

# LINEAR ALGEBRA AND MATRIX METHODS IN ECONOMETRICS

HENRI THEIL\*

*University of Florida*

## Contents

1. Introduction	5
2. Why are matrix methods useful in econometrics?	5
2.1. Linear systems and quadratic forms	5
2.2. Vectors and matrices in statistical theory	7
2.3. Least squares in the standard linear model	8
2.4. Vectors and matrices in consumption theory	10
3. Partitioned matrices	12
3.1. The algebra of partitioned matrices	12
3.2. Block-recursive systems	14
3.3. Income and price derivatives revisited	15
4. Kronecker products and the vectorization of matrices	16
4.1. The algebra of Kronecker products	16
4.2. Joint generalized least-squares estimation of several equations	17
4.3. Vectorization of matrices	19
5. Differential demand and supply systems	20
5.1. A differential consumer demand system	20
5.2. A comparison with simultaneous equation systems	22
5.3. An extension to the inputs of a firm: A singularity problem	23
5.4. A differential input demand system	23
5.5. Allocation systems	25
5.6. Extensions	25
6. Definite and semidefinite square matrices	27
6.1. Covariance matrices and Gauss-Markov further considered	27
6.2. Maxima and minima	29
6.3. Block-diagonal definite matrices	30
7. Diagonalizations	30
7.1. The standard diagonalization of a square matrix	30

\*Research supported in part by NSF Grant SOC76-82718. The author is indebted to Kenneth Clements (Reserve Bank of Australia, Sydney) and Michael Intriligator (University of California, Los Angeles) for comments on an earlier draft of this chapter.

7.2. Special cases	32
7.3. Aitken's theorem	33
7.4. The Cholesky decomposition	34
7.5. Vectors written as diagonal matrices	34
7.6. A simultaneous diagonalization of two square matrices	35
7.7. Latent roots of an asymmetric matrix	36
<b>8. Principal components and extensions</b>	<b>37</b>
8.1. Principal components	37
8.2. Derivations	38
8.3. Further discussion of principal components	40
8.4. The independence transformation in microeconomic theory	40
8.5. An example	43
8.6. A principal component interpretation	44
<b>9. The modeling of a disturbance covariance matrix</b>	<b>45</b>
9.1. Rational random behavior	46
9.2. The asymptotics of rational random behavior	47
9.3. Applications to demand and supply	49
<b>10. The Moore–Penrose inverse</b>	<b>51</b>
10.1. Proof of the existence and uniqueness	51
10.2. Special cases	52
10.3. A generalization of Aitken's theorem	53
10.4. Deleting an equation from an allocation model	56
<b>Appendix A: Linear independence and related topics</b>	<b>57</b>
<b>Appendix B: The independence transformation</b>	<b>58</b>
<b>Appendix C: Rational random behavior</b>	<b>61</b>
<b>References</b>	<b>64</b>

## 1. Introduction

Vectors and matrices played a minor role in the econometric literature published before World War II, but they have become an indispensable tool in the last several decades. Part of this development results from the importance of matrix tools for the statistical component of econometrics; another reason is the increased use of matrix algebra in the economic theory underlying econometric relations. The objective of this chapter is to provide a selective survey of both areas. Elementary properties of matrices and determinants are assumed to be known, including summation, multiplication, inversion, and transposition, but the concepts of linear dependence and orthogonality of vectors and the rank of a matrix are briefly reviewed in Appendix A. Reference is made to Dhrymes (1978), Graybill (1969), or Hadley (1961) for elementary properties not covered in this chapter.

Matrices are indicated by boldface italic upper case letters (such as  $A$ ), column vectors by boldface italic lower case letters ( $a$ ), and row vectors by boldface italic lower case letters with a prime added ( $a'$ ) to indicate that they are obtained from the corresponding column vector by transposition. The following abbreviations are used:

LS = least squares,

GLS = generalized least squares,

ML = maximum likelihood,

$\delta_{ij}$  = Kronecker delta ( $= 1$  if  $i = j$ ,  $0$  if  $i \neq j$ ).

## 2. Why are matrix methods useful in econometrics?

### 2.1. Linear systems and quadratic forms

A major reason why matrix methods are useful is that many topics in econometrics have a multivariate character. For example, consider a system of  $L$  simultaneous linear equations in  $L$  endogenous and  $K$  exogenous variables. We write  $y_{\alpha l}$  and  $x_{\alpha k}$  for the  $\alpha$ th observation on the  $l$ th endogenous and the  $k$ th exogenous variable. Then the  $j$ th equation for observation  $\alpha$  takes the form

$$\sum_{l=1}^L \gamma_{lj} y_{\alpha l} + \sum_{k=1}^K \beta_{kj} x_{\alpha k} = \varepsilon_{\alpha j}, \quad (2.1)$$

where  $\epsilon_{\alpha j}$  is a random disturbance and the  $\gamma$ 's and  $\beta$ 's are coefficients. We can write (2.1) for  $j=1, \dots, L$  in the form

$$y'_\alpha \Gamma + x'_\alpha \mathbf{B} = \epsilon'_\alpha, \quad (2.2)$$

where  $y'_\alpha = [y_{\alpha 1} \dots y_{\alpha L}]$  and  $x'_\alpha = [x_{\alpha 1} \dots x_{\alpha K}]$  are observation vectors on the endogenous and the exogenous variables, respectively,  $\epsilon'_\alpha = [\epsilon_{\alpha 1} \dots \epsilon_{\alpha L}]$  is a disturbance vector, and  $\Gamma$  and  $\mathbf{B}$  are coefficient matrices of order  $L \times L$  and  $K \times L$ , respectively:

$$\Gamma = \begin{bmatrix} \gamma_{11} & \gamma_{12} \dots \gamma_{1L} \\ \gamma_{21} & \gamma_{22} \dots \gamma_{2L} \\ \vdots & \vdots \quad \vdots \\ \gamma_{L1} & \gamma_{L2} \dots \gamma_{LL} \end{bmatrix}, \quad \mathbf{B} = \begin{bmatrix} \beta_{11} & \beta_{12} \dots \beta_{1L} \\ \beta_{21} & \beta_{22} \dots \beta_{2L} \\ \vdots & \vdots \quad \vdots \\ \beta_{K1} & \beta_{K2} \dots \beta_{KL} \end{bmatrix}.$$

When there are  $n$  observations ( $\alpha = 1, \dots, n$ ), there are  $Ln$  equations of the form (2.1) and  $n$  equations of the form (2.2). We can combine these equations compactly into

$$Y\Gamma + X\mathbf{B} = E, \quad (2.3)$$

where  $Y$  and  $X$  are observation matrices of the two sets of variables of order  $n \times L$  and  $n \times K$ , respectively:

$$Y = \begin{bmatrix} y_{11} & y_{12} \dots y_{1L} \\ y_{21} & y_{22} \dots y_{2L} \\ \vdots & \vdots \quad \vdots \\ y_{n1} & y_{n2} \dots y_{nL} \end{bmatrix}, \quad X = \begin{bmatrix} x_{11} & x_{12} \dots x_{1K} \\ x_{21} & x_{22} \dots x_{2K} \\ \vdots & \vdots \quad \vdots \\ x_{n1} & x_{n2} \dots x_{nK} \end{bmatrix},$$

and  $E$  is an  $n \times L$  disturbance matrix:

$$E = \begin{bmatrix} \epsilon_{11} & \epsilon_{12} \dots \epsilon_{1L} \\ \epsilon_{21} & \epsilon_{22} \dots \epsilon_{2L} \\ \vdots & \vdots \quad \vdots \\ \epsilon_{n1} & \epsilon_{n2} \dots \epsilon_{nL} \end{bmatrix}.$$

Note that  $\Gamma$  is square ( $L \times L$ ). If  $\Gamma$  is also non-singular, we can postmultiply (2.3) by  $\Gamma^{-1}$ :

$$Y = -X\mathbf{B}\Gamma^{-1} + E\Gamma^{-1}. \quad (2.4)$$

This is the *reduced form* for all  $n$  observations on all  $L$  endogenous variables, each of which is described linearly in terms of exogenous values and disturbances. By contrast, the equations (2.1) or (2.2) or (2.3) from which (2.4) is derived constitute the *structural form* of the equation system.

The previous paragraphs illustrate the convenience of matrices for linear systems. However, the expression “linear algebra” should not be interpreted in the sense that matrices are useful for linear systems only. The treatment of quadratic functions can also be simplified by means of matrices. Let  $g(z_1, \dots, z_k)$  be a three times differentiable function. A Taylor expansion yields

$$g(z_1, \dots, z_k) = g(\bar{z}_1, \dots, \bar{z}_k) + \sum_{i=1}^k (z_i - \bar{z}_i) \frac{\partial g}{\partial z_i} + \frac{1}{2} \sum_{i=1}^k \sum_{j=1}^k (z_i - \bar{z}_i) \frac{\partial^2 g}{\partial z_i \partial z_j} (z_j - \bar{z}_j) + O_3, \quad (2.5)$$

where  $O_3$  is a third-order remainder term, while the derivatives  $\partial g / \partial z_i$  and  $\partial^2 g / \partial z_i \partial z_j$  are all evaluated at  $z_1 = \bar{z}_1, \dots, z_k = \bar{z}_k$ . We introduce  $\mathbf{z}$  and  $\bar{\mathbf{z}}$  as vectors with  $i$ th elements  $z_i$  and  $\bar{z}_i$ , respectively. Then (2.5) can be written in the more compact form

$$g(\mathbf{z}) = g(\bar{\mathbf{z}}) + (\mathbf{z} - \bar{\mathbf{z}})' \frac{\partial g}{\partial \mathbf{z}} + \frac{1}{2} (\mathbf{z} - \bar{\mathbf{z}})' \frac{\partial^2 g}{\partial \mathbf{z} \partial \mathbf{z}'} (\mathbf{z} - \bar{\mathbf{z}}) + O_3, \quad (2.6)$$

where the column vector  $\partial g / \partial \mathbf{z} = [\partial g / \partial z_i]$  is the *gradient* of  $g(\cdot)$  at  $\bar{\mathbf{z}}$  (the vector of first-order derivatives) and the matrix  $\partial^2 g / \partial \mathbf{z} \partial \mathbf{z}' = [\partial^2 g / \partial z_i \partial z_j]$  is the *Hessian matrix* of  $g(\cdot)$  at  $\bar{\mathbf{z}}$  (the matrix of second-order derivatives). A Hessian matrix is always symmetric when the function is three times differentiable.

## 2.2. Vectors and matrices in statistical theory

Vectors and matrices are also important in the statistical component of econometrics. Let  $\mathbf{r}$  be a column vector consisting of the random variables  $r_1, \dots, r_n$ . The expectation  $\mathfrak{E}\mathbf{r}$  is defined as the column vector of expectations  $\mathfrak{E}r_1, \dots, \mathfrak{E}r_n$ . Next consider

$$(\mathbf{r} - \mathfrak{E}\mathbf{r})(\mathbf{r} - \mathfrak{E}\mathbf{r})' = \begin{bmatrix} r_1 - \mathfrak{E}r_1 \\ r_2 - \mathfrak{E}r_2 \\ \vdots \\ r_n - \mathfrak{E}r_n \end{bmatrix} \begin{bmatrix} r_1 - \mathfrak{E}r_1 & r_2 - \mathfrak{E}r_2 & \dots & r_n - \mathfrak{E}r_n \end{bmatrix}$$

and take the expectation of each element of this product matrix. When defining the expectation of a random matrix as the matrix of the expectations of the constituent elements, we obtain:

$$\mathfrak{E}[(\mathbf{r} - \mathfrak{E}\mathbf{r})(\mathbf{r} - \mathfrak{E}\mathbf{r})'] = \begin{bmatrix} \text{var } r_1 & \text{cov}(r_1, r_2) & \cdots & \text{cov}(r_1, r_n) \\ \text{cov}(r_2, r_1) & \text{var } r_2 & \cdots & \text{cov}(r_2, r_n) \\ \vdots & \vdots & \ddots & \vdots \\ \text{cov}(r_n, r_1) & \text{cov}(r_n, r_2) & \cdots & \text{var } r_n \end{bmatrix}.$$

This is the variance–covariance matrix (*covariance matrix*, for short) of the vector  $\mathbf{r}$ , to be written  $\mathfrak{V}(\mathbf{r})$ . The covariance matrix is always symmetric and contains the variances along the diagonal. If the elements of  $\mathbf{r}$  are pairwise uncorrelated,  $\mathfrak{V}(\mathbf{r})$  is a diagonal matrix. If these elements also have equal variances (equal to  $\sigma^2$ , say),  $\mathfrak{V}(\mathbf{r})$  is a *scalar matrix*,  $\sigma^2\mathbf{I}$ ; that is, a scalar multiple  $\sigma^2$  of the unit or identity matrix.

The multivariate nature of econometrics was emphasized at the beginning of this section. This will usually imply that there are several unknown parameters; we arrange these in a vector  $\boldsymbol{\theta}$ . The problem is then to obtain a “good” estimator  $\hat{\boldsymbol{\theta}}$  of  $\boldsymbol{\theta}$  as well as a satisfactory measure of how good the estimator is; the most popular measure is the covariance matrix  $\mathfrak{V}(\hat{\boldsymbol{\theta}})$ . Sometimes this problem is simple, but that is not always the case, in particular when the model is non-linear in the parameters. A general method of estimation is maximum likelihood (ML) which can be shown to have certain optimal properties for large samples under relatively weak conditions. The derivation of the ML estimates and their large-sample covariance matrix involves the *information matrix*, which is (apart from sign) the expectation of the matrix of second-order derivatives of the log-likelihood function with respect to the parameters. The prominence of ML estimation in recent years has greatly contributed to the increased use of matrix methods in econometrics.

### 2.3. Least squares in the standard linear model

We consider the model

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \quad (2.7)$$

where  $\mathbf{y}$  is an  $n$ -element column vector of observations on the dependent (or endogenous) variable,  $\mathbf{X}$  is an  $n \times K$  observation matrix of rank  $K$  on the  $K$  independent (or exogenous) variables,  $\boldsymbol{\beta}$  is a parameter vector, and  $\boldsymbol{\varepsilon}$  is a

disturbance vector. The *standard linear model* postulates that  $\boldsymbol{\varepsilon}$  has zero expectation and covariance matrix  $\sigma^2\mathbf{I}$ , where  $\sigma^2$  is an unknown positive parameter, and that the elements of  $\mathbf{X}$  are all non-stochastic. Note that this model can be viewed as a special case of (2.3) for  $\mathbf{\Gamma} = \mathbf{I}$  and  $L = 1$ .

The problem is to estimate  $\boldsymbol{\beta}$  and  $\sigma^2$ . The *least-squares* (LS) estimator of  $\boldsymbol{\beta}$  is

$$\mathbf{b} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} \quad (2.8)$$

which owes its name to the fact that it minimizes the residual sum of squares. To verify this proposition we write  $\mathbf{e} = \mathbf{y} - \mathbf{X}\mathbf{b}$  for the residual vector; then the residual sum of squares equals

$$\mathbf{e}'\mathbf{e} = \mathbf{y}'\mathbf{y} - 2\mathbf{y}'\mathbf{X}\mathbf{b} + \mathbf{b}'\mathbf{X}'\mathbf{X}\mathbf{b}, \quad (2.9)$$

which is to be minimized by varying  $\mathbf{b}$ . This is achieved by equating the gradient of (2.9) to zero. A comparison of (2.9) with (2.5) and (2.6), with  $\mathbf{z}$  interpreted as  $\mathbf{b}$ , shows that the gradient of (2.9) equals  $-2\mathbf{X}'\mathbf{y} + 2\mathbf{X}'\mathbf{X}\mathbf{b}$ , from which the solution (2.8) follows directly.

Substitution of (2.7) into (2.8) yields  $\mathbf{b} - \boldsymbol{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\varepsilon}$ . Hence, given  $\mathbb{E}\boldsymbol{\varepsilon} = \mathbf{0}$  and the non-randomness of  $\mathbf{X}$ ,  $\mathbf{b}$  is an unbiased estimator of  $\boldsymbol{\beta}$ . Its covariance matrix is

$$\mathbb{V}(\mathbf{b}) = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbb{V}(\boldsymbol{\varepsilon})\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} = \sigma^2(\mathbf{X}'\mathbf{X})^{-1} \quad (2.10)$$

because  $\mathbf{X}'\mathbb{V}(\boldsymbol{\varepsilon})\mathbf{X} = \sigma^2\mathbf{X}'\mathbf{X}$  follows from  $\mathbb{V}(\boldsymbol{\varepsilon}) = \sigma^2\mathbf{I}$ . The Gauss–Markov theorem states that  $\mathbf{b}$  is a best linear unbiased estimator of  $\boldsymbol{\beta}$ , which amounts to an optimum LS property within the class of  $\boldsymbol{\beta}$  estimators that are linear in  $\mathbf{y}$  and unbiased. This property implies that each element of  $\mathbf{b}$  has the smallest possible variance; that is, there exists no other linear unbiased estimator of  $\boldsymbol{\beta}$  whose elements have smaller variances than those of the corresponding elements of  $\mathbf{b}$ . A more general formulation of the Gauss–Markov theorem will be given and proved in Section 6.

Substitution of (2.8) into  $\mathbf{e} = \mathbf{y} - \mathbf{X}\mathbf{b}$  yields  $\mathbf{e} = \mathbf{M}\mathbf{y}$ , where  $\mathbf{M}$  is the symmetric matrix

$$\mathbf{M} = \mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' \quad (2.11)$$

which satisfies  $\mathbf{M}\mathbf{X} = \mathbf{0}$ ; therefore,  $\mathbf{e} = \mathbf{M}\mathbf{y} = \mathbf{M}(\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}) = \mathbf{M}\boldsymbol{\varepsilon}$ . Also,  $\mathbf{M}$  is *idempotent*, i.e.  $\mathbf{M}^2 = \mathbf{M}$ . The LS residual sum of squares equals  $\mathbf{e}'\mathbf{e} = \boldsymbol{\varepsilon}'\mathbf{M}'\mathbf{M}\boldsymbol{\varepsilon} = \boldsymbol{\varepsilon}'\mathbf{M}^2\boldsymbol{\varepsilon}$  and hence

$$\mathbf{e}'\mathbf{e} = \boldsymbol{\varepsilon}'\mathbf{M}\boldsymbol{\varepsilon}. \quad (2.12)$$

It is shown in the next paragraph that  $\mathfrak{E}(\boldsymbol{\varepsilon}'\mathbf{M}\boldsymbol{\varepsilon}) = \sigma^2(n - K)$  so that (2.12) implies that  $\sigma^2$  is estimated unbiasedly by  $\boldsymbol{e}'\boldsymbol{e}/(n - K)$ : the LS residual sum of squares divided by the excess of the number of observations ( $n$ ) over the number of coefficients adjusted ( $K$ ).

To prove  $\mathfrak{E}(\boldsymbol{\varepsilon}'\mathbf{M}\boldsymbol{\varepsilon}) = \sigma^2(n - K)$  we define the *trace* of a square matrix as the sum of its diagonal elements:  $\text{tr } \mathbf{A} = a_{11} + \dots + a_{nn}$ . We use  $\text{tr } \mathbf{A}\mathbf{B} = \text{tr } \mathbf{B}\mathbf{A}$  (if  $\mathbf{A}\mathbf{B}$  and  $\mathbf{B}\mathbf{A}$  exist) to write  $\boldsymbol{\varepsilon}'\mathbf{M}\boldsymbol{\varepsilon}$  as  $\text{tr } \mathbf{M}\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}'$ . Next we use  $\text{tr}(\mathbf{A} + \mathbf{B}) = \text{tr } \mathbf{A} + \text{tr } \mathbf{B}$  (if  $\mathbf{A}$  and  $\mathbf{B}$  are square of the same order) to write  $\text{tr } \mathbf{M}\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}'$  as  $\text{tr } \boldsymbol{\varepsilon}\boldsymbol{\varepsilon}' - \text{tr } \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}'$  [see (2.11)]. Thus, since  $\mathbf{X}$  is non-stochastic and the trace is a linear operator,

$$\begin{aligned}\mathfrak{E}(\boldsymbol{\varepsilon}'\mathbf{M}\boldsymbol{\varepsilon}) &= \text{tr } \mathfrak{E}(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}') - \text{tr } \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathfrak{E}(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}') \\ &= \sigma^2 \text{tr } \mathbf{I} - \sigma^2 \text{tr } \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' \\ &= \sigma^2 n - \sigma^2 \text{tr}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X},\end{aligned}$$

which confirms  $\mathfrak{E}(\boldsymbol{\varepsilon}'\mathbf{M}\boldsymbol{\varepsilon}) = \sigma^2(n - K)$  because  $(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X} = \mathbf{I}$  of order  $K \times K$ .

If, in addition to the conditions listed in the discussion following eq. (2.7), the elements of  $\boldsymbol{\varepsilon}$  are normally distributed, the LS estimator  $\mathbf{b}$  of  $\boldsymbol{\beta}$  is identical to the ML estimator; also,  $(n - K)s^2/\sigma^2$  is then distributed as  $\chi^2$  with  $n - K$  degrees of freedom and  $\mathbf{b}$  and  $s^2$  are independently distributed. For a proof of this result see, for example, Theil (1971, sec. 3.5).

If the covariance matrix of  $\boldsymbol{\varepsilon}$  is  $\sigma^2\mathbf{V}$  rather than  $\sigma^2\mathbf{I}$ , where  $\mathbf{V}$  is a non-singular matrix, we can extend the Gauss–Markov theorem to Aitken’s (1935) theorem. The best linear unbiased estimator of  $\boldsymbol{\beta}$  is now

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}\mathbf{y}, \quad (2.13)$$

and its covariance matrix is

$$\mathfrak{V}(\hat{\boldsymbol{\beta}}) = \sigma^2(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}. \quad (2.14)$$

The estimator  $\hat{\boldsymbol{\beta}}$  is the *generalized least-squares* (GLS) estimator of  $\boldsymbol{\beta}$ ; we shall see in Section 7 how it can be derived from the LS estimator  $\mathbf{b}$ .

#### 2.4. Vectors and matrices in consumption theory

It would be inappropriate to leave the impression that vectors and matrices are important in econometrics primarily because of problems of statistical inference. They are also important for the problem of how to specify economic relations. We shall illustrate this here for the analysis of consumer demand, which is one of the oldest topics in applied econometrics. References for the account which follows



include Barten (1977), Brown and Deaton (1972), Philips (1974), Theil (1975–76), and Deaton's chapter on demand analysis in this Handbook (Chapter 30).

Let there be  $N$  goods in the marketplace. We write  $\mathbf{p} = [p_i]$  and  $\mathbf{q} = [q_i]$  for the price and quantity vectors. The consumer's preferences are measured by a utility function  $u(\mathbf{q})$  which is assumed to be three times differentiable. His problem is to maximize  $u(\mathbf{q})$  by varying  $\mathbf{q}$  subject to the budget constraints  $\mathbf{p}'\mathbf{q} = M$ , where  $M$  is the given positive amount of total expenditure (to be called income for brevity's sake). Prices are also assumed to be positive and given from the consumer's point of view. Once he has solved this problem, the demand for each good becomes a function of income and prices. What can be said about the derivatives of demand,  $\partial q_i / \partial M$  and  $\partial q_i / \partial p_j$ ?

Neoclassical consumption theory answers this question by constructing the Lagrangian function  $u(\mathbf{q}) - \lambda(\mathbf{p}'\mathbf{q} - M)$  and differentiating this function with respect to the  $q_i$ 's. When these derivatives are equated to zero, we obtain the familiar proportionality of marginal utilities and prices:

$$\frac{\partial u}{\partial q_i} = \lambda p_i, \quad i = 1, \dots, N, \quad (2.15)$$

or, in vector notation,  $\partial u / \partial \mathbf{q} = \lambda \mathbf{p}$ : the gradient of the utility function at the optimal point is proportional to the price vector. The proportionality coefficient  $\lambda$  has the interpretation as the marginal utility of income.<sup>1</sup>

The proportionality (2.15) and the budget constraint  $\mathbf{p}'\mathbf{q} = M$  provide  $N + 1$  equations in  $N + 1$  unknowns:  $\mathbf{q}$  and  $\lambda$ . Since these equations hold identically in  $M$  and  $\mathbf{p}$ , we can differentiate them with respect to these variables. Differentiation of  $\mathbf{p}'\mathbf{q} = M$  with respect to  $M$  yields  $\sum_i p_i (\partial q_i / \partial M) = 1$  or

$$\mathbf{p}' \frac{\partial \mathbf{q}}{\partial M} = 1, \quad (2.16)$$

where  $\partial \mathbf{q} / \partial M = [\partial q_i / \partial M]$  is the vector of income derivatives of demand. Differentiation of  $\mathbf{p}'\mathbf{q} = M$  with respect to  $p_j$  yields  $\sum_i p_i (\partial q_i / \partial p_j) + q_j = 0$  ( $j = 1, \dots, N$ ) or

$$\mathbf{p}' \frac{\partial \mathbf{q}}{\partial \mathbf{p}'} = -\mathbf{q}', \quad (2.17)$$

where  $\partial \mathbf{q} / \partial \mathbf{p}' = [\partial q_i / \partial p_j]$  is the  $N \times N$  matrix of price derivatives of demand. Differentiation of (2.15) with respect to  $M$  and application of the chain rule

<sup>1</sup>Dividing both sides of (2.15) by  $p_i$  yields  $\partial u / \partial (p_i q_i) = \lambda$ , which shows that an extra dollar of income spent on any of the  $N$  goods raises utility by  $\lambda$ . This provides an intuitive justification for the interpretation. A more rigorous justification would require the introduction of the indirect utility function, which is beyond the scope of this chapter.

yields:

$$\sum_{k=1}^N \frac{\partial^2 u}{\partial q_i \partial q_k} \frac{\partial q_k}{\partial M} = p_i \frac{\partial \lambda}{\partial M}, \quad i = 1, \dots, N.$$

Similarly, differentiation of (2.15) with respect to  $p_j$  yields:

$$\sum_{k=1}^N \frac{\partial^2 u}{\partial q_i \partial q_k} \frac{\partial q_k}{\partial p_j} = p_i \frac{\partial \lambda}{\partial p_j} + \lambda \delta_{ij}, \quad i, j = 1, \dots, N,$$

where  $\delta_{ij}$  is the Kronecker delta ( $= 1$  if  $i = j$ ,  $0$  if  $i \neq j$ ). We can write the last two equations in matrix form as

$$U \frac{\partial \mathbf{q}}{\partial M} = \frac{\partial \lambda}{\partial M} \mathbf{p}, \quad U \frac{\partial \mathbf{q}}{\partial \mathbf{p}'} = \lambda \mathbf{I} + \mathbf{p} \frac{\partial \lambda}{\partial \mathbf{p}'}, \quad (2.18)$$

where  $U = \partial^2 u / \partial \mathbf{q} \partial \mathbf{q}'$  is the Hessian matrix of the consumer's utility function. We show at the end of Section 3 how the four equations displayed in (2.16)–(2.18) can be combined in partitioned matrix form and how they can be used to provide solutions for the income and price derivatives of demand under appropriate conditions.

### 3. Partitioned matrices

Partitioning a matrix into submatrices is one device for the exploitation of the mathematical structure of this matrix. This can be of considerable importance in multivariate situations.

#### 3.1. The algebra of partitioned matrices

We write the left-most matrix in (2.3) as  $\mathbf{Y} = [\mathbf{Y}_1 \quad \mathbf{Y}_2]$ , where

$$\mathbf{Y}_1 = \begin{bmatrix} y_{11} & y_{12} \\ y_{21} & y_{22} \\ \vdots & \vdots \\ y_{n1} & y_{n2} \end{bmatrix}, \quad \mathbf{Y}_2 = \begin{bmatrix} y_{13} & y_{14} \cdots y_{1L} \\ y_{23} & y_{24} \cdots y_{2L} \\ \vdots & \vdots \quad \vdots \\ y_{n3} & y_{n4} \cdots y_{nL} \end{bmatrix}.$$

The partitioning  $\mathbf{Y} = [\mathbf{Y}_1 \quad \mathbf{Y}_2]$  is by sets of columns, the observations on the first two endogenous variables being separated from those on the others. Partitioning

may take place by row sets and column sets. The addition rule for matrices can be applied in partitioned form,

$$\begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \\ \mathbf{A}_{31} & \mathbf{A}_{32} \end{bmatrix} + \begin{bmatrix} \mathbf{B}_{11} & \mathbf{B}_{12} \\ \mathbf{B}_{21} & \mathbf{B}_{22} \\ \mathbf{B}_{31} & \mathbf{B}_{32} \end{bmatrix} = \begin{bmatrix} \mathbf{A}_{11} + \mathbf{B}_{11} & \mathbf{A}_{12} + \mathbf{B}_{12} \\ \mathbf{A}_{21} + \mathbf{B}_{21} & \mathbf{A}_{22} + \mathbf{B}_{22} \\ \mathbf{A}_{31} + \mathbf{B}_{31} & \mathbf{A}_{32} + \mathbf{B}_{32} \end{bmatrix},$$

provided  $\mathbf{A}_{ij}$  and  $\mathbf{B}_{ij}$  have the same order for each  $(i, j)$ . A similar result holds for multiplication,

$$\begin{bmatrix} \mathbf{P}_{11} & \mathbf{P}_{12} \\ \mathbf{P}_{21} & \mathbf{P}_{22} \end{bmatrix} \begin{bmatrix} \mathbf{Q}_{11} & \mathbf{Q}_{12} \\ \mathbf{Q}_{21} & \mathbf{Q}_{22} \end{bmatrix} = \begin{bmatrix} \mathbf{P}_{11}\mathbf{Q}_{11} + \mathbf{P}_{12}\mathbf{Q}_{21} & \mathbf{P}_{11}\mathbf{Q}_{12} + \mathbf{P}_{12}\mathbf{Q}_{22} \\ \mathbf{P}_{21}\mathbf{Q}_{11} + \mathbf{P}_{22}\mathbf{Q}_{21} & \mathbf{P}_{21}\mathbf{Q}_{12} + \mathbf{P}_{22}\mathbf{Q}_{22} \end{bmatrix},$$

provided that the number of columns of  $\mathbf{P}_{11}$  and  $\mathbf{P}_{21}$  is equal to the number of rows of  $\mathbf{Q}_{11}$  and  $\mathbf{Q}_{12}$  (similarly for  $\mathbf{P}_{12}$ ,  $\mathbf{P}_{22}$ ,  $\mathbf{Q}_{21}$ ,  $\mathbf{Q}_{22}$ ).

The inverse of a symmetric partitioned matrix is frequently needed. Two alternative expressions are available:

$$\begin{bmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{B}' & \mathbf{C} \end{bmatrix}^{-1} = \begin{bmatrix} \mathbf{D} & -\mathbf{DBC}^{-1} \\ -\mathbf{C}^{-1}\mathbf{B}'\mathbf{D} & \mathbf{C}^{-1} + \mathbf{C}^{-1}\mathbf{B}'\mathbf{DBC}^{-1} \end{bmatrix}, \quad (3.1)$$

$$\begin{bmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{B}' & \mathbf{C} \end{bmatrix}^{-1} = \begin{bmatrix} \mathbf{A}^{-1} + \mathbf{A}^{-1}\mathbf{B}\mathbf{E}\mathbf{B}'\mathbf{A}^{-1} & -\mathbf{A}^{-1}\mathbf{B}\mathbf{E} \\ -\mathbf{E}\mathbf{B}'\mathbf{A}^{-1} & \mathbf{E} \end{bmatrix}, \quad (3.2)$$

where  $\mathbf{D} = (\mathbf{A} - \mathbf{BC}^{-1}\mathbf{B}')^{-1}$  and  $\mathbf{E} = (\mathbf{C} - \mathbf{B}'\mathbf{A}^{-1}\mathbf{B})^{-1}$ . The use of (3.1) requires that  $\mathbf{C}$  be non-singular; for (3.2) we must assume that  $\mathbf{A}$  is non-singular. The verification of these results is a matter of straightforward partitioned multiplication; for a constructive proof see Theil (1971, sec. 1.2).

The density function of the  $L$ -variate normal distribution with mean vector  $\boldsymbol{\mu}$  and non-singular covariance matrix  $\boldsymbol{\Sigma}$  is

$$f(\mathbf{x}) = \frac{1}{(2\pi)^{L/2} |\boldsymbol{\Sigma}|^{1/2}} \exp\left\{-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{x} - \boldsymbol{\mu})\right\}, \quad (3.3)$$

where  $|\boldsymbol{\Sigma}|$  is the determinant value of  $\boldsymbol{\Sigma}$ . Suppose that each of the first  $L'$  variates is uncorrelated with all  $L - L'$  other variates. Then  $\boldsymbol{\mu}$  and  $\boldsymbol{\Sigma}$  may be partitioned,

$$\boldsymbol{\mu} = \begin{bmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{bmatrix}, \quad \boldsymbol{\Sigma} = \begin{bmatrix} \boldsymbol{\Sigma}_1 & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Sigma}_2 \end{bmatrix}, \quad (3.4)$$

where  $(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1)$  contains the first- and second-order moments of the first  $L'$

variates and  $(\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_2)$  those of the last  $L - L'$ . The density function (3.3) can now be written as the product of

$$f_1(\mathbf{x}_1) = \frac{1}{(2\pi)^{L'/2} |\boldsymbol{\Sigma}_1|^{1/2}} \exp\left\{-\frac{1}{2}(\mathbf{x}_1 - \boldsymbol{\mu}_1)' \boldsymbol{\Sigma}_1^{-1} (\mathbf{x}_1 - \boldsymbol{\mu}_1)\right\}$$

and analogous function  $f_2(\mathbf{x}_2)$ . Clearly, the  $L$ -element normal vector consists of two subvectors which are independently distributed.

### 3.2. Block-recursive systems

We return to the equation system (2.3) and assume that the rows of  $\mathbf{E}$  are independent  $L$ -variate normal vectors with zero mean and covariance matrix  $\boldsymbol{\Sigma}$ , as shown in (2.4),  $\boldsymbol{\Sigma}_1$  being of order  $L' \times L'$ . We also assume that  $\boldsymbol{\Gamma}$  can be partitioned as

$$\boldsymbol{\Gamma} = \begin{bmatrix} \boldsymbol{\Gamma}_1 & \boldsymbol{\Gamma}_3 \\ \mathbf{0} & \boldsymbol{\Gamma}_2 \end{bmatrix}, \quad (3.5)$$

with  $\boldsymbol{\Gamma}_1$  of order  $L' \times L'$ . Then we can write (2.3) as

$$[\mathbf{Y}_1 \quad \mathbf{Y}_2] \begin{bmatrix} \boldsymbol{\Gamma}_1 & \boldsymbol{\Gamma}_3 \\ \mathbf{0} & \boldsymbol{\Gamma}_2 \end{bmatrix} + \mathbf{X}[\mathbf{B}_1 \quad \mathbf{B}_2] = [\mathbf{E}_1 \quad \mathbf{E}_2]$$

or

$$\mathbf{Y}_1 \boldsymbol{\Gamma}_1 + \mathbf{X} \mathbf{B}_1 = \mathbf{E}_1, \quad (3.6)$$

$$\mathbf{Y}_2 \boldsymbol{\Gamma}_2 + [\mathbf{X} \quad \mathbf{Y}_1] \begin{bmatrix} \mathbf{B}_2 \\ \boldsymbol{\Gamma}_3 \end{bmatrix} = \mathbf{E}_2, \quad (3.7)$$

where  $\mathbf{Y} = [\mathbf{Y}_1 \quad \mathbf{Y}_2]$ ,  $\mathbf{B} = [\mathbf{B}_1 \quad \mathbf{B}_2]$ , and  $\mathbf{E} = [\mathbf{E}_1 \quad \mathbf{E}_2]$  with  $\mathbf{Y}_1$  and  $\mathbf{E}_1$  of order  $n \times L'$  and  $\mathbf{B}_1$  of order  $K \times L'$ .

There is nothing special about (3.6), which is an equation system comparable to (2.3) but of smaller size. However, (3.7) is an equation system in which the  $L'$  variables whose observations are arranged in  $\mathbf{Y}_1$  can be viewed as exogenous rather than endogenous. This is indicated by combining  $\mathbf{Y}_1$  with  $\mathbf{X}$  in partitioned matrix form. There are two reasons why  $\mathbf{Y}_1$  can be viewed as exogenous in (3.7). First,  $\mathbf{Y}_1$  is obtained from the system (3.6) which does not involve  $\mathbf{Y}_2$ . Secondly, the random component  $\mathbf{E}_1$  in (3.6) is independent of  $\mathbf{E}_2$  in (3.7) because of the assumed normality with a block-diagonal  $\boldsymbol{\Sigma}$ . The case discussed here is that of a

block-recursive system, with a block-triangular  $\Gamma$  [see (3.5)] and a block-diagonal  $\Sigma$  [see (3.4)]. Under appropriate identification conditions, ML estimation of the unknown elements of  $\Gamma$  and  $B$  can be applied to the two subsystems (3.6) and (3.7) separately.

### 3.3. Income and price derivatives revisited

It is readily verified that eqs. (2.16)–(2.18) can be written in partitioned matrix form as

$$\begin{bmatrix} U & p \\ p' & 0 \end{bmatrix} \begin{bmatrix} \partial q / \partial M & \partial q / \partial p' \\ -\partial \lambda / \partial M & -\partial \lambda / \partial p' \end{bmatrix} = \begin{bmatrix} 0 & \lambda I \\ 1 & -q' \end{bmatrix}, \quad (3.8)$$

which is Barten's (1964) fundamental matrix equation in consumption theory. All three partitioned matrices in (3.8) are of order  $(N+1) \times (N+1)$ , and the left-most matrix is the Hessian matrix of utility function bordered by prices. If  $U$  is non-singular, we can use (3.2) for the inverse of this bordered matrix:

$$\begin{bmatrix} U & p \\ p' & 0 \end{bmatrix}^{-1} = \frac{1}{p'U^{-1}p} \begin{bmatrix} (p'U^{-1}p)U^{-1} - U^{-1}p(U^{-1}p)' & U^{-1}p \\ (U^{-1}p)' & -1 \end{bmatrix}.$$

Premultiplication of (3.8) by this inverse yields solutions for the income and price derivatives:

$$\frac{\partial q}{\partial M} = \frac{1}{p'U^{-1}p} U^{-1}p, \quad \frac{\partial \lambda}{\partial M} = \frac{1}{p'U^{-1}p} \quad (3.9)$$

$$\frac{\partial q}{\partial p'} = \lambda U^{-1} - \frac{\lambda}{p'U^{-1}p} U^{-1}p(U^{-1}p)' - \frac{1}{p'U^{-1}p} U^{-1}p q'. \quad (3.10)$$

It follows from (3.9) that we can write the income derivatives of demand as

$$\frac{\partial q}{\partial M} = \frac{\partial \lambda}{\partial M} U^{-1}p, \quad (3.11)$$

and from (3.9) and (3.10) that we can simplify the price derivatives to

$$\frac{\partial q}{\partial p'} = \lambda U^{-1} - \frac{\lambda}{\partial \lambda / \partial M} \frac{\partial q}{\partial M} \frac{\partial q'}{\partial M} - \frac{\partial q}{\partial M} q'. \quad (3.12)$$

The last matrix,  $-(\partial q / \partial M)q'$ , represents the *income effect* of the price changes

on demand. Note that this matrix has unit rank and is not symmetric. The two other matrices on the right in (3.12) are symmetric and jointly represent the substitution effect of the price changes. The first matrix,  $\lambda U^{-1}$ , gives the *specific substitution effect* and the second (which has unit rank) gives the *general substitution effect*. The latter effect describes the general competition of all goods for an extra dollar of income. The distinction between the two components of the substitution effect is from Houthakker (1960). We can combine these components by writing (3.12) in the form

$$\frac{\partial \mathbf{q}}{\partial \mathbf{p}'} = \lambda U^{-1} \left( \mathbf{I} - \mathbf{p} \frac{\partial \mathbf{q}'}{\partial M} \right) - \frac{\partial \mathbf{q}}{\partial M} \mathbf{q}', \quad (3.13)$$

which is obtained by using (3.11) for the first  $\partial \mathbf{q} / \partial M$  that occurs in (3.12).

#### 4. Kronecker products and the vectorization of matrices

A special form of partitioning is that in which all submatrices are scalar multiples of the same matrix  $\mathbf{B}$  of order  $p \times q$ . We write this as

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

and refer to  $\mathbf{A} \otimes \mathbf{B}$  as the Kronecker product of  $\mathbf{A} = [a_{ij}]$  and  $\mathbf{B}$ . The order of this product is  $mp \times nq$ . Kronecker products are particularly convenient when several equations are analyzed simultaneously.

##### 4.1. The algebra of Kronecker products

It is a matter of straightforward partitioned multiplication to verify that

$$(\mathbf{A} \otimes \mathbf{B})(\mathbf{C} \otimes \mathbf{D}) = \mathbf{AC} \otimes \mathbf{BD}, \quad (4.1)$$

provided  $\mathbf{AC}$  and  $\mathbf{BD}$  exist. Also, if  $\mathbf{A}$  and  $\mathbf{B}$  are square and non-singular, then

$$(\mathbf{A} \otimes \mathbf{B})^{-1} = \mathbf{A}^{-1} \otimes \mathbf{B}^{-1} \quad (4.2)$$

because (4.1) implies  $(\mathbf{A} \otimes \mathbf{B})(\mathbf{A}^{-1} \otimes \mathbf{B}^{-1}) = \mathbf{AA}^{-1} \otimes \mathbf{BB}^{-1} = \mathbf{I} \otimes \mathbf{I} = \mathbf{I}$ , where the three unit matrices will in general be of different order. We can obviously extend

(4.1) to

$$(A_1 \otimes B_1)(A_2 \otimes B_2)(A_3 \otimes B_3) = A_1 A_2 A_3 \otimes B_1 B_2 B_3$$

provided  $A_1 A_2 A_3$  and  $B_1 B_2 B_3$  exist.

Other useful properties of Kronecker products are:

$$(A \otimes B)' = A' \otimes B', \quad (4.3)$$

$$A \otimes (B + C) = A \otimes B + A \otimes C, \quad (4.4)$$

$$(B + C) \otimes A = B \otimes A + C \otimes A, \quad (4.5)$$

$$A \otimes (B \otimes C) = (A \otimes B) \otimes C. \quad (4.6)$$

Note the implication of (4.3) that  $A \otimes B$  is symmetric when  $A$  and  $B$  are symmetric. Other properties of Kronecker products are considered in Section 7.

#### 4.2. Joint generalized least-squares estimation of several equations

In (2.1) and (2.3) we considered a system of  $L$  linear equations in  $L$  endogenous variables. Here we consider the special case in which each equation describes one endogenous variable in terms of exogenous variables only. If the observations on all variables are  $\alpha = 1, \dots, n$ , we can write the  $L$  equations in a form similar to (2.7):

$$y_j = X_j \beta_j + \varepsilon_j, \quad j = 1, \dots, L, \quad (4.7)$$

where  $y_j = [y_{\alpha j}]$  is the observation vector on the  $j$ th endogenous variable,  $\varepsilon_j = [\varepsilon_{\alpha j}]$  is the associated disturbance vector with zero expectation,  $X_j$  is the observation matrix on the  $K_j$  exogenous variables in the  $j$ th equation, and  $\beta_j$  is the  $K_j$ -element parameter vector.

We can write (4.7) for all  $j$  in partitioned matrix form:

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_L \end{bmatrix} = \begin{bmatrix} X_1 & \mathbf{0} \dots \mathbf{0} \\ \mathbf{0} & X_2 \dots \mathbf{0} \\ \vdots & \vdots \quad \vdots \\ \mathbf{0} & \mathbf{0} \dots X_L \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_L \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_L \end{bmatrix} \quad (4.8)$$

or, more briefly, as

$$y = Z\beta + \varepsilon, \quad (4.9)$$

where  $y$  and  $\boldsymbol{\varepsilon}$  are  $Ln$ -element vectors and  $\mathbf{Z}$  contains  $Ln$  rows, while the number of columns of  $\mathbf{Z}$  and that of the elements of  $\boldsymbol{\beta}$  are both  $K_1 + \dots + K_L$ . The covariance matrix of  $\boldsymbol{\varepsilon}$  is thus of order  $Ln \times Ln$  and can be partitioned into  $L^2$  submatrices of the form  $\mathfrak{C}(\boldsymbol{\varepsilon}_j, \boldsymbol{\varepsilon}'_l)$ . For  $j = l$  this submatrix equals the covariance matrix  $\mathfrak{V}(\boldsymbol{\varepsilon}_j)$ . We assume that the  $n$  disturbances of each of the  $L$  equations have equal variance and are uncorrelated so that  $\mathfrak{V}(\boldsymbol{\varepsilon}_j) = \sigma_{jj}\mathbf{I}$ , where  $\sigma_{jj} = \text{var } \varepsilon_{\alpha j}$  (each  $\alpha$ ). For  $j \neq l$  the submatrix  $\mathfrak{C}(\boldsymbol{\varepsilon}_j, \boldsymbol{\varepsilon}'_l)$  contains the "contemporaneous" covariances  $\mathfrak{C}(\varepsilon_{\alpha j}, \varepsilon_{\alpha l})$  for  $\alpha = 1, \dots, n$  in the diagonal. We assume that these covariances are all equal to  $\sigma_{jl}$  and that all non-contemporaneous covariances vanish:  $\mathfrak{C}(\varepsilon_{\alpha j}, \varepsilon_{\alpha l}) = 0$  for  $\alpha \neq \eta$ . Therefore,  $\mathfrak{C}(\boldsymbol{\varepsilon}_j, \boldsymbol{\varepsilon}'_l) = \sigma_{jl}\mathbf{I}$ , which contains  $\mathfrak{V}(\boldsymbol{\varepsilon}_j) = \sigma_{jj}\mathbf{I}$  as a special case. The full covariance matrix of the  $Ln$ -element vector  $\boldsymbol{\varepsilon}$  is thus:

$$\mathfrak{V}(\boldsymbol{\varepsilon}) = \begin{bmatrix} \sigma_{11}\mathbf{I} & \sigma_{12}\mathbf{I} \dots \sigma_{1L}\mathbf{I} \\ \sigma_{21}\mathbf{I} & \sigma_{22}\mathbf{I} \dots \sigma_{2L}\mathbf{I} \\ \vdots & \vdots \quad \vdots \\ \sigma_{L1}\mathbf{I} & \sigma_{L2}\mathbf{I} \dots \sigma_{LL}\mathbf{I} \end{bmatrix} = \boldsymbol{\Sigma} \otimes \mathbf{I}, \quad (4.10)$$

where  $\boldsymbol{\Sigma} = [\sigma_{jl}]$  is the contemporaneous covariance matrix, i.e. the covariance matrix of  $[\varepsilon_{\alpha 1}, \dots, \varepsilon_{\alpha L}]$  for  $\alpha = 1, \dots, n$ .

Suppose that  $\boldsymbol{\Sigma}$  is non-singular so that  $\boldsymbol{\Sigma}^{-1} \otimes \mathbf{I}$  is the inverse of the matrix (4.10) in view of (4.2). Also, suppose that  $X_1, \dots, X_L$  and hence  $\mathbf{Z}$  have full column rank. Application of the GLS results (2.13) and (2.14) to (4.9) and (4.10) then yields

$$\hat{\boldsymbol{\beta}} = [\mathbf{Z}'(\boldsymbol{\Sigma}^{-1} \otimes \mathbf{I})\mathbf{Z}]^{-1} \mathbf{Z}'(\boldsymbol{\Sigma}^{-1} \otimes \mathbf{I})y \quad (4.11)$$

as the best linear unbiased estimator of  $\boldsymbol{\beta}$  with the following covariance matrix:

$$\mathfrak{V}(\hat{\boldsymbol{\beta}}) = [\mathbf{Z}'(\boldsymbol{\Sigma}^{-1} \otimes \mathbf{I})\mathbf{Z}]^{-1}. \quad (4.12)$$

In general,  $\hat{\boldsymbol{\beta}}$  is superior to LS applied to each of the  $L$  equations separately, but there are two special cases in which these estimation procedures are identical.

The first case is that in which  $X_1, \dots, X_L$  are all identical. We can then write  $X$  for each of these matrices so that the observation matrix on the exogenous variables in (4.8) and (4.9) takes the form

$$\mathbf{Z} = \begin{bmatrix} X & \mathbf{0} \dots \mathbf{0} \\ \mathbf{0} & X \dots \mathbf{0} \\ \vdots & \vdots \quad \vdots \\ \mathbf{0} & \mathbf{0} \dots X \end{bmatrix} = \mathbf{I} \otimes X. \quad (4.13)$$



This implies

$$\mathbf{Z}'(\boldsymbol{\Sigma}^{-1} \otimes \mathbf{I})\mathbf{Z} = (\mathbf{I} \otimes \mathbf{X}')(\boldsymbol{\Sigma}^{-1} \otimes \mathbf{I})(\mathbf{I} \otimes \mathbf{X}) = \boldsymbol{\Sigma}^{-1} \otimes \mathbf{X}'\mathbf{X}$$

and

$$\begin{aligned} [\mathbf{Z}'(\boldsymbol{\Sigma}^{-1} \otimes \mathbf{I})\mathbf{Z}]^{-1}\mathbf{Z}'(\boldsymbol{\Sigma}^{-1} \otimes \mathbf{I}) &= [\boldsymbol{\Sigma} \otimes (\mathbf{X}'\mathbf{X})^{-1}](\mathbf{I} \otimes \mathbf{X}')(\boldsymbol{\Sigma}^{-1} \otimes \mathbf{I}) \\ &= \mathbf{I} \otimes (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'. \end{aligned}$$

It is now readily verified from (4.11) that  $\hat{\boldsymbol{\beta}}$  consists of  $L$  subvectors of the LS form  $(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'y_j$ . The situation of identical matrices  $\mathbf{X}_1, \dots, \mathbf{X}_L$  occurs relatively frequently in applied econometrics; an example is the reduced form (2.4) for each of the  $L$  endogenous variables.

The second case in which (4.11) degenerates into subvectors equal to LS vectors is that of uncorrelated contemporaneous disturbances. Then  $\boldsymbol{\Sigma}$  is diagonal and it is easily verified that  $\hat{\boldsymbol{\beta}}$  consists of subvectors of the form  $(\mathbf{X}'_j\mathbf{X}_j)^{-1}\mathbf{X}'_jy_j$ . See Theil (1971, pp. 311–312) for the case in which  $\boldsymbol{\Sigma}$  is block-diagonal.

Note that the computation of the joint GLS estimator (4.11) requires  $\boldsymbol{\Sigma}$  to be known. This is usually not true and the unknown  $\boldsymbol{\Sigma}$  is then replaced by the sample moment matrix of the LS residuals [see Zellner (1962)]. This approximation is asymptotically (for large  $n$ ) acceptable under certain conditions; we shall come back to this matter in the opening paragraph of Section 9.

### 4.3. Vectorization of matrices

In eq. (2.3) we wrote  $Ln$  equations in matrix form with parameter matrices  $\boldsymbol{\Gamma}$  and  $\mathbf{B}$ , each consisting of several columns, whereas in (4.8) and (4.9) we wrote  $Ln$  equations in matrix form with a “long” parameter vector  $\boldsymbol{\beta}$ . If  $\mathbf{Z}$  takes the form (4.13), we can write (4.8) in the equivalent form  $\mathbf{Y} = \mathbf{X}\mathbf{B} + \mathbf{E}$ , where  $\mathbf{Y}$ ,  $\mathbf{B}$ , and  $\mathbf{E}$  are matrices consisting of  $L$  columns of the form  $y_j$ ,  $\boldsymbol{\beta}_j$ , and  $\boldsymbol{\varepsilon}_j$ . Thus, the elements of the parameter vector  $\boldsymbol{\beta}$  are then rearranged into the matrix  $\mathbf{B}$ . On the other hand, there are situations in which it is more attractive to work with vectors rather than matrices that consist of several columns. For example, if  $\hat{\boldsymbol{\beta}}$  is an unbiased estimator of the parameter vector  $\boldsymbol{\beta}$  with finite second moments, we obtain the covariance matrix of  $\hat{\boldsymbol{\beta}}$  by postmultiplying  $\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}$  by its transpose and taking the expectation, but this procedure does not work when the parameters are arranged in a matrix  $\mathbf{B}$  which consists of several columns. It is then appropriate to rearrange the parameters in vector form. This is a matter of designing an appropriate notation and evaluating the associated algebra.

Let  $\mathbf{A} = [\mathbf{a}_1 \dots \mathbf{a}_q]$  be a  $p \times q$  matrix,  $\mathbf{a}_i$  being the  $i$ th column of  $\mathbf{A}$ . We define  $\text{vec } \mathbf{A} = [\mathbf{a}'_1 \ \mathbf{a}'_2 \dots \mathbf{a}'_q]'$ , which is a  $pq$ -element column vector consisting of  $q$

subvectors, the first containing the  $p$  elements of  $\mathbf{a}_1$ , the second the  $p$  elements of  $\mathbf{a}_2$ , and so on. It is readily verified that  $\text{vec}(\mathbf{A} + \mathbf{B}) = \text{vec}\mathbf{A} + \text{vec}\mathbf{B}$ , provided that  $\mathbf{A}$  and  $\mathbf{B}$  are of the same order. Also, if the matrix products  $\mathbf{AB}$  and  $\mathbf{BC}$  exist,

$$\begin{aligned}\text{vec}\mathbf{AB} &= (\mathbf{I} \otimes \mathbf{A})\text{vec}\mathbf{B} = (\mathbf{B}' \otimes \mathbf{I})\text{vec}\mathbf{A}, \\ \text{vec}\mathbf{ABC} &= (\mathbf{I} \otimes \mathbf{AB})\text{vec}\mathbf{C} = (\mathbf{C}' \otimes \mathbf{A})\text{vec}\mathbf{B} = (\mathbf{C}'\mathbf{B}' \otimes \mathbf{I})\text{vec}\mathbf{A}.\end{aligned}$$

For proofs and extensions of these results see Dhrymes (1978, ch. 4).

## 5. Differential demand and supply systems

The differential approach to microeconomic theory provides interesting comparisons with equation systems such as (2.3) and (4.9). Let  $\mathbf{g}(\mathbf{z})$  be a vector of functions of a vector  $\mathbf{z}$ ; the approach uses the total differential of  $\mathbf{g}(\cdot)$ ,

$$d\mathbf{g} = \frac{\partial \mathbf{g}}{\partial \mathbf{z}'} d\mathbf{z}, \quad (5.1)$$

and it exploits what is known about  $\partial \mathbf{g} / \partial \mathbf{z}'$ . For example, the total differential of consumer demand is  $d\mathbf{q} = (\partial \mathbf{q} / \partial M) dM + (\partial \mathbf{q} / \partial \mathbf{p}') d\mathbf{p}$ . Substitution from (3.13) yields:

$$d\mathbf{q} = \frac{\partial \mathbf{q}}{\partial M} (dM - \mathbf{q}' d\mathbf{p}) + \lambda U^{-1} \left[ d\mathbf{p} - \left( \frac{\partial \mathbf{q}'}{\partial M} d\mathbf{p} \right) \mathbf{p} \right], \quad (5.2)$$

which shows that the income effect of the price changes is used to deflate the change in money income and, similarly, the general substitution effect to deflate the specific effect. Our first objective is to write the system (5.2) in a more attractive form.

### 5.1. A differential consumer demand system

We introduce the budget share  $w_i$  and the marginal share  $\theta_i$  of good  $i$ :

$$w_i = \frac{p_i q_i}{M}, \quad \theta_i = \frac{\partial (p_i q_i)}{\partial M}, \quad (5.3)$$

and also the Divisia (1925) volume index  $d(\log Q)$  and the Frisch (1932) price index  $d(\log P')$ :

$$d(\log Q) = \sum_{i=1}^N w_i d(\log q_i), \quad d(\log P') = \sum_{i=1}^N \theta_i d(\log p_i), \quad (5.4)$$

where  $\log$  (here and elsewhere) stands for natural logarithm. We prove in the next paragraph that (5.2) can be written in scalar form as

$$w_i d(\log q_i) = \theta_i d(\log Q) + \phi \sum_{j=1}^N \theta_{ij} d\left(\log \frac{p_j}{P'}\right), \quad (5.5)$$

where  $d[\log(p_j/P')]$  is an abbreviation of  $d(\log p_j) - d(\log P')$ , while  $\phi$  is the reciprocal of the income elasticity of the marginal utility of income:

$$\phi = \left( \frac{\partial \log \lambda}{\partial \log M} \right)^{-1}, \quad (5.6)$$

and  $\theta_{ij}$  is an element of the symmetric  $N \times N$  matrix

$$\Theta = \frac{\lambda}{\phi M} \mathbf{P} \mathbf{U}^{-1} \mathbf{P}, \quad (5.7)$$

with  $\mathbf{P}$  defined as the diagonal matrix with the prices  $p_1, \dots, p_N$  on the diagonal.

To verify (5.5) we apply (5.1) to  $M = \mathbf{p}'\mathbf{q}$ , yielding  $dM = \mathbf{q}'d\mathbf{p} + \mathbf{p}'d\mathbf{q}$  so that  $dM - \mathbf{q}'d\mathbf{p} = M d(\log Q)$  follows from (5.3) and (5.4). Therefore, premultiplication of (5.2) by  $(1/M)\mathbf{P}$  gives:

$$\frac{1}{M} \mathbf{P} d\mathbf{q} = \mathbf{P} \frac{\partial \mathbf{q}}{\partial M} d(\log Q) + \frac{\lambda}{M} \mathbf{P} \mathbf{U}^{-1} \mathbf{P} \left[ \mathbf{P}^{-1} d\mathbf{p} - \left( \frac{\partial \mathbf{q}'}{\partial M} d\mathbf{p} \right) \boldsymbol{\iota} \right], \quad (5.8)$$

where  $\boldsymbol{\iota} = \mathbf{P}^{-1}\mathbf{p}$  is a vector of  $N$  unit elements. The  $i$ th element of  $(1/M)\mathbf{P}d\mathbf{q}$  equals  $(p_i/M)dq_i = w_i d(\log q_i)$ , which confirms the left side of (5.5). The vector  $\mathbf{P}(\partial \mathbf{q}/\partial M)$  equals the marginal share vector  $\boldsymbol{\theta} = [\theta_i]$ , thus confirming the real-income term of (5.5). The  $j$ th element of the vector in brackets in (5.8) equals  $d(\log p_j) - d(\log P')$ , which agrees with the substitution term of (5.5). The verification of (5.5) is completed by  $(\lambda/M)\mathbf{P}\mathbf{U}^{-1}\mathbf{P} = \phi\Theta$  [see (5.7)]. Note that  $\Theta\boldsymbol{\iota} = (\lambda/\phi M)\mathbf{P}\mathbf{U}^{-1}\mathbf{p} = \mathbf{P}(\partial \mathbf{q}/\partial M)$  [see (3.11) and (5.6)]. Therefore,

$$\Theta\boldsymbol{\iota} = \boldsymbol{\theta}, \quad \boldsymbol{\iota}'\Theta\boldsymbol{\iota} = \boldsymbol{\iota}'\boldsymbol{\theta} = 1, \quad (5.9)$$

where  $\boldsymbol{\iota}'\boldsymbol{\theta} = \sum_i \theta_i = 1$  follows from (2.16). We conclude from  $\Theta\boldsymbol{\iota} = \boldsymbol{\theta}$  that the  $\theta_{ij}$ 's of the  $i$ th equation sum to the  $i$ th marginal share, and from  $\boldsymbol{\iota}'\Theta\boldsymbol{\iota} = 1$  that the  $\theta_{ij}$ 's of the entire system sum to 1. The latter property is expressed by referring to the  $\theta_{ij}$ 's as the *normalized price coefficients*.

### 5.2. A comparison with simultaneous equation systems

The  $N$ -equation system (5.5) describes the change in the demand for each good, measured by its contribution to the Divisia index [see (5.4)],<sup>2</sup> as the sum of a real-income component and a substitution component. This system may be compared with the  $L$ -equation system (2.1). There is a difference in that the latter system contains in principle more than one endogenous variable in each equation, whereas (5.5) has only one such variable if we assume the  $d(\log Q)$  and all price changes are exogenous.<sup>3</sup> Yet, the differential demand system is truly a system because of the cross-equation constraints implied by the symmetry of the normalized price coefficient matrix  $\Theta$ .

A more important difference results from the utility-maximizing theory behind (5.5), which implies that the coefficients are more directly interpretable than the  $\gamma$ 's and  $\beta$ 's of (2.1). Writing  $[\theta^{ij}] = \Theta^{-1}$  and inverting (5.7), we obtain:

$$\theta^{ij} = \frac{\phi M}{\lambda} \frac{\partial^2 u}{\partial(p_i q_i) \partial(p_j q_j)}, \quad (5.10)$$

which shows that  $\theta^{ij}$  measures (apart from  $\phi M/\lambda$  which does not involve  $i$  and  $j$ ) the change in the marginal utility of a dollar spent on  $i$  caused by an extra dollar spent on  $j$ . Equivalently, the normalized price coefficient matrix  $\Theta$  is inversely proportional to the Hessian matrix of the utility function in expenditure terms.

The relation (5.7) between  $\Theta$  and  $U$  allows us to analyze special preference structures. Suppose that the consumer's tastes can be represented by a utility function which is the sum of  $N$  functions, one for each good. Then the marginal utility of each good is independent of the consumption of all other goods, which we express by referring to this case as *preference independence*. The Hessian  $U$  is then diagonal and so is  $\Theta$  [see (5.7)], while  $\Theta u = \theta$  in (5.9) is simplified to  $\theta_{ii} = \theta_i$ . Thus, we can write (5.5) under preference independence as

$$w_i d(\log q_i) = \theta_i d(\log Q) + \phi \theta_i d\left(\log \frac{p_i}{P'}\right), \quad (5.11)$$

which contains only one Frisch-deflated price. The system (5.11) for  $i = 1, \dots, N$  contains only  $N$  unconstrained coefficients, namely  $\phi$  and  $N - 1$  unconstrained marginal shares.

The application of differential demand systems to data requires a parameterization which postulates that certain coefficients are constant. Several solutions have

<sup>2</sup>Note that this way of measuring the change in demand permits the exploitation of the symmetry of  $\Theta$ . When we have  $d(\log q_i)$  on the left, the coefficient of the Frisch-deflated price becomes  $\theta_{ij}/w_i$ , which is an element of an asymmetric matrix.

<sup>3</sup>This assumption may be relaxed; see Theil (1975-76, ch. 9-10) for an analysis of endogenous price changes.

been proposed, but these are beyond the scope of this chapter; see the references quoted in Section 2.4 above and also, for a further comparison with models of the type (2.1), Theil and Clements (1980).

### 5.3. An extension to the inputs of a firm: A singularity problem

Let the  $p_i$ 's and  $q_i$ 's be the prices and quantities of  $N$  inputs which a firm buys to make a product, the output of which is  $z$ . Let  $z = g(\mathbf{q})$  be the firm's production function,  $g(\cdot)$  being three times differentiable. Let the firm's objective be to minimize input expenditure  $\mathbf{p}'\mathbf{q}$  subject to  $z = g(\mathbf{q})$  for given output  $z$  and input prices  $\mathbf{p}$ . Our objective will be to analyze whether this minimum problem yields a differential input demand system similar to (5.5).

As in the consumer's case we construct a Lagrangian function, which now takes the form  $\mathbf{p}'\mathbf{q} - \rho[g(\mathbf{q}) - z]$ . By equating the derivative of this function with respect to  $\mathbf{q}$  to zero we obtain a proportionality of  $\partial g / \partial \mathbf{q}$  to  $\mathbf{p}$  [compare (2.15)]. This proportionality and the production function provide  $N + 1$  equations in  $N + 1$  unknowns:  $\mathbf{q}$  and  $\rho$ . Next we differentiate these equations with respect to  $z$  and  $\mathbf{p}$ , and we collect the derivatives in partitioned matrix form. The result is similar to the matrix equation (3.8) of consumption theory, and the Hessian  $U$  now becomes the Hessian  $\partial^2 g / \partial \mathbf{q} \partial \mathbf{q}'$  of the production function. We can then proceed as in (3.9) and following text if  $\partial^2 g / \partial \mathbf{q} \partial \mathbf{q}'$  is non-singular, but this is unfortunately not true when the firm operates under constant returns to scale. It is clearly unattractive to make an assumption which excludes this important case. In the account which follows<sup>4</sup> we solve this problem by formulating the production function in logarithmic form.

$$\log z = h(\mathbf{q}), \quad (5.12)$$

and using the following  $N \times N$  Hessian matrix:

$$\mathbf{H} = \left[ \frac{\partial^2 h}{\partial(\log q_i) \partial(\log q_j)} \right]. \quad (5.13)$$

### 5.4. A differential input demand system

The minimum of  $\mathbf{p}'\mathbf{q}$  subject to (5.12) for given  $z$  and  $\mathbf{p}$  will be a function of  $z$  and  $\mathbf{p}$ . We write  $C(z, \mathbf{p})$  for this minimum: the cost of producing output  $z$  at the input

<sup>4</sup>Derivations are omitted; the procedure is identical to that which is outlined above except that it systematically uses logarithms of output, inputs, and input prices. See Laitinen (1980), Laitinen and Theil (1978), and Theil (1977, 1980).

prices  $p$ . We define

$$\gamma = \frac{\partial \log C}{\partial \log z}, \quad \frac{1}{\psi} = 1 + \frac{1}{\gamma^2} \frac{\partial^2 \log C}{\partial (\log z)^2}, \quad (5.14)$$

so that  $\gamma$  is the output elasticity of cost and  $\psi < 1$  ( $> 1$ ) when this elasticity increases (decreases) with increasing output; thus,  $\psi$  is a curvature measure of the logarithmic cost function. It can be shown that the input demand equations may be written as

$$f_i d(\log q_i) = \gamma \theta_i d(\log z) - \psi \sum_{j=1}^N \theta_{ij} d\left(\log \frac{p_j}{P'}\right), \quad (5.15)$$

which should be compared with (5.5). In (5.15),  $f_i$  is the factor share of input  $i$  (its share in total cost) and  $\theta_i$  is its marginal share (the share in marginal cost),

$$f_i = \frac{p_i q_i}{C}, \quad \theta_i = \frac{\partial (p_i q_i) / \partial z}{\partial C / \partial z}, \quad (5.16)$$

which is the input version of (5.3). The Frisch price index on the far right in (5.15) is as shown in (5.4) but with  $\theta_i$  defined in (5.16). The coefficient  $\theta_{ij}$  in (5.15) is the  $(i, j)$ th element of the symmetric matrix

$$\Theta = \frac{1}{\psi} F(F - \gamma H)^{-1} F, \quad (5.17)$$

where  $H$  is given in (5.13) and  $F$  is the diagonal matrix with the factor shares  $f_1, \dots, f_N$  on the diagonal. This  $\Theta$  satisfies (5.9) with  $\theta = [\theta_i]$  defined in (5.16).

A firm is called *input independent* when the elasticity of its output with respect to each input is independent of all other inputs. It follows from (5.12) and (5.13) that  $H$  is then diagonal; hence,  $\Theta$  is also diagonal [see (5.17)] and  $\Theta \iota = \theta$  becomes  $\theta_{ii} = \theta_i$  so that we can simplify (5.15) to

$$f_i d(\log q_i) = \gamma \theta_i d(\log z) - \psi \theta_i d\left(\log \frac{p_i}{P'}\right), \quad (5.18)$$

which is to be compared with the consumer's equation (5.11) under preference independence. The Cobb–Douglas technology is a special case of input independence with  $H = \theta$ , implying that  $F(F - \gamma H)^{-1} F$  in (5.17) equals the diagonal matrix  $F$ . Since Cobb–Douglas may have constant returns to scale, this illustrates that the logarithmic formulation successfully avoids the singularity problem mentioned in the previous subsection.

## 5.5. Allocation systems

Summation of (5.5) over  $i$  yields the identity  $d(\log Q) = d(\log Q)$ , which means that (5.5) is an *allocation system* in the sense that it describes how the change in total expenditure is allocated to the  $N$  goods, given the changes in real income and relative prices. To verify this identity, we write (5.5) for  $i = 1, \dots, N$  in matrix form as

$$W\kappa = (\iota'W\kappa)\theta + \phi\Theta(I - \iota\theta')\pi, \quad (5.19)$$

where  $W$  is the diagonal matrix with  $w_1, \dots, w_N$  on the diagonal and  $\pi = [d(\log p_i)]$  and  $\kappa = [d(\log q_i)]$  are the vectors logarithmic price and quantity changes so that  $d(\log Q) = \iota'W\kappa$ ,  $d(\log P) = \theta'\pi$ . The proof is completed by premultiplying (5.19) by  $\iota'$ , which yields  $\iota'W\kappa = \iota'W\kappa$  in view of (5.9). Note that the substitution terms of the  $N$  demand equations have zero sum.

The input demand system (5.15) is not an allocation system because the firm does not take total input expenditure as given; rather, it minimizes this expenditure for given output  $z$  and given input prices  $p$ . Summation of (5.15) over  $i$  yields:

$$d(\log Q) = \gamma d(\log z), \quad (5.20)$$

where  $d(\log Q) = \sum_i f_i d(\log q_i) = \iota'F\kappa$  is the Divisia input volume index. Substitution of (5.20) into (5.15) yields:

$$f_i d(\log q_i) = \theta_i d(\log Q) - \psi \sum_{j=1}^N \theta_{ij} d\left(\log \frac{p_j}{P}\right). \quad (5.21)$$

We can interpret (5.20) as specifying the aggregate input change which is required to produce the given change in output, and (5.21) as an allocation system for the individual inputs given the aggregate input change and the changes in the relative input prices. It follows from (5.9) that we can write (5.19) and (5.21) for each  $i$  as

$$W\kappa = (\iota'W\kappa)\Theta\iota + \phi\Theta(I - \iota\theta')\pi, \quad (5.22)$$

$$F\kappa = (\iota'F\kappa)\Theta\iota - \psi\Theta(I - \iota\theta')\pi, \quad (5.23)$$

which shows that the normalized price coefficient matrix  $\Theta$  and the scalars  $\phi$  and  $\psi$  are the only coefficients in the two allocation systems.

## 5.6. Extensions

Let the firm adjust output  $z$  by maximizing its profit under competitive conditions, the price  $y$  of the product being exogenous from the firm's point of view.

Then marginal cost  $\partial C/\partial z$  equals  $y$ , while  $\theta_i$  of (5.16) equals  $\partial(p_i q_i)/\partial(yz)$ : the additional expenditure on input  $i$  resulting from an extra dollar of output revenue. Note that this is much closer to the consumer's  $\theta_i$  definition (5.3) than is (5.16).

If the firm sells  $m$  products with outputs  $z_1, \dots, z_m$  at exogenous prices  $y_1, \dots, y_m$ , total revenue equals  $R = \sum_r y_r z_r$  and  $g_r = y_r z_r / R$  is the revenue share of product  $r$ , while

$$d(\log Z) = \sum_{r=1}^m g_r d(\log z_r) \quad (5.24)$$

is the Divisia output volume index of the multiproduct firm. There are now  $m$  marginal costs,  $\partial C/\partial z_r$  for  $r=1, \dots, m$ , and each input has  $m$  marginal shares:  $\theta_i^r$  defined as  $\partial(p_i q_i)/\partial z_r$  divided by  $\partial C/\partial z_r$ , which becomes  $\theta_i^r = \partial(p_i q_i)/\partial(y_r z_r)$  under profit maximization. Multiproduct input demand equations can be formulated so that the substitution term in (5.15) is unchanged, but the output term becomes

$$\gamma \sum_{r=1}^m \theta_i^r g_r d(\log z_r), \quad (5.25)$$

which shows that input  $i$  can be of either more or less importance for product  $r$  than for product  $s$  depending on the values of  $\theta_i^r$  and  $\theta_i^s$ .

Maximizing profit by adjusting outputs yields an output supply system which will now be briefly described. The  $r$ th supply equation is

$$g_r d(\log z_r) = \psi^* \sum_{s=1}^m \theta_{rs}^* d\left(\log \frac{y_s}{P^{rs}}\right), \quad (5.26)$$

which describes the change<sup>5</sup> in the supply of product  $r$  in terms of all output price changes, each deflated by the corresponding Frisch input price index:

$$d(\log P^{rs}) = \sum_{i=1}^N \theta_i^r d(\log p_i). \quad (5.27)$$

Asterisks are added to the coefficients of (5.26) in order to distinguish output supply from input demand. The coefficient  $\psi^*$  is positive, while  $\theta_{rs}^*$  is a normalized price coefficient defined as

$$\theta_{rs}^* = \frac{1}{\psi^* R} y_r c^{rs} y_s, \quad (5.28)$$

<sup>5</sup>This change is measured by the contribution of product  $r$  to the Divisia output volume index (5.24). Note that this is similar to the left variables in (5.5) and (5.15).



where  $c^{rs}$  is an element of the inverse of the symmetric  $m \times m$  matrix  $[\partial^2 C / \partial z_r \partial z_s]$ . The similarity between (5.28) and (5.7) should be noted; we shall consider this matter further in Section 6. A multiproduct firm is called *output independent* when its cost function is the sum of  $m$  functions, one for each product.<sup>6</sup> Then  $[\partial^2 C / \partial z_r \partial z_s]$  and  $[\theta_{rs}^*]$  are diagonal [see (5.28)] so that the change in the supply of each product depends only on the change in its own deflated price [see (5.26)]. Note the similarity to preference and input independence [see (5.11) and (5.18)].

## 6. Definite and semidefinite square matrices

The expression  $\mathbf{x}'\mathbf{A}\mathbf{x}$  is a quadratic form in the vector  $\mathbf{x}$ . We met several examples in earlier sections: the second-order term in the Taylor expansion (2.6),  $\boldsymbol{\varepsilon}'\mathbf{M}\boldsymbol{\varepsilon}$  in the residual sum of squares (2.12), the expression in the exponent in the normal density function (3.3), the denominator  $\mathbf{p}'\mathbf{U}^{-1}\mathbf{p}$  in (3.9), and  $\mathbf{t}'\boldsymbol{\Theta}\mathbf{t}$  in (5.9). A more systematic analysis of quadratic forms is in order.

### 6.1. Covariance matrices and Gauss–Markov further considered

Let  $\mathbf{r}$  be a random vector with expectation  $\mathcal{E}\mathbf{r}$  and covariance matrix  $\boldsymbol{\Sigma}$ . Let  $\mathbf{w}'\mathbf{r}$  be a linear function of  $\mathbf{r}$  with non-stochastic weight vector  $\mathbf{w}$  so that  $\mathcal{E}(\mathbf{w}'\mathbf{r}) = \mathbf{w}'\mathcal{E}\mathbf{r}$ . The variance of  $\mathbf{w}'\mathbf{r}$  is the expectation of

$$[\mathbf{w}'(\mathbf{r} - \mathcal{E}\mathbf{r})]^2 = \mathbf{w}'(\mathbf{r} - \mathcal{E}\mathbf{r})(\mathbf{r} - \mathcal{E}\mathbf{r})'\mathbf{w},$$

so that  $\text{var } \mathbf{w}'\mathbf{r} = \mathbf{w}'\mathcal{V}(\mathbf{r})\mathbf{w} = \mathbf{w}'\boldsymbol{\Sigma}\mathbf{w}$ . Thus, the variance of any linear function of  $\mathbf{r}$  equals a quadratic form with the covariance matrix of  $\mathbf{r}$  as matrix.

If the quadratic form  $\mathbf{x}'\mathbf{A}\mathbf{x}$  is positive for any  $\mathbf{x} \neq \mathbf{0}$ ,  $\mathbf{A}$  is said to be *positive definite*. An example is a diagonal matrix  $\mathbf{A}$  with positive diagonal elements. If  $\mathbf{x}'\mathbf{A}\mathbf{x} \geq 0$  for any  $\mathbf{x}$ ,  $\mathbf{A}$  is called *positive semidefinite*. The covariance matrix  $\boldsymbol{\Sigma}$  of any random vector is always positive semidefinite because we just proved that  $\mathbf{w}'\boldsymbol{\Sigma}\mathbf{w}$  is the variance of a linear function and variances are non-negative. This covariance matrix is positive semidefinite but not positive definite if  $\mathbf{w}'\boldsymbol{\Sigma}\mathbf{w} = 0$  holds for some  $\mathbf{w} \neq \mathbf{0}$ , i.e. if there exists a non-stochastic linear function of the random vector. For example, consider the input allocation system (5.23) with a

<sup>6</sup>Hall (1973) has shown that the additivity of the cost function in the  $m$  outputs is a necessary and sufficient condition in order that the multiproduct firm can be broken up into  $m$  single-product firms in the following way: when the latter firms independently maximize profit by adjusting output, they use the same aggregate level of each input and produce the same level of output as the multiproduct firm.

disturbance vector  $\boldsymbol{\varepsilon}$  added:

$$\mathbf{F}\boldsymbol{\kappa} = (\boldsymbol{\iota}'\mathbf{F}\boldsymbol{\kappa})\boldsymbol{\Theta}\boldsymbol{\iota} - \psi\boldsymbol{\Theta}(\mathbf{I} - \boldsymbol{\iota}'\boldsymbol{\Theta})\boldsymbol{\pi} + \boldsymbol{\varepsilon}. \quad (6.1)$$

Premultiplication by  $\boldsymbol{\iota}'$  and use of (5.9) yields  $\boldsymbol{\iota}'\mathbf{F}\boldsymbol{\kappa} = \boldsymbol{\iota}'\mathbf{F}\boldsymbol{\kappa} + \boldsymbol{\iota}'\boldsymbol{\varepsilon}$ , or  $\boldsymbol{\iota}'\boldsymbol{\varepsilon} = 0$ , which means that the disturbances of the  $N$  equations sum to zero with unit probability. This property results from the allocation character of the system (6.1).

We return to the standard linear model described in the discussion following eq. (2.7). The Gauss–Markov theorem states that the LS estimator  $\mathbf{b}$  in (2.8) is best linear unbiased in the following sense: any other estimator  $\tilde{\boldsymbol{\beta}}$  of  $\boldsymbol{\beta}$  which is also linear in  $\mathbf{y}$  and unbiased has the property that  $\mathcal{V}(\tilde{\boldsymbol{\beta}}) - \mathcal{V}(\mathbf{b})$  is a positive semidefinite matrix. That is,

$$\mathbf{w}'[\mathcal{V}(\tilde{\boldsymbol{\beta}}) - \mathcal{V}(\mathbf{b})]\mathbf{w} \geq 0 \quad \text{for any } \mathbf{w},$$

or  $\mathbf{w}'\mathcal{V}(\tilde{\boldsymbol{\beta}})\mathbf{w} \geq \mathbf{w}'\mathcal{V}(\mathbf{b})\mathbf{w}$ . Since both sides of this inequality are the variance of an estimator of  $\mathbf{w}'\boldsymbol{\beta}$ , the implication is that within the class of linear unbiased estimators LS provides the estimator of any linear function of  $\boldsymbol{\beta}$  with the smallest possible variance. This is a stronger result than the statement in the discussion following eq. (2.10); that statement is confined to the estimation of elements rather than general linear functions of  $\boldsymbol{\beta}$ .

To prove the Gauss–Markov theorem we use the linearity of  $\tilde{\boldsymbol{\beta}}$  in  $\mathbf{y}$  to write  $\tilde{\boldsymbol{\beta}} = \mathbf{B}\mathbf{y}$ , where  $\mathbf{B}$  is a  $K \times n$  matrix consisting of non-stochastic elements. We define  $\mathbf{C} = \mathbf{B} - (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$  so that  $\tilde{\boldsymbol{\beta}} = \mathbf{B}\mathbf{y}$  can be written as

$$\begin{aligned} [\mathbf{C} + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}']\mathbf{y} &= [\mathbf{C} + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'](\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}) \\ &= (\mathbf{C}\mathbf{X} + \mathbf{I})\boldsymbol{\beta} + [\mathbf{C} + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}']\boldsymbol{\varepsilon}. \end{aligned}$$

The expectation of  $\tilde{\boldsymbol{\beta}}$  is thus  $(\mathbf{C}\mathbf{X} + \mathbf{I})\boldsymbol{\beta}$ , which must be identically equal to  $\boldsymbol{\beta}$  in order that the estimator be unbiased. Therefore,  $\mathbf{C}\mathbf{X} = \mathbf{0}$  and  $\tilde{\boldsymbol{\beta}} = \boldsymbol{\beta} + [\mathbf{C} + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}']\boldsymbol{\varepsilon}$  so that  $\mathcal{V}(\tilde{\boldsymbol{\beta}})$  equals

$$\begin{aligned} &[\mathbf{C} + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}']\mathcal{V}(\boldsymbol{\varepsilon})[\mathbf{C} + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}']' \\ &= \sigma^2\mathbf{C}\mathbf{C}' + \sigma^2(\mathbf{X}'\mathbf{X})^{-1} + \sigma^2\mathbf{C}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} + \sigma^2(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{C}'. \end{aligned}$$

It thus follows from (2.10) and  $\mathbf{C}\mathbf{X} = \mathbf{0}$  that  $\mathcal{V}(\tilde{\boldsymbol{\beta}}) - \mathcal{V}(\mathbf{b}) = \sigma^2\mathbf{C}\mathbf{C}'$ , which is a positive semidefinite matrix because  $\sigma^2\mathbf{w}'\mathbf{C}\mathbf{C}'\mathbf{w} = (\sigma\mathbf{C}'\mathbf{w})(\sigma\mathbf{C}'\mathbf{w})$  is the non-negative squared length of the vector  $\sigma\mathbf{C}'\mathbf{w}$ .

## 6.2. Maxima and minima

The matrix  $A$  is called *negative semidefinite* if  $x'Ax \leq 0$  holds for any  $x$  and *negative definite* if  $x'Ax < 0$  holds for any  $x \neq 0$ . If  $A$  is positive definite,  $-A$  is negative definite (similarly for semidefiniteness). If  $A$  is positive (negative) definite, all diagonal elements of  $A$  are positive (negative). This may be verified by considering  $x'Ax$  with  $x$  specified as a column of the unit matrix of appropriate order. If  $A$  is positive (negative) definite,  $A$  is non-singular because singularity would imply the existence of an  $x \neq 0$  so that  $Ax = 0$ , which is contradicted by  $x'Ax > 0$  ( $< 0$ ). If  $A$  is symmetric positive (negative) definite, so is  $A^{-1}$ , which is verified by considering  $x'Ax$  with  $x = A^{-1}y$  for  $y \neq 0$ .

For the function  $g(\cdot)$  of (2.6) to have a stationary value at  $z = \bar{z}$  it is necessary and sufficient that the gradient  $\partial g / \partial z$  at this point be zero. For this stationary value to be a local maximum (minimum) it is sufficient that the Hessian matrix  $\partial^2 g / \partial z \partial z'$  at this point be negative (positive) definite. We can apply this to the supply equation (5.26) which is obtained by adjusting the output vector  $z$  so as to maximize profit. We write profit as  $y'z - C$ , where  $y$  is the output price vector and  $C = \text{cost}$ . The gradient of profit as a function of  $z$  is  $y - \partial C / \partial z$  ( $y$  is independent of  $z$  because  $y$  is exogenous by assumption) and the Hessian matrix is  $-\partial^2 C / \partial z \partial z'$  so that a positive definite matrix  $\partial^2 C / \partial z \partial z'$  is a sufficient condition for maximum profit. Since  $\psi^*$  and  $R$  in (5.28) are positive, the matrix  $[\theta_{rs}^*]$  of the supply system (5.26) is positive definite. The diagonal elements of this matrix are therefore positive so that an increase in the price of a product raises its supply.

Similarly, a sufficient conditions for maximum utility is that the Hessian  $U$  be negative definite, implying  $\phi < 0$  [see (3.9) and (5.6)], and a sufficient condition for minimum cost is that  $F - \gamma H$  in (5.17) be positive definite. The result is that  $[\theta_{ij}]$  in both (5.5) and (5.15) is also positive definite. Since  $\phi$  and  $-\psi$  in these equations are negative, an increase in the Frisch-deflated price of any good (consumer good or input) reduces the demand for this good. For two goods,  $i$  and  $j$ , a positive (negative)  $\theta_{ij} = \theta_{ji}$  implies than an increase in the Frisch-deflated price of either good reduces (raises) the demand for the other; the two goods are then said to be *specific complements (substitutes)*. Under preference or input independence no good is a specific substitute or complement of any other good [see (5.11) and (5.18)]. The distinction between specific substitutes and complements is from Houthakker (1960); he proposed it for consumer goods, but it can be equally applied to a firm's inputs and also to outputs based on the sign of  $\theta_{rs}^* = \theta_{sr}^*$  in (5.26).

The assumption of a definite  $U$  or  $F - \gamma H$  is more than strictly necessary. In the consumer's case, when utility is maximized subject to the budget constraint  $p'q = M$ , it is sufficient to assume *constrained* negative definiteness, i.e.  $x'Ux < 0$  for all  $x \neq 0$  which satisfy  $p'x = 0$ . It is easy to construct examples of an indefinite

or singular semidefinite matrix  $U$  which satisfy this condition. Definiteness obviously implies constrained definiteness; we shall assume that  $U$  and  $F - \gamma H$  satisfy the stronger conditions so that the above analysis holds true.

### 6.3. Block-diagonal definite matrices

If a matrix is both definite and block-diagonal, the relevant principal submatrices are also definite. For example, if  $\Sigma$  of (3.4) is positive definite, then  $x'_1 \Sigma_1 x_1 + x'_2 \Sigma_2 x_2 > 0$  if either  $x_1 \neq 0$  or  $x_2 \neq 0$ , which would be violated if either  $\Sigma_1$  or  $\Sigma_2$  were not definite.

Another example is that of a logarithmic production function (5.12) when the inputs can be grouped into input groups so that the elasticity of output with respect to each input is independent of all inputs belonging to different groups. Then  $H$  of (5.13) is block-diagonal and so is  $\Theta$  [see (5.17)]. Thus, if  $i$  belongs to input group  $S_g$  ( $g = 1, 2, \dots$ ), the summation over  $j$  in the substitution term of (5.15) can be confined to  $j \in S_g$ ; equivalently, no input is then a specific substitute or complement of any input belonging to a different group. Also, summation of the input demand equations over all inputs of a group yields a composite demand equation for the input group which takes a similar form, while an appropriate combination of this composite equation with a demand equation for an individual input yields a conditional demand equation for the input within their group. These developments can also be applied to outputs and consumer goods, but they are beyond the scope of this chapter.

## 7. Diagonalizations

### 7.1. The standard diagonalization of a square matrix

For some  $n \times n$  matrix  $A$  we seek a vector  $x$  so that  $Ax$  equals a scalar multiple  $\lambda$  of  $x$ . This is trivially satisfied by  $x = 0$ , so we impose  $x'x = 1$  implying  $x \neq 0$ . Since  $Ax = \lambda x$  is equivalent to  $(A - \lambda I)x = 0$ , we thus have

$$(A - \lambda I)x = 0, \quad x'x = 1, \quad (7.1)$$

so that  $A - \lambda I$  is singular. This implies a zero determinant value,

$$|A - \lambda I| = 0, \quad (7.2)$$

which is known as the *characteristic equation* of  $A$ . For example, if  $A$  is diagonal

with  $d_1, \dots, d_n$  on the diagonal, (7.2) states that the product of  $d_i - \lambda$  over  $i$  vanishes so that each  $d_i$  is a solution of the characteristic equation. More generally, the characteristic equation of an  $n \times n$  matrix  $A$  is a polynomial of degree  $n$  and thus yields  $n$  solutions  $\lambda_1, \dots, \lambda_n$ . These  $\lambda_i$ 's are the *latent roots* of  $A$ ; the product of the  $\lambda_i$ 's equals the determinant of  $A$  and the sum of the  $\lambda_i$ 's equals the trace of  $A$ . A vector  $x_i$  which satisfies  $Ax_i = \lambda_i x_i$  and  $x_i' x_i = 1$  is called a *characteristic vector* of  $A$  corresponding to root  $\lambda_i$ .

Even if  $A$  consists of real elements, its roots need not be real, but these roots are all real when  $A$  is a real symmetric matrix. For suppose, to the contrary, that  $\lambda$  is a complex root and  $x + iy$  is a characteristic vector corresponding to this  $\lambda$ , where  $i = \sqrt{-1}$ . Then  $A(x + iy) = \lambda(x + iy)$ , which we premultiply by  $(x - iy)'$ :

$$x'Ax + y'Ay + i(x'Ay - y'Ax) = \lambda(x'x + y'y). \quad (7.3)$$

But  $x'Ay = y'Ax$  if  $A$  is symmetric, so that (7.3) shows that  $\lambda$  is the ratio of two real numbers,  $x'Ax + y'Ay$  and  $x'x + y'y$ . Roots of asymmetric matrices are considered at the end of this section.

Let  $\lambda_i$  and  $\lambda_j$  be two different roots ( $\lambda_i \neq \lambda_j$ ) of a symmetric matrix  $A$  and let  $x_i$  and  $x_j$  be corresponding characteristic vectors. We premultiply  $Ax_i = \lambda_i x_i$  by  $x_j'$  and  $Ax_j = \lambda_j x_j$  by  $x_i'$  and subtract:

$$x_j'Ax_i - x_i'Ax_j = (\lambda_i - \lambda_j)x_i'x_j.$$

Since the left side vanishes for a symmetric matrix  $A$ , we must have  $x_i'x_j = 0$  because  $\lambda_i \neq \lambda_j$ . This proves that characteristic vectors of a symmetric matrix are orthogonal when they correspond to different roots. When all roots of a symmetric  $n \times n$  matrix  $A$  are distinct, we thus have  $x_i'x_j = \delta_{ij}$  for all  $(i, j)$ . This is equivalent to

$$X'X = I, \quad \text{where } X = [x_1 \ x_2 \ \dots \ x_n]. \quad (7.4)$$

Also,

$$AX = [Ax_1 \ \dots \ Ax_n] = [\lambda_1 x_1 \ \dots \ \lambda_n x_n],$$

or

$$AX = X\Lambda, \quad (7.5)$$

where  $\Lambda$  is the diagonal matrix with  $\lambda_1, \dots, \lambda_n$  on the diagonal. Premultiplication of (7.5) by  $X'$  yields  $X'AX = X'X\Lambda$ , or

$$X'AX = \Lambda \quad (7.6)$$

in view of (7.4). Therefore, when we postmultiply a symmetric matrix  $A$  by a matrix  $X$  consisting of characteristic vectors of  $A$  and premultiply by  $X'$ , we obtain the diagonal matrix containing the latent roots of  $A$ . This double multiplication amounts to a *diagonalization* of  $A$ . Also, postmultiplication of (7.5) by  $X'$  yields  $AXX' = X\Lambda X'$  and hence, since (7.4) implies  $X' = X^{-1}$  or  $XX' = I$ ,

$$A = X\Lambda X' = \sum_{i=1}^n \lambda_i x_i x_i' \quad (7.7)$$

In the previous paragraph we assumed that the  $\lambda_i$ 's are distinct, but it may be shown that for any symmetric  $A$  there exists an  $X$  which satisfies (7.4)–(7.7), the columns of  $X$  being characteristic vectors of  $A$  and  $\Lambda$  being diagonal with the latent roots of  $A$  on the diagonal. The only difference is that in the case of multiple roots ( $\lambda_i = \lambda_j$  for  $i \neq j$ ) the associated characteristic vectors ( $x_i$  and  $x_j$ ) are not unique. Note that even when all  $\lambda$ 's are distinct, each  $x_i$  may be arbitrarily multiplied by  $-1$  because this affects neither  $Ax_i = \lambda_i x_i$  nor  $x_i' x_j = 0$  for any  $(i, j)$ ; however, this sign indeterminacy will be irrelevant for our purposes.

## 7.2. Special cases

Let  $A$  be square and premultiply  $Ax = \lambda x$  by  $A$  to obtain  $A^2 x = \lambda Ax = \lambda^2 x$ . This shows that  $A^2$  has the same characteristic vectors as  $A$  and latent roots equal to the squares of those of  $A$ . In particular, if a matrix is symmetric idempotent, such as  $M$  of (2.11), all latent roots are 0 or 1 because these are the only real numbers that do not change when squared. For a symmetric non-singular  $A$ , premultiply  $Ax = \lambda x$  by  $(\lambda A)^{-1}$  to obtain  $A^{-1}x = (1/\lambda)x$ . Thus,  $A^{-1}$  has the same characteristic vectors as those of  $A$  and latent roots equal to the reciprocals of those of  $A$ . If the symmetric  $n \times n$  matrix  $A$  is singular and has rank  $r$ , (7.2) is satisfied by  $\lambda = 0$  and this zero root has multiplicity  $n - r$ . It thus follows from (7.7) that  $A$  can then be written as the sum of  $r$  matrices of unit rank, each of the form  $\lambda_i x_i x_i'$ , with  $\lambda_i \neq 0$ .

Premultiplication of (7.7) by  $y'$  and postmultiplication by  $y$  yields  $y'Ay = \sum_i \lambda_i c_i^2$ , with  $c_i = y'x_i$ . Since  $y'Ay$  is positive (negative) for any  $y \neq 0$  if  $A$  is positive (negative) definite, this shows that all latent roots of a symmetric positive (negative) definite matrix are positive (negative). Similarly, all latent roots of a symmetric positive (negative) semidefinite matrix are non-negative (non-positive).

Let  $A_m$  be a symmetric  $m \times m$  matrix with roots  $\lambda_1, \dots, \lambda_m$  and characteristic vectors  $x_1, \dots, x_m$ ; let  $B_n$  be a symmetric  $n \times n$  matrix with roots  $\mu_1, \dots, \mu_n$  and characteristic vectors  $y_1, \dots, y_n$ . Hence,  $A_m \otimes B_n$  is of order  $mn \times mn$  and has  $mn$  latent roots and characteristic vectors. We use  $A_m x_i = \lambda_i x_i$  and  $B_n y_j = \mu_j y_j$  in

$$(A_m \otimes B_n)(x_i \otimes y_j) = (A_m x_i) \otimes (B_n y_j) = (\lambda_i x_i) \otimes (\mu_j y_j) = \lambda_i \mu_j (x_i \otimes y_j),$$

which shows that  $\mathbf{x}_i \otimes \mathbf{y}_j$  is a characteristic vector of  $\mathbf{A}_m \otimes \mathbf{B}_n$  corresponding to root  $\lambda_i \mu_j$ . It is easily verified that these characteristic vectors form an orthogonal matrix of order  $mn \times mn$ :

$$\begin{aligned}(\mathbf{x}_1 \otimes \mathbf{y}_1)'(\mathbf{x}_1 \otimes \mathbf{y}_1) &= (\mathbf{x}'_1 \otimes \mathbf{y}'_1)(\mathbf{x}_1 \otimes \mathbf{y}_1) = (\mathbf{x}'_1 \mathbf{x}_1) \otimes (\mathbf{y}'_1 \mathbf{y}_1) = 1, \\(\mathbf{x}_1 \otimes \mathbf{y}_1)'(\mathbf{x}_2 \otimes \mathbf{y}_1) &= (\mathbf{x}'_1 \mathbf{x}_2) \otimes (\mathbf{y}'_1 \mathbf{y}_1) = 0.\end{aligned}$$

Since the determinant of  $\mathbf{A}_m \otimes \mathbf{B}_n$  equals the product of the roots, we have

$$|\mathbf{A}_m \otimes \mathbf{B}_n| = \prod_{i=1}^m \prod_{j=1}^n \lambda_i \mu_j = \left( \prod_{i=1}^m \lambda_i^n \right) \left( \prod_{j=1}^n \mu_j^m \right) = |\mathbf{A}_m|^n |\mathbf{B}_n|^m.$$

It may similarly be verified that the rank (trace) of  $\mathbf{A}_m \otimes \mathbf{B}_n$  equals the product of the ranks (traces) of  $\mathbf{A}_m$  and  $\mathbf{B}_n$ .

### 7.3. Aitken's theorem

Any symmetric positive definite matrix  $\mathbf{A}$  can be written as  $\mathbf{A} = \mathbf{Q}\mathbf{Q}'$ , where  $\mathbf{Q}$  is some non-singular matrix. For example, we can use (7.7) and specify  $\mathbf{Q} = \mathbf{X}\mathbf{\Lambda}^{1/2}$ , where  $\mathbf{\Lambda}^{1/2}$  is the diagonal matrix which contains the positive square roots of the latent roots of  $\mathbf{A}$  on the diagonal. Since the roots of  $\mathbf{A}$  are all positive,  $\mathbf{\Lambda}^{1/2}$  is non-singular;  $\mathbf{X}$  is non-singular in view of (7.4); therefore,  $\mathbf{Q} = \mathbf{X}\mathbf{\Lambda}^{1/2}$  is also non-singular.

Consider in particular the disturbance covariance matrix  $\sigma^2\mathbf{V}$  in the discussion preceding eq. (2.13). Since  $\sigma^2 > 0$  and  $\mathbf{V}$  is non-singular by assumption, this covariance matrix is symmetric positive definite. Therefore,  $\mathbf{V}^{-1}$  is also symmetric positive definite and we can write  $\mathbf{V}^{-1} = \mathbf{Q}\mathbf{Q}'$  for some non-singular  $\mathbf{Q}$ . We premultiply (2.7) by  $\mathbf{Q}'$ :

$$\mathbf{Q}'\mathbf{y} = (\mathbf{Q}'\mathbf{X})\boldsymbol{\beta} + \mathbf{Q}'\boldsymbol{\varepsilon}. \quad (7.8)$$

The disturbance vector  $\mathbf{Q}'\boldsymbol{\varepsilon}$  has zero expectation and a covariance matrix equal to

$$\sigma^2\mathbf{Q}'\mathbf{V}\mathbf{Q} = \sigma^2\mathbf{Q}'(\mathbf{Q}\mathbf{Q}')^{-1}\mathbf{Q} = \sigma^2\mathbf{Q}'(\mathbf{Q}')^{-1}\mathbf{Q}^{-1}\mathbf{Q} = \sigma^2\mathbf{I},$$

so that the standard linear model and the Gauss–Markov theorem are now applicable. Thus, we estimate  $\boldsymbol{\beta}$  by applying LS to (7.8):

$$[(\mathbf{Q}'\mathbf{X})'\mathbf{Q}'\mathbf{X}]^{-1}(\mathbf{Q}'\mathbf{X})'\mathbf{Q}'\mathbf{y} = (\mathbf{X}'\mathbf{Q}\mathbf{Q}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Q}\mathbf{Q}'\mathbf{y} = (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}\mathbf{y},$$

which is the GLS estimator (2.13). The covariance matrix (2.14) is also easily verified.

#### 7.4. The Cholesky decomposition

The diagonalization (7.6) uses an orthogonal matrix  $X$  [see (7.4)], but it is also possible to use a triangular matrix. For example, consider a diagonal matrix  $D$  and an upper triangular matrix  $C$  with units in the diagonal,

$$C = \begin{bmatrix} 1 & c_{12} & c_{13} \\ 0 & 1 & c_{23} \\ 0 & 0 & 1 \end{bmatrix}, \quad D = \begin{bmatrix} d_1 & 0 & 0 \\ 0 & d_2 & 0 \\ 0 & 0 & d_3 \end{bmatrix},$$

yielding

$$C'DC = \begin{bmatrix} d_1 & d_1c_{12} & d_1c_{13} \\ d_1c_{12} & d_1c_{12}^2 + d_2 & d_1c_{12}c_{13} + d_2c_{23} \\ d_1c_{13} & d_1c_{12}c_{13} + d_2c_{23} & d_1c_{13}^2 + d_2c_{23}^2 + d_3 \end{bmatrix}.$$

It is readily verified that any  $3 \times 3$  symmetric positive definite matrix  $A = [a_{ij}]$  can be uniquely written as  $C'DC$  ( $d_1 = a_{11}$ ,  $c_{12} = a_{12}/a_{11}$ , etc.). This is the so-called *Cholesky decomposition* of a matrix; for applications to demand analysis see Barten and Geyskens (1975) and Theil and Laitinen (1979). Also, note that  $D = (C')^{-1}AC^{-1}$  and that  $C^{-1}$  is upper triangular with units in the diagonal.

#### 7.5. Vectors written as diagonal matrices

On many occasions we want to write a vector in the form of a diagonal matrix. An example is the price vector  $p$  which occurs as a diagonal matrix  $P$  in (5.8). An alternative notation is  $\hat{p}$ , with the hat indicating that the vector has become a diagonal matrix. However, such notations are awkward when the vector which we want to write as a diagonal matrix is itself the product of one or several matrices and a vector. For example, in Section 8 we shall meet a nonsingular  $N \times N$  matrix  $X$  and the vector  $X^{-1}t$ . We write this vector in diagonal matrix form as

$$(X^{-1}t)_\Delta = \begin{bmatrix} \sum_j x^{1j} & 0 & \dots & 0 \\ 0 & \sum_j x^{2j} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sum_j x^{Nj} \end{bmatrix}, \quad (7.9)$$

where  $x^{ij}$  is an element of  $X^{-1}$  and all summations are over  $j = 1, \dots, N$ . It is easily



verified that

$$(X^{-1}t)_\Delta t = X^{-1}t, \quad t'(X^{-1}t)_\Delta = t'(X')^{-1}. \quad (7.10)$$

### 7.6. A simultaneous diagonalization of two square matrices

We extend (7.1) to

$$(A - \lambda B)x = 0, \quad x'Bx = 1, \quad (7.11)$$

where  $A$  and  $B$  are symmetric  $n \times n$  matrices,  $B$  being positive definite. Thus,  $B^{-1}$  is symmetric positive definite so that  $B^{-1} = QQ'$  for some non-singular  $Q$ . It is easily seen that (7.11) is equivalent to

$$(Q'AQ - \lambda I)y = 0, \quad y'y = 1, \quad y = Q^{-1}x. \quad (7.12)$$

This shows that (7.11) can be reduced to (7.1) with  $A$  interpreted as  $Q'AQ$ . If  $A$  is symmetric, so is  $Q'AQ$ . Therefore, all results for symmetric matrices described earlier in this section are applicable. In particular, (7.11) has  $n$  solutions,  $\lambda_1, \dots, \lambda_n$  and  $x_1, \dots, x_n$ , the  $x_i$ 's being unique when the  $\lambda_i$ 's are distinct. From  $y_i'y_j = \delta_{ij}$  and  $y_i = Q^{-1}x_i$  we have  $x_i'Bx_j = \delta_{ij}$  and hence  $X'BX = I$ , where  $X$  is the  $n \times n$  matrix with  $x_i$  as the  $i$ th column. We write  $(A - \lambda B)x = 0$  as  $Ax_i = \lambda_i Bx_i$  for the  $i$ th solution and as  $AX = BX\Lambda$  for all solutions jointly, where  $\Lambda$  is diagonal with  $\lambda_1, \dots, \lambda_n$  on the diagonal. Premultiplication of  $AX = BX\Lambda$  by  $X'$  then yields  $X'AX = X'BX\Lambda = \Lambda$ . Therefore,

$$X'AX = \Lambda, \quad X'BX = I, \quad (7.13)$$

which shows that both matrices are simultaneously diagonalized,  $A$  being transformed into the latent root matrix  $\Lambda$ , and  $B$  into the unit matrix.

It is noteworthy that (7.11) can be interpreted in terms of a constrained extremum problem. Let us seek the maximum of the quadratic form  $x'Ax$  for variations in  $x$  subject to  $x'Bx = 1$ . So we construct the Lagrangian function  $x'Ax - \mu(x'Bx - 1)$ , which we differentiate with respect to  $x$ , yielding  $2Ax - 2\mu Bx$ . By equating this derivative to zero we obtain  $Ax = \mu Bx$ , which shows that  $\mu$  must be a root  $\lambda_i$  of (7.11). Next, we premultiply  $Ax = \mu Bx$  by  $x'$ , which gives  $x'Ax = \mu x'Bx = \mu$  and shows that the largest root  $\lambda_i$  is the maximum of  $x'Ax$  subject to  $x'Bx = 1$ . Similarly, the smallest root is the minimum of  $x'Ax$  subject to  $x'Bx = 1$ , and all  $n$  roots are stationary values of  $x'Ax$  subject to  $x'Bx = 1$ .

### 7.7. Latent roots of an asymmetric matrix

Some or all latent roots of an asymmetric square matrix  $A$  may be complex. If (7.2) yields complex roots, they occur in conjugate complex pairs of the form  $a \pm bi$ . The absolute value of such a root is defined as  $\sqrt{a^2 + b^2}$ , which equals  $|a|$  if  $b = 0$ , i.e. if the root is real. If  $A$  is asymmetric, the latent roots of  $A$  and  $A'$  are still the same but a characteristic vector of  $A$  need not be a characteristic vector of  $A'$ . If  $A$  is asymmetric and has multiple roots, it may have fewer characteristic vectors than the number of its rows and columns. For example,

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

is an asymmetric  $2 \times 2$  matrix with a double unit root, but it has only one characteristic vector,  $[1 \ 0]'$ . A further analysis of this subject involves the Jordan canonical form, which is beyond the scope of this chapter; see Bellman (1960, ch. 11).

Latent roots of asymmetric matrices play a role in the stability analysis of dynamic equations systems. Consider the reduced form

$$y_t = a + Ay_{t-1} + A^*x_t + u_t, \quad (7.14)$$

where  $y_t$  is an  $L$ -element observation vector on endogenous variables at time  $t$ ,  $a$  is a vector of constant terms,  $A$  is a square coefficient matrix,  $A^*$  is an  $L \times K$  coefficient matrix,  $x_t$  is a  $K$ -element observation vector on exogenous variables at  $t$ , and  $u_t$  is a disturbance vector. Although  $A$  is square, there is no reason why it should be symmetric so that its latent roots may include conjugate complex pairs. In the next paragraph we shall be interested in the limit of  $A^s$  as  $s \rightarrow \infty$ . Recall that  $A^2$  has roots equal to the squares of those of  $A$ ; this also holds for the complex roots of an asymmetric  $A$ . Therefore,  $A^s$  has latent roots equal to the  $s$ th power of those of  $A$ . If the roots of  $A$  are all less than 1 in absolute value, those of  $A^s$  will all converge to zero as  $s \rightarrow \infty$ , which means that the limit of  $A^s$  for  $s \rightarrow \infty$  is a zero matrix. Also, let  $S = I + A + \dots + A^s$ ; then, by subtracting  $AS = A + A^2 + \dots + A^{s+1}$  we obtain  $(I - A)S = I - A^{s+1}$  so that we have the following result for the limit of  $S$  when all roots of  $A$  are less than 1 in absolute value:

$$\lim_{s \rightarrow \infty} (I + A + \dots + A^s) = (I - A)^{-1}. \quad (7.15)$$

We proceed to apply these results to (7.14) by lagging it by one period,  $y_{t-1} = a + Ay_{t-2} + A^*x_{t-1} + u_{t-1}$ , and substituting this into (7.14):

$$y_t = (I + A)a + A^2y_{t-2} + A^*x_t + AA^*x_{t-1} + u_t + Au_{t-1}$$

When we do this  $s$  times, we obtain:

$$\begin{aligned} y_t &= (\mathbf{I} + \mathbf{A} + \cdots + \mathbf{A}^s) \mathbf{a} + \mathbf{A}^{s+1} y_{t-s-1} \\ &\quad + \mathbf{A}^s \mathbf{x}_t + \mathbf{A} \mathbf{A}^s \mathbf{x}_{t-1} + \cdots + \mathbf{A}^s \mathbf{A}^s \mathbf{x}_{t-s} \\ &\quad + \mathbf{u}_t + \mathbf{A} \mathbf{u}_{t-1} + \cdots + \mathbf{A}^s \mathbf{u}_{t-s}. \end{aligned} \quad (7.16)$$

If all roots of  $\mathbf{A}$  are less than 1 in absolute value, so that  $\mathbf{A}^s$  converges to zero as  $s \rightarrow \infty$  and (7.15) holds, the limit of (7.16) for  $s \rightarrow \infty$  becomes

$$y_t = (\mathbf{I} - \mathbf{A})^{-1} \mathbf{a} + \sum_{s=0}^{\infty} \mathbf{A}^s \mathbf{A}^s \mathbf{x}_{t-s} + \sum_{s=0}^{\infty} \mathbf{A}^s \mathbf{u}_{t-s}, \quad (7.17)$$

which is the *final form* of the equation system. This form expresses each current endogenous variable in terms of current and lagged exogenous variables as well as current and lagged disturbances; all lagged endogenous variables are eliminated from the reduced form (7.14) by successive lagged application of (7.14). The coefficient matrices  $\mathbf{A}^s \mathbf{A}^s$  of  $\mathbf{x}_{t-s}$  for  $s = 0, 1, 2, \dots$  in (7.17) may be viewed as matrix multipliers; see Goldberger (1959). The behavior of the elements of  $\mathbf{A}^s \mathbf{A}^s$  as  $s \rightarrow \infty$  is dominated by the root of  $\mathbf{A}$  with the largest absolute value. If this root is a conjugate complex pair, the behavior of these elements for increasing  $s$  is of the damped oscillatory type.

Endogenous variables occur in (7.14) only with a one-period lag. Suppose that  $\mathbf{A} y_{t-1}$  in (7.14) is extended to  $\mathbf{A}_1 y_{t-1} + \cdots + \mathbf{A}_\tau y_{t-\tau}$ , where  $\tau$  is the largest lag which occurs in the equation system. It may be shown that the relevant determinantal equation is now

$$|\lambda^\tau (-\mathbf{I}) + \lambda^{\tau-1} \mathbf{A}_1 + \cdots + \mathbf{A}_\tau| = 0. \quad (7.18)$$

When there are  $L$  endogenous variables, (7.18) yields  $L\tau$  solutions which may include conjugate complex pairs. All these solutions should be less than 1 in absolute value in order that the system be stable, i.e. in order that the coefficient matrix of  $\mathbf{x}_{t-s}$  in the final form converges to zero as  $s \rightarrow \infty$ . It is readily verified that for  $\tau = 1$  this condition refers to the latent roots of  $\mathbf{A}_1$ , in agreement with the condition underlying (7.17).

## 8. Principal components and extensions

### 8.1. Principal components

Consider an  $n \times K$  observation matrix  $\mathbf{Z}$  on  $K$  variables. Our objective is to approximate  $\mathbf{Z}$  by a matrix of unit rank,  $\mathbf{v} \mathbf{c}'$ , where  $\mathbf{v}$  is an  $n$ -element vector of

values taken by some variable (to be constructed below) and  $c$  is a  $K$ -element coefficient vector. Thus, the approximation describes each column of  $Z$  as proportional to  $v$ . It is obvious that if the rank of  $Z$  exceeds 1, there will be a non-zero  $n \times K$  discrepancy matrix  $Z - vc'$  no matter how we specify  $v$  and  $c$ ; we select  $v$  and  $c$  by minimizing the sum of the squares of all  $Kn$  discrepancies. Also, since  $vc'$  remains unchanged when  $v$  is multiplied by  $k \neq 0$  and  $c$  by  $1/k$ , we shall impose  $v'v = 1$ . It is shown in the next subsection that the solution is  $v = v_1$  and  $c = c_1$ , defined by

$$(ZZ' - \lambda_1 I) v_1 = 0, \quad (8.1)$$

$$c_1 = Z'v_1, \quad (8.2)$$

where  $\lambda_1$  is the largest latent root of the symmetric positive semidefinite matrix  $ZZ'$ . Thus, (8.1) states that  $v_1$  is a characteristic vector of  $ZZ'$  corresponding to the largest latent root. (We assume that the non-zero roots of  $ZZ'$  are distinct.) Note that  $v_1$  may be arbitrarily multiplied by  $-1$  in (8.1) but that this changes  $c_1$  into  $-c_1$  in (8.2) so that the product  $v_1 c_1'$  remains unchanged.

Our next objective is to approximate the discrepancy matrix  $Z - v_1 c_1'$  by a matrix of unit rank  $v_2 c_2'$ , so that  $Z$  is approximated by  $v_1 c_1' + v_2 c_2'$ . The criterion is again the residual sum of squares, which we minimize by varying  $v_2$  and  $c_2$  subject to the constraints  $v_2'v_2 = 1$  and  $v_2'v_1 = 0$ . It is shown in the next subsection that the solution is identical to (8.1) and (8.2) except that the subscript 1 becomes 2 with  $\lambda_2$  interpreted as the second largest latent root of  $ZZ'$ . The vectors  $v_1$  and  $v_2$  are known as the first and second *principal components* of the  $K$  variables whose observations are arranged in the  $n \times K$  matrix  $Z$ .

More generally, the  $i$ th principal component  $v_i$  and the associated coefficient vector  $c_i$  are obtained from

$$(ZZ' - \lambda_i I) v_i = 0, \quad (8.3)$$

$$c_i = Z'v_i, \quad (8.4)$$

where  $\lambda_i$  is the  $i$ th largest root of  $ZZ'$ . This solution is obtained by approximating  $Z - v_1 c_1' - \dots - v_{i-1} c_{i-1}'$  by a matrix  $v_i c_i'$ , the criterion being the sum of the squared discrepancies subject to the unit length condition  $v_i'v_i = 1$  and the orthogonality conditions  $v_i'v_j = 0$  for  $j = 1, \dots, i-1$ .

## 8.2. Derivations

It is easily verified that the sum of the squares of all elements of any matrix  $A$  (square or rectangular) is equal to  $\text{tr} A'A$ . Thus, the sum of the squares of the

elements of the discrepancy matrix  $\mathbf{Z} - \mathbf{v}\mathbf{c}'$  equals

$$\begin{aligned}\text{tr}(\mathbf{Z} - \mathbf{v}\mathbf{c}')'(\mathbf{Z} - \mathbf{v}\mathbf{c}') &= \text{tr}\mathbf{Z}'\mathbf{Z} - \text{tr}\mathbf{c}\mathbf{v}'\mathbf{Z} - \text{tr}\mathbf{Z}'\mathbf{v}\mathbf{c}' + \text{tr}\mathbf{c}\mathbf{v}'\mathbf{c}' \\ &= \text{tr}\mathbf{Z}'\mathbf{Z} - 2\mathbf{v}'\mathbf{Z}\mathbf{c} + (\mathbf{v}'\mathbf{v})(\mathbf{c}'\mathbf{c}),\end{aligned}$$

which can be simplified to

$$\text{tr}\mathbf{Z}'\mathbf{Z} - 2\mathbf{v}'\mathbf{Z}\mathbf{c} + \mathbf{c}'\mathbf{c} \quad (8.5)$$

in view of  $\mathbf{v}'\mathbf{v} = 1$ . The derivative of (8.5) with respect to  $\mathbf{c}$  is  $-2\mathbf{Z}'\mathbf{v} + 2\mathbf{c}$  so that minimizing (8.5) by varying  $\mathbf{c}$  for given  $\mathbf{v}$  yields  $\mathbf{c} = \mathbf{Z}'\mathbf{v}$ , in agreement with (8.2). By substituting  $\mathbf{c} = \mathbf{Z}'\mathbf{v}$  into (8.5) we obtain  $\text{tr}\mathbf{Z}'\mathbf{Z} - \mathbf{v}'\mathbf{Z}\mathbf{Z}'\mathbf{v}$ ; hence, our next step is to maximize  $\mathbf{v}'\mathbf{Z}\mathbf{Z}'\mathbf{v}$  for variations in  $\mathbf{v}$  subject to  $\mathbf{v}'\mathbf{v} = 1$ . So we construct the Lagrangian function  $\mathbf{v}'\mathbf{Z}\mathbf{Z}'\mathbf{v} - \mu(\mathbf{v}'\mathbf{v} - 1)$  and differentiate it with respect to  $\mathbf{v}$  and equate the derivative to zero. This yields  $\mathbf{Z}\mathbf{Z}'\mathbf{v} = \mu\mathbf{v}$  so that  $\mathbf{v}$  must be a characteristic vector of  $\mathbf{Z}\mathbf{Z}'$  corresponding to root  $\mu$ . This confirms (8.1) if we can prove that  $\mu$  equals the largest root  $\lambda_1$  of  $\mathbf{Z}\mathbf{Z}'$ . For this purpose we premultiply  $\mathbf{Z}\mathbf{Z}'\mathbf{v} = \mu\mathbf{v}$  by  $\mathbf{v}'$ , which gives  $\mathbf{v}'\mathbf{Z}\mathbf{Z}'\mathbf{v} = \mu\mathbf{v}'\mathbf{v} = \mu$ . Since we seek the maximum of  $\mathbf{v}'\mathbf{Z}\mathbf{Z}'\mathbf{v}$ , this shows that  $\mu$  must be the largest root of  $\mathbf{Z}\mathbf{Z}'$ .

To verify (8.3) and (8.4) for  $i = 2$ , we consider

$$\begin{aligned}\text{tr}(\mathbf{Z} - \mathbf{v}_1\mathbf{c}'_1 - \mathbf{v}_2\mathbf{c}'_2)'(\mathbf{Z} - \mathbf{v}_1\mathbf{c}'_1 - \mathbf{v}_2\mathbf{c}'_2) \\ &= \text{tr}(\mathbf{Z} - \mathbf{v}_1\mathbf{c}'_1)'(\mathbf{Z} - \mathbf{v}_1\mathbf{c}'_1) - 2\text{tr}(\mathbf{Z} - \mathbf{v}_1\mathbf{c}'_1)'\mathbf{v}_2\mathbf{c}'_2 + \text{tr}\mathbf{c}_2\mathbf{v}'_2\mathbf{v}_2\mathbf{c}'_2 \\ &= \text{tr}(\mathbf{Z} - \mathbf{v}_1\mathbf{c}'_1)'(\mathbf{Z} - \mathbf{v}_1\mathbf{c}'_1) - 2\mathbf{c}'_2\mathbf{Z}'\mathbf{v}_2 + \mathbf{c}'_2\mathbf{c}_2,\end{aligned} \quad (8.6)$$

where the last step is based on  $\mathbf{v}'_1\mathbf{v}_2 = 0$  and  $\mathbf{v}'_2\mathbf{v}_2 = 1$ . Minimization of (8.6) with respect to  $\mathbf{c}_2$  for given  $\mathbf{v}_2$  thus yields  $\mathbf{c}_2 = \mathbf{Z}'\mathbf{v}_2$ , in agreement with (8.4). Substitution of  $\mathbf{c}_2 = \mathbf{Z}'\mathbf{v}_2$  into (8.6) shows that the function to be minimized with respect to  $\mathbf{v}_2$  takes the form of a constant [equal to the trace in the last line of (8.6)] minus  $\mathbf{v}'_2\mathbf{Z}\mathbf{Z}'\mathbf{v}_2$ . So we maximize  $\mathbf{v}'_2\mathbf{Z}\mathbf{Z}'\mathbf{v}_2$  by varying  $\mathbf{v}_2$  subject to  $\mathbf{v}'_1\mathbf{v}_2 = 0$  and  $\mathbf{v}'_2\mathbf{v}_2 = 1$ , using the Lagrangian function  $\mathbf{v}'_2\mathbf{Z}\mathbf{Z}'\mathbf{v}_2 - \mu_1\mathbf{v}'_1\mathbf{v}_2 - \mu_2(\mathbf{v}'_2\mathbf{v}_2 - 1)$ . We take the derivative of this function with respect to  $\mathbf{v}_2$  and equate it to zero:

$$2\mathbf{Z}\mathbf{Z}'\mathbf{v}_2 - \mu_1\mathbf{v}_1 - 2\mu_2\mathbf{v}_2 = \mathbf{0}. \quad (8.7)$$

We premultiply this by  $\mathbf{v}'_1$ , which yields  $2\mathbf{v}'_1\mathbf{Z}\mathbf{Z}'\mathbf{v}_2 = \mu_1\mathbf{v}'_1\mathbf{v}_1 = \mu_1$  because  $\mathbf{v}'_1\mathbf{v}_2 = 0$ . But  $\mathbf{v}'_1\mathbf{Z}\mathbf{Z}'\mathbf{v}_2 = 0$  and hence  $\mu_1 = 0$  because (8.1) implies  $\mathbf{v}'_2\mathbf{Z}\mathbf{Z}'\mathbf{v}_1 = \lambda_1\mathbf{v}'_2\mathbf{v}_1 = 0$ . We can thus simplify (8.7) to  $\mathbf{Z}\mathbf{Z}'\mathbf{v}_2 = \mu_2\mathbf{v}_2$  so that  $\mathbf{v}_2$  is a characteristic vector of  $\mathbf{Z}\mathbf{Z}'$  corresponding to root  $\mu_2$ . This vector must be orthogonal to the characteristic vector  $\mathbf{v}_1$  corresponding to the largest root  $\lambda_1$ , while the root  $\mu_2$  must be as large as possible because the objective is to maximize  $\mathbf{v}'_2\mathbf{Z}\mathbf{Z}'\mathbf{v}_2 = \mu_2\mathbf{v}'_2\mathbf{v}_2 = \mu_2$ . Therefore,

$\mu_2$  must be the second largest root  $\lambda_2$  of  $\mathbf{Z}\mathbf{Z}'$ , which completes the proof of (8.3) and (8.4) for  $i = 2$ . The extension to larger values of  $i$  is left as an exercise for the reader.

### 8.3. Further discussion of principal components

If the rank of  $\mathbf{Z}$  is  $r$ , (8.3) yields  $r$  principal components corresponding to positive roots  $\lambda_1, \dots, \lambda_r$ . In what follows we assume that  $\mathbf{Z}$  has full column rank so that there are  $K$  principal components corresponding to  $K$  positive roots,  $\lambda_1, \dots, \lambda_K$ .

By premultiplying (8.3) by  $\mathbf{Z}'$  and using (8.4) we obtain:

$$(\mathbf{Z}'\mathbf{Z} - \lambda_i\mathbf{I})\mathbf{c}_i = \mathbf{0}, \quad (8.8)$$

so that the coefficient vector  $\mathbf{c}_i$  is a characteristic vector of  $\mathbf{Z}'\mathbf{Z}$  corresponding to root  $\lambda_i$ . The vectors  $\mathbf{c}_1, \dots, \mathbf{c}_K$  are orthogonal, but they do not have unit length. To prove this we introduce the matrix  $\mathbf{V}$  of all principal components and the associated coefficient matrix  $\mathbf{C}$ :

$$\mathbf{V} = [\mathbf{v}_1 \dots \mathbf{v}_K], \quad \mathbf{C} = [\mathbf{c}_1 \dots \mathbf{c}_K], \quad (8.9)$$

so that (8.3) and (8.4) for  $i = 1, \dots, K$  can be written as

$$\mathbf{Z}\mathbf{Z}'\mathbf{V} = \mathbf{V}\mathbf{\Lambda}, \quad (8.10)$$

$$\mathbf{C} = \mathbf{Z}'\mathbf{V}, \quad (8.11)$$

where  $\mathbf{\Lambda}$  is the diagonal matrix with  $\lambda_1, \dots, \lambda_K$  on the diagonal. By premultiplying (8.10) by  $\mathbf{V}'$  and using (8.11) and  $\mathbf{V}'\mathbf{V} = \mathbf{I}$  we obtain:

$$\mathbf{C}'\mathbf{C} = \mathbf{\Lambda}, \quad (8.12)$$

which shows that  $\mathbf{c}_1, \dots, \mathbf{c}_K$  are orthogonal vectors and that the squared length of  $\mathbf{c}_i$  equals  $\lambda_i$ .

If the observed variables are measured as deviations from their means,  $\mathbf{Z}'\mathbf{Z}$  in (8.8) equals their sample covariance matrix multiplied by  $n$ . Since  $\mathbf{Z}'\mathbf{Z}$  need not be diagonal, the observed variables may be correlated. But the principal components are all uncorrelated because  $\mathbf{v}'_i\mathbf{v}_j = 0$  for  $i \neq j$ . Therefore, these components can be viewed as *uncorrelated linear combinations of correlated variables*.

### 8.4. The independence transformation in microeconomic theory

The principal component technique can be extended so that two square matrices are simultaneously diagonalized. An attractive way of discussing this extension is

in terms of the differential demand and supply equations of Section 5. Recall that under preference independence the demand equation (5.5) takes the form (5.11) with only one relative price. Preference independence amounts to additive utility and is thus quite restrictive. But if the consumer is not preference independent with respect to the  $N$  observed goods, we may ask whether it is possible to transform these goods so that the consumer is preference independent with respect to the transformed goods. Similarly, if a firm is not input independent, can we derive transformed inputs so that the firm is input independent with respect to these? An analogous question can be asked for the outputs of a multiproduct firm; below we consider the inputs of a single-product firm in order to fix the attention.

Consider the input allocation equation (5.21) and divide by  $f_i$ :

$$d(\log q_i) = \frac{\theta_i}{f_i} d(\log Q) - \frac{\psi}{f_i} \sum_{j=1}^N \theta_{ij} d\left(\log \frac{p_j}{P'}\right). \quad (8.13)$$

This shows that  $\theta_i/f_i$  is the elasticity of the demand for input  $i$  with respect to the Divisia input volume index; we shall express this by referring to  $\theta_i/f_i$  as the *Divisia elasticity* of input  $i$ , which is the firm's input version of the consumer's income elasticity of the demand for a good.<sup>7</sup> Also, (8.13) shows that  $-\psi\theta_{ij}/f_i$  is the elasticity of input  $i$  with respect to the Frisch-deflated price of input  $j$ . Under input independence the substitution term is simplified [see (5.15) and (5.18)] so that (8.13) becomes

$$d(\log q_i) = \frac{\theta_i}{f_i} d(\log Q) - \frac{\psi\theta_i}{f_i} d\left(\log \frac{p_i}{P'}\right). \quad (8.14)$$

Hence, all price elasticities vanish except the own-price elasticities; the latter take the form  $-\psi\theta_i/f_i$  and are thus proportional to the Divisia elasticities with  $-\psi$  as the (negative) proportionality coefficient.

Next consider the input allocation system in the form (5.23):

$$F\kappa = (\iota'F\kappa)\Theta\iota - \psi\Theta(I - \iota'\Theta)\pi. \quad (8.15)$$

Our objective is to define transformed inputs which diagonalize  $\Theta$ . We perform this under the condition that total input expenditure and its Divisia decomposition are invariant under the transformation. The derivation is given in Appendix B and the result may be described by means of a simultaneous diagonalization

<sup>7</sup>The consumer's version of  $\theta_i/f_i$  is  $\theta_i/w_i$ ; it is easily verified [see (5.3)] that  $\theta_i/w_i$  equals the elasticity of  $q_i$  with respect to  $M$ .

similar to (7.13):

$$X'\Theta X = \Lambda, \quad X'FX = I, \quad (8.16)$$

where  $\Lambda$  is the diagonal matrix with the roots  $\lambda_1, \dots, \lambda_N$  on the diagonal. These roots are the Divisia elasticities of the transformed inputs. The allocation equation for transformed input  $i$  takes the form

$$d(\log q_{Ti}) = \lambda_i d(\log Q) - \psi \lambda_i d\left(\log \frac{p_{Ti}}{P'}\right), \quad (8.17)$$

where the subscript T stands for “transformed”. A comparison of (8.17) and (8.14) shows that the Divisia volume and Frisch price indexes and  $\psi$  are all invariant under the transformation.

Recall from (7.11) and (7.13) that any column  $x_i$  of the matrix  $X$  in (8.16) satisfies

$$(\Theta - \lambda_i F)x_i = \mathbf{0}. \quad (8.18)$$

We premultiply this by  $-\psi F^{-1}$ :

$$[-\psi F^{-1}\Theta - (-\psi \lambda_i)I]x_i = \mathbf{0}. \quad (8.19)$$

Since  $-\psi F^{-1}\Theta = [-\psi \theta_{ij}/f_i]$  is the price elasticity matrix of the observed inputs [see (8.13)] and  $-\psi \lambda_i$  is the own-price elasticity of transformed input  $i$  [see (8.17)], (8.19) shows that the latter elasticity is a latent root of the price elasticity matrix  $-\psi F^{-1}\Theta$  of the observed inputs. This is an asymmetric matrix, but the  $\lambda_i$ 's are nevertheless real. To prove this we premultiply (8.18) by  $F^{-1/2}$  and write the result as

$$(F^{-1/2}\Theta F^{-1/2} - \lambda_i I)F^{1/2}x_i = \mathbf{0}. \quad (8.20)$$

Since  $F^{-1/2}\Theta F^{-1/2}$  is symmetric positive definite, the  $\lambda_i$ 's are all real and positive. Hence, all transformed inputs have positive Divisia elasticities. The diagonalization (8.20) is unique when the  $\lambda_i$ 's are distinct. This means that the transformed inputs are identified by their Divisia elasticities.

These elasticities can be used as a tool for the interpretation of the transformed inputs. Another tool is the so-called *composition matrix*

$$T = (X^{-1}t)_\Delta X^{-1}, \quad (8.21)$$

where  $(X^{-1}t)_\Delta$  is defined in (7.9). The column sums of  $T$  are the factor shares  $f_1, \dots, f_N$  of the observed inputs and the row sums are the factor shares of the



transformed inputs. Each row of  $T$  gives the composition of the factor share of a transformed input in terms of observed inputs; each column of  $T$  shows the composition of the factor share of an observed input in terms of transformed inputs. For proofs of these results we refer to Appendix B; below we consider an example illustrating these results, after which a comparison with principal components will follow at the end of this section.

### 8.5. An example

We consider a two-input translog specification of (5.12):

$$\log z = \text{constant} + \alpha \log K + \beta \log L + \xi \sqrt{\alpha\beta} \log K \log L, \quad (8.22)$$

where  $K$  is capital,  $L$  is labor, and  $\alpha$ ,  $\beta$ , and  $\xi$  are constants satisfying  $\alpha > 0$ ,  $\beta > 0$ , and  $-1 < \xi < 1$ ; units are chosen so that  $K = L = 1$  holds at the point of minimum input expenditure. Then it may be shown that the  $2 \times 2$  price coefficient matrix  $-\psi\Theta = [-\psi\theta_{ij}]$  of (8.15) equals

$$-\psi\Theta = \frac{-1}{1-\xi^2} \begin{bmatrix} f_K & \xi\sqrt{f_K f_L} \\ \xi\sqrt{f_K f_L} & f_L \end{bmatrix} \begin{array}{l} \text{capital} \\ \text{labor} \end{array}, \quad (8.23)$$

where  $f_K$  is the factor share of capital and  $f_L$  that of labor ( $f_K + f_L = 1$ ). Recall from Section 6 that inputs  $i$  and  $j$  are called specific complements (substitutes) when  $\theta_{ij} = \theta_{ji}$  is positive (negative). Thus, (8.23) combined with  $\psi > 0$  shows that capital and labor are specific complements (substitutes) when  $\xi$  is positive (negative), i.e. when the elasticity of output with respect to either input is an increasing (decreasing) function of the other input [see (8.22)].

The input independence transformation eliminates all specific substitutability and complementarity relations. The mathematical tool is the simultaneous diagonalization (8.16). It may be verified that, for  $-\psi\Theta$  given in (8.23), the matrix

$$X = \frac{1}{\sqrt{2f_K f_L}} \begin{bmatrix} \sqrt{f_L} & -\sqrt{f_L} \\ \sqrt{f_K} & \sqrt{f_K} \end{bmatrix} \quad (8.24)$$

satisfies  $X'FX = I$  and that  $X'(\psi\Theta)X$  is a diagonal matrix whose diagonal elements are  $1/(1-\xi)$  and  $1/(1+\xi)$ . A comparison with (8.16) and (8.17) shows that the own-price elasticities of the transformed inputs are

$$-\psi\lambda_1 = \frac{-1}{1-\xi}, \quad -\psi\lambda_2 = \frac{-1}{1+\xi}. \quad (8.25)$$

Multiple roots occur when  $\xi = 0$ , but this is the uninteresting case in which (8.22) becomes Cobb–Douglas, which is in input independent form and thus requires no transformation.

Substitution of (8.24) into (8.21) yields the composition matrix

$$\mathbf{T} = \frac{1}{2} \begin{bmatrix} f_K + \sqrt{f_K f_L} & f_L + \sqrt{f_K f_L} \\ f_K - \sqrt{f_K f_L} & f_L - \sqrt{f_K f_L} \end{bmatrix} \begin{matrix} (\text{T}_1) \\ (\text{T}_2) \end{matrix} \quad (8.26)$$

(capital)                      (labor)

The column sums are the factor shares of the observed inputs:  $f_K$  for capital and  $f_L$  for labor. The row sums are the factor shares of the transformed inputs:  $\frac{1}{2} + \sqrt{f_K f_L}$  for the input  $\text{T}_1$  corresponding to root  $\lambda_1$  and  $\frac{1}{2} - \sqrt{f_K f_L}$  for  $\text{T}_2$  corresponding to  $\lambda_2$ . The following is a numerical specification of (8.26), bordered by row and column sums, for  $f_K = 0.2$  and  $f_L = 0.8$ :

0.3	0.6	0.9	( $\text{T}_1$ )
-0.1	0.2	0.1	( $\text{T}_2$ )
0.2	0.8	1	
(capital)	(labor)		

Both observed inputs are positively represented in  $\text{T}_1$ , whereas  $\text{T}_2$  is a *contrast* between labor and capital. When the firm buys more  $\text{T}_2$ , its operation becomes more labor-intensive, each dollar spent on  $\text{T}_2$  being equivalent to two dollars worth of labor compensated by one dollar worth of capital services which is given up.

### 8.6. A principal component interpretation

We return to (8.8) with  $\mathbf{Z}'\mathbf{Z}$  interpreted as  $n$  times the matrix of mean squares and products of the values taken by the observed variables. In many applications of the principal component technique, the observed variables have different dimensions (dollars, dollars per year, gallons, etc.). This causes a problem, because principal components change in value when we change the units in which the observed variables are measured. To solve this problem, statisticians frequently standardize these variables by using their standard deviations as units. This amounts to replacing  $\mathbf{Z}'\mathbf{Z}$  in (8.8) by  $\mathbf{D}^{-1/2}\mathbf{Z}'\mathbf{Z}\mathbf{D}^{-1/2}$ , where  $\mathbf{D}$  is the diagonal matrix whose diagonal is identical to that of  $\mathbf{Z}'\mathbf{Z}$ . Thus,  $\lambda_i$  of (8.8) is now obtained from the characteristic equation

$$|\mathbf{D}^{-1/2}\mathbf{Z}'\mathbf{Z}\mathbf{D}^{-1/2} - \lambda_i\mathbf{I}| = 0. \quad (8.27)$$

It is of interest to compare this with

$$|\mathbf{F}^{-1/2}\boldsymbol{\Theta}\mathbf{F}^{-1/2} - \lambda_i\mathbf{I}| = 0, \quad (8.28)$$

which is the characteristic equation associated with (8.20). In both cases, (8.27) and (8.28), we determine a latent root of a symmetric positive definite matrix ( $\mathbf{Z}'\mathbf{Z}$  or  $\boldsymbol{\Theta}$ ) pre- and postmultiplied by a diagonal matrix. However, the diagonal elements of  $\mathbf{F}$  are not identical to those of  $\boldsymbol{\Theta}$ , which is in contrast to  $\mathbf{D}$  and  $\mathbf{Z}'\mathbf{Z}$  in (8.27). The diagonal elements of  $\mathbf{F}$  describe the expenditure *levels* of the inputs (measured as fractions of total expenditure), whereas each diagonal element of  $\boldsymbol{\Theta}$  describes the *change* in the demand for an input caused by a change in its Frisch-deflated price.

Thus, while  $\mathbf{D}$  in (8.27) is directly obtained from  $\mathbf{Z}'\mathbf{Z}$ , the analogous matrix  $\mathbf{F}$  in (8.28) is unrelated to  $\boldsymbol{\Theta}$ . Why do we have this unrelated  $\mathbf{F}$ , which describes expenditure levels, in (8.28) and in the simultaneous diagonalization (8.16)? The answer is that the input independence transformation is subject to the constraint that total input expenditure and its Divisia decomposition remain invariant. We may view this transformation as a *cost-constrained* principal component transformation. Similarly, when the transformation is applied to the consumer demand system (5.5) or to the output supply system (5.26), it is budget-constrained in the first case and revenue-constrained in the second. Such constraints are more meaningful from an economic point of view than the standardization procedure in (8.27).

## 9. The modeling of a disturbance covariance matrix

We mentioned in Section 4 that the disturbance covariance matrix  $\boldsymbol{\Sigma}$  which occurs in the GLS estimator (4.11) is normally unknown and that it is then usually replaced by the sample moment matrix of the LS residuals. Although this approximation is acceptable under certain conditions when the sample is sufficiently large, it is less satisfactory when the number of equations,  $L$  in (4.7) and (4.8), is also large. The reason is that  $\boldsymbol{\Sigma}$  contains many unknowns when  $L$  is large or even moderately large. In fact, the sample moment matrix  $\mathbf{S}$  of the residuals may be singular so that  $\boldsymbol{\Sigma}^{-1}$  in (4.11) cannot be approximated by  $\mathbf{S}^{-1}$ . This situation often occurs in applied econometrics, e.g. in the estimation of a fairly large system of demand equations. One way of solving this problem is by modeling the matrix  $\boldsymbol{\Sigma}$ . Below we describe how this can be performed when the equations are behavioral equations of some decision-maker.

### 9.1. Rational random behavior

Let  $\mathbf{x} = [x_1 \dots x_k]'$  be the vector of variables controlled by this decision-maker. We write  $J$  for the feasible region of  $\mathbf{x}$ ;  $\bar{\mathbf{x}}$  for the optimal value of  $\mathbf{x}$  ( $\bar{\mathbf{x}} \in J$ ); and  $l(\mathbf{x}, \bar{\mathbf{x}})$  for the loss incurred when the decision is  $\mathbf{x}$  rather than  $\bar{\mathbf{x}}$ :

$$l(\mathbf{x}, \bar{\mathbf{x}}) = \begin{cases} 0 & \text{if } \mathbf{x} = \bar{\mathbf{x}}, \\ > 0 & \text{if } \mathbf{x} \neq \bar{\mathbf{x}}. \end{cases} \quad (9.1)$$

We assume that the optimal decision  $\bar{\mathbf{x}}$  depends on numerous factors, some of which are unknown, so that  $\bar{\mathbf{x}}$  is only *theoretically optimal* in the sense that it is optimal under perfect knowledge. The decision-maker can improve on his ignorance by acquiring information. If he does not do this, we describe the decision made as random with a differentiable density function  $p_0(\mathbf{x})$ , to be called the *prior density function*. (The assumption of randomness is justified by the decision-maker's uncertainty as to the factors determining  $\bar{\mathbf{x}}$ .) If he does acquire information,  $p_0(\cdot)$  is transformed into some other density function  $p(\cdot)$  and the amount of information received is defined as

$$I = \int_J p(\mathbf{x}) \log \frac{p(\mathbf{x})}{p_0(\mathbf{x})} dx_1 \dots dx_k, \quad (9.2)$$

which is a concept from statistical information theory [see Theil (1967)]. We write  $c(I)$  for the cost of information and

$$\bar{l} = \int_J l(\mathbf{x}, \bar{\mathbf{x}}) p(\mathbf{x}) dx_1 \dots dx_k \quad (9.3)$$

for the expected loss. If  $c(I)$  and  $\bar{l}$  are measured in the same unit (dollars or any other unit), the natural solution is the decision distribution with density function  $p(\cdot)$  which minimizes  $c(I) + \bar{l}$ . This  $p(\cdot)$  was derived by Barbosa (1975) and the result (see Appendix C) is

$$p(\mathbf{x}) \propto p_0(\mathbf{x}) \exp \left\{ - \frac{l(\mathbf{x}, \bar{\mathbf{x}})}{c'} \right\} \quad \text{if } \mathbf{x} \in J, \quad (9.4)$$

where  $\propto$  means "is proportional to", the proportionality coefficient being independent of  $\mathbf{x}$ , and  $c'$  is the marginal cost of information  $dc/dI$  at the solution (9.4).

Behavior generated by the distribution which density function (9.4) is called *rational random behavior*. This distribution is determined by three factors: the prior density function  $p_0(\mathbf{x})$ , the loss function  $l(\mathbf{x}, \bar{\mathbf{x}})$ , and the marginal cost of

information  $c'$ . For example, let  $x$  be a scalar and the loss function quadratic:  $l(x, \bar{x}) = \frac{1}{2}(x - \bar{x})^2$ . Then, if  $p_0(x)$  equals a constant independent of  $x$  for each  $x \in J$ , (9.4) becomes

$$p(x) \propto \exp\left\{-\frac{1}{2} \frac{(x - \bar{x})^2}{c'}\right\} \quad \text{if } x \in J, \quad (9.5)$$

which is the density function of a truncated normal decision distribution over the interval  $J$ .

## 9.2. The asymptotics of rational random behavior

The case of a small marginal cost of information is of particular interest. Imagine that the prices of the goods and services which the decision-maker buys in order to acquire information decline so that  $c'$  converges to zero; it is shown in Appendix C that the random decision with density function (9.4) then converges in probability to the theoretically optimal decision  $\bar{x}$ . Also, if the loss function has a zero gradient and a symmetric positive definite Hessian matrix  $A$  at  $x = \bar{x}$ ,

$$\frac{\partial}{\partial x} l(x, \bar{x}) = 0 \quad \text{and} \quad \frac{\partial^2}{\partial x \partial x'} l(x, \bar{x}) = A \quad \text{at } x = \bar{x}, \quad (9.6)$$

then as  $c' \rightarrow 0$  the density function  $p(x)$  of (9.4) converges to

$$\frac{1}{(2\pi)^{k/2}} \left| \frac{1}{c'} A \right|^{1/2} \exp\left\{-\frac{1}{2}(x - \bar{x})' \left(\frac{1}{c'} A\right) (x - \bar{x})\right\}, \quad (9.7)$$

which is the density function of the multinormal decision distribution [see (3.3)] with mean vector  $\bar{x}$  and covariance matrix  $c' A^{-1}$ . Note that (9.7) is completely determined by  $c'$  and two characteristics of the loss function: the theoretically optimal decision  $\bar{x}$  at which the loss vanishes [see (9.1)] and the Hessian matrix  $A$  of this function at this point. The relationship between the covariance matrix  $c' A^{-1}$  and the Hessian matrix  $A$  of the loss function enables us to model the disturbance covariance matrix of the decision-maker's behavioral equations; examples will follow in the next subsection.

The prior density function  $p_0(\cdot)$  does not occur in the asymptotic result (9.7). This reflects the fact that when information is cheap in the sense that its marginal cost  $c'$  is small, the decision-maker acquires information to such an extent that his behavior becomes independent of his prior notions. Thus, whereas we obtained (9.5) under the assumption that  $p_0(x)$  is a constant independent of  $x$ , this assumption is unnecessary in the asymptotic case  $c' \rightarrow 0$ . Also, (9.5) is the density

function of a truncated normal distribution, but the truncation becomes irrelevant as  $c' \rightarrow 0$ . The asymptotic version of (9.5) is the univariate normal density function with mean equal to the theoretically optimal decision  $\bar{x}$  and variance equal to the marginal cost of information  $c'$ . The declining variance as  $c' \rightarrow 0$  reflects the attractiveness of a more extensive search for the theoretical optimum when information becomes cheaper.

It is of interest to compare the density function (9.7) of the asymptotic normal decision distribution with the asymptotic normal density of a statistical estimator. In fact, it is not difficult to prove that rational random behavior for small  $c'$  is equivalent to large-sample ML estimation of the theoretically optimal decision  $\bar{x}$ , with a large sample interpreted as a small marginal cost of information. The clue for this equivalence is the similarity of the roles played by the Hessian matrix of the loss function and the information matrix in ML theory.

A second statistical comparison is that with Kadane's (1971) small- $\sigma$  asymptotics, which consists of an asymptotic series of a multiple  $\sigma$  of the variance of the disturbance of a structural equation. If this equation is a behavioral equation of a decision-maker, Kadane's approach is equivalent to the asymptotic version of rational random behavior when we identify  $\sigma$  with  $c'$ .

Another statistical comparison of interest is that with the theorem which states that out of all distributions with range  $(-\infty, \infty)$  and a given mean vector and a given covariance matrix, the multinormal distribution has the largest entropy. The link between this theorem and the normal density function (9.7) is the information definition (9.2); both (9.2) and the entropy are measures from information theory. However, note that the normal density (9.7) is not obtained by imposing a given mean vector and covariance matrix *a priori*. The mean vector and covariance matrix (9.7) are determined by the loss function, apart from the scalar  $c'$ .

Yet another statistical comparison is with Bayesian inference. There is considerable similarity between the exact (i.e., non-asymptotic) result (9.4) and the Bayesian derivation of the posterior density function of a parameter vector. The occurrence of the prior density function on the right in (9.4) provides one similarity. Another is the presence of  $c'$  (which depends on the information acquired) in the exponent of (9.4); this should be compared with the role of the likelihood function (representing the information obtained from the sample) in the Bayesian formula. A third similarity is the disappearance of the prior density function from the asymptotic result (9.7). In Bayesian analysis, too, the posterior density function is dominated by the likelihood function and is no longer affected by the prior density function when the sample is large. All these similarities reflect the fact that rational random behavior and Bayesian inference both describe learning processes based on acquiring information. Nevertheless, the two theories are not equivalent because of the occurrence of the unknown constant  $\bar{x}$  in (9.4). The likelihood function in Bayesian analysis involves no unknown constants; this function is determined by the parameter vector, which is viewed as random, and

the sample, which is viewed as a set of known constants for the derivation of the posterior density function.

### 9.3. Applications to demand and supply

When we apply the theory of rational random behavior to the utility-maximizing consumer or the cost-minimizing firm, we must take into consideration that the criterion function is subject to a constraint (a budget or technology constraint). This can be solved by using the constraint to eliminate one of the goods. The consumer's loss function in (9.4) then involves  $N - 1$  quantities and its derivation from an algebraically specified utility function is straightforward. However, the differential approach provides no such specification so that (9.4) cannot be used; the same holds for the firm because the approach provides no algebraic specification of the production function. But it is possible to use the asymptotic result (9.7) which requires only the theoretically optimal decision and the Hessian matrix of the loss function. The account which follows is therefore based on the asymptotic decision distribution (9.7) rather than (9.4); this also has the advantage of not requiring a specification of the prior density function  $p_0(\cdot)$ .

Consider the input allocation system in the form (6.1),

$$F\kappa = (t'F\kappa)\Theta t - \psi\Theta(I - u'\Theta)\pi + \varepsilon, \quad (9.8)$$

or in scalar form, using  $\Theta t = \theta = [\theta_i]$ ,

$$f_i d(\log q_i) = \theta_i d(\log Q) - \psi \sum_{j=1}^N \theta_{ij} d\left(\log \frac{P_j}{P'}\right) + \varepsilon_i, \quad (9.9)$$

where  $[\varepsilon_i] = \varepsilon$ . The left variable in (9.9) is the  $i$ th decision variable of the firm. The right side, excluding  $\varepsilon_i$ , is the theoretical optimum of this variable, while  $\varepsilon_i$  is the random deviation from this optimum which is predicted by the theory of rational random behavior. Since (9.7) implies normality with a mean equal to the theoretically optimal decision, the  $\varepsilon_i$ 's are multinormal with zero mean. Their covariance matrix (see Appendix C) is

$$\text{cov}(\varepsilon) = \sigma^2(\Theta - \Theta u'\Theta), \quad (9.10)$$

or in scalar form,

$$\text{cov}(\varepsilon_i, \varepsilon_j) = \sigma^2(\theta_{ij} - \theta_i\theta_j), \quad (9.11)$$

where  $\sigma^2$  is a coefficient which is proportional to the marginal cost of information

$c'$ . The covariance matrix (9.10) equals a scalar multiple of the coefficient matrix of  $\pi = [d(\log p_i)]$  in (9.8) so that the covariance (9.11) is proportional to the substitution effect (specific plus general) of a change in the price of  $j$  on the demand for  $i$ .

The above result also holds when (9.7) is applied to the consumer; the only modification required is that  $f_i$  and  $\psi$  in (9.9) become  $w_i$  and  $-\phi$ , respectively [see (5.5)]. Note in particular that the disturbance covariance matrix (9.10) involves only unknown ( $\sigma^2$ ) in addition to  $\Theta$  which already occurs in the systematic part of the equations. Thus, the implications of rational random behavior are quite strong. We obtain even stronger implications for the demand and supply equations of the multiproduct firm; a brief account follows below.

Recall that when the firm makes  $m$  products, the output term of the input demand system (5.15) takes the form (5.25). So, by adding a disturbance  $\varepsilon_i$  we obtain:

$$f_i d(\log q_i) = \gamma \sum_{r=1}^m \theta_r' g_r d(\log z_r) - \psi \sum_{j=1}^N \theta_{ij} d\left(\log \frac{p_j}{p'}\right) + \varepsilon_i. \quad (9.12)$$

Application of (9.7) yields the result that the  $\varepsilon_i$ 's of this input demand system are multinormal with zero means and that their variances and covariances take the same form (9.11) which also holds for the consumer and the single-product firm. Next, by assuming that the firm adjusts its outputs so as to maximize profit, we obtain the output supply system (5.26) which we reproduce with a disturbance  $\varepsilon_r^*$  added:

$$g_r d(\log z_r) = \psi^* \sum_{s=1}^m \theta_{rs}^* d\left(\log \frac{y_s}{p'^s}\right) + \varepsilon_r^*. \quad (9.13)$$

By applying (9.7) we find that the  $\varepsilon_r^*$ 's are multinormal with zero means and the following variance-covariance structure:

$$\text{cov}(\varepsilon_r^*, \varepsilon_s^*) = \frac{\sigma^2 \psi^*}{\gamma \psi} \theta_{rs}^*. \quad (9.14)$$

Since  $\sigma^2$  already occurs in (9.11) for the input demand disturbances, (9.14) provides no further unknowns. In addition, (9.7) implies that *the input demand disturbances* (the  $\varepsilon_i$ 's) *are stochastically independent of the output supply disturbances* (the  $\varepsilon_r^*$ 's). This independence has important implications for statistical inference in demand and supply models; it implies that  $g_r d(\log z_r)$  can be viewed as predetermined in the input demand system (9.12). It is also important for the problem of how to organize the firm in terms of its input and output management, but such matters are beyond the scope of this chapter.



## 10. The Moore–Penrose inverse

A matrix has an inverse only if it is square and nonsingular, but any  $m \times n$  matrix  $A$  of rank  $r$  has a unique Moore–Penrose inverse, written  $A^+$ , which is determined by the following four conditions:

$$AA^+A = A, \quad (10.1)$$

$$A^+AA^+ = A^+, \quad (10.2)$$

$$AA^+ \text{ and } A^+A \text{ are symmetric.} \quad (10.3)$$

It may be verified that these conditions are satisfied by  $A^+ = A^{-1}$  in the special case  $m = n = r$ . Our first objective is to prove that  $A^+$  exists and is unique.<sup>8</sup>

### 10.1. Proof of the existence and uniqueness

The uniqueness of  $A^+$  is established by assuming that conditions (10.1)–(10.3) have two solutions,  $A^+ = B$  and  $A^+ = C$ , and verifying the following 16 steps based on (10.1)–(10.3):

$$\begin{aligned} B &= BAB = B(AB)' = BB'A' = BB'(ACA)' = BB'A'C'A' \\ &= B(AB)'(AC)' = BABAC = BAC = BACAC = (BA)'(CA)'C \\ &= A'B'A'C'C = (ABA)'C'C = A'C'C = (CA)'C = CAC = C. \end{aligned}$$

Therefore,  $B = C$ , which proves that  $A^+$  is unique when it exists.

To prove the existence of  $A^+$  we consider first a zero matrix  $A$  of order  $m \times n$ ; then  $A^+$  equals the  $n \times m$  zero matrix, which may be verified by checking (10.1)–(10.3). Next consider a non-zero matrix  $A$  so that its rank  $r$  is positive. Then  $A'A$  is a symmetric positive semidefinite matrix of order  $n \times n$  and rank  $r$ , and it is possible to express  $A^+$  in terms of the positive latent roots of  $A'A$  and the characteristic vectors associated with these roots. Write  $D$  for the diagonal  $r \times r$  matrix which contains the positive roots of  $A'A$  on the diagonal and  $H$  for an  $n \times r$  matrix whose columns are characteristic vectors corresponding to these roots. Then (7.7) applied to  $A'A$  yields

$$A'A = HDH', \quad (10.4)$$

<sup>8</sup>There are other generalized inverses besides the Moore–Penrose inverse, most of which are obtained by deleting one or more of the four conditions. For example, using (10.1) and (10.2) but deleting (10.3) yields the reflexive generalized inverse, which is not unique; see Laitinen and Theil (1979) for an application of this inverse to consumption theory. Monographs on applications of generalized inverses to statistics include Albert (1972), Ben-Israel and Greville (1974), Boullion and Odell (1971), Pringle and Rayner (1971), and Rao and Mitra (1971).

and the result for  $A^+$  is

$$A^+ = HD^{-1}H'A', \quad (10.5)$$

which is an  $n \times m$  matrix of rank  $r$ .

To verify (10.5) we introduce an  $n \times (n-r)$  matrix  $K$  whose columns are characteristic vectors of  $A'A$  corresponding to the zero roots:

$$A'AK = 0. \quad (10.6)$$

The  $n \times n$  matrix  $[H \ K]$  consists of characteristic vectors of  $A'A$  corresponding to all roots and is therefore an orthogonal matrix, which can be expressed in two ways. Premultiplying  $[H \ K]$  by its transpose and equating the product to the unit matrix yields

$$H'H = I, \quad K'K = I, \quad H'K = 0, \quad (10.7)$$

while postmultiplying  $[H \ K]$  by its transpose and equating the product to the unit matrix gives

$$HH' + KK' = I. \quad (10.8)$$

The verification of (10.5) is now a matter of checking conditions (10.1)–(10.3). Premultiplying (10.5) by  $A$  yields  $AA^+ = AHD^{-1}H'A'$ , which is symmetric. Next we postmultiply (10.5) by  $A$ ,  $A^+A = HD^{-1}H'A'A$ , and hence in view of (10.4) and (10.7),  $A^+A = HD^{-1}H'HDH' = HH'$ , which is also symmetric. We postmultiply this by (10.5):

$$A^+AA^+ = HH'HD^{-1}H'A' = HD^{-1}H'A' = A^+,$$

which confirms (10.2). Finally, we postmultiply  $AA^+ = AHD^{-1}H'A'$  by  $A$ :

$$AA^+A = AHD^{-1}H'A'A = AHD^{-1}H'HDH' = AHH' = A.$$

To verify the last step,  $AHH' = A$ , we premultiply (10.6) by  $K'$ , which gives  $(AK)'AK = 0$  or  $AK = 0$ . Therefore,  $AKK' = 0$  so that premultiplication of (10.8) by  $A$  yields  $AHH' = A$ .

## 10.2. Special cases

If  $A$  has full column rank so that  $(A'A)^{-1}$  exists,  $A^+ = (A'A)^{-1}A'$ , which may either be verified from (10.4) and (10.5) for  $r = n$  or by checking (10.1)–(10.3). We may thus write the LS coefficient vector in (2.8) as  $b = X^+y$ , which may be

viewed as an extension of  $\mathbf{b} = \mathbf{X}^{-1}\mathbf{y}$  in the special case of a square non-singular  $\mathbf{X}$  (as many regressors as observations).

If  $\mathbf{A}$  is a symmetric  $n \times n$  matrix of rank  $r$ , then

$$\mathbf{A} = \sum_{i=1}^r \lambda_i \mathbf{x}_i \mathbf{x}_i', \quad \mathbf{A}^+ = \sum_{i=1}^r \frac{1}{\lambda_i} \mathbf{x}_i \mathbf{x}_i', \quad (10.9)$$

where  $\lambda_1, \dots, \lambda_r$  are the non-zero latent roots of  $\mathbf{A}$  and  $\mathbf{x}_1, \dots, \mathbf{x}_r$  are characteristic vectors associated with these roots. Also,  $\mathbf{A}\mathbf{x}_i = \mathbf{0}$  and  $\mathbf{A}^+\mathbf{x}_i = \mathbf{0}$  for  $i = r+1, \dots, n$ , where  $\mathbf{x}_{r+1}, \dots, \mathbf{x}_n$  are characteristic vectors of  $\mathbf{A}$  corresponding to the zero roots. Thus, if  $\mathbf{A}$  is symmetric,  $\mathbf{A}^+$  has characteristic vectors identical to those of  $\mathbf{A}$ , the same number of zero roots, and non-zero roots equal to the reciprocals of the non-zero roots of  $\mathbf{A}$ . The verification of these results is again a matter of checking (10.1)–(10.3) and using  $\mathbf{x}_i' \mathbf{x}_j = \delta_{ij}$ . Since a symmetric idempotent matrix such as  $\mathbf{M}$  in (2.11) has only zero and unit roots, it thus follows as a corollary that such a matrix is equal to its own Moore–Penrose inverse.

### 10.3. A generalization of Aitken's theorem

We return to the linear model (2.7), reproduced here:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}. \quad (10.10)$$

As before, we assume that  $\mathbf{X}$  is an  $n \times K$  matrix of rank  $K$  consisting of non-stochastic elements and that  $\boldsymbol{\varepsilon}$  has zero expectation, but we now assume that the covariance matrix of  $\boldsymbol{\varepsilon}$  takes the singular form  $\sigma^2 \mathbf{V}$ , the  $n \times n$  matrix  $\mathbf{V}$  having rank  $r < n$ . Hence, the Aitken estimator (2.13) does not exist, but it seems reasonable to ask whether

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{V}^+\mathbf{X})^{-1} \mathbf{X}'\mathbf{V}^+ \mathbf{y} \quad (10.11)$$

exists and is a best linear unbiased estimator of  $\boldsymbol{\beta}$ . It will appear that each of these properties (the existence and the best linear unbiasedness) requires a special condition involving both  $\mathbf{V}$  and  $\mathbf{X}$ .

The matrix  $\mathbf{V}$  is comparable to  $\mathbf{A}'\mathbf{A}$  in (10.4) and (10.6) in that both are symmetric positive semidefinite  $n \times n$  matrices of rank  $r$ . Therefore, we can apply (10.4) and (10.6) to  $\mathbf{V}$  rather than  $\mathbf{A}'\mathbf{A}$ :

$$\mathbf{V} = \mathbf{H}\mathbf{D}\mathbf{H}', \quad (10.12)$$

$$\mathbf{V}\mathbf{K} = \mathbf{0}, \quad (10.13)$$

where  $\mathbf{D}$  is now the  $r \times r$  diagonal matrix with the positive latent roots of  $\mathbf{V}$  on the diagonal,  $\mathbf{H}$  is an  $n \times r$  matrix whose columns are characteristic vectors of  $\mathbf{V}$  corresponding to these roots, and  $\mathbf{K}$  is an  $n \times (n-r)$  matrix consisting of characteristic vectors of  $\mathbf{V}$  that correspond to the zero roots. The results (10.7) and (10.8) are also valid in the present interpretation. In addition, (10.9) and (10.12) imply

$$\mathbf{V}^+ = \mathbf{H}\mathbf{D}^{-1}\mathbf{H}'. \quad (10.14)$$

Our strategy, similar to that of the proof of Aitken's theorem in Section 7, will be to premultiply (10.10) by an appropriate matrix so that the transformed disturbance vector has a scalar covariance matrix. We select  $\mathbf{D}^{-1/2}\mathbf{H}'$ , where  $\mathbf{D}^{-1/2}$  is the diagonal matrix with the reciprocals of the positive square roots of the diagonal elements of  $\mathbf{D}$  in the diagonal:

$$\mathbf{D}^{-1/2}\mathbf{H}'\mathbf{y} = (\mathbf{D}^{-1/2}\mathbf{H}'\mathbf{X})\boldsymbol{\beta} + \mathbf{D}^{-1/2}\mathbf{H}'\boldsymbol{\varepsilon}. \quad (10.15)$$

The covariance matrix of  $\mathbf{D}^{-1/2}\mathbf{H}'\boldsymbol{\varepsilon}$  is

$$\mathfrak{E}(\mathbf{D}^{-1/2}\mathbf{H}'\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}'\mathbf{H}\mathbf{D}^{-1/2}) = \sigma^2\mathbf{D}^{-1/2}\mathbf{H}'\mathbf{V}\mathbf{H}\mathbf{D}^{-1/2} = \sigma^2\mathbf{I},$$

where the last step is based on  $\mathbf{H}'\mathbf{V}\mathbf{H} = \mathbf{D}$ , which is obtained by premultiplying (10.12) by  $\mathbf{H}'$  and postmultiplying by  $\mathbf{H}$  and using  $\mathbf{H}'\mathbf{H} = \mathbf{I}$  [see (10.7)]. Since  $\mathbf{D}^{-1/2}\mathbf{H}'\boldsymbol{\varepsilon}$  thus has a scalar covariance matrix, let us apply LS to (10.15). Assuming that  $\mathbf{H}'\mathbf{X}$  and hence  $\mathbf{D}^{-1/2}\mathbf{H}'\mathbf{X}$  have full column rank, we find the following estimator of  $\boldsymbol{\beta}$ :

$$(\mathbf{D}^{-1/2}\mathbf{H}'\mathbf{X})^+ \mathbf{D}^{-1/2}\mathbf{H}'\mathbf{y} = (\mathbf{X}'\mathbf{H}\mathbf{D}^{-1}\mathbf{H}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{H}\mathbf{D}^{-1}\mathbf{H}'\mathbf{y}. \quad (10.16)$$

This is indeed identical to (10.11) in view of (10.14).

Two considerations are important for the appraisal of this procedure. First, we assumed that  $\mathbf{H}'\mathbf{X}$  has full column rank; if the rank is smaller, the matrix product in parentheses on the right in (10.16) is singular so that (10.11) does not exist. Therefore, a necessary and sufficient condition for the existence of the estimator (10.11) is that  $\mathbf{H}'\mathbf{X}$  have maximum rank, where  $\mathbf{H}$  consists of  $r$  characteristic vectors of  $\mathbf{V}$  corresponding to the positive roots. Secondly, we obtained (10.15) by premultiplying (10.10) by  $\mathbf{D}^{-1/2}\mathbf{H}'$ , which reduces the number of observations from  $n$  to  $r$ . We can recover the "missing"  $n-r$  observations by premultiplication by  $\mathbf{K}'$ , yielding  $\mathbf{K}'\mathbf{y} = \mathbf{K}'\mathbf{X}\boldsymbol{\beta} + \mathbf{K}'\boldsymbol{\varepsilon}$ . The covariance matrix of  $\mathbf{K}'\boldsymbol{\varepsilon}$  is  $\sigma^2\mathbf{K}'\mathbf{V}\mathbf{K} = \mathbf{0}$  [see (10.13)] so that  $\mathbf{K}'\boldsymbol{\varepsilon}$  vanishes with unit probability. Therefore,

$$\mathbf{K}'\mathbf{y} = \mathbf{K}'\mathbf{X}\boldsymbol{\beta}, \quad (10.17)$$

which amounts to a linear constraint on  $\boldsymbol{\beta}$  unless  $\mathbf{K}'\mathbf{X} = \mathbf{0}$ .

To clarify this situation, consider the following example for  $K = 1$ ,  $n = 3$ , and  $r = 2$ :

$$\mathbf{X} = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}, \quad \mathbf{V} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \quad \mathbf{H} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix}, \quad \mathbf{K} = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}. \quad (10.18)$$

Here  $\mathbf{X}$  has full column rank but  $\mathbf{H}'\mathbf{X} = \mathbf{0}$  so that the matrix product in parentheses on the right in (10.16) is singular; in fact, the underlying equation (10.15) does not contain  $\beta$  at all when  $\mathbf{H}'\mathbf{X} = \mathbf{0}$ . Thus, the estimator (10.11) does not exist, but in the case of (10.18) it is nevertheless possible to determine  $\beta$  (a scalar in this case) exactly! The reason is that (10.18) implies  $\mathbf{K}'\mathbf{y} = y_3$  and  $\mathbf{K}'\mathbf{X} = 1$  so that (10.17) states that  $y_3$  equals the parameter. Ultimately, this results from the zero value of the third diagonal element of  $\mathbf{V}$  in (10.18) and the non-zero third element of  $\mathbf{X}$ .

Under the assumptions stated in the discussion following eq. (10.10), the estimator (10.11) exists when  $\mathbf{H}'\mathbf{X}$  has full column rank and it is a best linear unbiased estimator of  $\beta$  when  $\mathbf{K}'\mathbf{X} = \mathbf{0}$  [so that (10.17) is not a real constraint on  $\beta$ ]. A proof of the latter statement follows in the next paragraph. If  $\mathbf{K}'\mathbf{X}$  is a non-zero matrix, (10.17) is a linear constraint on  $\beta$  which should be incorporated in the estimation procedure; see Theil (1971, sec. 6.8).

We can write any linear estimator of  $\beta$  as

$$\tilde{\beta} = [\mathbf{A} + (\mathbf{X}'\mathbf{V}^+\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^+] \mathbf{y}, \quad (10.19)$$

where  $\mathbf{A}$  is some  $K \times n$  matrix consisting of non-stochastic elements. By substituting  $\mathbf{X}\beta + \varepsilon$  for  $\mathbf{y}$  in (10.19) and taking the expectation we find that the unbiasedness of  $\tilde{\beta}$  requires

$$\mathbf{A}\mathbf{X} = \mathbf{0}, \quad (10.20)$$

so that  $\tilde{\beta} - \beta = [\mathbf{A} + (\mathbf{X}'\mathbf{V}^+\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^+] \varepsilon$  and the covariance matrix of  $\tilde{\beta}$  equals

$$\mathcal{V}(\tilde{\beta}) = \sigma^2 [\mathbf{A} + (\mathbf{X}'\mathbf{V}^+\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^+] \mathbf{V} [\mathbf{A}' + \mathbf{V}^+\mathbf{X}(\mathbf{X}'\mathbf{V}^+\mathbf{X})^{-1}]. \quad (10.21)$$

For  $\mathbf{A} = \mathbf{0}$  we have  $\tilde{\beta} = \hat{\beta}$  in view of (10.19). Thus, using  $\mathbf{V}^+\mathbf{V}\mathbf{V}^+ = \mathbf{V}^+$  and (10.21), we obtain:

$$\mathcal{V}(\hat{\beta}) = \sigma^2 (\mathbf{X}'\mathbf{V}^+\mathbf{X})^{-1}, \quad (10.22)$$

which is a generalization of (2.14). The excess of (10.21) over (10.22) equals a multiple  $\sigma^2$  of  $\mathbf{A}\mathbf{V}\mathbf{A}' + \mathbf{A}\mathbf{V}\mathbf{V}^+\mathbf{X}(\mathbf{X}'\mathbf{V}^+\mathbf{X})^{-1} + (\mathbf{X}'\mathbf{V}^+\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^+\mathbf{V}\mathbf{A}'$ . But  $\mathbf{A}\mathbf{V}\mathbf{V}^+\mathbf{X} = \mathbf{0}$  so that  $\mathcal{V}(\tilde{\beta}) - \mathcal{V}(\hat{\beta}) = \sigma^2 \mathbf{A}\mathbf{V}\mathbf{A}'$ , which is positive semidefinite and thus

establishes that  $\hat{\beta}$  is best linear unbiased. To verify that  $AVV^+X$  is a zero matrix we use (10.12) and (10.14) in

$$VV^+ = HDH'HD^{-1}H' = HH' = I - KK',$$

where the last two steps are based on (10.7) and (10.8). So, using (10.20) and  $K'X = \mathbf{0}$  also, we have

$$AVV^+X = AX - AKK'X = \mathbf{0} - \mathbf{0} = \mathbf{0}.$$

The matrix  $\mathcal{V}(\hat{\beta}) - \mathcal{V}(\tilde{\beta}) = \sigma^2AVA'$  is obviously zero when we select  $A = \mathbf{0}$ , but it may also be zero for  $A \neq \mathbf{0}$  when  $V$  is singular, which suggests that there is no unique best linear unbiased estimator of  $\beta$ . This is not true, however; if the estimator (10.11) exists, i.e. if  $H'X$  has full column rank, it is the unique best linear unbiased estimator of  $\beta$  when  $K'X = \mathbf{0}$ . The reason is that  $AVA' = \mathbf{0}$  is equivalent to  $\mathcal{E}[A\epsilon(A\epsilon)'] = \mathbf{0}$  so that  $A\epsilon$  is a zero vector with unit probability. Using (10.20) also, we obtain  $Ay = A(X\beta + \epsilon) = \mathbf{0}$ , which in conjunction with (10.19) shows that the best linear unbiased estimator of  $\beta$  must be of the form (10.11), even though  $A$  may be a non-zero matrix.

#### 10.4. Deleting an equation from an allocation model

The Moore–Penrose inverse can also be conveniently used to prove that when we estimate an  $N$ -equation allocation system such as (6.1), we can simply delete one of the  $N$  equations (it does not matter which). The clue is the fact that each equation can be obtained by adding the  $N - 1$  others. We prove this below for an allocation system which is linear in the parameters. The strategy of the proof will be to start with GLS estimation of  $N - 1$  equations with a non-singular disturbance covariance matrix, followed by adding the deleted equation (so that the disturbance covariance matrix becomes singular), and then proving that the resulting estimator (10.11) is identical to the original GLS estimator.

We can formulate the problem in the following more general way. Let  $y = X\beta + \epsilon$  have a non-singular covariance matrix  $\mathcal{V}(\epsilon) = \sigma^2V$  of order  $n \times n$ . We premultiply by a matrix  $B$  of order  $(n + n') \times n$  and rank  $n$ :

$$By = BX\beta + B\epsilon. \tag{10.23}$$

For example, take  $B' = [I \ C]$ , which means that we add to the original  $n$  observations  $n'$  linear combinations of these observations. The covariance matrix of  $B\epsilon$  takes the singular form  $\sigma^2BVB'$ . Thus, the matrix  $V$  of the previous subsection becomes  $BVB'$  here, while  $X$  becomes  $BX$ . We conclude that condition

$K'X = \mathbf{0}$  is now  $K'(BX) = \mathbf{0}$ , where  $K$  is a matrix whose  $n'$  columns are characteristic vectors of  $BVB'$  corresponding to the zero roots:  $(BVB')K = \mathbf{0}$  and  $K'K = I$ . Evidently, a sufficient condition for  $K$  is  $B'K = \mathbf{0}$  and  $K'K = I$ . Such a  $K$  can be obtained as a matrix whose columns are characteristic vectors of the idempotent matrix  $I - B(B'B)^{-1}B'$  corresponding to the unit roots:

$$\left[ I - B(B'B)^{-1}B' \right] K = K.$$

The GLS estimator (10.11) of  $\beta$  in (10.23) is then

$$\left[ X'B'(BVB')^+ BX \right]^{-1} X'B'(BVB')^+ By. \quad (10.24)$$

This is identical to  $(X'V^{-1}X)^{-1}X'V^{-1}y$ , and hence to the GLS estimator obtained from the original  $n$  observations, because

$$B'(BVB')^+ B = V^{-1},$$

which follows from  $BVB'(BVB')^+BVB' = BVB'$  [see (10.1)] premultiplied by  $V^{-1}(B'B)^{-1}B'$  and postmultiplied by  $B(B'B)^{-1}V^{-1}$ . It is unnecessary to check the condition that  $H'(BX)$  has full column rank,  $H$  being a matrix whose  $n$  columns are characteristic vectors of  $BVB'$  corresponding to the positive roots. The reason is that the estimator (10.24) would not exist if the condition were not satisfied, whereas we know that (10.24) equals  $(X'V^{-1}X)^{-1}X'V^{-1}y$ .

### Appendix A: Linear independence and related topics

Consider a matrix  $V = [v_1 \dots v_n]$  and a linear combination  $Vc$  of its  $n$  columns. The vectors  $v_1, \dots, v_n$  are said to be *linearly independent* if  $Vc = \mathbf{0}$  implies  $c = \mathbf{0}$ , i.e. if there exists no non-trivial linear combination of the  $v_i$ 's which is a zero vector. For example, the columns of the  $2 \times 2$  unit matrix are linearly independent because

$$c_1 \begin{bmatrix} 1 \\ 0 \end{bmatrix} + c_2 \begin{bmatrix} 0 \\ 1 \end{bmatrix} = \begin{bmatrix} c_1 \\ c_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \quad \text{implies} \quad c_1 = c_2 = 0,$$

but  $v_1 = [1 \ 0]'$  and  $v_2 = [2 \ 0]'$  are not linearly independent because  $c_1 v_1 + c_2 v_2 = \mathbf{0}$  if (for example)  $c_1 = 2$  and  $c_2 = -1$ .

For any  $m \times n$  matrix  $A$  the column rank is defined as the largest number of linearly independent columns, and the row rank as the largest number of linearly independent rows. It can be shown that these two ranks are always equal; we can thus speak about the *rank*  $r$  of  $A$ , which obviously satisfies  $r \leq m, n$ . If all

columns (rows) of  $A$  are linearly independent,  $A$  is said to have full column (row) rank. For any  $A$ , the ranks of  $A$ ,  $A'$ ,  $A'A$ , and  $AA'$  are all equal. Also, the rank of  $AB$  is at most equal to the rank of  $A$  and that of  $B$ . For example,

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

which illustrates that the rank of  $AB$  may be smaller than both that of  $A$  and that of  $B$ . (A zero matrix has zero rank.) If  $A$  is square ( $n \times n$ ) and has full rank ( $r = n$ ), it is called *non-singular* and its inverse  $A^{-1}$  exists.

For any vector  $v = [v_i]$ , its *length* is defined as the positive square root of  $v'v = \sum_i v_i^2$ . If  $v'v = 1$ ,  $v$  is said to have *unit length*. The *inner product* of two vectors  $v = [v_i]$  and  $w = [w_i]$  consisting of the same number of elements is defined as  $v'w = \sum_i v_i w_i$ . If  $v'w = 0$ ,  $v$  and  $w$  are called *orthogonal vectors*.

A square matrix  $X$  which satisfies  $X' = X^{-1}$  is called an *orthogonal matrix*. Premultiplication of  $X' = X^{-1}$  by  $X$  gives  $XX' = I$ , which shows that each row of  $X$  has unit length and that any two rows of  $X$  are orthogonal vectors. Postmultiplication of  $X' = X^{-1}$  by  $X$  gives  $X'X = I$  so that each column of  $X$  (each row of  $X'$ ) also has unit length and any two columns of  $X$  are also orthogonal vectors.

## Appendix B: The independence transformation

The independence transformation is based on three axioms, the first being the invariance of total expenditure. Let a dollar spent on observed input  $j$  result in  $r_{ij}$  dollars spent on transformed input  $i$ , so that the expenditure on  $i$  equals  $\sum_j r_{ij} p_j q_j$  and the total expenditure on all transformed inputs equals  $\sum_j (\sum_i r_{ij}) p_j q_j$ , which must be identical to  $\sum_j p_j q_j$  because of the invariance postulated. Therefore,  $\sum_i r_{ij} = 1$  for each  $j$ , or

$$t'R = t', \quad (\text{B.1})$$

where  $R = [r_{ij}]$ . By dividing the expenditure  $\sum_j r_{ij} p_j q_j$  on transformed input  $i$  by total expenditure  $C$  (which is invariant) we obtain the factor share  $f_{Ti}$  of this input. Therefore,  $f_{Ti} = \sum_j r_{ij} f_j$ , or

$$F_T t = R F t, \quad (\text{B.2})$$

where  $F_T$  is the diagonal factor share matrix of the transformed inputs.

The second axiom states that the logarithmic price and quantity changes of the transformed inputs are linear combinations of their observed counterparts,  $\pi_T = S_1 \pi$  and  $\kappa_T = S_2 \kappa$ , so that the associated Divisia indexes are invariant. The Divisia



volume index is  $d(\log Q) = \iota' F \kappa$  and its transformed counterpart is  $\iota' F_T \kappa_T = \iota' F_T S_2 \kappa = \iota' F (R' S_2) \kappa$  [see (B.2)]. Thus, the invariance of this index requires  $R' S_2 = I$  or  $S_2 = (R')^{-1}$ . We can proceed similarly for the price index  $\iota' F \pi$ , which yields the same result for  $S_1$ , so that the price and quantity transformations use the same matrix,  $\pi_T = S \pi$  and  $\kappa_T = S \kappa$ , where  $S = (R')^{-1}$ . See remark (3) below for the case of a singular  $R$ .

The third axiom diagonalizes  $\Theta$ . We premultiply (8.15) by  $R$ , which yields  $R F \kappa = R F R' S \kappa = R F R' \kappa_T$  on the left because  $R' S = I$  and  $S \kappa = \kappa_T$ . When we proceed similarly on the right and use (B.1) also, we obtain:

$$R F R' \kappa_T = (\iota' F \kappa) (R \Theta R') \iota - \psi R \Theta R' [I - \iota' (R \Theta R')] \pi_T, \quad (\text{B.3})$$

which is an allocation system of the same form as (8.15), with logarithmic price and quantity changes  $\pi_T$  and  $\kappa_T$ , provided  $R F R'$  on the left equals the diagonal factor share matrix  $F_T$ . The new normalized price coefficient matrix is  $R \Theta R'$ , which occurs in the same three places in (B.3) as  $\Theta$  does in (8.15). [The matrix  $R \Theta R'$  is indeed normalized because  $\iota' R \Theta R' \iota = \iota' \Theta \iota = 1$  follows from (B.1).] Therefore,  $R F R' = F_T$  and  $R \Theta R' = \text{diagonal}$  are the conditions under which (B.3) is an input independent allocation system. These are two conditions on  $R$ , which must satisfy (B.1) also.

We proceed to prove that

$$R = (X^{-1} \iota)_\Delta X' \quad (\text{B.4})$$

satisfies these three conditions, with  $X$  defined in (8.16) and  $(X^{-1} \iota)_\Delta$  in (7.9). First,  $\iota' R = \iota'$  is true for (B.4) in view of (7.10). Secondly,  $R F R' = (X^{-1} \iota)_\Delta X' F X (X^{-1} \iota)_\Delta = (X^{-1} \iota)_\Delta^2$  [see (8.16)] so that

$$R F R' = F_T = (X^{-1} \iota)_\Delta^2 = \text{diagonal}. \quad (\text{B.5})$$

Thirdly, using  $\Theta = (X')^{-1} \Lambda X^{-1}$  [see (8.16)], we have  $R \Theta R' = (X^{-1} \iota)_\Delta^2 \Lambda$ , which is diagonal. So, using (B.5) also and premultiplying (B.3) by  $(R F R')^{-1} = (X^{-1} \iota)_\Delta^{-2}$ , we obtain:

$$\kappa_T = (\iota' F \kappa) \Lambda \iota - \psi \Lambda (I - \iota' R \Theta R') \pi_T, \quad (\text{B.6})$$

which is the matrix version of (8.17). The expression which is subtracted in parentheses in the substitution term of (B.6) represents the deflation by the Frisch price index, which is invariant. To prove this we note that the marginal share vector of the transformed inputs equals  $R \Theta \iota = R \theta$  in view of the real-income term in (B.3) and  $R' \iota = \iota$ ; the invariance of the Frisch index then follows from  $(R \theta)' \pi_T = \theta' R' S \pi = \theta' \pi$ .

The expenditure on transformed input  $i$  equals  $r_{ij}p_jq_j$  dollars insofar as it originates with observed input  $j$ . By dividing this amount by total expenditure  $C$  we obtain  $r_{ij}f_j$ , which is thus the factor share of transformed input  $i$  insofar as it originates with observed input  $j$ . This  $r_{ij}f_j$  is an element of the matrix  $\mathbf{RF}$ , to be written  $\mathbf{T}$ :

$$\mathbf{T} = \mathbf{RF} = (\mathbf{X}^{-1}\iota)_{\Delta}\mathbf{X}^{-1}, \quad (\text{B.7})$$

where the last step is based on (B.4) and  $\mathbf{F} = (\mathbf{X}')^{-1}\mathbf{X}^{-1}$  [see (8.16)]. Postmultiplication of (B.7) by  $\iota$  gives  $\mathbf{T}\iota = \mathbf{RF}\iota = \mathbf{F}_{\tau}\iota$  [see (B.2)]; hence the row sums of  $\mathbf{T}$  are the factor shares of the transformed inputs. Also,  $\iota'\mathbf{T} = \iota'\mathbf{RF} = \iota'\mathbf{F}$ , so that the column sums of  $\mathbf{T}$  are the factor shares of the observed inputs. Note that (B.7) and its row and column sums confirm the results on the composition matrix. Note further that  $\mathbf{F} = (\mathbf{X}')^{-1}\mathbf{X}^{-1}$  and  $\Theta = (\mathbf{X}')^{-1}\Lambda\mathbf{X}^{-1}$  [see (8.16)] imply that the price elasticity matrix  $-\psi\mathbf{F}^{-1}\Theta$  in (8.19) equals  $-\psi\mathbf{X}\Lambda\mathbf{X}^{-1}$ . So, using (B.7) also, we have

$$\mathbf{T}(-\psi\mathbf{F}^{-1}\Theta) = -\psi(\mathbf{X}^{-1}\iota)_{\Delta}\Lambda\mathbf{X}^{-1} = -\psi\Lambda(\mathbf{X}^{-1}\iota)_{\Delta}\mathbf{X}^{-1} = -\psi\Lambda\mathbf{T}.$$

Combining the first and last member yields  $\iota'_i(-\psi\mathbf{F}^{-1}\Theta) = -\psi\lambda_i\iota'_i$ , where  $\iota'_i$  is the  $i$ th row of  $\mathbf{T}$ , or

$$\iota'_i[-\psi\mathbf{F}^{-1}\Theta - (-\psi\lambda_i)\mathbf{I}] = 0.$$

Therefore, each row of the composition matrix is a characteristic row vector of the (asymmetric) price elasticity matrix of the observed inputs.

We conclude with the following remarks.

(1) Although the solution (B.4) satisfies all three conditions, it is not unique. However, it may be shown that this solution is unique up to premultiplication by an arbitrary permutation matrix; such a multiplication affects only the order in which the transformed inputs are listed.

(2) We proved in the second paragraph that the price and quantity transformations take the form  $\pi_{\tau} = \mathbf{S}\pi$  and  $\kappa_{\tau} = \mathbf{S}\kappa$ , where  $\mathbf{S} = (\mathbf{R}')^{-1}$ . It thus follows from (B.1) that  $\mathbf{S}^{-1}\iota = \iota$  or  $\mathbf{S}\iota = \iota$ . Therefore, when the prices of the observed inputs change proportionately,  $\pi$  being a scalar multiple  $k$  of  $\iota$ , the price of each transformed input changes in the same proportion:  $\pi_{\tau} = \mathbf{S}(k\iota) = k\mathbf{S}\iota = k\iota$ . The quantities have the same desirable property.

(3) It follows from (B.4) that  $\mathbf{R}$  is singular when  $(\mathbf{X}^{-1}\iota)_{\Delta}$  contains a zero diagonal element, and from (B.5) that this implies a zero factor share of one of the transformed inputs. In that case  $\mathbf{S} = (\mathbf{R}')^{-1}$  does not exist. The simplest way to interpret this situation is by means of a perturbation of the firm's technology so

that the  $i$ th element of  $X^{-1}\mathbf{t}$  converges from a small non-zero value to zero. It may be shown that  $d(\log p_{Ti})$  then increases beyond bounds. If the increase is toward  $\infty$ , transformed input  $i$  is priced out of the market; if it is toward  $-\infty$ ,  $i$  becomes a free good; in both cases no money is spent on  $i$  in the limit. In particular, if (5.12) represents a homothetic technology,  $N-1$  elements of  $X^{-1}\mathbf{t}$  are zero and all observed inputs collectively behave as one transformed input with unitary Divisia elasticity; no money is spent on any transformed input whose Divisia elasticity differs from 1. For proofs of these results see Theil (1977).

(4) The independence transformation was first formulated by Brooks (1970) and axiomatically justified by Theil (1975–76, ch. 12) for a finite-change version of the consumer demand system (5.22). The  $\lambda_i$ 's are then income elasticities of transformed consumer goods. Rossi (1979a) proved that when all observed goods are specific substitutes, the transformed good with the smallest income elasticity represents all observed goods positively and that all other transformed goods are contrasts between observed goods similar to  $T_2$  in (8.26). The former transformed good serves to satisfy the consumer's wants associated with the observed goods in the least luxurious manner; this result is of particular interest when the transformation is applied to a group of goods which satisfy similar wants such as different brands of the same type of commodity.<sup>9</sup> For an integrated exposition of the independence transformation in consumption and production theory see Theil (1980, ch. 10–11).

### Appendix C: Rational random behavior

To verify (9.4) we write  $p^*(\mathbf{x}) = p(\mathbf{x}) + \delta f(\mathbf{x})$  for some density function other than the  $p(\mathbf{x})$  of (9.4), where  $\delta$  is independent of  $\mathbf{x}$  so that  $f(\cdot)$  must satisfy

$$\int_J f(\mathbf{x}) dx_1 \dots dx_k = 0. \quad (\text{C.1})$$

The information  $I^*$  and the expected loss  $\bar{l}^*$  associated with  $p^*(\cdot)$  are

$$I^* = \int_J [p(\mathbf{x}) + \delta f(\mathbf{x})] \log \frac{p(\mathbf{x}) + \delta f(\mathbf{x})}{p_0(\mathbf{x})} dx_1 \dots dx_k, \quad (\text{C.2})$$

$$\bar{l}^* = \bar{l} + \delta \int_J l(\mathbf{x}, \bar{\mathbf{x}}) f(\mathbf{x}) dx_1 \dots dx_k, \quad (\text{C.3})$$

<sup>9</sup>When  $\Theta$  is block-diagonal, so is  $X$  in (8.16), which means that the independence transformation can be applied to each block separately. We have a block-diagonal  $\Theta$  under block independence. See the end of Section 6 for block independent inputs; the extension to block independent consumer goods is straightforward.

where  $\bar{l}$  is the expected loss (9.3) associated with the  $p(\cdot)$  of (9.4). We apply a Taylor expansion to (C.2) as a function of  $\delta$ :

$$I^* = I + k_1\delta + \frac{1}{2}k_2\delta^2 + O(\delta^3), \quad (\text{C.4})$$

where  $I$  is the information (9.2) associated with (9.4) and

$$k_1 = \int_J f(\mathbf{x}) \log \frac{p(\mathbf{x})}{p_0(\mathbf{x})} dx_1 \dots dx_k, \quad (\text{C.5})$$

$$k_2 = \int_J \frac{[f(\mathbf{x})]^2}{p(\mathbf{x})} dx_1 \dots dx_k. \quad (\text{C.6})$$

Next we apply a Taylor expansion to  $c(I^*)$ , writing  $c' = dc/dI$  and  $c'' = d^2c/dI^2$  for the derivatives of  $c(\cdot)$  at the  $I$  of (9.4):

$$c(I^*) = c(I) + \delta k_1 c' + \frac{1}{2} \delta^2 (k_2 c' + k_1^2 c'') + O(\delta^3)$$

and we add this to (C.3):

$$\begin{aligned} c(I^*) + \bar{l}^* &= c(I) + \bar{l} + \delta \left[ k_1 c' + \int_J l(\mathbf{x}, \bar{\mathbf{x}}) f(\mathbf{x}) dx_1 \dots dx_k \right] \\ &\quad + \frac{1}{2} \delta^2 (k_2 c' + k_1^2 c'') + O(\delta^3) \end{aligned} \quad (\text{C.7})$$

For  $c(I) + \bar{l}$  to be minimal we require the coefficient of  $\delta$  in (C.7) to vanish for any  $f(\cdot)$  satisfying (C.1) and that of  $\delta^2$  to be positive. The latter condition is satisfied when  $c' > 0$  and  $c'' \geq 0$  (a positive nondecreasing marginal cost of information) because (C.6) implies  $k_2 > 0$  when  $f(\mathbf{x}) \neq 0$  for some  $\mathbf{x}$ . It follows from (C.5) that the former condition amounts to a zero value of

$$\int_J \left[ c' \log \frac{p(\mathbf{x})}{p_0(\mathbf{x})} + l(\mathbf{x}, \bar{\mathbf{x}}) \right] f(\mathbf{x}) dx_1 \dots dx_k.$$

This integral vanishes, given (C.1), when the expression in brackets is a constant independent of  $\mathbf{x}$ , which yields (9.4) directly.

To prove the asymptotic results for small  $c'$  we take the logarithm of (9.4):

$$\log p(\mathbf{x}) = \text{constant} + \log p_0(\mathbf{x}) - \frac{l(\mathbf{x}, \bar{\mathbf{x}})}{c'}, \quad (\text{C.8})$$

and substitute  $\bar{\mathbf{x}}$  for  $\mathbf{x}$ , using (9.1):

$$\log p(\bar{\mathbf{x}}) = \text{constant} + \log p_0(\bar{\mathbf{x}}).$$

Since the constants in these equations are equal, subtraction yields

$$\log \frac{p(\bar{x})}{p(x)} = \log \frac{p_0(\bar{x})}{p_0(x)} + \frac{l(x, \bar{x})}{c'}. \quad (\text{C.9})$$

It follows from (9.1) that as  $c' \rightarrow 0$  the last term increases beyond bounds for any  $x \neq \bar{x}$ , so that the same holds for  $p(\bar{x})/p(x)$  on the left. Hence, as  $c' \rightarrow 0$  the density  $p(x)$  becomes zero for each  $x \neq \bar{x}$  and the random decision with density function (9.4) thus converges in probability to  $\bar{x}$ .

To verify the asymptotic distribution (9.7), we define

$$v = \frac{1}{\sqrt{c'}}(x - \bar{x}), \quad (\text{C.10})$$

so that  $l(x, \bar{x}) = l(\bar{x} + \sqrt{c'}v, \bar{x})$ . We apply a Taylor expansion to  $l(x, \bar{x})/c'$ , using (9.6):

$$\begin{aligned} \frac{l(x, \bar{x})}{c'} &= \frac{1}{c'} \left[ \frac{1}{2}(\sqrt{c'}v)'A(\sqrt{c'}v) + O(c'^{3/2}) \right] \\ &= \frac{1}{2}v'Av + O(\sqrt{c'}). \end{aligned} \quad (\text{C.11})$$

We assume that  $p_0(x)$  is positive and differentiable around  $\bar{x}$ . Hence, we can apply a Taylor expansion to  $\log p_0(x)$  and write it as  $\log p_0(\bar{x})$  plus a linear remainder term in  $x - \bar{x}$ . Therefore, in view of (C.10),

$$\log p_0(x) = \log p_0(\bar{x}) + O(\sqrt{c'}),$$

which in conjunction with (C.8) and (C.11) shows that  $\log p(x)$  equals a constant minus  $\frac{1}{2}v'Av$  plus two remainder terms which both converge to zero as  $c' \rightarrow 0$ . The result (9.7) is then obtained by substitution from (C.10) for  $v$  in  $\frac{1}{2}v'Av$ .

We obtain (9.11) from (9.7) by using the budget or technology constraint to eliminate one of the decision variables from the criterion function. Let these variables be the quantities bought by the consumer; it was shown by Theil (1975–76, sec. 2.6–2.7) that (9.7) then yields variances and covariances of the form

$$\text{cov}(q_i, q_j) = -k \left( \lambda u^{ij} - \frac{\lambda}{\partial \lambda / \partial M} \frac{\partial q_i}{\partial M} \frac{\partial q_j}{\partial M} \right), \quad (\text{C.12})$$

where  $k > 0$  is proportional to the marginal cost of information  $c'$ . A comparison of (C.12) with (3.12) shows that  $\text{cov}(q_i, q_j)$  is proportional to the substitution

component (specific plus general) of  $\partial q_i / \partial p_j$ . We obtain (9.11) from (C.12) by rearrangements required by the left variable in (5.5).

The results (9.11) and (9.14) for the multiproduct firm and the stochastic independence of the input demand disturbances and the output supply disturbances were derived by Laitinen and Theil (1978). Reference should also be made to Bowman et al. (1979) and to Rossi (1979b, 1979c) for a comparison of rational random behavior and search theory.

## References

- Aitken, A. C. (1935) "On Least Squares and Linear Combination of Observations", *Proceedings of the Royal Society of Edinburgh*, 55, 42–48.
- Albert, A. (1972) *Regression and the Moore–Penrose Pseudoinverse*. New York: Academic Press.
- Barbosa, F. de H. (1975) "Rational Random Behavior: Extensions and Applications", Doctoral dissertation, The University of Chicago.
- Barten, A. P. (1964) "Consumer Demand Functions under Conditions of Almost Additive Preferences", *Econometrica*, 32, 1–38.
- Barten, A. P. (1977) "The Systems of Consumer Demand Functions Approach: A Review", *Econometrica*, 45, 23–51.
- Barten, A. P. and E. Geyskens (1975) "The Negativity Condition in Consumer Demand", *European Economic Review*, 6, 227–260.
- Bellman, R. (1960) *Introduction to Matrix Analysis*. New York: McGraw-Hill Book Company.
- Ben-Israel, A. and T. N. E. Greville (1974) *Generalized Inverses: Theory and Applications*. New York: John Wiley and Sons.
- Boullion, T. L. and P. L. Odell (1971) *Generalized Inverse Matrices*. New York: John Wiley and Sons.
- Bowman, J. P., K. Laitinen and H. Theil (1979) "New Results on Rational Random Behavior", *Economics Letters*, 2, 201–204.
- Brooks, R. B. (1970) "Diagonalizing the Hessian Matrix of the Consumer's Utility Function", Doctoral dissertation, The University of Chicago.
- Brown, A. and A. Deaton (1972) "Surveys in Applied Economics: Models of Consumer Behaviour", *Economic Journal*, 82, 1145–1236.
- Dhrymes, P. J. (1978) *Mathematics for Econometrics*. New York: Springer-Verlag.
- Divisia, F. (1925) "L'indice monétaire et la théorie de la monnaie", *Revue d'Economie Politique*, 39, 980–1008.
- Frisch, R. (1932) *New Methods of Measuring Marginal Utility*. Tübingen: J. C. B. Mohr.
- Goldberger, A. S. (1959) *Impact Multipliers and Dynamic Properties of the Klein–Goldberger Model*. Amsterdam: North-Holland Publishing Company.
- Graybill, F. A. (1969) *Introduction to Matrices with Applications in Statistics*. Belmont, Cal.: Wadsworth Publishing Company.
- Hadley, G. (1961) *Linear Algebra*. Reading, Mass.: Addison-Wesley Publishing Company.
- Hall, R. E. (1973) "The Specification of Technology with Several Kinds of Output", *Journal of Political Economy*, 81, 878–892.
- Houthakker, H. S. (1960) "Additive Preferences", *Econometrica*, 28, 244–257; Errata, 30 (1962) 633.
- Kadane, J. B. (1971) "Comparison of  $k$ -Class Estimators when the Disturbances are Small", *Econometrica*, 39, 723–737.
- Laitinen, K. (1980) *A Theory of the Multiproduct Firm*. Amsterdam: North-Holland Publishing Company.
- Laitinen, K. and H. Theil (1978) "Supply and Demand of the Multiproduct Firm", *European Economic Review*, 11, 107–154.
- Laitinen, K. and H. Theil (1979) "The Antonelli Matrix and the Reciprocal Slutsky Matrix", *Economics Letters*, 3, 153–157.

- Phlips, L. (1974) *Applied Consumption Analysis*. Amsterdam: North-Holland Publishing Company.
- Pringle, R. M. and A. A. Rayner (1971) *Generalized Inverse Matrices with Applications to Statistics*. London: Charles Griffin and Co.
- Rao, C. R. and S. K. Mitra (1971) *Generalized Inverse of Matrices and Its Applications*. New York: John Wiley and Sons.
- Rossi, P. E. (1979a) "The Independence Transformation of Specific Substitutes and Specific Complements", *Economics Letters*, 2, 299–301.
- Rossi, P. E. (1979b) "The Cost of Search and Rational Random Behavior", *Economics Letters*, 3, 5–8.
- Rossi, P. E. (1979c) "Asymptotic Search Behavior Based on the Weibull Distribution", *Economics Letters*, 3, 211–213.
- Theil, H. (1967) *Economics and Information Theory*. Amsterdam: North-Holland Publishing Company.
- Theil, H. (1971) *Principles of Econometrics*. New York: John Wiley and Sons.
- Theil, H. (1975–76) *Theory and Measurement of Consumer Demand*, 2 vols. Amsterdam: North-Holland Publishing Company.
- Theil, H. (1977) "The Independent Inputs of Production", *Econometrica*, 45, 1303–1327.
- Theil, H. (1980) *The System-Wide Approach to Microeconomics*. Chicago: University of Chicago Press.
- Theil, H. and K. W. Clements (1980) "Recent Methodological Advances in Economic Equation Systems", *American Behavioral Scientist*, 23, 789–809.
- Theil, H. and K. Laitinen (1979) "Maximum Likelihood Estimation of the Rotterdam Model under Two Different Conditions", *Economics Letters*, 2, 239–244.
- Zellner, A. (1962) "An Efficient Method of Estimating Seemingly Unrelated Regressions and Tests for Aggregation Bias", *Journal of the American Statistical Association*, 57, 348–368.