

Matrix Calculus & Zero-One Matrices

STATISTICAL AND
ECONOMETRIC
APPLICATIONS



Darrell A. Turkington

0037422

Matrix Calculus and Zero-One Matrices

Statistical and Econometric Applications

DARRELL A. TURKINGTON

University of Western Australia



CAMBRIDGE
UNIVERSITY PRESS

PUBLISHED BY THE PRESS SYNDICATE OF THE UNIVERSITY OF CAMBRIDGE
The Pitt Building, Trumpington Street, Cambridge, United Kingdom

CAMBRIDGE UNIVERSITY PRESS

The Edinburgh Building, Cambridge CB2 2RU, UK
40 West 20th Street, New York, NY 10011-4211, USA
10 Stamford Road, Oakleigh, VIC 3166, Australia
Ruiz de Alarcón 13, 28014 Madrid, Spain
Dock House, The Waterfront, Cape Town 8001, South Africa

<http://www.cambridge.org>

© Darrell A. Turkington 2002

This book is in copyright. Subject to statutory exception
and to the provisions of relevant collective licensing agreements,
no reproduction of any part may take place without
the written permission of Cambridge University Press.

First published 2002

Printed in the United States of America

Typeface Times New Roman 10/12 pt. System L^AT_EX 2_ε [TB]

A catalog record for this book is available from the British Library.

Library of Congress Cataloging in Publication Data

Turkington, Darrell A.

Matrix calculus and zero-one matrices : statistical and econometric applications /
Darrell A. Turkington.

p. cm.

Includes bibliographical references and index.

ISBN 0-521-80788-3

1. Matrices. 2. Mathematical statistics. I. Title.

QA188 .T865 2001

512.9'434 – dc21

2001025614

ISBN 0 521 80788 3 hardback

Preface

This book concerns itself with the mathematics behind the application of classical statistical procedures to econometric models. I first tried to apply such procedures in 1983 when I wrote a book with Roger Bowden on instrumental variable estimation. I was impressed with the amount of differentiation involved and the difficulty I had in recognizing the end product of this process. I thought there must be an easier way of doing things. Of course at the time, like most econometricians, I was blissfully unaware of matrix calculus and the existence of zero-one matrices. Since then several books have been published in these areas showing us the power of these concepts. See, for example Graham (1981), Magnus (1988), Magnus and Neudecker (1999), and Lutkepohl (1996).

This present book arose when I set myself two tasks: first, to make myself a list of rules of matrix calculus that were most useful in applying classical statistical procedures to econometrics; second, to work out the basic building blocks of such procedures – the score vector, the information matrix, and the Cramer–Rao lower bound – for a sequence of econometric models of increasing statistical complexity. I found that the mathematics involved working with operators that were generalizations of the well-known vec operator, and that a very simple zero-one matrix kept cropping up. I called the matrix a shifting matrix for reasons that are obvious in the book. Its basic nature is illustrated by the fact that all Toeplitz circulant matrices can be written as linear combinations of shifting matrices.

The book falls naturally into two parts. The first part outlines the classical statistical procedures used throughout the work and aims at providing the reader with the mathematical tools needed to apply these procedures to econometric models. The statistical procedures are dealt with in Chap. 1. Chapter 2 deals with elements of matrix algebra. In this chapter, generalized vec and devec operators are defined and their basic properties investigated. Chapter 3 concerns itself with zero-one matrices. Well-known zero-one matrices such as commutation matrices, elimination matrices, and duplication matrices are defined and their properties listed. Several new zero-one matrices are introduced in this

chapter. Explicit expressions are given for the generalized vec and devec of the commutation matrix, and the properties of these matrices are investigated in several theorems. Shifting matrices are defined and the connection among these matrices and Toeplitz and circulant matrices is explained. Moreover, the essential role they play in time-series processes is demonstrated. Chapter 4 is devoted to matrix calculus. The approach taken in this chapter is to derive the matrix calculus results from a few basic rules that are generalizations of the chain rule and product rule of ordinary calculus. Some of these results are new, involving as they do generalized vecs of commutation matrices. A list of useful rules is given at the end of the chapter.

The second part of the book is designed to illustrate how the mathematical tools discussed in the preceding chapters greatly facilitate the application of classical statistical procedures to econometric models in that they speed up the difficult differentiation involved and help in the required asymptotic work.

In all, nine linear statistical models are considered. The first three models (Chap. 5) are based on the linear-regression model: the basic model, the linear-regression model with autoregressive disturbances, and the linear-regression model with moving-average disturbances. The next three models (Chap. 6) are based on the seemingly unrelated regression equations (SURE) model: the basic model, the SURE model with vector autoregressive disturbances, and the SURE model with vector moving-average disturbances. The final three models (Chap. 7) are based on the linear simultaneous equations (LSE) model. We consider the basic LSE model and the two variations that come about when we assume vector autoregressive or vector moving-average disturbances.

For each model considered, the basic building blocks of classical statistics are obtained: the score vector, the information matrix, and the Cramer-Rao lower bound. Statistical analysis is then conducted with these concepts. Where possible, econometric estimators of the parameters of primary interest that achieve the Cramer-Rao lower bound are discussed. Iterative interpretations of the maximum-likelihood estimators that link them with the econometric estimators are presented. Classical test statistics for hypotheses of interest are obtained.

The models were chosen in such a way as to form a sequence of models of increasing statistical complexity. The reader can then see, for example, how the added complication changes the information matrix or the Cramer-Rao lower bound. There are, in fact, two such sequences in operation. We have in Chap. 5, for example, the basic linear-regression model followed by versions of this model with more complicated disturbance structures. Second, between chapters, we have sequences of models with the same characteristics assigned to the disturbances: for example, the linear-regression model with autoregressive disturbances followed by the SURE model and the LSE model with vector autoregressive disturbances.

It is assumed that the reader has a good working knowledge of matrix algebra, basic statistics, and classical econometrics and is familiar with standard asymptotic theory. As such, the book should be useful for graduate students in econometrics and for practicing econometricians. Statisticians interested in how their procedures apply to other fields may also be attracted to this work.

Several institutions should be mentioned in this preface: first, my home university, the University of Western Australia, for allowing me time off from teaching to concentrate on the manuscript; second, the University of Warwick and the University of British Columbia for providing me with stimulating environments at which to spend my sabbaticals. At Warwick I first became interested in matrix calculus; at British Columbia I put the finishing touches to the manuscript.

Several individuals must also be thanked: my teacher Tom Rothenberg, to whom I owe an enormous debt; Adrian Pagan, for his sound advice; Jan Magnus, for introducing me to the intricacies of zero-one matrices; my colleagues Les Jennings, Michael McAleer, Shiqing Ling, and Jakob Madsen for their helpful suggestions and encouragement; Helen Reidy for her great patience and skill in typing the many drafts of this work; finally, my family, Sonia, Joshua, and Nikola, for being there for me.

Matrix Calculus and Zero-One Matrices

This book presents the reader with mathematical tools taken from matrix calculus and zero-one matrices and demonstrates how these tools greatly facilitate the application of classical statistical procedures to econometric models. The matrix calculus results are derived from a few basic rules that are generalizations of the rules of ordinary calculus. These results are summarized in a useful table. Well-known zero-one matrices, together with some new ones, are defined, their mathematical roles explained, and their useful properties presented.

The basic building blocks of classical statistics, namely, the score vector, the information matrix, and the Cramer-Rao lower bound, are obtained for a sequence of linear econometric models of increasing statistical complexity. From these are obtained interactive interpretations of maximum likelihood estimators, linking them with efficient econometric estimators. Classical test statistics are also derived and compared for hypotheses of interest.

Darrell A. Turkington is Professor in the Department of Economics at the University of Western Australia, Perth. He is coauthor with Roger J. Bowden of the widely cited work *Instrumental Variables* (1984) in the Econometric Society Monographs series published by Cambridge University Press. Professor Turkington has published in leading international journals such as the *Journal of the American Statistical Association*, the *International Economic Review*, and the *Journal of Econometrics*. He has held visiting positions at the University of California, Berkeley, the University of Warwick, and the University of British Columbia.

Contents

1. Classical Statistical Procedures	1
1.1 Introduction	1
1.2 The Fisher Vector, the Information Matrix, and the Cramer-Rao Lower Bound	2
1.3 Maximum Likelihood and Estimational Test Procedures	7
1.4 Nullness Parameters	1
1.5 Differentiation and Asymptotics	3
2. Elements of Matrix Algebra	7
2.1 Introduction	7
2.2 Kronecker Products	7
2.3 The Vec and the Dvec Operators	10
2.4 Kronecker Vec and Dvec Operators	14
2.5 Kronecker Matrices and Block Matrices	23
3. Zero-One Matrices	28
3.1 Introduction	28
3.2 Selection Matrices and Permutation Matrices	29
3.3 The Commutation Matrix and Generalized Vec and Dvec of the Commutation Matrix	29
3.4 Elimination Matrices E_i	43
3.5 Duplication Matrices D_i	43
3.6 Results Concerning Zero-One Matrices Associated with \mathbf{A}	44
3.7 Shifting Matrices	46
4. Matrix Calculus	67
4.1 Introduction	67
4.2 The \mathbf{C} Polynomials	68
4.3 Some Simple Matrix Calculus Results	69

Contents

<i>Preface</i>	ix
1 Classical Statistical Procedures	1
1.1 Introduction	1
1.2 The Score Vector, the Information Matrix, and the Cramer–Rao Lower Bound	2
1.3 Maximum Likelihood Estimators and Test Procedures	2
1.4 Nuisance Parameters	4
1.5 Differentiation and Asymptotics	5
2 Elements of Matrix Algebra	7
2.1 Introduction	7
2.2 Kronecker Products	7
2.3 The Vec and the Devec Operators	10
2.4 Generalized Vec and Devec Operators	14
2.5 Triangular Matrices and Band Matrices	23
3 Zero-One Matrices	28
3.1 Introduction	28
3.2 Selection Matrices and Permutation Matrices	29
3.3 The Commutation Matrix and Generalized Vecs and Devecs of the Commutation Matrix	29
3.4 Elimination Matrices L , \bar{L}	43
3.5 Duplication Matrices D , \bar{L}'	43
3.6 Results Concerning Zero-One Matrices Associated with an $n \times n$ Matrix	44
3.7 Shifting Matrices	46
4 Matrix Calculus	67
4.1 Introduction	67
4.2 Basic Definitions	68
4.3 Some Simple Matrix Calculus Results	69

4.4	Matrix Calculus and Zero-One Matrices	69
4.5	The Chain Rule and the Product Rule for Matrix Calculus	71
4.6	Rules for Vecs of Matrices	73
4.7	Rules Developed from the Properties of K_{Gn} , $K_{Gn}^{\tau_n}$, and $K_{nG}^{\tau_n}$	78
4.8	Rules for Scalar Functions of a Matrix	80
4.9	Tables of Results	83
5	Linear-Regression Models	87
5.1	Introduction	87
5.2	The Basic Linear-Regression Model	88
5.3	The Linear-Regression Model with Autoregressive Disturbances	90
5.4	Linear-Regression Model with Moving-Average Disturbances	100
	Appendix 5.A Probability Limits Associated with the Information Matrix for the Autoregressive Disturbances Model	107
	Appendix 5.B Probability Limits Associated with the Information Matrix for the Moving-Average Disturbances Model	108
6	Seemingly Unrelated Regression Equations Models	110
6.1	Introduction	110
6.2	The Standard SURE Model	110
6.3	The SURE Model with Vector Autoregressive Disturbances	117
6.4	The SURE Model with Vector Moving-Average Disturbances	132
	Appendix 6.A Probability Limits Associated with the Information Matrix of the Model with Moving-Average Disturbances	142
7	Linear Simultaneous Equations Models	146
7.1	Introduction	146
7.2	The Standard Linear Simultaneous Equations Model	148
7.3	The Linear Simultaneous Equations Model with Vector Autoregressive Disturbances	161
7.4	The Linear Simultaneous Equations Model with Vector Moving-Average Disturbances	184
	Appendix 7.A Consistency of \tilde{R} and $\tilde{\Sigma}$	195
	<i>References and Suggested Readings</i>	197
	<i>Index</i>	201

1 Classical Statistical Procedures

1.1. INTRODUCTION

An alternative title to this book could have been *The Application of Classical Statistical Procedures to Econometrics* or something along these lines. What it purports to do is provide the reader with mathematical tools that facilitate the application of classical statistical procedures to the complicated statistical models that we are confronted with in econometrics. It then demonstrates how these procedures can be applied to a sequence of linear econometric models, each model being more complicated statistically than the previous one. The statistical procedures I have in mind are these centered around the likelihood function: procedures that involve the score vector, the information matrix, and the Cramer-Rao lower bound, together with maximum-likelihood estimation and classical test statistics.

Until recently, such procedures were little used by econometricians. The likelihood function in most econometric models is complicated, and the first-order conditions for maximizing this function usually give rise to a system of nonlinear equations that is not easily solved. As a result, econometricians developed their own class of estimators, instrumental variable estimators, that had the same asymptotic properties as those of maximum-likelihood estimators (MLEs) but were far more tractable mathematically [see Bowden and Turkington (1990)]. Nor did econometricians make much use of the prescribed classical statistical procedures for obtaining test statistics for the hypotheses of interest in econometric models; rather, test statistics were developed on an ad hoc basis.

All that changed in the last couple of decades, when there was renewed interest by econometricians in maximum-likelihood procedures and in developing Lagrangian multiplier test (LMT) statistics. One reason for this change was the advent of large, fast computers. A complicated system of nonlinear equations could now be solved so we would have in hand the maximum-likelihood estimates even though we had no algebraic expression for the underlying estimators. Another more recent explanation for this change in attitude is the

advent of results on zero-one matrices and matrix calculus. Works by Graham (1981), Magnus (1988), Magnus and Neudecker (1988), and Lutkepohl (1996) have shown us the importance of zero-one matrices, their connection to matrix calculus, and the power of matrix calculus particularly with respect to applying classical statistical procedures.

In this introductory chapter, I have a brief and nonrigorous summary of the classical statistical procedures that are used extensively in the latter part of this book.

1.2. THE SCORE VECTOR, THE INFORMATION MATRIX, AND THE CRAMER-RAO LOWER BOUND

Let θ be a $k \times 1$ vector of unknown parameters associated with a statistical model and let $l(\theta)$ be the log-likelihood function that satisfies certain regularity conditions and is twice differentiable. Let $\partial l / \partial \theta$ denote the $k \times 1$ vector of partial derivatives of l . Then $\partial l / \partial \theta$ is called the **score vector**. Let $\partial^2 l / \partial \theta \partial \theta'$ denote the $k \times k$ Hessian matrix of $l(\theta)$. Then the (asymptotic) **information matrix** is defined as

$$I(\theta) = - \lim_{n \rightarrow \infty} \frac{1}{n} E(\partial^2 l / \partial \theta \partial \theta'),$$

where n denotes the sample size. Now the limit of the expectation need not be the same as the probability limit. However, for the models we consider in this book, based as they are on the multivariate normal distribution, the two concepts will be the same. As a result it is often more convenient to regard the information matrix as

$$I(\theta) = -p \lim_{n \rightarrow \infty} \frac{1}{n} \partial^2 l / \partial \theta \partial \theta'.$$

The inverse of this matrix, $I^{-1}(\theta)$, is called the (asymptotic) **Cramer-Rao lower bound**. Let $\hat{\theta}$ be a consistent estimator of θ such that

$$\sqrt{n}(\hat{\theta} - \theta) \xrightarrow{d} N(0, V).$$

The matrix V is called the asymptotic covariance matrix of $\hat{\theta}$. Then V exceeds the Cramer-Rao lower bound $I^{-1}(\theta)$ in the sense that $V - I^{-1}(\theta)$ is a positive-semidefinite matrix. If $V = I^{-1}(\theta)$, then $\hat{\theta}$ is called a best asymptotically normally distributed estimator (which is shortened to BAN estimator).

1.3. MAXIMUM LIKELIHOOD ESTIMATORS AND TEST PROCEDURES

Classical statisticians prescribed a procedure for obtaining a BAN estimator, namely the maximum-likelihood procedure. Let \oplus denote the parameter space. Then any value of θ that maximizes $l(\theta)$ over \oplus is called a maximum-likelihood

estimate, and the underlying estimator is called the MLE. The first-order conditions for this maximization are given by

$$\frac{\partial l(\theta)}{\partial \theta} = 0.$$

Let $\tilde{\theta}$ denote the MLE of θ . Then $\tilde{\theta}$ is consistent, and $\tilde{\theta}$ is the BAN estimator so

$$\sqrt{n}(\tilde{\theta} - \theta) \xrightarrow{d} N[0, I^{-1}(\theta)].$$

Let h be a $G \times 1$ vector whose elements are functions of the elements of θ . We denote this by $h(\theta)$. Suppose we are interested in developing test statistics for the null hypothesis

$$H_O : h(\theta) = 0$$

against the alternative

$$H_A : h(\theta) \neq 0.$$

Let $\tilde{\theta}$ denote the MLE of θ and $\bar{\theta}$ denote the constrained MLE of θ ; that is, $\bar{\theta}$ is the MLE of θ we obtain after we impose H_O on our statistical model. Let $\partial h(\theta)/\partial \theta$ denote the $k \times G$ matrix whose (ij) element is $\partial h_j/\partial \theta_i$. Then classical statisticians prescribed three competing procedures for obtaining a test statistic for H_O . These are as follows.

LAGRANGIAN MULTIPLIER TEST STATISTIC

$$T_1 = \frac{1}{n} \frac{\partial l(\bar{\theta})'}{\partial \theta} I^{-1}(\bar{\theta}) \frac{\partial l(\bar{\theta})}{\partial \theta}.$$

Note that the LMT statistic uses the constrained MLE of θ . If H_O is true, $\bar{\theta}$ should be close to $\tilde{\theta}$ and as, by the first-order conditions, $\partial l(\tilde{\theta})/\partial \theta = 0$, the derivative $\partial l(\theta)/\partial \theta$ evaluated at $\bar{\theta}$ should also be close to the null vector. The test statistic is a measure of the distance $\partial l(\bar{\theta})/\partial \theta$ is from the null vector.

WALD TEST STATISTIC

$$T_2 = nh(\tilde{\theta})' \left[\frac{\partial h(\tilde{\theta})'}{\partial \theta} I^{-1}(\tilde{\theta}) \frac{\partial h(\tilde{\theta})}{\partial \theta} \right]^{-1} h(\tilde{\theta}).$$

Note that the Wald test statistic uses the (unconstrained) MLE of θ . Essentially it is based on the asymptotic distribution of $\sqrt{nh}(\tilde{\theta})$ under H_O , the statistic itself measuring the distance $h(\tilde{\theta})$ is from the null vector.

LIKELIHOOD RATIO TEST STATISTIC

$$T_3 = 2[l(\tilde{\theta}) - l(\bar{\theta})].$$

Note that the likelihood ratio test (LRT) statistic uses both the unconstrained MLE $\tilde{\theta}$ and the constrained MLE $\bar{\theta}$. If H_O is indeed true, it should not matter whether we impose it or not, so $l(\bar{\theta})$ should be approximately the same as $l(\tilde{\theta})$. The test statistic T_3 measures the difference between $l(\tilde{\theta})$ and $l(\bar{\theta})$.

All three test statistics are asymptotically equivalent in the sense that, under H_0 , they all have the same limiting χ^2 distribution and under H_A , with local alternatives, they have the same limiting noncentral χ^2 distribution. Usually imposing the null hypothesis on our model leads to a simpler statistical model, and thus the constrained MLEs $\bar{\theta}$ are more obtainable than the $\tilde{\theta}$ MLEs. For this reason the LMT statistic is often the easiest statistic to form. Certainly it is the one that has been most widely used in econometrics.

1.4. NUISANCE PARAMETERS

Let us now partition θ into $\theta = (\alpha' \beta')'$, where α is a $k_1 \times 1$ vector of parameters of primary interest and β is a $k_2 \times 1$ vector of nuisance parameters, $k_1 + k_2 = k$. The terms used here do not imply that the parameters in β are unimportant to our statistical model. Rather, they indicate that the purpose of our analysis is to make statistical inference about the parameters in α instead of those in β .

In this situation, two approaches can be taken. First, we can derive the information matrix $I(\theta)$ and the Cramer–Rao lower bound $I^{-1}(\theta)$. Let

$$I(\theta) = \begin{bmatrix} I_{\alpha\alpha} & I_{\alpha\beta} \\ I_{\beta\alpha} & I_{\beta\beta} \end{bmatrix},$$

$$I^{-1}(\theta) = \begin{pmatrix} I^{\alpha\alpha} & I^{\alpha\beta} \\ I^{\beta\alpha} & I^{\beta\beta} \end{pmatrix}$$

be these matrices partitioned according to our partition of θ . As far as α is concerned we can now work with $I_{\alpha\alpha}$ and $I^{\alpha\alpha}$ in place of $I(\theta)$ and $I^{-1}(\theta)$, respectively. For example, $I^{\alpha\alpha}$ is the Cramer–Rao lower bound for the asymptotic covariance matrix of a consistent estimator of α . If $\tilde{\alpha}$ is the MLE of α , then

$$\sqrt{n}(\tilde{\alpha} - \alpha) \xrightarrow{d} N(0, I^{\alpha\alpha}),$$

and so on.

A particular null hypothesis that has particular relevance for us is

$$H_0 : \alpha = 0$$

against

$$H_A : \alpha \neq 0.$$

Under this first approach, the classical test statistics for this null hypothesis would be the following test statistics.

LAGRANGIAN TEST STATISTIC

$$T_1 = \frac{1}{n} \frac{\partial l(\bar{\theta})}{\partial \alpha} I^{\alpha\alpha}(\bar{\theta}) \frac{\partial l(\bar{\theta})}{\partial \alpha}.$$

WALD TEST STATISTIC

$$T_2 = n\tilde{\alpha}' I^{\alpha\alpha}(\tilde{\theta})^{-1} \tilde{\alpha}.$$

LIKELIHOOD RATIO TEST STATISTIC

$$T_3 = 2[l(\tilde{\theta}) - l(\bar{\theta})].$$

Under H_0 all three test statistics would have a limiting χ^2 distribution with k_1 degrees of freedom, and the nature of the tests insists that we use the upper tail of this distribution to find the appropriate critical region.

The second approach is to work with the concentrated log-likelihood function. Here we undertake a stepwise maximization of the log-likelihood function. We first maximize $l(\theta)$ with respect to the nuisance parameters β to obtain $\beta = \bar{\beta}(\alpha)$, say. The vector $\bar{\beta}$ is then placed back in the log-likelihood function to obtain

$$\bar{l}(\alpha) = l[\alpha, \bar{\beta}(\alpha)].$$

The function $\bar{l}(\alpha)$ is called the concentrated likelihood function. Our analysis can now be reworked with $\bar{l}(\alpha)$ in place of $l(\theta)$.

For example, let

$$\bar{l} = -p \lim \frac{1}{n} \frac{\partial \bar{l}}{\partial \alpha \partial \alpha'},$$

and let $\hat{\alpha}$ be any consistent estimator of α such that

$$\sqrt{n}(\hat{\alpha} - \alpha) \xrightarrow{d} N(0, V_\alpha).$$

Then $V_\alpha \geq \bar{l}^{-1}$ in the sense that their difference is a positive-semidefinite matrix. If $\tilde{\alpha}$ is the MLE of α , then $\tilde{\alpha}$ is obtained from

$$\frac{\partial \bar{l}}{\partial \alpha} = 0,$$

$$\sqrt{n}(\tilde{\alpha} - \alpha) \xrightarrow{d} N(0, \bar{l}^{-1}),$$

and so on. As far as test procedures go for the null hypothesis $H_0 : h(\alpha) = 0$, under this second approach we rewrite the test statistics by using \bar{l} and \bar{l} in place of $l(\theta)$ and $I(\theta)$, respectively. In this book, I largely use the first approach as one of my expressed aims is to achieve the complete information matrix $I(\theta)$ for a sequence of econometric models.

1.5. DIFFERENTIATION AND ASYMPTOTICS

Before we leave this brief chapter, note that classical statistical procedures involve us in much differentiation. The score vector $\partial l / \partial \theta$, the Hessian matrix $\partial^2 l / \partial \theta \partial \theta'$, and $\partial h / \partial \theta$ all involve working out partial derivatives. It is at this stage that difficulties can arise in applying these procedures to econometric

models. As hinted at in Section 1.2, the log-likelihood function $l(\theta)$ for most econometric models is a complicated function, and it is no trivial matter to obtain the derivatives required in our application. Usually it is too great a task for ordinary calculus. Although in some cases it can be done, [see, for example, Rothenberg and Leenders (1964)], what often happens when one attempts to do the differentiation by using ordinary calculus is that one is confronted with a hopeless mess. It is precisely this problem that has motivated the writing of this book. I hope that it will go some way toward alleviating it.

It is assumed that the reader is familiar with standard asymptotic theory. Every attempt has been made to make the rather dull but necessary asymptotic analysis in this book as readable as possible. Only the probability limits of the information matrices that are required in our statistical analysis are worked out in full. The probability limits themselves are assumed to exist – a more formal mathematical analysis would give a list of sufficient conditions needed to ensure this. Finally, as already noted, use is made of the shortcut notation

$$\sqrt{n}(\tilde{\beta} - \beta) \xrightarrow{d} N(0, V)$$

rather than the more formally correct notation

$$\sqrt{n}(\tilde{\beta} - \beta) \xrightarrow{d} x \sim N(0, V).$$

2 Elements of Matrix Algebra

2.1. INTRODUCTION

In this chapter, we consider matrix operators that are used throughout the book and special square matrices, namely triangular matrices and band matrices, that will crop up continually in our future work. From the elements of an $m \times n$ matrix, $A = (a_{ij})$ and a $p \times q$ matrix, $B = (b_{ij})$, the Kronecker product forms an $mp \times nq$ matrix. The vec operator forms a column vector out of a given matrix by stacking its columns one underneath the other. The devec operator forms a row vector out of a given matrix by stacking its rows one alongside the other. In like manner, a generalized vec operator forms a new matrix from a given matrix by stacking a certain number of its columns under each other and a generalized devec operator forms a new matrix by stacking a certain number of rows alongside each other. It is well known that the Kronecker product is intimately connected with the vec operator, but we shall see that this connection also holds for the devec and generalized operators as well. Finally we look at special square matrices with zeros above or below the main diagonal or whose nonzero elements form a band surrounded by zeros. The approach I have taken in this chapter, as indeed in several other chapters, is to list, without proof, well-known properties of the mathematical concept, in hand. If, however, I want to present a property in a different light or if I have something new to say about the concept, then I will give a proof.

2.2. KRONECKER PRODUCTS

Let $A = (a_{ij})$ be an $m \times n$ matrix and B a $p \times q$ matrix. The $mp \times nq$ matrix given by

$$\begin{bmatrix} a_{11}B & \dots & a_{1n}B \\ \vdots & & \vdots \\ a_{m1}B & \dots & a_{mn}B \end{bmatrix}$$