Appendix D
Summary of Matrix Algebra

This appendix summarizes the matrix algebra concepts, including the algebra of probability, needed for the study of multiple linear regression models using matrices in Appendix E. None of this material is used in the main text.

D.1 BASIC DEFINITIONS

DEFINITION D.1 (Matrix)
A **matrix** is a rectangular array of numbers. More precisely, an $m \times n$ matrix has m rows and n columns. The positive integer m is called the *row dimension*, and n is called the *column dimension*.

We use uppercase boldface letters to denote matrices. We can write an $m \times n$ matrix generically as

$$\mathbf{A} = [a_{ij}] = \begin{bmatrix} a_{11} & a_{12} & a_{13} & \cdots & a_{1n} \\ a_{21} & a_{22} & a_{23} & \cdots & a_{2n} \\ \vdots & & & & \\ a_{m1} & a_{m2} & a_{m3} & \cdots & a_{mn} \end{bmatrix},$$

where a_{ij} represents the element in the i^{th} row and the j^{th} column. For example, a_{25} stands for the number in the second row and the fifth column of $\mathbf{A}$. A specific example of a 2×3 matrix is

$$\mathbf{A} = \begin{bmatrix} 2 & -1 & 7 \\ -4 & 5 & 0 \end{bmatrix}, \qquad \textbf{(D.1)}$$

where $a_{13} = 7$. The shorthand $\mathbf{A} = [a_{ij}]$ is often used to define matrix operations.

DEFINITION D.2 (Square Matrix)
A **square matrix** has the same number of rows and columns. The dimension of a square matrix is its number of rows and columns.

DEFINITION D.3 (Vectors)
(i) A $1 \times m$ matrix is called a **row vector** (of dimension m) and can be written as $\boldsymbol{x} \equiv (x_1, x_2, \ldots, x_m)$.

(ii) An $n \times 1$ matrix is called a **column vector** and can be written as

$$\mathbf{y} \equiv \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}.$$

DEFINITION D.4 (Diagonal Matrix)
A square matrix $\boldsymbol{A}$ is a **diagonal matrix** when all of its off-diagonal elements are zero, that is, $a_{ij} = 0$ for all $i \neq j$. We can always write a diagonal matrix as

$$\mathbf{A} = \begin{bmatrix} a_{11} & 0 & 0 & \cdots & 0 \\ 0 & a_{22} & 0 & \cdots & 0 \\ \vdots & & & & \\ 0 & 0 & 0 & \cdots & a_{nn} \end{bmatrix}.$$

DEFINITION D.5 (Identity and Zero Matrices)
(i) The $n \times n$ **identity matrix**, denoted $\mathbf{I}$, or sometimes $\mathbf{I}_n$ to emphasize its dimension, is the diagonal matrix with unity (one) in each diagonal position, and zero elsewhere:

$$\mathbf{I} \equiv \mathbf{I}_n \equiv \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ \vdots & & & & \\ 0 & 0 & 0 & \cdots & 1 \end{bmatrix}.$$

(ii) The $m \times n$ **zero matrix**, denoted $\mathbf{0}$, is the $m \times n$ matrix with zero for all entries. This need not be a square matrix.

D.2 MATRIX OPERATIONS

Matrix Addition

Two matrices $\mathbf{A}$ and $\mathbf{B}$, each having dimension $m \times n$, can be added element by element: $\mathbf{A} + \mathbf{B} = [a_{ij} + b_{ij}]$. More precisely,

$$\mathbf{A} + \mathbf{B} = \begin{bmatrix} a_{11}+b_{11} & a_{12}+b_{12} & \cdots & a_{1n}+b_{1n} \\ a_{21}+b_{21} & a_{22}+b_{22} & \cdots & a_{2n}+b_{2n} \\ \vdots & & & \\ a_{m1}+b_{m1} & a_{m2}+b_{m2} & \cdots & a_{mn}+b_{mn} \end{bmatrix}.$$

For example,

$$\begin{bmatrix} 2 & -1 & 7 \\ -4 & 5 & 0 \end{bmatrix} + \begin{bmatrix} 1 & 0 & -4 \\ 4 & 2 & 3 \end{bmatrix} = \begin{bmatrix} 3 & -1 & 3 \\ 0 & 7 & 3 \end{bmatrix}.$$

Matrices of different dimensions cannot be added.

Scalar Multiplication

Given any real number γ (often called a scalar), **scalar multiplication** is defined as $\gamma\mathbf{A} \equiv [\gamma a_{ij}]$, or

$$\gamma\mathbf{A} = \begin{bmatrix} \gamma a_{11} & \gamma a_{12} & \dots & \gamma a_{1n} \\ \gamma a_{21} & \gamma a_{22} & \dots & \gamma a_{2n} \\ \vdots & & & \\ \gamma a_{m1} & \gamma a_{m2} & \dots & \gamma a_{mn} \end{bmatrix}.$$

For example, if $\gamma = 2$ and $\mathbf{A}$ is the matrix in equation (D.1), then

$$\gamma\mathbf{A} = \begin{bmatrix} 4 & -2 & 14 \\ -8 & 10 & 0 \end{bmatrix}.$$

Matrix Multiplication

To multiply matrix $\mathbf{A}$ by matrix $\mathbf{B}$ to form the product $\mathbf{AB}$, the *column* dimension of $\mathbf{A}$ must equal the *row* dimension of $\mathbf{B}$. Therefore, let $\mathbf{A}$ be an $m \times n$ matrix and let $\mathbf{B}$ be an $n \times p$ matrix. Then, **matrix multiplication** is defined as

$$\mathbf{AB} = \left[\sum_{k=1}^{n} a_{ik}b_{kj} \right].$$

In other words, the $(i,j)^{\text{th}}$ element of the new matrix $\mathbf{AB}$ is obtained by multiplying each element in the i^{th} row of $\mathbf{A}$ by the corresponding element in the j^{th} column of $\mathbf{B}$ and adding these n products together. A schematic may help make this process more transparent:

$$\begin{array}{cccccc} & \mathbf{A} & \mathbf{B} & & \mathbf{AB} & \\ i^{\text{th}}\text{ row} \rightarrow & \begin{bmatrix} \\ a_{i1}\, a_{i2}\, a_{i3} \dots a_{in} \\ \\ \end{bmatrix} & \begin{bmatrix} b_{1j} \\ b_{2j} \\ b_{3j} \\ \vdots \\ b_{nj} \end{bmatrix} & = & \begin{bmatrix} \\ \sum_{k=1}^{n} a_{ik}b_{kj} \\ \uparrow \\ \end{bmatrix}, & \\ & & \uparrow & & & \\ & & j^{\text{th}}\text{ column} & & (i,j)^{\text{th}}\text{ element} & \end{array}$$

where, by the definition of the summation operator in Appendix A,

$$\sum_{k=1}^{n} a_{ik}b_{kj} = a_{i1}b_{1j} + a_{i2}b_{2j} + \ldots + a_{in}b_{nj}.$$

For example,

$$\begin{bmatrix} 2 & -1 & 0 \\ -4 & 1 & 0 \end{bmatrix} \begin{bmatrix} 0 & 1 & 6 & 0 \\ -1 & 2 & 0 & 1 \\ 3 & 0 & 0 & 0 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 12 & -1 \\ -1 & -2 & -24 & 1 \end{bmatrix}.$$

We can also multiply a matrix and a vector. If $\mathbf{A}$ is an $n \times m$ matrix and $\mathbf{y}$ is an $m \times 1$ vector, then $\mathbf{Ay}$ is an $n \times 1$ vector. If $\mathbf{x}$ is a $1 \times n$ vector, then $\mathbf{xA}$ is a $1 \times m$ vector.

Matrix addition, scalar multiplication, and matrix multiplication can be combined in various ways, and these operations satisfy several rules that are familiar from basic operations on numbers. In the following list of properties, $\mathbf{A}$, $\mathbf{B}$, and $\mathbf{C}$ are matrices with appropriate dimensions for applying each operation, and α and β are real numbers. Most of these properties are easy to illustrate from the definitions.

PROPERTIES OF MATRIX MULTIPLICATION: (1) $(\alpha + \beta)\mathbf{A} = \alpha\mathbf{A} + \beta\mathbf{A}$; (2) $\alpha(\mathbf{A} + \mathbf{B}) = \alpha\mathbf{A} + \alpha\mathbf{B}$; (3) $(\alpha\beta)\mathbf{A} = \alpha(\beta\mathbf{A})$; (4) $\alpha(\mathbf{AB}) = (\alpha\mathbf{A})\mathbf{B}$; (5) $\mathbf{A} + \mathbf{B} = \mathbf{B} + \mathbf{A}$; (6) $(\mathbf{A} + \mathbf{B}) + \mathbf{C} = \mathbf{A} + (\mathbf{B} + \mathbf{C})$; (7) $(\mathbf{AB})\mathbf{C} = \mathbf{A}(\mathbf{BC})$; (8) $\mathbf{A}(\mathbf{B} + \mathbf{C}) = \mathbf{AB} + \mathbf{AC}$; (9) $(\mathbf{A} + \mathbf{B})\mathbf{C} = \mathbf{AC} + \mathbf{BC}$; (10) $\mathbf{IA} = \mathbf{AI} = \mathbf{A}$; (11) $\mathbf{A} + \mathbf{0} = \mathbf{0} + \mathbf{A} = \mathbf{A}$; (12) $\mathbf{A} - \mathbf{A} = \mathbf{0}$; (13) $\mathbf{A0} = \mathbf{0A} = \mathbf{0}$; and (14) $\mathbf{AB} \neq \mathbf{BA}$, even when both products are defined.

The last property deserves further comment. If $\mathbf{A}$ is $n \times m$ and $\mathbf{B}$ is $m \times p$, then $\mathbf{AB}$ is defined, but $\mathbf{BA}$ is defined only if $n = p$ (the row dimension of $\mathbf{A}$ equals the column dimension of $\mathbf{B}$). If $\mathbf{A}$ is $m \times n$ and $\mathbf{B}$ is $n \times m$, then $\mathbf{AB}$ and $\mathbf{BA}$ are both defined, but they are not usually the same; in fact, they have different dimensions, unless $\mathbf{A}$ and $\mathbf{B}$ are both square matrices. Even when $\mathbf{A}$ and $\mathbf{B}$ are both square, $\mathbf{AB} \neq \mathbf{BA}$, except under special circumstances.

Transpose

DEFINITION D.6 (Transpose)

Let $\mathbf{A} = [a_{ij}]$ be an $m \times n$ matrix. The **transpose** of $\mathbf{A}$, denoted $\mathbf{A}'$ (called $\mathbf{A}$ *prime*), is the $n \times m$ matrix obtained by interchanging the rows and columns of $\mathbf{A}$. We can write this as $\mathbf{A}' \equiv [a_{ji}]$.

For example,

$$\mathbf{A} = \begin{bmatrix} 2 & -1 & 7 \\ -4 & 5 & 0 \end{bmatrix}, \quad \mathbf{A}' = \begin{bmatrix} 2 & -4 \\ -1 & 5 \\ 7 & 0 \end{bmatrix}.$$

PROPERTIES OF TRANSPOSE: (1) $(\mathbf{A}')' = \mathbf{A}$; (2) $(\alpha\mathbf{A})' = \alpha\mathbf{A}'$ for any scalar α; (3) $(\mathbf{A} + \mathbf{B})' = \mathbf{A}' + \mathbf{B}'$; (4) $(\mathbf{AB})' = \mathbf{B}'\mathbf{A}'$, where $\mathbf{A}$ is $m \times n$ and $\mathbf{B}$ is $n \times k$; (5) $\mathbf{x}'\mathbf{x} = \sum_{i=1}^{n} x_i^2$, where $\boldsymbol{x}$ is an $n \times 1$ vector; and (6) If $\mathbf{A}$ is an $n \times k$ matrix with *rows* given by the $1 \times k$ vectors $\boldsymbol{a}_1, \boldsymbol{a}_2, \ldots, \boldsymbol{a}_n$, so that we can write

$$\mathbf{A} = \begin{bmatrix} \boldsymbol{a}_1 \\ \boldsymbol{a}_2 \\ \vdots \\ \boldsymbol{a}_n \end{bmatrix},$$

then $\mathbf{A}' = (\boldsymbol{a}_1' \, \boldsymbol{a}_2' \ldots \boldsymbol{a}_n')$.

DEFINITION D.7 (Symmetric Matrix)
A square matrix $\mathbf{A}$ is a **symmetric matrix** if and only if $\mathbf{A}' = \mathbf{A}$.

If $\mathbf{X}$ is any $n \times k$ matrix, then $\mathbf{X}'\mathbf{X}$ is always defined and is a symmetric matrix, as can be seen by applying the first and fourth transpose properties (see Problem D.3).

Partitioned Matrix Multiplication

Let $\mathbf{A}$ be an $n \times k$ matrix with rows given by the $1 \times k$ vectors $\boldsymbol{a}_1, \boldsymbol{a}_2, \ldots, \boldsymbol{a}_n$, and let $\mathbf{B}$ be an $n \times m$ matrix with rows given by $1 \times m$ vectors $\boldsymbol{b}_1, \boldsymbol{b}_2, \ldots, \boldsymbol{b}_n$:

$$\mathbf{A} = \begin{bmatrix} \boldsymbol{a}_1 \\ \boldsymbol{a}_2 \\ \vdots \\ \boldsymbol{a}_n \end{bmatrix}, \quad \mathbf{B} = \begin{bmatrix} \boldsymbol{b}_1 \\ \boldsymbol{b}_2 \\ \vdots \\ \boldsymbol{b}_n \end{bmatrix}.$$

Then,

$$\mathbf{A}'\mathbf{B} = \sum_{i=1}^{n} \boldsymbol{a}_i' \boldsymbol{b}_i,$$

where for each i, $\boldsymbol{a}_i'\boldsymbol{b}_i$ is a $k \times m$ matrix. Therefore, $\mathbf{A}'\mathbf{B}$ can be written as the sum of n matrices, each of which is $k \times m$. As a special case, we have

$$\mathbf{A}'\mathbf{A} = \sum_{i=1}^{n} \boldsymbol{a}_i' \boldsymbol{a}_i,$$

where $\boldsymbol{a}_i'\boldsymbol{a}_i$ is a $k \times k$ matrix for all i.

Trace

The trace of a matrix is a very simple operation defined only for *square* matrices.

DEFINITION D.8 (Trace)
For any $n \times n$ matrix $\mathbf{A}$, the **trace of a matrix** $\mathbf{A}$, denoted $\mathrm{tr}(\mathbf{A})$, is the sum of its diagonal elements. Mathematically,

$$\mathrm{tr}(\mathbf{A}) = \sum_{i=1}^{n} a_{ii}.$$

PROPERTIES OF TRACE: (1) $\mathrm{tr}(\mathbf{I}_n) = n$; (2) $\mathrm{tr}(\mathbf{A}') = \mathrm{tr}(\mathbf{A})$; (3) $\mathrm{tr}(\mathbf{A} + \mathbf{B}) = \mathrm{tr}(\mathbf{A}) + \mathrm{tr}(\mathbf{B})$; (4) $\mathrm{tr}(\alpha\mathbf{A}) = \alpha\mathrm{tr}(\mathbf{A})$, for any scalar α; and (5) $\mathrm{tr}(\mathbf{AB}) = \mathrm{tr}(\mathbf{BA})$, where $\mathbf{A}$ is $m \times n$ and $\mathbf{B}$ is $n \times m$.

Inverse

The notion of a matrix inverse is very important for square matrices.

DEFINITION D.9 (Inverse)
An $n \times n$ matrix $\mathbf{A}$ has an **inverse**, denoted $\mathbf{A}^{-1}$, provided that $\mathbf{A}^{-1}\mathbf{A} = \mathbf{I}_n$ and $\mathbf{A}\mathbf{A}^{-1} = \mathbf{I}_n$. In this case, $\mathbf{A}$ is said to be *invertible* or *nonsingular.* Otherwise, it is said to be *noninvertible* or *singular.*

PROPERTIES OF INVERSE: (1) If an inverse exists, it is unique; (2) $(\alpha\mathbf{A})^{-1} = (1/\alpha)\mathbf{A}^{-1}$, if $\alpha \neq 0$ and $\mathbf{A}$ is invertible; (3) $(\mathbf{AB})^{-1} = \mathbf{B}^{-1}\mathbf{A}^{-1}$, if $\mathbf{A}$ and $\mathbf{B}$ are both $n \times n$ and invertible; and (4) $(\mathbf{A}')^{-1} = (\mathbf{A}^{-1})'$.

We will not be concerned with the mechanics of calculating the inverse of a matrix. Any matrix algebra text contains detailed examples of such calculations.

D.3 LINEAR INDEPENDENCE. RANK OF A MATRIX

For a set of vectors having the same dimension, it is important to know whether one vector can be expressed as a linear combination of the remaining vectors.

DEFINITION D.10 (Linear Independence)
Let $\{\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_r\}$ be a set of $n \times 1$ vectors. These are **linearly independent vectors** if and only if

$$\alpha_1\boldsymbol{x}_1 + \alpha_2\boldsymbol{x}_2 + \ldots + \alpha_r\boldsymbol{x}_r = \mathbf{0} \qquad \textbf{(D.2)}$$

implies that $\alpha_1 = \alpha_2 = \ldots = \alpha_r = 0$. If (D.2) holds for a set of scalars that are not all zero, then $\{\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_r\}$ is *linearly dependent.*

The statement that $\{\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_r\}$ is linearly dependent is equivalent to saying that at least one vector in this set can be written as a linear combination of the others.

DEFINITION D.11 (Rank)
(i) Let $\mathbf{A}$ be an $n \times m$ matrix. The **rank of a matrix** $\mathbf{A}$, denoted $\mathrm{rank}(\mathbf{A})$, is the maximum number of linearly independent columns of $\mathbf{A}$.

(ii) If $\mathbf{A}$ is $n \times m$ and $\mathrm{rank}(\mathbf{A}) = m$, then $\mathbf{A}$ has *full column rank.*

If $\mathbf{A}$ is $n \times m$, its rank can be at most m. A matrix has full column rank if its columns form a linearly independent set. For example, the 3×2 matrix

$$\begin{bmatrix} 1 & 3 \\ 2 & 6 \\ 0 & 0 \end{bmatrix}$$

can have at most rank two. In fact, its rank is only one because the second column is three times the first column.

PROPERTIES OF RANK: (1) $\text{rank}(\mathbf{A}') = \text{rank}(\mathbf{A})$; (2) If $\mathbf{A}$ is $n \times k$, then $\text{rank}(\mathbf{A}) \leq \min(n,k)$; and (3) If $\mathbf{A}$ is $k \times k$ and $\text{rank}(\mathbf{A}) = k$, then $\mathbf{A}$ is nonsingular.

D.4 QUADRATIC FORMS AND POSITIVE DEFINITE MATRICES

DEFINITION D.12 (Quadratic Form)
Let $\mathbf{A}$ be an $n \times n$ *symmetric* matrix. The **quadratic form** associated with the matrix $\mathbf{A}$ is the real-valued function defined for all $n \times 1$ vectors $\boldsymbol{x}$:

$$f(\boldsymbol{x}) = \boldsymbol{x}'\mathbf{A}\boldsymbol{x} = \sum_{i=1}^{n} a_{ii}x_i^2 + 2\sum_{i=1}^{n}\sum_{j>i} a_{ij}x_ix_j.$$

DEFINITION D.13 (Positive Definite and Positive Semi-Definite)
(i) A symmetric matrix $\mathbf{A}$ is said to be **positive definite (p.d.)** if

$$\boldsymbol{x}'\mathbf{A}\boldsymbol{x} > 0 \text{ for all } n \times 1 \text{ vectors } \boldsymbol{x} \text{ except } \boldsymbol{x} = \mathbf{0}.$$

(ii) A symmetric matrix $\mathbf{A}$ is **positive semi-definite (p.s.d.)** if

$$\boldsymbol{x}'\mathbf{A}\boldsymbol{x} \geq 0 \text{ for all } n \times 1 \text{ vectors.}$$

If a matrix is positive definite or positive semi-definite, it is automatically assumed to be symmetric.

PROPERTIES OF POSITIVE DEFINITE AND POSITIVE SEMI-DEFINITE MATRICES: (1) A positive definite matrix has diagonal elements that are strictly positive, while a p.s.d. matrix has nonnegative diagonal elements; (2) If $\mathbf{A}$ is p.d., then $\mathbf{A}^{-1}$ exists and is p.d.; (3) If $\mathbf{X}$ is $n \times k$, then $\mathbf{X}'\mathbf{X}$ and $\mathbf{X}\mathbf{X}'$ are p.s.d.; and (4) If $\mathbf{X}$ is $n \times k$ and $\text{rank}(\mathbf{X}) = k$, then $\mathbf{X}'\mathbf{X}$ is p.d. (and therefore nonsingular).

D.5 IDEMPOTENT MATRICES

DEFINITION D.14 (Idempotent Matrix)
Let $\mathbf{A}$ be an $n \times n$ symmetric matrix. Then $\mathbf{A}$ is said to be an **idempotent matrix** if and only if $\mathbf{AA} = \mathbf{A}$.

For example,

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

is an idempotent matrix, as direct multiplication verifies.

PROPERTIES OF IDEMPOTENT MATRICES: Let $\mathbf{A}$ be an $n \times n$ idempotent matrix. (1) $\text{rank}(\mathbf{A}) = \text{tr}(\mathbf{A})$, and (2) $\mathbf{A}$ is positive semi-definite.

We can construct idempotent matrices very generally. Let $\mathbf{X}$ be an $n \times k$ matrix with $\text{rank}(\mathbf{X}) = k$. Define

$$\mathbf{P} \equiv \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$$
$$\mathbf{M} \equiv \mathbf{I}_n - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' = \mathbf{I}_n - \mathbf{P}.$$

Then $\mathbf{P}$ and $\mathbf{M}$ are symmetric, idempotent matrices with $\text{rank}(\mathbf{P}) = k$ and $\text{rank}(\mathbf{M}) = n - k$. The ranks are most easily obtained by using Property 1: $\text{tr}(\mathbf{P}) = \text{tr}[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}]$ (from Property 5 for trace) $= \text{tr}(\mathbf{I}_k) = k$ (by Property 1 for trace). It easily follows that $\text{tr}(\mathbf{M}) = \text{tr}(\mathbf{I}_n) - \text{tr}(\mathbf{P}) = n - k$.

D.6 DIFFERENTIATION OF LINEAR AND QUADRATIC FORMS

For a given $n \times 1$ vector $\boldsymbol{a}$, consider the linear function defined by

$$f(\boldsymbol{x}) = \boldsymbol{a}'\boldsymbol{x},$$

for all $n \times 1$ vectors $\boldsymbol{x}$. The derivative of f with respect to $\boldsymbol{x}$ is the $1 \times n$ vector of partial derivatives, which is simply

$$\partial f(\boldsymbol{x})/\partial \boldsymbol{x} = \boldsymbol{a}'.$$

For an $n \times n$ symmetric matrix $\mathbf{A}$, define the quadratic form

$$g(\boldsymbol{x}) = \boldsymbol{x}'\mathbf{A}\boldsymbol{x}.$$

Then,

$$\partial g(\boldsymbol{x})/\partial \boldsymbol{x} = 2\boldsymbol{x}'\mathbf{A},$$

which is a $1 \times n$ vector.

D.7 MOMENTS AND DISTRIBUTIONS OF RANDOM VECTORS

In order to derive the expected value and variance of the OLS estimators using matrices, we need to define the expected value and variance of a **random vector**. As its name suggests, a random vector is simply a vector of random variables. We also need to define the multivariate normal distribution. These concepts are simply extensions of those covered in Appendix B.

Expected Value

DEFINITION D.15 (Expected Value)

(i) If y is an $n \times 1$ random vector, the **expected value** of y, denoted $E(y)$, is the vector of expected values: $E(y) = [E(y_1), E(y_2), \ldots, E(y_n)]'$.

(ii) If $\mathbf{Z}$ is an $n \times m$ random matrix, $E(\mathbf{Z})$ is the $n \times m$ matrix of expected values: $E(\mathbf{Z}) = [E(z_{ij})]$.

PROPERTIES OF EXPECTED VALUE: (1) If $\mathbf{A}$ is an $m \times n$ matrix and b is an $n \times 1$ vector, where both are nonrandom, then $E(\mathbf{A}y + b) = \mathbf{A}E(y) + b$ and (2) If $\mathbf{A}$ is $p \times n$ and $\mathbf{B}$ is $m \times k$, where both are nonrandom, then $E(\mathbf{AZB}) = \mathbf{A}E(\mathbf{Z})\mathbf{B}$.

Variance-Covariance Matrix

DEFINITION D.16 (Variance-Covariance Matrix)

If y is an $n \times 1$ random vector, its **variance-covariance matrix**, denoted $\mathrm{Var}(y)$, is defined as

$$\mathrm{Var}(y) = \begin{bmatrix} \sigma_1^2 & \sigma_{12} & \cdots & \sigma_{1n} \\ \sigma_{21} & \sigma_2^2 & \cdots & \sigma_{2n} \\ \vdots & & & \\ \sigma_{n1} & \sigma_{n2} & \cdots & \sigma_n^2 \end{bmatrix},$$

where $\sigma_j^2 = \mathrm{Var}(y_j)$ and $\sigma_{ij} = \mathrm{Cov}(y_i, y_j)$. In other words, the variance-covariance matrix has the variances of each element of y down its diagonal, with covariance terms in the off diagonals. Because $\mathrm{Cov}(y_i, y_j) = \mathrm{Cov}(y_j, y_i)$, it immediately follows that a variance-covariance matrix is symmetric.

PROPERTIES OF VARIANCE: (1) If a is an $n \times 1$ nonrandom vector, then $\mathrm{Var}(a'y) = a'[\mathrm{Var}(y)]a \geq 0$; (2) If $\mathrm{Var}(a'y) > 0$ for all $a \neq \mathbf{0}$, $\mathrm{Var}(y)$ is positive definite; (3) $\mathrm{Var}(y) = E[(y - \boldsymbol{\mu})(y - \boldsymbol{\mu})']$, where $\boldsymbol{\mu} = E(y)$; (4) If the elements of y are uncorrelated, $\mathrm{Var}(y)$ is a diagonal matrix. If, in addition, $\mathrm{Var}(y_j) = \sigma^2$ for $j = 1,2,\ldots,n$, then $\mathrm{Var}(y) = \sigma^2 I_n$ and (5) If A is an $m \times n$ nonrandom matrix and b is an $n \times 1$ nonrandom vector, then $\mathrm{Var}(\mathbf{A}y + b) = A[\mathrm{Var}(y)]A'$.

Multivariate Normal Distribution

The normal distribution for a random variable was discussed at some length in Appendix B. We need to extend the normal distribution to random vectors. We will not provide an expression for the probability distribution function, as we do not need it. It is important to know that a multivariate normal random vector is completely characterized by its mean and its variance-covariance matrix. Therefore, if y is an $n \times 1$ multivariate normal random vector with mean $\boldsymbol{\mu}$ and variance-covariance matrix $\boldsymbol{\Sigma}$, we write $y \sim \mathrm{Normal}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. We now state several useful properties of the **multivariate normal distribution**.

PROPERTIES OF THE MULTIVARIATE NORMAL DISTRIBUTION: (1) If $y \sim \mathrm{Normal}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, then each element of y is normally distributed; (2) If $y \sim \mathrm{Normal}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, then y_i and y_j, any two elements of y, are independent if and only if they are uncorrelated, that is, $\sigma_{ij} = 0$; (3) If $y \sim \mathrm{Normal}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, then $\mathbf{A}y + b \sim \mathrm{Normal}(\mathbf{A}\boldsymbol{\mu} + b, \mathbf{A}\boldsymbol{\Sigma}\mathbf{A}')$, where $\mathbf{A}$ and b are nonrandom; (4) If $y \sim \mathrm{Normal}(\mathbf{0}, \boldsymbol{\Sigma})$, then, for nonrandom matrices $\mathbf{A}$ and $\mathbf{B}$, $\mathbf{A}y$ and $\mathbf{B}y$ are independent if and only if $\mathbf{A}\boldsymbol{\Sigma}\mathbf{B}' = \mathbf{0}$. In particular, if $\boldsymbol{\Sigma} = \sigma^2\mathbf{I}_n$, then $\mathbf{AB}' = \mathbf{0}$ is necessary and sufficient for independence of $\mathbf{A}y$ and $\mathbf{B}y$; (5) If $y \sim \mathrm{Normal}(\mathbf{0}, \sigma^2\mathbf{I}_n)$, $\mathbf{A}$ is a $k \times n$ nonrandom matrix, and $\mathbf{B}$ is an $n \times n$ symmetric, idempotent matrix, then $\mathbf{A}y$ and $y'\mathbf{B}y$ are independent if and only if $\mathbf{AB} = \mathbf{0}$; and (6) If $y \sim \mathrm{Normal}(\mathbf{0}, \sigma^2\mathbf{I}_n)$ and $\mathbf{A}$ and $\mathbf{B}$ are nonrandom symmetric, idempotent matrices, then $y'\mathbf{A}y$ and $y'\mathbf{B}y$ are independent if and only if $\mathbf{AB} = \mathbf{0}$.

Chi-Square Distribution

In Appendix B, we defined a **chi-square random variable** as the sum of *squared* independent standard normal random variables. In vector notation, if $u \sim \mathrm{Normal}(\mathbf{0}, \mathbf{I}_n)$, then $u'u \sim \chi_n^2$.

PROPERTIES OF THE CHI-SQUARE DISTRIBUTION: (1) If $u \sim \mathrm{Normal}(\mathbf{0}, \mathbf{I}_n)$ and $\mathbf{A}$ is an $n \times n$ symmetric, idempotent matrix with $\mathrm{rank}(\mathbf{A}) = q$, then $u'\mathbf{A}u \sim \chi_q^2$; (2) If $u \sim \mathrm{Normal}(\mathbf{0}, \mathbf{I}_n)$ and $\mathbf{A}$ and $\mathbf{B}$ are $n \times n$ symmetric, idempotent matrices such that $\mathbf{AB} = \mathbf{0}$, then $u'\mathbf{A}u$ and $u'\mathbf{B}u$ are independent, chi-square random variables; and (3) If $z \sim \mathrm{Normal}\ (\mathbf{0}, \mathbf{C})$ where $\mathbf{C}$ is an $m \times m$ nonsingular matrix, then $z'\mathbf{C}^{-1}z \sim \chi_m^2$.

t Distribution

We also defined the ***t* distribution** in Appendix B. Now we add an important property.

PROPERTY OF THE T DISTRIBUTION: If $u \sim \mathrm{Normal}(\mathbf{0}, \mathbf{I}_n)$, c is an $n \times 1$ nonrandom vector, $\mathbf{A}$ is a nonrandom $n \times n$ symmetric, idempotent matrix with rank q, and $\mathbf{A}c = \mathbf{0}$, then $\{c'u/(c'c)^{1/2}\}/(u'\mathbf{A}u)^{1/2} \sim t_q$.

F Distribution

Recall that an ***F* random variable** is obtained by taking two *independent* chi-square random variables and finding the ratio of each, standardized by degrees of freedom.

PROPERTY OF THE F DISTRIBUTION: If $u \sim \mathrm{Normal}(\mathbf{0}, \mathbf{I}_n)$ and $\mathbf{A}$ and $\mathbf{B}$ are $n \times n$ nonrandom symmetric, idempotent matrices with $\mathrm{rank}(\mathbf{A}) = k_1$, $\mathrm{rank}(\mathbf{B}) = k_2$, and $\mathbf{AB} = \mathbf{0}$, then $(u'\mathbf{A}u/k_1)/(u'\mathbf{B}u/k_2) \sim F_{k_1,k_2}$.

SUMMARY

This appendix contains a condensed form of the background information needed to study the classical linear model using matrices. While the material here is self-contained, it is primarily intended as a review for readers who are familiar with matrix algebra and multivariate statistics, and it will be used extensively in Appendix E.

KEY TERMS

Chi-Square Random Variable
Column Vector
Diagonal Matrix
Expected Value
F Random Variable
Idempotent Matrix
Identity Matrix
Inverse
Linearly Independent Vectors
Matrix
Matrix Multiplication
Multivariate Normal Distribution
Positive Definite (p.d.)
Positive Semi-Definite (p.s.d.)
Quadratic Form
Random Vector
Rank of a Matrix
Row Vector
Scalar Multiplication
Square Matrix
Symmetric Matrix
t Distribution
Trace of a Matrix
Transpose
Variance-Covariance Matrix
Zero Matrix

PROBLEMS

D.1 (i) Find the product **AB** using

$$\mathbf{A} = \begin{bmatrix} 2 & -1 & 7 \\ -4 & 5 & 0 \end{bmatrix}, \quad \mathbf{B} = \begin{bmatrix} 0 & 1 & 6 \\ 1 & 8 & 0 \\ 3 & 0 & 0 \end{bmatrix}.$$

(ii) Does **BA** exist?

D.2 If **A** and **B** are $n \times n$ diagonal matrices, show that $\mathbf{AB} = \mathbf{BA}$.

D.3 Let **X** be any $n \times k$ matrix. Show that $\mathbf{X}'\mathbf{X}$ is a symmetric matrix.

D.4 (i) Use the properties of trace to argue that $\text{tr}(\mathbf{A}'\mathbf{A}) = \text{tr}(\mathbf{A}\mathbf{A}')$ for any $n \times m$ matrix **A**.

(ii) For $\mathbf{A} = \begin{bmatrix} 2 & 0 & -1 \\ 0 & 3 & 0 \end{bmatrix}$, verify that $\text{tr}(\mathbf{A}'\mathbf{A}) = \text{tr}(\mathbf{A}\mathbf{A}')$.

D.5 (i) Use the definition of inverse to prove the following: if **A** and **B** are $n \times n$ nonsingular matrices, then $(\mathbf{AB})^{-1} = \mathbf{B}^{-1}\mathbf{A}^{-1}$.

(ii) If **A**, **B**, and **C** are all $n \times n$ nonsingular matrices, find $(\mathbf{ABC})^{-1}$ in terms of $\mathbf{A}^{-1}$, $\mathbf{B}^{-1}$, and $\mathbf{C}^{-1}$.

D.6 (i) Show that if **A** is an $n \times n$ symmetric, positive definite matrix, then **A** must have strictly positive diagonal elements.

(ii) Write down a 2×2 symmetric matrix with strictly positive diagonal elements that is *not* positive definite.

D.7 Let **A** be an $n \times n$ symmetric, positive definite matrix. Show that if **P** is any $n \times n$ nonsingular matrix, then $\mathbf{P}'\mathbf{AP}$ is positive definite.

D.8 Prove Property 5 of variances for vectors, using Property 3.

Appendix E
The Linear Regression Model in Matrix Form

This appendix derives various results for ordinary least squares estimation of the multiple linear regression model using matrix notation and matrix algebra (see Appendix D for a summary). The material presented here is much more advanced than that in the text.

E.1 THE MODEL AND ORDINARY LEAST SQUARES ESTIMATION

Throughout this appendix, we use the t subscript to index observations and an n to denote the sample size. It is useful to write the multiple linear regression model with k parameters as follows:

$$y_t = \beta_1 + \beta_2 x_{t2} + \beta_3 x_{t3} + \ldots + \beta_k x_{tk} + u_t, \; t = 1,2,\ldots,n, \tag{E.1}$$

where y_t is the dependent variable for observation t, and x_{tj}, $j = 2,3,\ldots,k$, are the independent variables. Notice how our labeling convention here differs from the text: we call the intercept β_1 and let $\beta_2, \ldots, \beta_k$ denote the slope parameters. This relabeling is not important, but it simplifies the notation for the matrix approach to multiple regression.

For each t, define a $1 \times k$ vector, $\boldsymbol{x}_t = (1, x_{t2}, \ldots, x_{tk})$, and let $\boldsymbol{\beta} = (\beta_1, \beta_2, \ldots, \beta_k)'$ be the $k \times 1$ vector of all parameters. Then, we can write (E.1) as

$$y_t = \boldsymbol{x}_t\boldsymbol{\beta} + u_t, \; t = 1,2,\ldots,n. \tag{E.2}$$

[Some authors prefer to define $\boldsymbol{x}_t$ as a column vector, in which case, $\boldsymbol{x}_t$ is replaced with $\boldsymbol{x}_t'$ in (E.2). Mathematically, it makes more sense to define it as a row vector.] We can write (E.2) in full matrix notation by appropriately defining data vectors and matrices. Let $\boldsymbol{y}$ denote the $n \times 1$ vector of observations on y: the t^{th} element of $\boldsymbol{y}$ is y_t. Let **X** be the $n \times k$ vector of observations on the explanatory variables. In other words, the t^{th} row of **X** consists of the vector $\boldsymbol{x}_t$. Equivalently, the $(t,j)^{\text{th}}$ element of **X** is simply x_{tj}: