial starting vectors are obtained as

$$\begin{bmatrix} y_0 \\ \vdots \\ y_{-p+1} \end{bmatrix} = Q \begin{bmatrix} v_1 \\ \vdots \\ v_{Kp} \end{bmatrix} + \begin{bmatrix} \mu \\ \vdots \\ \mu \end{bmatrix},$$

where the v_i are independent variates with mean zero and unit variance.

D.2 Evaluating Distributions of Functions of Multiple Time Series by Simulation

Suppose we are interested in the function $q_T = q(y_1, \ldots, y_T)$ of some VAR(p) process y_t, where q_T is of dimension $(M \times 1)$. The quantity q_T may be some

estimator or test statistic. To investigate the distribution F_T of q_T, we generate a large number, say N, of independent multiple time series of length T and compute the corresponding values of q_T, say $q_T(n)$, $n = 1, \ldots, N$. The properties of F_T are then estimated from the empirical distribution of the $q_T(n)$. For instance, the mean vector of q_T is estimated as

$$\frac{1}{N} \sum_{n=1}^{N} q_T(n).$$

Analogously, we may estimate the variances, standard deviations, quantiles or other characteristics of F_T.

D.3 Resampling Methods

If the distribution of the disturbances of a VAR model under consideration is unknown, so-called *bootstrap* or *resampling* methods may be applied to investigate the distributions of functions of stochastic processes or multiple time series. Suppose a time series $y_1, \ldots, y_T$ and the presample values required for estimation are available. Fitting a VAR(p) model to this time series, we get coefficient estimates $\widehat{\nu}, \widehat{A}_1, \ldots, \widehat{A}_p$, and a series of residuals $\widehat{u}_1, \ldots, \widehat{u}_T$. An estimator of a quantity of interest, say $q = q(A_1, \ldots, A_p)$, is then obtained as

$$\widehat{q} = q(\widehat{A}_1, \ldots, \widehat{A}_p). \tag{D.3.1}$$

The properties of $\widehat{q}$ follow from those of $\widehat{A}_1, \ldots, \widehat{A}_p$. To assess the sampling uncertainty of $\widehat{q}$, confidence intervals are often established, based on the asymptotic distribution of $\widehat{q}$. Alternatively, if q is a test statistic, its p-value may be of interest which can be approximated on the basis of the asymptotic distribution. Unfortunately, this distribution is often a rather poor approximation of the actual distribution for a given finite sample. In some of these cases, bootstrap methods provide a better small sample approximation. The theoretical justification for the bootstrap also rests on asymptotic theory, however. In particular, it can usually be justified if the quantity of interest has a normal limiting distribution (Horowitz (2001)).

A *residual based bootstrap* is often used in this context. Assuming that a sample $y_1, \ldots, y_T$ plus presample values as required are available, it proceeds as follows:

(1) The parameters of the model under consideration are estimated. Let $\widehat{u}_t$, $t = 1, \ldots, T$, be the estimation residuals.
(2) Centered residuals $\widehat{u}_1 - \overline{u}_\cdot, \ldots, \widehat{u}_T - \overline{u}_\cdot$ are computed. Here $\overline{u}_\cdot = T^{-1} \sum \widehat{u}_t$ denotes the usual average. Bootstrap residuals $u_1^*, \ldots, u_T^*$ are then obtained by randomly drawing with replacement from the centered residuals.

(3) Bootstrap time series are computed recursively as

$$y_t^* = \widehat{\nu} + \widehat{A}_1 y_{t-1}^* + \cdots + \widehat{A}_p y_{t-p}^* + u_t^*, \quad t = 1, \ldots, T,$$

where the same initial values may be used for each generated series, $(y_{-p+1}^*, \ldots, y_0^*) = (y_{-p+1}, \ldots, y_0)$.

(4) Based on the bootstrap time series, the parameters $A_1, \ldots, A_p$ are reestimated.

(5) Using the parameter estimates obtained in the previous stage, a bootstrap version of the statistic of interest, say $\widehat{q}^*$, is calculated.

(6) These steps are repeated N times, where N is a large number.

There is now a range of other bootstrap methods which may have advantages in certain situations. For example, rather than using a residual-based bootstrap, a *block bootstrap* may be applied which is based on the original observations rather than the model residuals (see, e.g., Li & Maddala (1996) for details). It may be preferable if there is uncertainty regarding specific aspects of the model like, for instance, the VAR order. These methods are not discussed here because residual based bootstraps are still the most popular methods in the present context.

In the following, the symbol q denotes the quantity of interest for which a confidence interval is desired. Its estimator implied by the estimators of the model coefficients and the corresponding bootstrap estimator are denoted by $\widehat{q}$ and $\widehat{q}^*$, respectively. The following bootstrap confidence intervals are examples that have been considered in the literature in the context of impulse response analysis (see, e.g., Benkwitz, Lütkepohl & Wolters (2001)):

- *Standard percentile interval*
  Denoting by $s_{\gamma/2}^*$ and $s_{(1-\gamma/2)}^*$ the $\gamma/2$- and $(1-\gamma/2)$-quantiles, respectively, of the N bootstrap versions of $\widehat{q}^*$, the interval

$$CI_S = \left[s_{\gamma/2}^*, s_{(1-\gamma/2)}^* \right],$$

  may be set up. It is the percentile confidence interval discussed, e.g., by Efron & Tibshirani (1993).
- *Hall's percentile interval*
  Hall (1992) uses the result that asymptotically the distribution of $\sqrt{T}(\widehat{q}-q)$ corresponds to that of $\sqrt{T}(\widehat{q}^* - \widehat{q})$, to derive the interval

$$CI_H = \left[\widehat{q} - t_{(1-\gamma/2)}^*, \widehat{q} - t_{\gamma/2}^* \right].$$

  Here $t_{\gamma/2}^*$ and $t_{(1-\gamma/2)}^*$ are the $\gamma/2$- and $(1-\gamma/2)$-quantiles, respectively, of $(\widehat{q}^* - \widehat{q})$ and the interval is obtained by pretending that these are the quantiles of $(\widehat{q} - q)$.
- *Hall's studentized interval*
  A studentized statistic $(\widehat{q} - q)/(\widehat{\mathrm{Var}}(\widehat{q}))^{1/2}$ often results in more precise

confidence intervals at least in theory. Using bootstrap quantiles $t^{**}_{\gamma/2}$ and $t^{**}_{(1-\gamma/2)}$ from the distribution of $(\widehat{q}^* - \widehat{q})/(\widehat{\mathrm{Var}}(\widehat{q}^*))^{1/2}$, an interval

$$CI_{SH} = \left[\widehat{q} - t^{**}_{(1-\gamma/2)}\sqrt{\widehat{\mathrm{Var}}(\widehat{q})},\ \widehat{q} - t^{**}_{\gamma/2}\sqrt{\widehat{\mathrm{Var}}(\widehat{q})}\right]$$

can be constructed by using these quantities in conjunction with $(\widehat{q} - q)/(\widehat{\mathrm{Var}}(\widehat{q}))^{1/2}$. Here the variance $\mathrm{Var}(\widehat{q})$ may be estimated from the bootstrap estimates of q,

$$\widehat{\mathrm{Var}}(\widehat{q}) = \frac{1}{N-1}\sum_{i=1}^{N}\left(\widehat{q}^{*,i} - \overline{\widehat{q}^*}\right)^2,$$

where N is the number of bootstrap replications and $\widehat{q}^{*,i}$ denotes the value of the statistic of interest obtained in the i-th bootstrap replication. Moreover, the variances $\widehat{\mathrm{Var}}(\widehat{q}^*)$ may be estimated by a bootstrap within each bootstrap replication. In other words,

$$\widehat{\mathrm{Var}}(\widehat{q}^*) = \frac{1}{N^*-1}\sum_{i=1}^{N^*}\left(\widehat{q}^{**,i} - \overline{\widehat{q}^{**}}\right)^2,$$

where $\widehat{q}^{**,i}$ is obtained by a double bootstrap, that is, pseudo-data are generated according to a process obtained on the basis of the bootstrap systems parameters and N^* is the number of bootstrap replications within each bootstrap replication.

A number of refinements and modifications of these intervals exist (see Hall (1992)).

The bootstrap confidence intervals have the property that they attain the nominal confidence content at least asymptotically under general conditions. Roughly speaking, if $\sqrt{T}(\widehat{q} - q)$ converges as $T \to \infty$, $\sqrt{T}(\widehat{q}^* - \widehat{q})$ converges to the same limit distribution under suitable conditions (e.g., Hall (1992)). Therefore CI_H has the correct size asymptotically, that is, $\Pr(q \in CI_H) \to 1 - \gamma$ as $T \to \infty$, under general conditions, and, hence, Hall's percentile method is asymptotically precise. The same holds for the CI_{SH} interval. On the other hand, to obtain such a result for the standard percentile interval CI_S, the limiting distribution of $\sqrt{T}(\widehat{q} - q)$ has to be symmetric about zero. For example, this result holds if it is zero mean normal. Roughly speaking, CI_S works with an implicit asymptotic unbiasedness assumption for $\widehat{q}$. If the distribution of $\widehat{q}$ is not centered at q, CI_S will generally not have the desired confidence content even asymptotically (see also Efron & Tibshirani (1993) and Benkwitz et al. (2000) for a more detailed discussion of this point).

If $\widehat{q}$ is a statistic for which a p-value is desired, the following method may be used. Recall that the p-value of a test is the probability of obtaining a value of the test statistic greater than the observed one, if the null hypothesis holds.

Hence, the p-value may be estimated by the proportion of bootstrap values $\widehat{q}^*$ exceeding the value of the test statistic $\widehat{q}$. Again, under general assumptions, this estimator is consistent.

References

Abraham, B. (1980). Intervention analysis and multiple time series, *Biometrika* **67**: 73–78.

Ahn, S. K. (1988). Distribution for residual autocovariances in multivariate autoregressive models with structured parameterization, *Biometrika* **75**: 590–593.

Ahn, S. K. & Reinsel, G. C. (1988). Nested reduced-rank autoregressive models for multiple time series, *Journal of the American Statistical Association* **83**: 849–856.

Ahn, S. K. & Reinsel, G. C. (1990). Estimation of partially nonstationary multivariate autoregressive models, *Journal of the American Statistical Association* **85**: 813–823.

Akaike, H. (1969). Fitting autoregressive models for prediction, *Annals of the Institute of Statistical Mathematics* **21**: 243–247.

Akaike, H. (1971). Autoregressive model fitting for control, *Annals of the Institute of Statistical Mathematics* **23**: 163–180.

Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle, *in* B. N. Petrov & F. Csáki (eds), *2nd International Symposium on Information Theory*, Académiai Kiadó, Budapest, pp. 267–281.

Akaike, H. (1974). A new look at the statistical model identification, *IEEE Transactions on Automatic Control* **AC-19**: 716–723.

Akaike, H. (1976). Canonical correlation analysis of time series and the use of an information criterion, *in* R. K. Mehra & D. G. Lainiotis (eds), *Systems Identification: Advances and Case Studies*, Academic Press, New York, pp. 27–96.

Amisano, G. & Giannini, C. (1997). *Topics in Structural VAR Econometrics*, 2nd edn, Springer, Berlin.

Anděl, J. (1983). Statistical analysis of periodic autoregression, *Aplikace Mathematiky* **28**: 364–385.

Anděl, J. (1987). On multiple periodic autoregression, *Aplikace Mathematiky* **32**: 63–80.

Anděl, J. (1989). On nonlinear models for time series, *Statistics* **20**: 615–632.

Anderson, B. D. O. & Moore, J. B. (1979). *Optimal Filtering*, Prentice-Hall, Englewood Cliffs, NJ.

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

Anderson, T. W. (1984). *An Introduction to Multivariate Statistical Analysis*, 2nd edn, John Wiley, New York.

Anderson, T. W. (1999). Asymptotic distribution of the reduced rank regression estimator under general conditions, *Annals of Statistics* **27**: 1141–1154.

Anderson, T. W. (2002). Canonical correlation analysis and reduced rank regression in autoregressive models, *Annals of Statistics* **30**: 1134–1154.

Andrews, D. W. K. (1987). Asymptotic results for generalized Wald tests, *Econometric Theory* **3**: 348–358.

Ansley, C. F. & Kohn, R. (1983). Exact likelihood of vector autoregressive-moving average process with missing or aggregated data, *Biometrika* **70**: 275–278.

Aoki, M. (1987). *State Space Modeling of Time Series*, Springer, Berlin.

Baba, Y., Engle, R. F., Kraft, D. F. & Kroner, K. F. (1990). Multivariate simultaneous generalized ARCH, *mimeo*, Department of Economics, University of California, San Diego.

Bai, J. (1994). Least squares estimation of a shift in linear processes, *Journal of Time Series Analysis* **15**: 453–472.

Bai, J., Lumsdaine, R. L. & Stock, J. H. (1998). Testing for and dating common breaks in multivariate time series, *Review of Economic Studies* **65**: 395–432.

Baillie, R. T. (1981). Prediction from the dynamic simultaneous equation model with vector autoregressive errors, *Econometrica* **49**: 1331–1337.

Baillie, R. T. & Bollerslev, T. (1992). Prediction in dynamic models with time-dependent conditional variances, *Journal of Econometrics* **52**: 91–113.

Banerjee, A., Dolado, J. J., Galbraith, J. W. & Hendry, D. F. (1993). *Co-integration, Error-Correction, and the Econometric Analysis of Non-stationary Data*, Oxford University Press, Oxford.

Baringhaus, L. & Henze, N. (1988). A consistent test for multivariate normality based on the empirical characteristic function, *Metrika* **35**: 339–348.

Barone, P. (1987). A method for generating independent realizations of a multivariate normal stationary and invertible ARMA(p, q) process, *Journal of Time Series Analysis* **8**: 125–130.

Bartel, H. & Lütkepohl, H. (1998). Estimating the Kronecker indices of cointegrated echelon form VARMA models, *Econometrics Journal* **1**: C76–C99.

Basu, A. K. & Sen Roy, S. (1986). On some asymptotic results for multivariate autoregressive models with estimated parameters, *Calcutta Statistical Association Bulletin* **35**: 123–132.

Basu, A. K. & Sen Roy, S. (1987). On asymptotic prediction problems for multivariate autoregressive models in the unstable nonexplosive case, *Calcutta Statistical Association Bulletin* **36**: 29–37.

Bauer, D. & Wagner, M. (2002). Estimating cointegrated systems using subspace algorithms, *Journal of Econometrics* **111**: 47–84.

Bauer, D. & Wagner, M. (2003). A canonical form for unit root processes in the state space framework, Diskussionsschriften 03-12, Universität Bern.

Bauwens, L., Laurent, S. & Rombouts, J. V. K. (2004). Multivariate GARCH models: A survey, CORE Discussion Paper 2003/31.

Benkwitz, A., Lütkepohl, H. & Neumann, M. (2000). Problems related to bootstrapping impulse responses of autoregressive processes, *Econometric Reviews* **19**: 69–103.

Benkwitz, A., Lütkepohl, H. & Wolters, J. (2001). Comparison of bootstrap confidence intervals for impulse responses of German monetary systems, *Macroeconomic Dynamics* **5**: 81–100.

Bera, A. K. & Higgins, M. L. (1993). ARCH models: Properties, estimation and testing, *Journal of Economic Surveys* **7**: 305–366.

Berk, K. N. (1974). Consistent autoregressive spectral estimates, *Annals of Statistics* **2**: 489–502.

Bernanke, B. (1986). Alternative explanations of the money–income correlation, *Carnegie-Rochester Conference Series on Public Policy*, North-Holland, Amsterdam.

Bewley, R. & Yang, M. (1995). Tests for cointegration based on canonical correlation analysis, *Journal of the American Statistical Association* **90**: 990–996.

Bhansali, R. J. (1978). Linear prediction by autoregressive model fitting in the time domain, *Annals of Statistics* **6**: 224–231.

Bierens, H. J. (1981). *Robust Methods and Asymptotic Theory in Nonlinear Econometrics*, Springer, Berlin.

Bierens, H. J. (1997). Nonparametric cointegration analysis, *Journal of Econometrics* **77**: 379–404.

Black, F. (1976). Studies in stock price volatility changes, *Proceedings of the 1976 Meeting of the Business and Economic Statistics Section*, American Statistical Association, pp. 177–181.

Blanchard, O. & Quah, D. (1989). The dynamic effects of aggregate demand and supply disturbances, *American Economic Review* **79**: 655–673.

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity, *Journal of Econometrics* **31**: 307–327.

Bollerslev, T. (1990). Modeling the coherence in short-run nominal exchange rates: A multivariate generalized ARCH model, *Review of Economics and Statistics* **72**: 498–505.

Bollerslev, T., Chou, R. Y. & Kroner, K. F. (1992). ARCH modelling in finance: A review of the theory and empirical evidence, *Journal of Econometrics* **52**: 5–59.

Bollerslev, T., Engle, R. F. & Nelson, D. B. (1994). ARCH models, *in* R. Engle & D. McFadden (eds), *Handbook of Econometrics*, Vol. 4, Elsevier, Amsterdam, pp. 2959–3038.

Bollerslev, T., Engle, R. F. & Wooldridge, J. M. (1988). A capital asset pricing model with time-varying covariances, *Journal of Political Economy* **96**: 116–131.

Bossaerts, P. (1988). Common nonstationary components of asset prices, *Journal of Economic Dynamics and Control* **12**: 347–364.

Boswijk, H. P. & Doornik, J. A. (2002). Identifying, estimating and testing restricted cointegrated systems: An overview, *Paper presented at the Henri Theil Memorial Conference*, Universiteit van Amsterdam.

Boswijk, H. P. & Franses, P. H. (1995). Periodic cointegration: Representation and inference, *Review of Economics and Statistics* **77**: 436–454.

Boswijk, H. P. & Franses, P. H. (1996). Unit roots in periodic autoregressions, *Journal of Time Series Analysis* **17**: 221–245.

Boswijk, H. P., Franses, P. H. & Haldrup, N. (1997). Multiple unit roots in periodic autoregression, *Journal of Econometrics* **80**: 167–193.

Boswijk, H. P. & Lucas, A. (2002). Semi-nonparametric cointegration testing, *Journal of Econometrics* **108**: 253–280.

Box, G. E. P. & Jenkins, G. M. (1976). *Time Series Analysis: Forecasting and Control*, Holden-Day, San Francisco.

Box, G. E. P. & Tiao, G. C. (1975). Intervention analysis with applications to economic and environmental problems, *Journal of the American Statistical Association* **70**: 70–79.

Braun, P. A., Nelson, D. B. & Sunier, A. M. (1995). Good news, bad news, volatility, and betas, *Journal of Finance* **50**: 1575–1603.

Breitung, J., Brüggemann, R. & Lütkepohl, H. (2004). Structural vector autoregressive modeling and impulse responses, *in* H. Lütkepohl & M. Krätzig (eds), *Applied Time Series Econometrics*, Cambridge University Press, Cambridge, pp. 159–196.

Breitung, J. & Hassler, U. (2002). Inference on the cointegration rank in fractionally integrated processes, *Journal of Econometrics* **110**: 167–185.

Breusch, T. S. (1978). Testing for autocorrelation in dynamic linear models, *Australian Economic Papers* **17**: 334–355.

Breusch, T. S. & Schmidt, P. (1988). Alternative forms of the Wald test: How long is a piece of string, *Communications in Statistics, Theory and Methods* **17**: 2789–2795.

Brockwell, P. J. & Davis, R. A. (1987). *Time Series: Theory and Methods*, Springer, New York.

Brooks, C. & Burke, S. (2003). Information criteria for GARCH model selection, *European Journal of Finance* **9**: 557–580.

Brooks, C., Burke, S. & Persand, G. (2003). Multivariate GARCH models: Software choice and estimation issues, *Journal of Applied Econometrics* **18**: 725–734.

Brüggemann, R. (2004). *Model Reduction Methods for Vector Autoregressive Processes*, Springer, Berlin.

Brüggemann, R. & Lütkepohl, H. (2001). Lag selection in subset VAR models with an application to a U.S. monetary system, *in* R. Friedmann, L. Knüppel & H. Lütkepohl (eds), *Econometric Studies: A Festschrift in Honour of Joachim Frohn*, LIT Verlag, Münster, pp. 107–128.

Brüggemann, R. & Lütkepohl, H. (2004). Practical problems with reduced rank ML estimators for cointegration parameters and a simple alternative, *Discussion paper*, European University Institute, Florence.

Brüggemann, R., Lütkepohl, H. & Saikkonen, P. (2004). Residual autocorrelation testing for vector error correction models, *Discussion paper*, European University Institute, Florence.

Caines, P. E. (1988). *Linear Stochastic Systems*, John Wiley, New York.

Candelon, B. & Lütkepohl, H. (2001). On the reliability of Chow-type tests for parameter constancy in multivariate dynamic models, *Economics Letters* **73**: 155–160.

Caner, M. (1998). Tests for cointegration with infinite variance errors, *Journal of Econometrics* **86**: 155–175.

Caporale, G. M. & Pittis, N. (2004). Estimator choice and Fisher's paradox: A Monte Carlo study, *Econometric Reviews* **23**: 25–52.

Cheung, Y.-W. & Ng, L. K. (1996). A causality-in-variance test and its application to financial market prices, *Journal of Econometrics* **72**: 33–48.

Chitturi, R. V. (1974). Distribution of residual autocorrelations in multiple autoregressive schemes, *Journal of the American Statistical Association* **69**: 928–934.

Choi, I. (1994). Residual-based tests for the null of stationarity with applications to U.S. macroeconomic time series, *Econometric Theory* **10**: 720–749.

Chow, G. C. (1975). *Analysis and Control of Dynamic Economic Systems*, John Wiley, New York.

Chow, G. C. (1981). *Econometric Analysis by Control Methods*, John Wiley, New York.

Chow, G. C. (1984). Random and changing coefficient models, *in* Z. Griliches & M. D. Intriligator (eds), *Handbook of Econometrics*, Vol. 2, North-Holland, Amsterdam, pp. 1213–1245.

Christiano, L. J., Eichenbaum, M. & Evans, C. (1996). The effects of monetary policy shocks: Evidence from the flow of funds, *Review of Economics and Statistics* **78**: 16–34.

Cipra, T. (1985). Periodic moving average process, *Aplikace Matematiky* **30**: 218–229.

Cleveland, W. P. & Tiao, G. C. (1979). Modeling seasonal time series, *Economie Appliquée* **32**: 107–129.

Comte, F. & Lieberman, O. (2000). Second-order noncausality in multivariate GARCH processes, *Journal of Time Series Analysis* **21**: 535–557.

Comte, F. & Lieberman, O. (2003). Asymptotic theory for multivariate GARCH processes, *Journal of Multivariate Analysis* **84**: 61–84.

Cooley, T. F. & Prescott, E. (1973). Varying parameter regression: A theory and some applications, *Annals of Economic and Social Measurement* **2**: 463–474.

Cooley, T. F. & Prescott, E. (1976). Estimation in the presence of stochastic parameter variation, *Econometrica* **44**: 167–184.

Cooper, D. M. & Wood, E. F. (1982). Identifying multivariate time series models, *Journal of Time Series Analysis* **3**: 153–164.

Crowder, M. J. (1976). Maximum-likelihood estimation for dependent observations, *Journal of the Royal Statistical Society* **B38**: 45–53.

Davidson, J. (1994). *Stochastic Limit Theory*, Oxford University Press, Oxford.

Davidson, J. (2000). *Econometric Theory*, Blackwell, Oxford.

Davidson, J. E. H., Hendry, D. F., Srba, F. & Yeo, S. (1978). Econometric modelling of the aggregate time-series relationship between consumer's expenditure and income in the United Kingdom, *Economic Journal* **88**: 661–692.

Davies, N., Triggs, C. M. & Newbold, P. (1977). Significance levels of the Box-Pierce portmanteau statistic in finite samples, *Biometrika* **64**: 517–522.

Deistler, M. & Hannan, E. J. (1981). Some properties of the parameterization of ARMA systems with unknown order, *Journal of Multivariate Analysis* **11**: 474–484.

Deistler, M. & Pötscher, B. M. (1984). The behaviour of the likelihood function for ARMA models, *Advances in Applied Probability* **16**: 843–865.

Dempster, A. P., Laird, N. M. & Rubin, D. B. (1977). Maximum-likelihood from incomplete data via the EM-algorithm, *Journal of the Royal Statistical Society* **B39**: 1–38.

Diebold, F. X. & Nerlove, M. (1989). Dynamic exchange rate volatility: A multivariate latent factor ARCH model, *Journal of Applied Econometrics* **4**: 1–21.

Doan, T., Litterman, R. B. & Sims, C. A. (1984). Forecasting and conditional projection using realistic prior distributions, *Econometric Reviews* **3**: 1–144.

Dolado, J. J. & Lütkepohl, H. (1996). Making Wald tests work for cointegrated VAR systems, *Econometric Reviews* **15**: 369–386.

Doornik, J. A. (1996). Testing vector error autocorrelation and heteroscedasticity, unpublished paper, Nuffield College.

Doornik, J. A. & Hansen, H. (1994). A practical test of multivariate normality, unpublished paper, Nuffield College.

Doornik, J. A. & Hendry, D. F. (1997). *Modelling Dynamic Systems Using PcFiml 9.0 for Windows*, International Thomson Business Press, London.

Doornik, J. A., Hendry, D. F. & Nielsen, B. (1998). Inference in cointegrating models: UK M1 revisited, *Journal of Economic Surveys* **12**: 533–572.

Dufour, J.-M. (1985). Unbiasedness of predictions from estimated vector autoregressions, *Econometric Theory* **1**: 387–402.

Dufour, J.-M. & Renault, E. (1998). Short run and long run causality in time series: Theory, *Econometrica* **66**: 1099–1125.

Dufour, J.-M. & Roy, R. (1985). Some robust exact results on sample autocorrelations and tests of randomness, *Journal of Econometrics* **29**: 257–273.

Dunsmuir, W. T. M. & Hannan, E. J. (1976). Vector linear time series models, *Advances in Applied Probability* **8**: 339–364.

Edgerton, D. & Shukur, G. (1999). Testing autocorrelation in a system perspective, *Econometric Reviews* **18**: 343–386.

Efron, B. & Tibshirani, R. J. (1993). *An Introduction to the Bootstrap*, Chapman & Hall, New York.

Eichenbaum, M. & Evans, C. (1995). Some empirical evidence on the effects of shocks to monetary policy on exchange rates, *Quarterly Journal of Economics* **110**: 975–1009.

Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation, *Econometrica* **50**: 987–1007.

Engle, R. F. (2002). Dynamic conditional correlation: A simple class of multivariate generalized autoregressive conditional heteroscedasticity models, *Journal of Business & Economic Statistics* **20**: 339–350.

Engle, R. F. & Granger, C. W. J. (1987). Co-integration and error correction: Representation, estimation and testing, *Econometrica* **55**: 251–276.

Engle, R. F., Granger, C. W. J. & Kraft, D. (1986). Combining competing forecasts of inflation using a bivariate ARCH model, *Journal of Economic Dynamics and Control* **8**: 151–165.

Engle, R. F., Hendry, D. F. & Richard, J. F. (1983). Exogeneity, *Econometrica* **51**: 277–304.

Engle, R. F. & Kroner, K. F. (1995). Multivariate simultaneous generalized GARCH, *Econometric Theory* **11**: 122–150.

Engle, R. F., Lilien, D. M. & Robins, R. P. (1987). Estimating time varying risk premia in the term structure: The ARCH-M model, *Econometrica* **55**: 391–407.

Engle, R. F. & Ng, V. G. (1993). Measuring and testing the impact of news on volatility, *Journal of Finance* **48**: 1749–1778.

Engle, R. F. & Watson, M. (1981). A one-factor multivariate time series model of metropolitan wage rates, *Journal of the American Statistical Association* **76**: 774–781.

Engle, R. F. & Yoo, B. S. (1987). Forecasting and testing in co-integrated systems, *Journal of Econometrics* **35**: 143–159.

Ericsson, N. R. (1994). Testing exogeneity: An introduction, *in* N. R. Ericsson (ed.), *Testing Exogeneity*, Oxford University Press, Oxford, pp. 3–38.

Fiorentini, G., Sentana, E. & Calzolari, G. (2004). On the validity of the Jarque-Bera normality test in conditionally heteroskedastic dynamic regression models, *Economics Letters* **83**: 307–312.

Fisher, L. A. & Huh, H. (1999). Weak exogeneity and long-run and contemporaneous identifying restrictions in VEC models, *Economics Letters* **63**: 159–165.

Forni, M., Hallin, M., Lippi, M. & Reichlin, L. (2000). The generalized dynamic factor model: Identification and estimation, *Review of Economics and Statistics* **82**: 540–552.

Friedmann, R. (1981). The reliability of policy recommendations and forecasts from linear econometric models, *International Economic Review* **22**: 415–428.

Fuller, W. A. (1976). *Introduction to Statistical Time Series*, John Wiley, New York.

Galí, J. (1992). How well does the IS-LM model fit postwar U.S. data?, *Quarterly Journal of Economics* **107**: 709–738.

Galí, J. (1999). Technology, employment, and the business cycle: Do technology shocks explain aggregate fluctuations?, *American Economic Review* **89**: 249–271.

Gallant, A. R. (1987). *Nonlinear Statistical Models*, John Wiley, New York.

Gallant, A. R., Rossi, P. E. & Tauchen, G. (1993). Nonlinear dynamic structures, *Econometrica* **61**: 871–907.

Gallant, A. R. & White, H. (1988). *A Unified Theory of Estimation and Inference for Nonlinear Dynamic Models*, Blackwell, Oxford.

Gasser, T. (1975). Goodness-of-fit tests for correlated data, *Biometrika* **62**: 563–570.

Geweke, J. (1977). The dynamic factor analysis of economic time-series models, *in* D. J. Aigner & A. S. Goldberger (eds), *Latent Variables in Socio-Economic Models*, North-Holland, New York, pp. 365–383.

Geweke, J. (1982). Causality, exogeneity, and inference, *in* W. Hildenbrand (ed.), *Advances in Econometrics*, Cambridge University Press, Cambridge, pp. 209–235.

Geweke, J. (1984). Inference and causality in economic time series models, *in* Z. Griliches & M. D. Intriligator (eds), *Handbook of Econometrics*, Vol. 2, North-Holland, Amsterdam, pp. 1101–1144.

Geweke, J., Meese, R. & Dent, W. (1983). Comparing alternative tests of causality in temporal systems: Analytic results and experimental evidence, *Journal of Econometrics* **21**: 161–194.

Ghysels, E., Harvey, A. & Renault, E. (1996). Stochastic volatility, *in* G. S. Maddala & C. R. Rao (eds), *Handbook of Statistics*, Elsevier, Amsterdam, pp. 119–191.

Ghysels, E. & Osborn, D. R. (2001). *The Econometric Analysis of Seasonal Time Series*, Cambridge University Press, Cambridge.

Giannini, C. (1992). *Topics in Structural VAR Econometrics*, Springer, Heidelberg.

Gladyshev, E. G. (1961). Periodically correlated random sequences, *Soviet Mathematics* **2**: 385–388.

Glosten, L., Jagannathan, R. & Runkle, D. (1993). Relationship between the expected value and the volatility of the nominal excess return on stocks, *Journal of Finance* **48**: 1779–1801.

Godfrey, L. G. (1978). Testing for higher order serial correlation in regression equations when the regressors include lagged dependent variables, *Econometrica* **46**: 1303–1313.

Godfrey, L. G. (1988). *Misspecification Tests in Econometrics*, Cambridge University Press, Cambridge.

Gonzalo, J. (1994). Five alternative methods of estimating long-run equilibrium relationships, *Journal of Econometrics* **60**: 203–233.

Gonzalo, J. & Lee, T.-H. (1998). Pitfalls in testing for long run relationships, *Journal of Econometrics* **86**: 129–154.

Gonzalo, J. & Ng, S. (2001). A systematic framework for analyzing the dynamic effects of permanent and transitory shocks, *Journal of Economic Dynamics and Control* **25**: 1527–1546.

Granger, C. W. J. (1969a). Investigating causal relations by econometric models and cross-spectral methods, *Econometrica* **37**: 424–438.

Granger, C. W. J. (1969b). Prediction with a generalized cost of error function, *Operations Research Quarterly* **20**: 199–207.

Granger, C. W. J. (1981). Some properties of time series data and their use in econometric model specification, *Journal of Econometrics* **16**: 121–130.

Granger, C. W. J. (1982). Generating mechanisms, models, and causality, *in* W. Hildenbrand (ed.), *Advances in Econometrics*, Cambridge University Press, Cambridge, pp. 237–253.

Granger, C. W. J. & Andersen, A. P. (1978). *An Introduction to Bilinear Time Series Models*, Vandenhoeck & Ruprecht, Göttingen.

Granger, C. W. J. & Newbold, P. (1986). *Forecasting Economic Time Series*, 2nd edn, Academic Press, New York.

Granger, C. W. J., Robins, R. P. & Engle, R. F. (1986). Wholesale and retail prices: Bivariate time series modeling with forecastable error variances, *in* D. A. Belsley & E. Kuh (eds), *Model Reliability*, MIT Press, Cambridge, pp. 1–17.

Granger, C. W. J. & Teräsvirta, T. (1993). *Modelling Nonlinear Economic Relationships*, Oxford University Press, Oxford.

Graybill, F. A. (1969). *Introduction to Matrices with Applications in Statistics*, Wadsworth, Belmont, CA.

Gregory, A. W. & Veall, M. R. (1985). Formulating Wald tests of nonlinear restrictions, *Econometrica* **53**: 1465–1468.

Guo, L., Huang, D. W. & Hannan, E. J. (1990). On ARX(∞) approximation, *Journal of Multivariate Analysis* **32**: 17–47.

Hafner, C. M. & Herwartz, H. (1998a). Structural analysis of portfolio risk using beta impulse response functions, *Statistica Neerlandica* **52**: 336–355.

Hafner, C. M. & Herwartz, H. (1998b). Time-varying market price of risk in the CAPM. Approaches, empirical evidence and implications, *Finance* **19**: 93–112.

Haggan, V. & Ozaki, T. (1980). Amplitude-dependent exponential AR model fitting for non-linear random vibrations, *in* O. D. Anderson (ed.), *Time Series*, North-Holland, Amsterdam, pp. 57–71.

Hall, P. (1992). *The Bootstrap and Edgeworth Expansion*, Springer, New York.

Hamilton, J. D. (1986). A standard error for the estimated state vector of a state space model, *Journal of Econometrics* **33**: 387–397.

Hamilton, J. D. (1994). *Time Series Analysis*, Princeton University Press, Princeton, New Jersey.

Hannan, E. J. (1969). The identification of vector mixed autoregressive moving average systems, *Biometrika* **56**: 223–225.

Hannan, E. J. (1970). *Multiple Time Series*, John Wiley, New York.

Hannan, E. J. (1976). The identification and parameterization of ARMAX and state space forms, *Econometrica* **44**: 713–723.

Hannan, E. J. (1979). The statistical theory of linear systems, *in* P. R. Krishnaiah (ed.), *Developments in Statistics*, Academic Press, New York, pp. 83–121.

Hannan, E. J. (1981). Estimating the dimension of a linear system, *Journal of Multivariate Analysis* **11**: 459–473.

Hannan, E. J. & Deistler, M. (1988). *The Statistical Theory of Linear Systems*, John Wiley, New York.

Hannan, E. J. & Kavalieris, L. (1984). Multivariate linear time series models, *Advances in Applied Probability* **16**: 492–561.

Hannan, E. J. & Quinn, B. G. (1979). The determination of the order of an autoregression, *Journal of the Royal Statistical Society* **B41**: 190–195.

Hannan, E. J. & Rissanen, J. (1982). Recursive estimation of mixed autoregressive moving average order, *Biometrika* **69**: 81–94.

Hansen, G., Kim, J. R. & Mittnik, S. (1998). Testing cointegrating coefficients in vector autoregressive error correction models, *Economics Letters* **58**: 1–5.

Hansen, H. & Johansen, S. (1999). Some tests for parameter constancy in cointegrated VAR-models, *Econometrics Journal* **2**: 306–333.

Hansen, P. R. (2003). Structural changes in the cointegrated vector autoregressive model, *Journal of Econometrics* **114**: 261–295.

Harbo, I., Johansen, S., Nielsen, B. & Rahbek, A. (1998). Asymptotic inference on cointegrating rank in partial systems, *Journal of Business & Economic Statistics* **16**: 388–399.

Harvey, A. C. (1981). *The Econometric Analysis of Time Series*, Allan, Oxford.

Harvey, A. C. (1984). A unified view of statistical forecasting procedures, *Journal of Forecasting* **3**: 245–275.

Harvey, A. C. (1987). Applications of the Kalman filter in econometrics, *in* T. F. Bewley (ed.), *Advances in Econometrics – Fifth World Congress*, Vol. I, Cambridge University Press, Cambridge, pp. 285–313.

Harvey, A. C. (1989). *Forecasting, Structural Time Series Models, and the Kalman Filter*, Cambridge University Press, Cambridge.

Harvey, A. C. & Pierse, R. G. (1984). Estimating missing observations in economic time series, *Journal of the American Statistical Association* **79**: 125–131.

Harvey, A. C. & Todd, P. H. J. (1983). Forecasting economic time series with structural and Box-Jenkins models: A case study, *Journal of Business & Economic Statistics* **1**: 299–315.

Haug, A. A. (1996). Tests for cointegration: A Monte Carlo comparison, *Journal of Econometrics* **71**: 89–115.

Haugh, L. D. & Box, G. E. P. (1977). Identification of dynamic regression (distributed lag) models connecting two time series, *Journal of the American Statistical Association* **72**: 121–130.

Hausman, J. A. (1983). Specification and estimation of simultaneous equation models, *in* Z. Griliches & M. D. Intriligator (eds), *Handbook of Econometrics*, Vol. 1, North-Holland, Amsterdam, pp. 391–448.

Hayashi, F. (2000). *Econometrics*, Princeton University Press, Princeton.

Hendry, D. F. (1995). *Dynamic Econometrics*, Oxford University Press, Oxford.

Hendry, D. F. & Krolzig, H.-M. (2001). *Automatic Econometric Model Selection with PcGets*, Timberlake Consultants Press, London.

Hendry, D. F. & von Ungern-Sternberg, T. (1981). Liquidity and inflation effects on consumer's expenditure, *in* A. S. Deaton (ed.), *Essays in the Theory and Measurement of Consumer's Behavior*, Cambridge University Press, Cambridge.

Herwartz, H. (1995). *Analyse saisonaler Zeitreihen mit Hilfe periodischer Zeitreihenmodelle*, Josef Eul, Bergisch Gladbach.

Herwartz, H. (2004). Conditional heteroskedasticity, *in* H. Lütkepohl & M. Krätzig (eds), *Applied Time Series Econometrics*, Cambridge University Press, Cambridge, pp. 197–221.

Herwartz, H. & Lütkepohl, H. (2000). Multivariate volatility analysis of VW stock prices, *International Journal of Intelligent Systems in Accounting, Finance & Management* **9**: 35–54.

Hildreth, C. & Houck, J. P. (1968). Some estimators for a linear model with random coefficients, *Journal of the American Statistical Association* **63**: 584–595.

Hillmer, S. C. & Tiao, G. C. (1979). Likelihood function of stationary multiple autoregressive moving average models, *Journal of the American Statistical Association* **74**: 652–660.

Ho, M. S. & Sørensen, B. E. (1996). Finding cointegration rank in high dimensional systems using the Johansen test: An illustration using data based Monte Carlo simulations, *Review of Economics and Statistics* **78**: 726–732.

Hogg, R. V. & Craig, A. T. (1978). *Introduction to Mathematical Statistics*, Macmillan, New York.

Hong, Y. (2001). A test for volatility spillover with application to exchange rates, *Journal of Econometrics* **103**: 183–224.

Horowitz, J. L. (2001). The bootstrap, *in* J. J. Heckman & E. Leamer (eds), *Handbook of Econometrics*, Vol. 5, North-Holland, Amsterdam, pp. 3159–3228.

Hosking, J. R. M. (1980). The multivariate portmanteau statistic, *Journal of the American Statistical Association* **75**: 602–608.

Hosking, J. R. M. (1981a). Equivalent forms of the multivariate portmanteau statistic, *Journal of the Royal Statistical Society* **B43**: 261–262.

Hosking, J. R. M. (1981b). Lagrange-multiplier tests of multivariate time series models, *Journal of the Royal Statistical Society* **B43**: 219–230.

Hsiao, C. (1979). Autoregressive modeling of Canadian money and income data, *Journal of the American Statistical Association* **74**: 553–560.

Hsiao, C. (1982). Time series modelling and causal ordering of Canadian money, income and interest rates, *in* O. D. Anderson (ed.), *Time Series Analysis: Theory and Practice 1*, North-Holland, Amsterdam, pp. 671–699.

Huang, D. W. & Guo, L. (1990). Estimation of nonstationary ARMAX models based on the Hannan-Rissanen method, *Annals of Statistics* **18**: 1729–1756.

Hubrich, K., Lütkepohl, H. & Saikkonen, P. (2001). A review of systems cointegration tests, *Econometric Reviews* **20**: 247–318.

Hylleberg, S. (1986). *Seasonality in Regression*, Academic Press, Orlando.

Inder, B. (1993). Estimating long run relationships in economics: A comparison of different approaches, *Journal of Econometrics* **57**: 53–68.

Jacobson, T., Vredin, A. & Warne, A. (1997). Common trends and hysteresis in Scandinavian unemployment, *European Economic Review* **41**: 1781–1816.

Jarque, C. M. & Bera, A. K. (1987). A test for normality of observations and regression residuals, *International Statistical Review* **55**: 163–172.

Jazwinski, A. H. (1970). *Stochastic Processes and Filtering Theory*, Academic Press, New York.

Jeantheau, T. (1998). Strong consistency of estimators for multivariate ARCH models, *Econometric Theory* **14**: 70–86.

Jenkins, G. M. & Alavi, A. S. (1981). Some aspects of modelling and forecasting multivariate time series, *Journal of Time Series Analysis* **2**: 1–47.

Johansen, S. (1988). Statistical analysis of cointegration vectors, *Journal of Economic Dynamics and Control* **12**: 231–254.

Johansen, S. (1991). Estimation and hypothesis testing of cointegration vectors in Gaussian vector autoregressive models, *Econometrica* **59**: 1551–1580.

Johansen, S. (1992). Cointegration in partial systems and the efficiency of single-equation analysis, *Journal of Econometrics* **52**: 389–402.

Johansen, S. (1994). The role of the constant and linear terms in cointegration analysis of nonstationary time series, *Econometric Reviews* **13**: 205–231.

Johansen, S. (1995). *Likelihood-based Inference in Cointegrated Vector Autoregressive Models*, Oxford University Press, Oxford.

Johansen, S. & Juselius, K. (1990). Maximum likelihood estimation and inference on cointegration–with applications to the demand for money, *Oxford Bulletin of Economics and Statistics* **52**: 169–210.

Johansen, S., Mosconi, R. & Nielsen, B. (2000). Cointegration analysis in the presence of structural breaks in the deterministic trend, *Econometrics Journal* **3**: 216–249.

Jones, R. H. (1980). Maximum likelihood fitting of ARMA models to time series with missing observations, *Technometrics* **22**: 389–395.

Jones, R. H. & Brelsford, W. M. (1967). Time series with periodic structure, *Biometrika* **54**: 403–408.

Judge, G. G., Griffiths, W. E., Hill, R. C., Lütkepohl, H. & Lee, T.-C. (1985). *The Theory and Practice of Econometrics*, 2nd edn, John Wiley, New York.

Kalman, R. E. (1960). A new approach to linear filtering and prediction problems, *Journal of Basic Engineering* **82**: 35–45.

Kalman, R. E. & Bucy, R. S. (1961). New results in linear filtering and prediction theory, *Journal of Basic Engineering* **83**: 95–108.

Kapetanios, G. (2003). A note on an iterative least-squares estimation method for ARMA and VARMA models, *Economics Letters* **79**: 305–312.

Kauppi, H. (2004). On the robustness of hypothesis testing based on fully modified vector autoregression when some roots are almost one, *Econometric Theory* **20**: 341–359.

Kilian, L. (1998). Small-sample confidence intervals for impulse response functions, *Review of Economics and Statistics* **80**: 218–230.

Kilian, L. (1999). Finite-sample properties of percentile and percentile-t bootstrap confidence intervals for impulse responses, *Review of Economics and Statistics* **81**: 652–660.

Kilian, L. & Demiroglu, U. (2000). Residual-based tests for normality in autoregressions: Asymptotic theory and simulation evidence, *Journal of Business & Economic Statistics* **18**: 40–50.

King, R. G., Plosser, C. I., Stock, J. H. & Watson, M. W. (1991). Stochastic trends and economic fluctuations, *American Economic Review* **81**: 819–840.

Kitagawa, G. (1981). A nonstationary time series model and its fitting by a recursive filter, *Journal of Time Series Analysis* **2**: 103–116.

Kleibergen, F. & Paap, R. (2002). Priors, posteriors and bayes factors for a Bayesian analysis of cointegration, *Journal of Econometrics* **111**: 223–249.

Kleibergen, F. & van Dijk, H. K. (1994). On the shape of the likelihood/posterior in cointegration models, *Econometric Theory* **10**: 514–551.

Kohn, R. (1979). Asymptotic estimation and hypothesis testing results for vector linear time series models, *Econometrica* **47**: 1005–1030.

Kohn, R. (1981). A note on an alternative derivation of the likelihood of an autoregressive moving average process, *Economics Letters* **7**: 233–236.

Koop, G. (1992). Aggregate shocks and macroeconomic fluctuations: A Bayesian approach, *Journal of Applied Econometrics* **7**: 395–411.

Koop, G., Pesaran, M. H. & Potter, S. M. (1996). Impulse response analysis in nonlinear multivariate models, *Journal of Econometrics* **74**: 119–147.

Koop, G., Strachan, R. W., van Dijk, H. K. & Villani, M. (2005). Bayesian approaches to cointegration, *in* K. Patterson & T. C. Mills (eds), *Palgrave Handbook of Econometrics, Volume 1: Econometric Theory*, Palgrave Macmillan, Houndmills.

Koreisha, S. G. & Pukkila, T. M. (1987). Identification of nonzero elements in the polynomial matrices of mixed VARMA processes, *Journal of the Royal Statistical Society* **B49**: 112–126.

Lee, K. C., Pesaran, M. H. & Pierse, R. G. (1992). Persistence of shocks and its sources in a multisectoral model of UK output growth, *Economic Journal* **102**: 342–356.

Lewis, R. & Reinsel, G. C. (1985). Prediction of multivariate time series by autoregressive model fitting, *Journal of Multivariate Analysis* **16**: 393–411.

Li, H. & Maddala, G. S. (1996). Bootstrapping time series models, *Econometric Reviews* **15**: 115–158.

Li, W. K. & Hui, Y. V. (1988). An algorithm for the exact likelihood of periodic autoregressive moving average models, *Communications in Statistics–Simulation and Computation* **17**: 1483–1494.

Li, W. K. & McLeod, A. I. (1981). Distribution of the residual autocorrelations in multivariate ARMA time series models, *Journal of the Royal Statistical Society* **B43**: 231–239.

Lin, W.-L. (1992). Alternative estimators of factor GARCH models – A Monte Carlo comparison, *Journal of Applied Econometrics* **7**: 259–279.

Lin, W.-L. (1997). Impulse response functions for conditional volatility in GARCH models, *Journal of Business & Economic Statistics* **15**: 15–25.

Ling, S. & McAleer, M. (2003). Asymptotic theory for a new vector ARMA GARCH model, *Econometric Theory* **19**: 280–310.

Litterman, R. B. (1986). Forecasting with Bayesian vector autoregressions - five years of experience, *Journal of Business & Economic Statistics* **4**: 25–38.

Liu, J. (1989). On the existence of a general multiple bilinear time series, *Journal of Time Series Analysis* **10**: 341–355.

Liu, L.-M. & Hanssens, D. M. (1982). Identification of multiple-input transfer function models, *Communications in Statistics – Theory and Methods* **11**: 297–314.

Ljung, G. M. & Box, G. E. P. (1978). On a measure of lack of fit in time series models, *Biometrika* **65**: 297–303.

Lomnicki, Z. A. (1961). Tests for departure from normality in the case of linear stochastic processes, *Metrika* **4**: 37–62.

Lucas, A. (1997). Cointegration testing using pseudo likelihood ratio tests, *Econometric Theory* **13**: 149–169.

Lucas, A. (1998). Inference on cointegrating ranks using LR and LM tests based on pseudo-likelihoods, *Econometric Reviews* **17**: 185–214.

Lucas, Jr., R. E. & Sargent, T. J. (1981). *Rational Expectations and Econometric Practice*, Vol. 1 and 2, University of Minnesota Press, Minneapolis.

Lundbergh, S. & Teräsvirta, T. (2002). Evaluating GARCH models, *Journal of Econometrics* **110**: 417–435.

Lütkepohl, H. (1984). Linear transformations of vector ARMA processes, *Journal of Econometrics* **26**: 283–293.

Lütkepohl, H. (1985). Comparison of criteria for estimating the order of a vector autoregressive process, *Journal of Time Series Analysis* **6**: 35–52, Correction, **8** (1987), 373.

Lütkepohl, H. (1986). *Prognose aggregierter Zeitreihen*, Vandenhoeck & Ruprecht, Göttingen.

Lütkepohl, H. (1987). *Forecasting Aggregated Vector ARMA Processes*, Springer, Berlin.

Lütkepohl, H. (1988a). Asymptotic distribution of the moving average coefficients of an estimated vector autoregressive process, *Econometric Theory* **4**: 77–85.

Lütkepohl, H. (1988b). Prediction tests for structural stability, *Journal of Econometrics* **39**: 267–296.

Lütkepohl, H. (1989a). A note on the asymptotic distribution of impulse response functions of estimated VAR models with orthogonal residuals, *Journal of Econometrics* **42**: 371–376.

Lütkepohl, H. (1989b). Prediction tests for structural stability of multiple time series, *Journal of Business & Economic Statistics* **7**: 129–135.

Lütkepohl, H. (1990). Asymptotic distributions of impulse response functions and forecast error variance decompositions of vector autoregressive models, *Review of Economics and Statistics* **72**: 116–125.

Lütkepohl, H. (1991). *Introduction to Multiple Time Series Analysis*, Springer, Berlin.

Lütkepohl, H. (1992). Testing for time varying parameters in vector autoregressive models, *in* W. E. Griffiths, H. Lütkepohl & M. E. Bock (eds), *Readings in Econometric Theory and Practice*, North-Holland, Amsterdam, pp. 243–264.

Lütkepohl, H. (1993). Testing for causation between two variables in higher dimensional VAR models, *in* H. Schneeweiß & K. F. Zimmermann (eds), *Studies in Applied Econometrics*, Physica, Heidelberg, pp. 75–91.

Lütkepohl, H. (1996a). *Handbook of Matrices*, John Wiley, Chichester.

Lütkepohl, H. (1996b). Testing for nonzero impulse responses in vector autoregressive processes, *Journal of Statistical Planning and Inference* **50**: 1–20.

Lütkepohl, H. (1997). Statistische Modellierung von Volatilitäten, *Allgemeines Statistisches Archiv* **81**: 62–84.

Lütkepohl, H. (2004). Vector autoregressive and vector error correction models, *in* H. Lütkepohl & M. Krätzig (eds), *Applied Time Series Econometrics*, Cambridge University Press, Cambridge, pp. 86–158.

Lütkepohl, H. & Burda, M. M. (1997). Modified Wald tests under nonregular conditions, *Journal of Econometrics* **78**: 315–332.

Lütkepohl, H. & Claessen, H. (1997). Analysis of cointegrated VARMA processes, *Journal of Econometrics* **80**: 223–239.

Lütkepohl, H. & Krätzig, M. (eds) (2004). *Applied Time Series Econometrics*, Cambridge University Press, Cambridge.

Lütkepohl, H. & Poskitt, D. S. (1991). Estimating orthogonal impulse responses via vector autoregressive models, *Econometric Theory* **7**: 487–496.

Lütkepohl, H. & Poskitt, D. S. (1996). Testing for causation using infinite order vector autoregressive processes, *Econometric Theory* **12**: 61–87.

Lütkepohl, H. & Poskitt, D. S. (1998). Consistent estimation of the number of cointegration relations in a vector autoregressive model, *in* R. Galata & H. Küchenhoff (eds), *Econometrics in Theory and Practice. Festschrift for Hans Schneeweiß*, Physica, Heidelberg, pp. 87–100.

Lütkepohl, H. & Reimers, H.-E. (1992a). Granger-causality in cointegrated VAR processes: The case of the term structure, *Economics Letters* **40**: 263–268.

Lütkepohl, H. & Reimers, H.-E. (1992b). Impulse response analysis of cointegrated systems, *Journal of Economic Dynamics and Control* **16**: 53–78.

Lütkepohl, H. & Saikkonen, P. (1999a). A lag augmentation test for the cointegrating rank of a VAR process, *Economics Letters* **63**: 23–27.

Lütkepohl, H. & Saikkonen, P. (1999b). Order selection in testing for the cointegrating rank of a VAR process, *in* R. F. Engle & H. White (eds), *Cointegration, Causality, and Forecasting. A Festschrift in Honour of Clive W.J. Granger*, Oxford University Press, Oxford, pp. 168–199.

Lütkepohl, H. & Saikkonen, P. (2000). Testing for the cointegrating rank of a VAR process with a time trend, *Journal of Econometrics* **95**: 177–198.

Lütkepohl, H., Saikkonen, P. & Trenkler, C. (2001). Maximum eigenvalue versus trace tests for the cointegrating rank of a VAR process, *Econometrics Journal* **4**: 287–310.

Lütkepohl, H., Saikkonen, P. & Trenkler, C. (2004). Testing for the cointegrating rank of a VAR process with level shift at unknown time, *Econometrica* **72**: 647–662.

Lütkepohl, H. & Schneider, W. (1989). Testing for nonnormality of autoregressive time series, *Computational Statistics Quarterly* **5**: 151–168.

MacKinnon, J. G., Haug, A. A. & Michelis, L. (1999). Numerical distribution functions of likelihood ratio tests for cointegration, *Journal of Applied Econometrics* **14**: 563–577.

Magnus, J. R. (1988). *Linear Structures*, Charles Griffin, London.

Magnus, J. R. & Neudecker, H. (1988). *Matrix Differential Calculus with Applications in Statistics and Econometrics*, John Wiley, Chichester.

Mann, H. B. & Wald, A. (1943). On the statistical treatment of linear stochastic difference equations, *Econometrica* **11**: 173–220.

Mardia, K. V. (1980). Tests for univariate and multivariate normality, *in* P. R. Krishnaiah (ed.), *Handbook of Statistics*, Vol. 1, North-Holland, Amsterdam, pp. 279–320.

Meinhold, R. J. & Singpurwalla, N. D. (1983). Understanding the Kalman-filter, *American Statistician* **37**: 123–127.

Mittnik, S. (1990). Computation of theoretical autocovariance matrices of multivariate autoregressive moving average time series, *Journal of the Royal Statistical Society* **B52**: 151–155.

Mood, A. M., Graybill, F. A. & Boes, D. C. (1974). *Introduction to the Theory of Statistics*, 3rd edn, McGraw-Hill, Auckland.

Morrison, D. F. (1976). *Multivariate Statistical Methods*, 2nd edn, McGraw-Hill, New York.

Murata, Y. (1982). *Optimal Control Methods for Linear Discrete-Time Economic Systems*, Springer, New York.

Muth, J. F. (1961). Rational expectations and the theory of price movements, *Econometrica* **29**: 315–335.

Nankervis, J. C. & Savin, N. E. (1988). The Student's t approximation in a stationary first order autoregressive model, *Econometrica* **56**: 119–145.

Nelson, D. (1991). Conditional heteroskedasticity in asset returns: A new approach, *Econometrica* **59**: 347–370.

Nicholls, D. F. & Pagan, A. R. (1985). Varying coefficient regression, *in* E. J. Hannan, P. R. Krishnaiah & M. M. Rao (eds), *Handbook of Statistics*, Vol. 5, North-Holland, Amsterdam, pp. 413–449.

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

Nicholls, D. F. & Quinn, B. C. (1982). *Random Coefficient Autoregressive Models: An Introduction*, Springer, New York.

Nijman, T. E. (1985). *Missing Observations in Dynamic Macroeconomic Modeling*, Free University Press, Amsterdam.

Osborn, D. R. & Smith, J. P. (1989). The performance of periodic autoregressive models in forecasting seasonal U.K. consumption, *Journal of Business & Economic Statistics* **7**: 117–127.

Ozaki, T. (1980). Non-linear time series models for nonlinear random vibrations, *Journal of Applied Probability* **17**: 84–93.

Pagan, A. (1980). Some identification and estimation results for regression models with stochastically varying coefficients, *Journal of Econometrics* **13**: 341–363.

Pagan, A. (1995). Three econometric methodologies: An update, *in* L. Oxley, D. A. R. George, C. J. Roberts & S. Sayer (eds), *Surveys in Econometrics*, Blackwell, Oxford.

Pagan, A. (1996). The econometrics of financial markets, *Journal of Empirical Finance* **3**: 15–102.

Pagano, M. (1978). On periodic and multiple autoregressions, *Annals of Statistics* **6**: 1310–1317.

Pantelidis, T. & Pittis, N. (2004). Testing for Granger causality in variance in the presence of causality in mean, *Economics Letters* **85**: 201–207.

Park, J. Y. & Phillips, P. C. B. (1988). Statistical inference in regressions with integrated processes: Part 1, *Econometric Theory* **4**: 468–497.

Park, J. Y. & Phillips, P. C. B. (1989). Statistical inference in regressions with integrated processes: Part 2, *Econometric Theory* **5**: 95–131.

Paulsen, J. (1984). Order determination of multivariate autoregressive time series with unit roots, *Journal of Time Series Analysis* **5**: 115–127.

Paulsen, J. & Tjøstheim, D. (1985). On the estimation of residual variance and order in autoregressive time series, *Journal of the Royal Statistical Society* **B47**: 216–228.

Penm, J. H. W., Brailsford, T. J. & Terrell, R. D. (2000). A robust algorithm in sequentially selecting subset time series systems using neural networks, *Journal of Time Series Analysis* **21**: 389–412.

Penm, J. H. W. & Terrell, R. D. (1982). On the recursive fitting of subset autoregressions, *Journal of Time Series Analysis* **3**: 43–59.

Penm, J. H. W. & Terrell, R. D. (1984). Multivariate subset autoregressive modelling with zero constraints for detecting 'overall causality', *Journal of Econometrics* **24**: 311–330.

Penm, J. H. W. & Terrell, R. D. (1986). The 'derived' moving average model and its role in causality, *in* J. Gani & M. B. Priestley (eds), *Essays in Time Series and Allied Processes*, Applied Probability Trust, Sheffield, pp. 99–111.

Pesaran, M. H. (1987). *The Limits to Rational Expectations*, Blackwell, Oxford.

Pesaran, M. H. & Shin, Y. (1996). Cointegration and speed of convergence to equilibrium, *Journal of Econometrics* **71**: 117–143.

Phillips, P. C. B. (1987). Time series regression with a unit root, *Econometrica* **55**: 277–301.

Phillips, P. C. B. (1988). Multiple regression with integrated time series, *Contemporary Mathematics* **80**: 79–105.

Phillips, P. C. B. (1991). Optimal inference in cointegrated systems, *Econometrica* **59**: 283–306.

Phillips, P. C. B. (1994). Some exact distribution theory for maximum likelihood estimators of cointegrating coefficients in error correction models, *Econometrica* **62**: 73–93.

Phillips, P. C. B. (1995). Fully modified least squares and vector autoregression, *Econometrica* **63**: 1023–1078.

Phillips, P. C. B. & Durlauf, S. N. (1986). Multiple time series regression with integrated processes, *Review of Economic Studies* **53**: 473–495.

Phillips, P. C. B. & Hansen, B. E. (1990). Statistical inference in instrumental variables regression with I(1) processes, *Review of Economic Studies* **57**: 99–125.

Phillips, P. C. B. & Ouliaris, S. (1990). Asymptotic properties of residual based tests for cointegration, *Econometrica* **58**: 165–193.

Phillips, P. C. B. & Park, J. Y. (1988). Asymptotic equivalence of ordinary least squares and generalized least squares in regressions with integrated regressors, *Journal of the American Statistical Association* **83**: 111–115.

Phillips, P. C. B. & Solo, V. (1992). Asymptotics for linear processes, *Annals of Statistics* **20**: 971–1001.

Poskitt, D. S. (1987). A modified Hannan-Rissanen strategy for mixed autoregressive-moving average order determination, *Biometrika* **74**: 781–790.

Poskitt, D. S. (1989). A method for the estimation of transfer function models, *Journal of the Royal Statistical Society* **B51**: 29–46.

Poskitt, D. S. (1992). Identification of echelon canonical forms for vector linear processes using least squares, *Annals of Statistics* **20**: 196–215.

Poskitt, D. S. (2003). On the specification of cointegrated autoregressive moving-average forecasting systems, *International Journal of Forecasting* **19**: 503–519.

Poskitt, D. S. (2004). On the identification and estimation of partially nonstationary ARMAX systems, *Discussion paper*, Monash University, Australia.

Poskitt, D. S. & Lütkepohl, H. (1995). Consistent specification of cointegrated autoregressive moving average systems, *Discussion Paper 54*, SFB 373, Humboldt-Universität zu Berlin.

Poskitt, D. S. & Tremayne, A. R. (1981). An approach to testing linear time series models, *Annals of Statistics* **9**: 974–986.

Poskitt, D. S. & Tremayne, A. R. (1982). Diagnostic tests for multiple time series models, *Annals of Statistics* **10**: 114–120.

Priestley, M. B. (1980). State-dependent models: A general approach to non-linear time series analysis, *Journal of Time Series Analysis* **1**: 47–71.

Priestley, M. B. (1988). *Non-Linear and Non-Stationary Time Series Analysis*, Academic Press, London.

Proietti, T. (2002). Forecasting with structural time series models, *in* M. P. Clements & D. F. Hendry (eds), *A Companion to Economic Forecasting*, Blackwell, Oxford, pp. 105–132.

Quenouille, M. H. (1957). *The Analysis of Multiple Time-Series*, Griffin, London.

Quinn, B. (1980). Order determination for a multivariate autoregression, *Journal of the Royal Statistical Society* **B42**: 182–185.

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

Reimers, H.-E. (1991). *Analyse kointegrierter Variablen mittels vektorautoregressiver Modelle*, Physica, Heidelberg.

Reinsel, G. C. (1983). Some results on multivariate autoregressive index models, *Biometrika* **70**: 145–156.

Reinsel, G. C. (1993). *Elements of Multivariate Time Series Analysis*, Springer, New York.

Reinsel, G. C. & Ahn, S. K. (1992). Vector autoregressive models with unit roots and reduced rank structure: Estimation, likelihood ratio test, and forecasting, *Journal of Time Series Analysis* **13**: 353–375.

Reinsel, G. C. & Velu, R. P. (1998). *Multivariate Reduced-Rank Regression, Theory and Applications*, Springer, New York.

Rohatgi, V. K. (1976). *An Introduction to Probability Theory and Mathematical Statistics*, John Wiley, New York.

Rothenberg, T. J. (1971). Identification in parametric models, *Econometrica* **39**: 577–591.

Roussas, G. G. (1973). *A First Course in Mathematical Statistics*, Addison-Wesley, Reading, MA.

Runkle, D. E. (1987). Vector autoregressions and reality, *Journal of Business & Economic Statistics* **5**: 437–442.

Saikkonen, P. (1992). Estimation and testing of cointegrated systems by an autoregressive approximation, *Econometric Theory* **8**: 1–27.

Saikkonen, P. (2005). Stability results for nonlinear error correction models, *Journal of Econometrics*.

Saikkonen, P. & Lütkepohl, H. (1996). Infinite order cointegrated vector autoregressive processes: Estimation and inference, *Econometric Theory* **12**: 814–844.

Saikkonen, P. & Lütkepohl, H. (1999). Local power of likelihood ratio tests for the cointegrating rank of a VAR process, *Econometric Theory* **15**: 50–78.

Saikkonen, P. & Lütkepohl, H. (2000a). Asymptotic inference on nonlinear functions of the coefficients of infinite order cointegrated VAR processes, *in* W. A. Barnett, D. F. Hendry, S. Hylleberg, T. Teräsvirta, D. Tjøstheim & A. Würtz (eds), *Nonlinear Econometric Modeling in Time Series Analysis*, Cambridge University Press, Cambridge, pp. 165–201.

Saikkonen, P. & Lütkepohl, H. (2000b). Testing for the cointegrating rank of a VAR process with an intercept, *Econometric Theory* **16**: 373–406.

Saikkonen, P. & Lütkepohl, H. (2000c). Testing for the cointegrating rank of a VAR process with structural shifts, *Journal of Business & Economic Statistics* **18**: 451–464.

Saikkonen, P. & Lütkepohl, H. (2000d). Trend adjustment prior to testing for the cointegrating rank of a vector autoregressive process, *Journal of Time Series Analysis* **21**: 435–456.

Saikkonen, P. & Luukkonen, R. (1997). Testing cointegration in infinite order vector autoregressive processes, *Journal of Econometrics* **81**: 93–129.

Salmon, M. (1982). Error correction mechanisms, *Economic Journal* **92**: 615–629.

Samaranayake, V. A. & Hasza, D. P. (1988). Properties of predictors for multivariate autoregressive models with estimated parameters, *Journal of Time Series Analysis* **9**: 361–383.

Sargent, T. J. & Sims, C. A. (1977). Business cycle modeling without pretending to have too much a priori economic theory, *in* C. A. Sims (ed.), *New Methods in Business Cycle Research: Proceedings from a Conference*, Federal Reserve Bank of Minneapolis, Minneapolis, pp. 45–109.

Schmidt, P. (1973). The asymptotic distribution of dynamic multipliers, *Econometrica* **41**: 161–164.

Schmidt, P. (1976). *Econometrics*, Marcel Dekker, New York.

Schneider, W. (1988). Analytical uses of Kalman filtering in econometrics – a survey, *Statistical Papers* **29**: 3–33.

Schneider, W. (1991). Stability analysis using Kalman filtering, scoring, EM, and an adaptive EM method, *in* P. Hackl & A. Westlund (eds), *Economic Structural Change, Analysis and Forecasting*, Springer, Berlin, pp. 191–221.

Schneider, W. (1992). Systems of seemingly unrelated regression equations with time-varying coefficients – an interplay of Kalman filtering, scoring, EM- and MINQUE-method, *Computers and Mathematics with Applications* **24**: 1–16.

Schwarz, G. (1978). Estimating the dimension of a model, *Annals of Statistics* **6**: 461–464.

Searle, S. R. (1982). *Matrix Algebra Useful for Statistics*, John Wiley, New York.

Serfling, R. J. (1980). *Approximation Theorems of Mathematical Statistics*, John Wiley, New York.

Shapiro, M. & Watson, M. W. (1988). Sources of business cycle fluctuations, *NBER Macroeconomics Annual* **3**: 111–156.

Shibata, R. (1980). Asymptotically efficient selection of the order of the model for estimating parameters of a linear process, *Annals of Statistics* **8**: 147–164.

Shin, Y. (1994). A residual-based test of the null of cointegration against the alternative of no cointegration, *Econometric Theory* **10**: 91–115.

Sims, C. A. (1980). Macroeconomics and reality, *Econometrica* **48**: 1–48.

Sims, C. A. (1981). An autoregressive index model for the U.S. 1948-1975, *in* J. Kmenta & J. B. Ramsey (eds), *Large-Scale Macro-Econometric Models*, North-Holland, Amsterdam, pp. 283–327.

Sims, C. A. (1986). Are forecasting models usable for policy analysis?, *Quarterly Review, Federal Reserve Bank of Minneapolis* **10**: 2–16.

Sims, C. A., Stock, J. H. & Watson, M. W. (1990). Inference in linear time series models with some unit roots, *Econometrica* **58**: 113–144.

Sims, C. A. & Zha, T. (1999). Error bands for impulse responses, *Econometrica* **67**: 1113–1155.

Snell, A. (1999). Testing for r versus $r-1$ cointegrating vectors, *Journal of Econometrics* **88**: 151–191.

Solo, V. (1984). The exact likelihood for a multivariate ARMA model, *Journal of Multivariate Analysis* **15**: 164–173.

Stensholt, B. K. & Tjøstheim, D. (1987). Multiple bilinear time series models, *Journal of Time Series Analysis* **8**: 221–233.

Stock, J. H. (1987). Asymptotic properties of least squares estimators of cointegrating vectors, *Econometrica* **55**: 1035–1056.

Stock, J. H. & Watson, M. W. (1988). Testing for common trends, *Journal of the American Statistical Association* **83**: 1097–1107.

Stock, J. H. & Watson, M. W. (1993). A simple estimator of cointegrating vectors in higher order integrated systems, *Econometrica* **61**: 783–820.

Stock, J. H. & Watson, M. W. (2002a). Forecasting using principal components from a large number of predictors, *Journal of the American Statistical Association* **97**: 1167–1179.

Stock, J. H. & Watson, M. W. (2002b). Macroeconomic forecasting using diffusion indexes, *Journal of Business & Economic Statistics* **20**: 147–162.

Strachan, R. W. (2003). Valid Bayesian estimation of the cointegrating error correction model, *Journal of Business & Economic Statistics* **21**: 185–195.

Strachan, R. W. & Inder, B. (2004). Bayesian analysis of the error correction model, *Journal of Econometrics* **123**: 307–325.

Subba Rao, T. & Gabr, M. M. (1984). *An Introduction to Bispectral Analysis and Bilinear Time Series Models*, Springer, New York.

Swamy, P. A. V. B. (1971). *Statistical Inference in Random Coefficient Regression Models*, Springer, Berlin.

Swamy, P. A. V. B. & Tavlas, G. S. (2001). Random coefficient models, *in* B. H. Baltagi (ed.), *A Companion to Theoretical Econometrics*, Blackwell, Oxford, pp. 410–428.

Tay, A. S. & Wallis, K. F. (2002). Density forecasting: A survey, *in* M. P. Clements & D. F. Hendry (eds), *Companion to Economic Forecasting*, Blackwell, Oxford, pp. 45–68.

Taylor, S. J. (1986). *Modelling Financial Time Series*, John Wiley, Chichester.

Theil, H. (1971). *Principles of Econometrics*, Wiley, Santa Barbara.

Tiao, G. C. & Box, G. E. P. (1981). Modeling multiple time series with applications, *Journal of the American Statistical Association* **76**: 802–816.

Tiao, G. C. & Grupe, M. R. (1980). Hidden periodic autoregressive-moving average models in time series data, *Biometrika* **67**: 365–373.

Tiao, G. C. & Tsay, R. S. (1989). Model specification in multivariate time series (with discussion), *Journal of the Royal Statistical Society* **B51**: 157–213.

Tjøstheim, D. & Paulsen, J. (1983). Bias of some commonly-used time series estimates, *Biometrika* **70**: 389–399.

Toda, H. Y. & Phillips, P. C. B. (1993). Vector autoregressions and causality, *Econometrica* **61**: 1367–1393.

Toda, H. Y. & Yamamoto, T. (1995). Statistical inference in vector autoregressions with possibly integrated processes, *Journal of Econometrics* **66**: 225–250.

Tong, H. (1983). *Threshold Models in Non-linear Time Series Analysis*, Springer, New York.

Tsay, R. S. (1984). Order selection in nonstationary autoregressive models, *Annals of Statistics* **12**: 1425–1433.

Tsay, R. S. (1985). Model identification in dynamic regression (distributed lag) models, *Journal of Business & Economic Statistics* **3**: 228–237.

Tsay, R. S. (1988). Outliers, level shifts and variance changes in time series, *Journal of Forecasting* **7**: 1–20.

Tsay, R. S. (1989a). Identifying multivariate time series models, *Journal of Time Series Analysis* **10**: 357–372.

Tsay, R. S. (1989b). Parsimonious parameterization of vector autoregressive moving average models, *Journal of Business & Economic Statistics* **7**: 327–341.

Tse, Y. K. & Tsui, A. K. C. (1999). A note on diagnosing multivariate conditional heteroscedasticity models, *Journal of Time Series Analysis* **20**: 679–691.

Tse, Y. K. & Tsui, A. K. C. (2002). A multivariate generalized autoregressive conditional heteroscedasticity model with time-varying correlations, *Journal of Business & Economic Statistics* **20**: 351–362.

Tso, M. K.-S. (1981). Reduced-rank regression and canonical analysis, *Journal of the Royal Statistical Society* **B43**: 183–189.

Uhlig, H. (1994). What are the effects of monetary policy on output? Results from an agnostic identification procedure, mimeo. Department of Economics, Humboldt University Berlin.

van der Weide, R. (2002). GO-GARCH: A multivariate generalized orthogonal GARCH model, *Journal of Applied Econometrics* **17**: 549–564.

van Overschee, P. & DeMoor, B. (1994). N4sid: Subspace algorithms for the identification of combined deterministic-stochastic systems, *Automatica* **30**: 75–93.

Velu, R. P., Reinsel, G. C. & Wichern, D. W. (1986). Reduced rank models for multiple time series, *Biometrika* **73**: 105–118.

Vlaar, P. J. G. (2004). On the asymptotic distribution of impulse response functions with long-run restrictions, *Econometric Theory* **20**: 891–903.

Vrontos, I. D., Dellaportas, P. & Politis, D. N. (2003). A full-factor multivariate GARCH model, *Econometrics Journal* **6**: 312–334.

Wallis, K. F. (1977). Multiple time series analysis and the final form of econometric models, *Econometrica* **45**: 1481–1497.

Watanabe, N. (1985). Note on the Kalman filter with estimated parameters, *Journal of Time Series Analysis* **6**: 269–278.

Watson, M. W. & Engle, R. F. (1983). Alternative algorithms for the estimation of dynamic factor, MIMIC and varying coefficient regression models, *Journal of Econometrics* **23**: 385–400.

Weber, C. E. (1995). Cyclical output, cyclical unemployment, and Okun's coefficient: A new approach, *Journal of Applied Econometrics* **10**: 433–445.

Wei, W. W. S. (1990). *Time Series Analysis: Univariate and Multivariate Methods*, Addison-Wesley, Redwood City, CA.

White, H. & MacDonald, G. M. (1980). Some large-sample tests for nonnormality in the linear regression model, *Journal of the American Statistical Association* **75**: 16–28.

Wickens, M. R. & Breusch, T. S. (1988). Dynamic specification, the long run and the estimation of transformed regression models, *Economic Journal* **98**: 189–205.

Wold, H. (1938). *A Study in the Analysis of Stationary Time-Series*, Almqvist and Wiksells, Uppsala.

Wong, H. & Li, W. K. (1997). On a multivariate conditional heteroscedastic model, *Biometrika* **84**: 111–123.

Yamamoto, T. (1980). On the treatment of autocorrelated errors in the multiperiod prediction of dynamic simultaneous equation models, *International Economic Review* **21**: 735–748.

Yang, M. & Bewley, R. (1996). On cointegration tests for VAR models with drift, *Economics Letters* **51**: 45–50.

Yap, S. F. & Reinsel, G. C. (1995). Estimation and testing for unit roots in a partially nonstationary vector autoregressive moving average model, *Journal of the American Statistical Association* **90**: 253–267.

Young, P. C., Jakeman, A. J. & McMurtrie, R. (1980). An instrumental variable method for model order identification, *Automatica* **16**: 281–294.

Zellner, A. (1962). An efficient method of estimating seemingly unrelated regressions and tests of aggregation bias, *Journal of the American Statistical Association* **57**: 348–368.

Zellner, A. & Palm, F. (1974). Time series analysis and simultaneous equation econometric models, *Journal of Econometrics* **2**: 17–54.

Index of Notation

Most of the notation is clearly defined in the text where it is used. The following list is meant to provide some general guidelines. Occasionally, in the text a symbol has a meaning which differs from the one specified in this list when confusion is unlikely. For instance, A usually stands for a VAR coefficient matrix whereas in the Appendix it is often a general matrix.

General Symbols

$=$	equals
$:=$	equals by definition
$\Rightarrow$	implies
$\Leftrightarrow$	is equivalent to
$\sim$	is distributed as
$\in$	element of
$\subset$	subset of
$\cup$	union
$\cap$	intersection
$\sum$	summation sign
$\prod$	product sign
$\rightarrow$	converges to, approaches
$\xrightarrow{p}$	converges in probability to
$\xrightarrow{a.s.}$	converges almost surely to
$\xrightarrow{q.m.}$	converges in quadratic mean to
$\xrightarrow{d}$	converges in distribution to
i.i.d.	independently, identically distributed
lim	limit
plim	probability limit
max	maximum
min	minimum
sup	supremum, least upper bound
ln	natural logarithm
exp	exponential function
$\|z\|$	absolute value or modulus of z
K	dimension of a stochastic process or time series
T	sample size, time series length
$\mathbb{R}$	real numbers
$\mathbb{R}^m$	m-dimensional Euclidean space
$\mathbb{C}$	complex numbers
$\mathbb{Z}$	integers
$\mathbb{N}$	positive integers
$\mathbb{I}(\cdot)$	indicator function
L	lag operator
Δ	differencing operator
E	expectation
Var	variance
Cov	covariance, covariance matrix
MSE	mean squared error (matrix)
Pr	probability

$l(\cdot)$	likelihood function
$\ln l$	log-likelihood function
$l_0(\cdot)$	approximate likelihood function
$\ln l_0$	approximate log-likelihood function
λ_{LM}	Lagrange multiplier statistic
λ_{LR}	likelihood ratio statistic
λ_W	Wald statistic
Q_h	portmanteau statistic
$\bar{Q}_h$	modified portmanteau statistic
d.f.	degrees of freedom
AIC	Akaike information criterion
FPE	final prediction error (criterion)
HQ	Hannan-Quinn (criterion)
SC	Schwarz criterion

Distributions and Stochastic Processes

$\mathcal{N}(\mu, \Sigma)$	(multivariate) normal distribution with mean (vector) μ and variance (covariance matrix) Σ
$\chi^2(m)$	χ^2-distribution with m degrees of freedom
$F(m, n)$	F-distribution with m numerator and n denominator degrees of freedom
$t(m)$	t-distribution with m degrees of freedom
AR	autoregressive (process)
AR(p)	autoregressive process of order p
ARCH	autoregressive conditional heteroskedasticity
ARMA	autoregressive moving average (process)
ARMA(p, q)	autoregressive moving average process of order (p, q)
ARMA$_E$	echelon form VARMA model
ARMA$_E(p_1, \ldots, p_K)$	echelon form VARMA model with Kronecker indices $(p_1, \ldots, p_K)$
EC-ARMA$_{RE}$	error correction echelon form VARMA model
GARCH	generalized autoregressive conditional heteroskedasticity
MA	moving average (process)
MA(q)	moving average process of order q
MGARCH	multivariate generalized autoregressive conditional heteroskedasticity
PAR	periodic (vector) autoregression
VAR	vector autoregressive (process)
VAR(p)	vector autoregressive process of order p
VARMA	vector autoregressive moving average (process)
VARMA(p, q)	vector autoregressive moving average process of order (p, q)
VECM	vector error correction model

Vector and Matrix Operations

M'	transpose of M
M^{adj}	adjoint of M
M^{-1}	inverse of M
M^{+}	Moore-Penrose generalized inverse of M
$M_{\perp}$	orthogonal complement of M
$M^{1/2}$	square root of M
M^{k}	k-th power of M
MN	matrix product of M and N
$+$	plus
$-$	minus
$\otimes$	Kronecker product
$\det(M)$, $\det M$	determinant of M
$\lvert M\rvert$	determinant of M
$\lVert M\rVert$	Euclidean norm of M
$\mathrm{rk}(M)$, $\mathrm{rk}\, M$	rank of M
$\mathrm{tr}(M)$, $\mathrm{tr}\, M$	trace of M
vec	column stacking operator
vech	column stacking operator for symmetric matrices (stacks the elements on and below the main diagonal only)
$\dfrac{\partial \varphi}{\partial \beta'}$	vector or matrix of first order partial derivatives of φ with respect to β
$\dfrac{\partial^2 \varphi}{\partial \beta \partial \beta'}$	Hessian matrix of φ, matrix of second order partial derivatives of φ with respect to β

General Matrices

$\mathbf{D}_m$	$(m^2 \times \frac{1}{2}m(m+1))$ duplication matrix
I_m	$(m \times m)$ unit or identity matrix
$\mathcal{I}(\cdot)$	information matrix
$\mathcal{I}_a(\cdot)$	asymptotic information matrix
J	$:= [I_K : 0 : \cdots : 0]$
$\mathbf{K}_{mn}$	$(mn \times mn)$ commutation matrix
$\mathbf{L}_m$	$(\frac{1}{2}m(m+1) \times m^2)$ elimination matrix
0	zero or null matrix or vector

Vectors and Matrices Related to Stochastic Processes and Multiple Time Series

u_t K-dimensional white noise process

u_{kt} k-th element of u_t

$u_{(k)} := \begin{bmatrix} u_{k1} \\ \vdots \\ u_{kT} \end{bmatrix}$

$U := [u_1, \ldots, u_T]$

$\mathbf{u} := \mathrm{vec}(U)$

$U_t := \begin{bmatrix} u_t \\ 0 \\ \vdots \\ 0 \end{bmatrix}$ or $\begin{bmatrix} u_t \\ 0 \\ \vdots \\ 0 \\ u_t \\ 0 \\ \vdots \\ 0 \end{bmatrix}$

y_t K-dimensional stochastic process

y_{kt} k-th element of y_t

$y_{(k)} := \begin{bmatrix} y_{k1} \\ \vdots \\ y_{kT} \end{bmatrix}$

$\overline{y} := \sum_{t=1}^{T} y_t/T$, sample mean (vector)

$y_t(h)$ h-step forecast of y_{t+h} at origin t

$Y := [y_1, \ldots, y_T]$

$\mathbf{y} := \mathrm{vec}(Y)$

$Y_t := \begin{bmatrix} y_t \\ \vdots \\ y_{t-p+1} \end{bmatrix}$ or $\begin{bmatrix} y_t \\ \vdots \\ y_{t-p+1} \\ u_t \\ \vdots \\ u_{t-q+1} \end{bmatrix}$ or $\begin{bmatrix} y_t \\ \vdots \\ y_{t-p+1} \\ x_t \\ \vdots \\ x_{t-s+1} \end{bmatrix}$

$Z_t := \begin{bmatrix} 1 \\ y_t \\ \vdots \\ y_{t-p+1} \end{bmatrix}$

Matrices and Vectors Related to VAR and VARMA Representations and VECMs (Parts I, II, III, IV)

A_i VAR coefficient matrix

$A := [A_1, \ldots, A_p]$

$\boldsymbol{\alpha} := \mathrm{vec}(A)$

$$\mathbf{A} := \begin{bmatrix} A_1 & \ldots & A_{p-1} & A_p \\ I_K & & 0 & 0 \\ & \ddots & \vdots & \vdots \\ 0 & \ldots & I_K & 0 \end{bmatrix} \text{ or } \begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{bmatrix}$$

$$\mathbf{A}_{11} := \begin{bmatrix} A_1 & \ldots & A_{p-1} & A_p \\ I_K & & 0 & 0 \\ & \ddots & \vdots & \vdots \\ 0 & \ldots & I_K & 0 \end{bmatrix} \quad (Kp \times Kp)$$

$$\mathbf{A}_{12} := \begin{bmatrix} M_1 & \ldots & M_{q-1} & M_q \\ 0 & \ldots & 0 & 0 \\ \vdots & & \vdots & \vdots \\ 0 & \ldots & 0 & 0 \end{bmatrix} \quad (Kp \times Kq)$$

$\mathbf{A}_{21} := 0 \quad (Kq \times Kp)$

$$\mathbf{A}_{22} := \begin{bmatrix} 0 & 0 \\ I_{K(q-1)} & 0 \end{bmatrix} \quad (Kq \times Kq)$$

M_i MA coefficient matrix

$\mathbf{m} := \mathrm{vec}[M_1, \ldots, M_q]$

$$\mathbf{M} := \begin{bmatrix} \mathbf{M}_{11} & \mathbf{M}_{12} \\ \mathbf{M}_{21} & \mathbf{M}_{22} \end{bmatrix} \quad (K(p+q) \times K(p+q))$$

$$\mathbf{M}_{11} := \begin{bmatrix} -M_1 & \ldots & -M_{q-1} & -M_q \\ I_K & & 0 & 0 \\ & \ddots & \vdots & \vdots \\ 0 & \ldots & I_K & 0 \end{bmatrix} \quad (Kq \times Kq)$$

$$\mathbf{M}_{12} := \begin{bmatrix} -A_1 & \dots & -A_{p-1} & -A_p \\ 0 & \dots & 0 & 0 \\ \vdots & & \vdots & \vdots \\ 0 & \dots & 0 & 0 \end{bmatrix} \quad (Kq \times Kp)$$

$$\mathbf{M}_{21} := 0 \quad (Kp \times Kq)$$

$$\mathbf{M}_{22} := \begin{bmatrix} 0 & 0 \\ I_{K(p-1)} & 0 \end{bmatrix} \quad (Kp \times Kp)$$

Φ_i coefficient matrix of canonical MA representation
Π_i coefficient matrix of pure VAR representation
α loading matrix of VECM
β cointegration matrix
$\mathbf{\Pi}$ $:= \alpha\beta'$
$\mathbf{\Gamma}_i$ short-run coefficient matrix of VECM

Impulse Responses and Related Quantities

Φ_i matrix of forecast error impulse responses
Ψ_m $:= \sum_{i=0}^{m} \Phi_i$, matrix of accumulated forecast error impulse responses
Ψ_∞ $:= \sum_{i=0}^{\infty} \Phi_i$, matrix of total or long-run forecast error impulse responses
Θ_i matrix of orthogonalized impulse responses
Ξ_m $:= \sum_{i=0}^{m} \Theta_i$, matrix of accumulated orthogonalized impulse responses
Ξ_∞ $:= \sum_{i=0}^{\infty} \Theta_i$, matrix of total or long-run orthogonalized impulse responses
$\omega_{jk,h}$ proportion of h-step forecast error variance of variable j, accounted for by innovations in variable k
$\mathbf{\Xi}$ matrix of long-run effects
$\mathbf{\Xi}_j^*$ matrix of transitory effects

Moment Matrices

Γ	$:= \text{plim } ZZ'/T$
$\Gamma_y(h)$	$:= \text{Cov}(y_t, y_{t-h})$ for a stationary process y_t
$R_y(h)$	correlation matrix corresponding to $\Gamma_y(h)$
Σ_u	$:= E(u_t u_t') = \text{Cov}(u_t)$, white noise covariance matrix
Σ_y	$:= E[(y_t - \mu)(y_t - \mu)'] = \text{Cov}(y_t)$, covariance matrix of a stationary process y_t
P	lower triangular Choleski decomposition of Σ_u
$\Sigma_{\widehat{\boldsymbol{\alpha}}}$	covariance matrix of the asymptotic distribution of $\sqrt{T}(\widehat{\boldsymbol{\alpha}} - \boldsymbol{\alpha})$
$\Omega(h)$	correction term for MSE matrix of h-step forecast
$\Sigma_y(h)$	MSE or forecast error covariance matrix of h-step forecast of y_t
$\Sigma_{\widehat{y}}(h)$	approximate MSE matrix of h-step forecast of estimated process y_t

Author Index

Subject Index