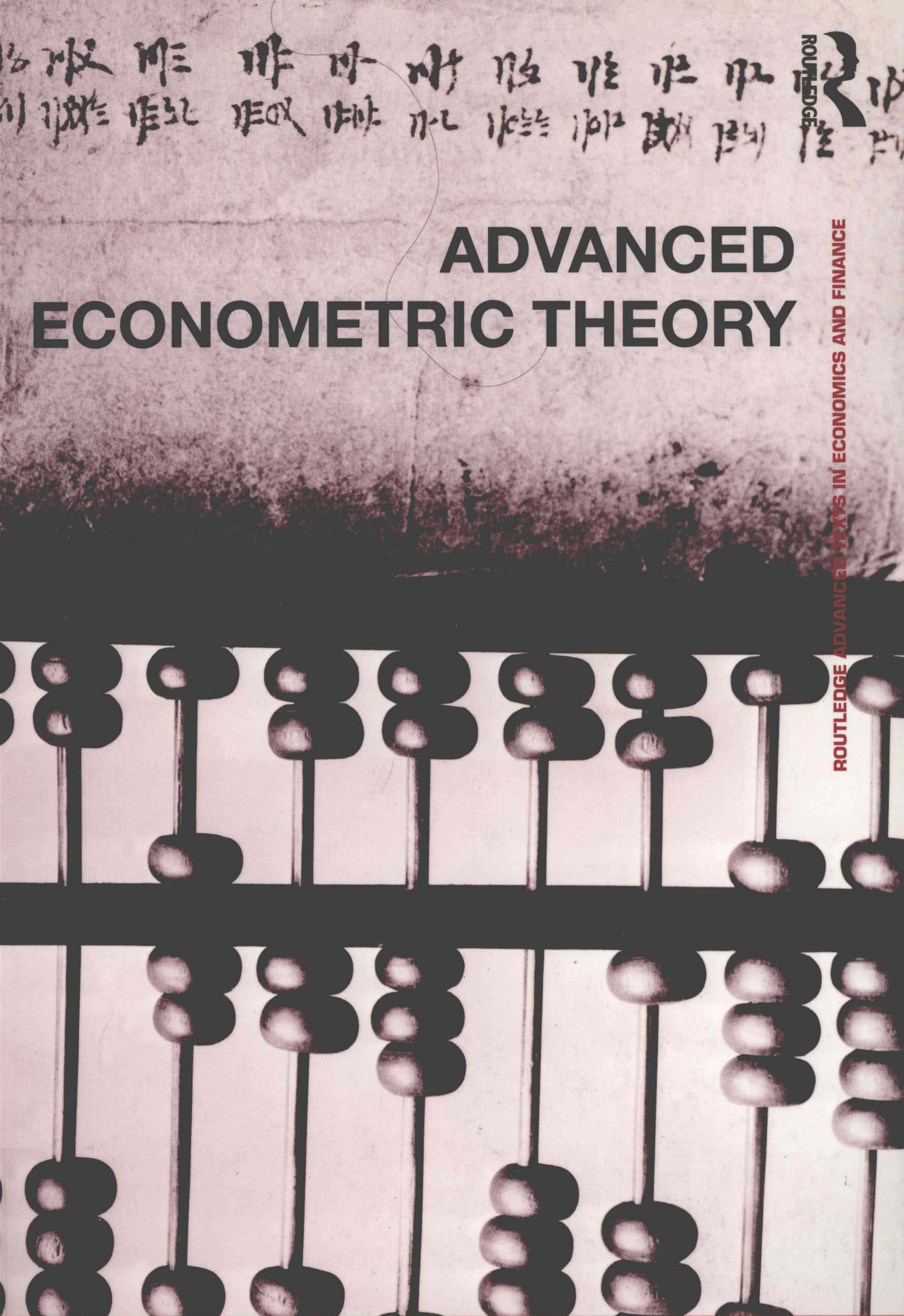

ADVANCED ECONOMETRIC THEORY

ROUTLEDGE ADVANCED TEXTS IN ECONOMICS AND FINANCE

ROUTLEDGE

John S. Chipman

First published 2011
by Routledge
2 Park Square, Milton Park, Abingdon, Oxon, OX14 4RN

Simultaneously published in the USA and Canada
by Routledge
711 Third Avenue, New York, NY 10017

Routledge is an imprint of the Taylor & Francis Group, an informa business

British Library Cataloguing in Publication Data
A catalogue record for this book is available from the British Library

Library of Congress Cataloging in Publication Data
Chipman, John Somerset, 1926-
Advanced econometric theory/John S. Chipman
p. cm.
1. Econometrics. I. Title.
HB139.C479 2010
330.01'5195–dc22
2010038672

ISBN: 978-0-415-32629-2 (hbk)
ISBN: 978-0-415-32630-8 (pbk)
ISBN: 978-0-203-50507-6 (ebk)

Typeset in Times New Roman
by Sunrise Setting Ltd, Torquay, UK

MIX
Paper from responsible sources
FSC www.fsc.org FSC® C004839

Printed and bound in Great Britain by
TJ International Ltd, Padstow, Cornwall

Advanced Econometric Theory

Routledge Advanced Texts in Economics and Finance

1 **Financial Econometrics**
Peijie Wang

2 **Macroeconomics for Developing Countries (Second Edition)**
Raghbendra Jha

3 **Advanced Mathematical Economics**
Rakesh Vohra

4 **Understanding Macroeconomic Theory**
John M. Barron, Bradley T. Ewing, and Gerald J. Lynch

5 **Regional Economics**
Roberta Capello

6 **Mathematical Finance**
Core theory, problems and statistical algorithms
Nikolai Dokuchaev

7 **Applied Health Economics**
Andrew M. Jones, Nigel Rice, Teresa Bago d'Uva, and Silvia Balia

8 **Information Economics**
Urs Birchler and Monika Bütler

9 **Financial Econometrics (Second Edition)**
Peijie Wang

10 **Development Finance**
Debates, dogmas and new directions
Stephen Spratt

11 **Culture and Economics**
On values, economics and international business
Eelke de Jong

12 **Modern Public Economics (Second Edition)**
Raghbendra Jha

13 **Introduction to Estimating Economic Models**
Atsushi Maki

14 **Advanced Econometric Theory**
John S. Chipman

To my former students and colleagues in econometrics

Preface

Some years ago I decided that the lecture notes I have been distributing to my graduate students in econometrics for many years, being quite different in many ways from existing textbooks, might be worth publishing as a book. This is the result.

I am indebted to many colleagues and students. I benefited in Chapter 1 from helpful suggestions from Simo Puntanen, George Styan, and John Geweke, as well as my then colleague Christopher Sims. Chapters 2 and 4 build on my early (1964) collaboration with Malempati M. Rao. In Chapter 3, I greatly benefited from stimulating conversations with Paul Garrett of the University of Minnesota Mathematics Department, as well as the assistance of his colleague Joel Roberts who supplied me the proofs of Lemma 3.4.3 and Theorem 3.4.1 (but I do not hold him responsible for my diagrammatic interpretations). Regarding Chapter 5, I am especially indebted to Morris L. Eaton with whom I spent many hours working out proofs of two lemmas underlying extensions of James–Stein estimation. Chapter 6 builds on a 1968 paper undertaken in collaboration with my former student K. Rao Kadiyala as well as with the two referees (and subsequent coauthors) of that paper, on estimation of a mean with autoregressive residuals; this was followed up by my 1979 paper on least-squares estimation of linear trend. Chapter 6 adds an analysis of alternative methods of estimation due to Prais and Winsten as well as to that of Cochrane and Orcutt. In this chapter (as well as the next), I benefited from interactions with my late colleague Clifford Hildreth. Chapter 7 builds on an unpublished 1965 paper of mine cited by Judge et al. (1985, pp. 282–6) and others. Elizabeth Dolan did the initial work for the three diagrams, which was converted to TeX (the software invented by Donald E. Knuth, author of *The TeXbook*) by Augustine Mok, using PICTeX (authored by Michael J. Wichura). This software has been used for diagrams throughout the book. In Section 8.6 of Chapter 8 I was greatly helped by Lei (Nick) Guo, an economics graduate student, in working out in great detail solutions to equations containing matching terms, needed to derive explicit expressions for the Gram–Charlier and Edgeworth series. I later received a nice expository paper from Brian Shea, a statistics graduate student, also partly covering this ground. Stephen Stigler kindly sent me a copy of Anders Hald's 2002 monograph on the history of asymptotic expansions.

One way in which this book departs from existing textbooks is that the coverage is less exhaustive and more selective. An unfortunate result is that some topics such as nonlinear regression, Granger and Sims causality, and unit-root tests, as well as three-stage least squares, tests for overidentification, specification error in sumultaneous-equations models, etc., are omitted entirely. In exchange, some topics that seldom get covered in econometrics textbooks, such as linear aggregation (in Chapter 2), reduced-rank regression (in Chapter 3), ridge regression (in Chapter 4), Stein estimation (in Chapter 5), asymptotic expansions of and asymptotic approximations to single-equation simultaneous-equations estimators, and recursive models (both in Chapter 8), are covered in detail.

The book also covers special topics such as computation of percentage points of the noncentral F distribution (Appendix to Chapter 4) and the beta approximation to the Durbin–Watson statistic (Section 7.3 of Chapter 7).

Chapter 9 contains solutions to all the exercises. While this may cause problems for some instructors, it is designed to be of help to those who choose to use the book for self-instruction.

I have tried to maintain a uniform notation throughout. Unfortunately, the traditional Cowles-Commission notation for simultaneous-equations models is inconsistent with the standard notation for linear regression models, and this has necessitated departing somewhat from the traditional simultaneous-equations notation that is still in use today.

The book was originally typeset with $\mathcal{A}\mathcal{M}\mathcal{S}$-TeX. Thanks are due to Barbara Beeton of the American Mathematical Society and Michael Spivak (author of $\mathcal{A}\mathcal{M}\mathcal{S}$-TeX) for their advice. Thanks are also due to my former assistant Hwikwon Ham.

Early drafts of the book were gone over carefully by my then research assistant Daniel Rodriguez Delgado. At the final stages I have been particularly helped in the typesetting of Chapter 7 by Augustine Mok and Chin-Tung Grace Chan, as well as by the assistance of the former in preparing the Table of Contents and proofreading the entire book, and of the latter in gaining permission for the reproduction of passages from two previous publications, namely the references Chipman 1997 and 1999 listed in the Bibliography. The Index was prepared by Augustine Mok and Chin-Tung Grace Chan, using the LaTeX document-preparation system (authored by Leslie Lamport).

John S. Chipman
Department of Economics
University of Minnesota
Minneapolis, MN 55455
USA

January 2011

Contents

Figures and tables

Figures

Tables

1 Multivariate analysis and the linear regression model

1.1 Introduction

This book is mainly concerned with the linear model

$$Y = X\mathrm{B} + \mathrm{E}, \quad \text{i.e.,} \quad y_{tj} = \sum_{i=1}^{k} x_{ti}\beta_{ij} + \varepsilon_{tj} \quad (t = 1, 2, \ldots, n), \tag{1.1.1}$$

where $X = [x_{ti}]$ is an $n \times k$ matrix of n observations on k independent or exogenous variables, $Y = [y_{tj}]$ is an $n \times m$ matrix of n observations on m jointly dependent or endogenous variables, and $\mathrm{E} = [\varepsilon_{tj}]$ is an $n \times m$ matrix of random errors with zero means and specified variances and covariances, where $t = 1, 2, \ldots, n$. To allow for a constant term, the first column of X may be specified to be a column of 1s. We shall in fact for the most part in Chapters 2–7 concentrate on the special univariate case $m = 1$, returning to the multivariate case in Chapter 8. The purpose of this chapter is to embed the above model in a multivariate model in which the rows $x_{t\cdot} = (x_{t1}, x_{t2}, \ldots, x_{tk})$ of X and $y_{t\cdot} = (y_{t1}, y_{t2}, \ldots, y_{tm})$ of Y are specified to have a joint distribution, and to consider the problem of the best predictor of $y_{t\cdot}$ given $x_{t\cdot}$. The linearity of the above model emerges as a practical aspect of optimal prediction.

We shall start with an analysis of the multivariate statistical model of two jointly distributed random row vectors (of which (1.1.1) is a sample of size n)

$$x = (x_1, x_2, \ldots, x_k) \quad \text{and} \quad y = (y_1, y_2, \ldots, y_m) \tag{1.1.2}$$

with first and second moments

$$\xi = E\{x\}, \quad \eta = E\{y\}, \quad \mathrm{M} = E\{(x, y)'(x, y)\} = \begin{pmatrix} \mathrm{M}_{xx} & \mathrm{M}_{xy} \\ \mathrm{M}_{yx} & \mathrm{M}_{yy} \end{pmatrix}. \tag{1.1.3}$$

It is desired to predict y given knowledge of x.

We shall think of the model (1.1.1) as consisting of n observations on the best linear predictor of y given x, or equivalently (as we shall see) as the best linear approximation of the conditional expectation of y given x. Thus, estimation consists of a two-stage process: (1) that of finding, out of the class of linear functions $y = x\mathrm{B}$, the formula for the best linear predictor of y given x, namely $y = x\hat{\mathrm{B}}$; (2) that of estimating the $k \times m$ matrix $\hat{\mathrm{B}}$ from data. The optimal formula we obtain for B simply gives us an interpretation of the model; for it may be assumed that the means and moment matrix given by (1.1.3) are themselves unknown parameters. The estimation problem consists in the problem of finding a numerical estimate B of B from data on the xs and ys. In order to define precisely what is meant by a "best linear predictor" we will need the following two definitions.

DEFINITION 1.1.1 Let A be any real $n \times n$ matrix, and let x denote any $1 \times n$ row vector with real components. Then

$A \succcurlyeq 0$ ("A is nonnegative-definite") means: $xAx' \geqq 0$ for all x;
$A \approx 0$ ("A is zero-definite") means: $xAx' = 0$ for all x;
$A \succ 0$ ("A is positive-semidefinite") means: $xAx' \geqq 0$ for all x and $xAx' > 0$ for some x; i.e., $A \succcurlyeq 0$ and not $A \approx 0$;
$A \succ\succ 0$ ("A is positive-definite") means: $xAx' > 0$ for all $x \neq 0$.

We further define

$$A \succcurlyeq B \quad \text{if and only if} \quad A - B \succcurlyeq 0, \tag{1.1.4}$$

or $B \preccurlyeq A$, and similarly for the other listed relations. The relation $\succcurlyeq$ is known as the Löwner ordering.[1]

THEOREM 1.1.1 *The Löwner ordering* (1.1.4) *is a partial ordering of symmetric matrices, i.e., it is transitive, reflexive, and anti-symmetric.*

Proof: Transitivity of $\succcurlyeq$ means that $A \succcurlyeq B$ and $B \succcurlyeq C$ imply $A \succcurlyeq C$. This follows from the identity

$$z(A - C)z' = z(A - B)z' + z(B - C)z',$$

since if both terms on the right are nonnegative, so is the term on the left. *Reflexivity* of $\succcurlyeq$ means that $A = B$ implies $A \succcurlyeq B$; this follows immediately from the definition. *Anti-symmetry* of $\succcurlyeq$ means that $A \succcurlyeq B$ and $B \succcurlyeq A$ imply $A = B$ (cf. Birkhoff 1948, p. 1); i.e., setting $C = A - B$, $C \approx 0$ implies $C = 0$. This is true if and only if C is symmetric, and is proved as follows.

Let $C \approx 0$. (i) Choose z to be the coordinate vector with 1 in ith place and zeros elsewhere; then $zCz' = 0$ implies $c_{ii} = 0$ for all i, i.e., all the diagonal elements of C are zero. (ii) Choose z so that its ith and jth components ($i \neq j$) are equal to 1 and the remaining components zero; then $c_{ii} + c_{jj} + c_{ij} + c_{ji} = 0$. From (i) this implies $c_{ij} = -c_{ji}$ for $i \neq j$. Thus from (i) and (ii) we have $C = -C'$, i.e., C is *skew-symmetric* (cf. Wedderburn 1934, p. 8). But a skew-symmetric matrix vanishes if and only if it is symmetric, i.e., $C = -C' = -C$ implies $2C = 0$ hence $C = 0$, and the converse is trivial. □

The result is illustrated by the counterexample

$$zCz' = [z_1 \quad z_2] \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} z_1 \\ z_2 \end{bmatrix} = z_2 z_1 - z_1 z_2 = 0.$$

In what follows we shall only be concerned with the Löwner ordering among symmetric matrices.

It follows from Theorem 1.1.1 that if $\mathcal{A}$ is a collection of square symmetric matrices, and a matrix $B \in \mathcal{A}$ is a minimum (i.e., $A \succcurlyeq B$ for all $A \in \mathcal{A}$), then this minimum is unique. For should there be two such matrices B_1, B_2, then $A \succcurlyeq B_1$ for all $A \in \mathcal{A}$, hence in particular it is true that $B_2 \succcurlyeq B_1$; likewise, $B_1 \succcurlyeq B_2$, hence from the anti-symmetry of $\succcurlyeq$, $B_1 = B_2$.

Coming back to our problem of optimal prediction, a widely accepted criterion for goodness of prediction is that of mean-square error: Out of a set of mappings $F: \mathcal{R}^k \to \mathcal{R}^m$ one chooses the mapping $\hat{F}$ which minimizes the matrix mean-square error

$$R(F) = E\{[y - F(x)]'[y - F(x)]\} \tag{1.1.5}$$

in terms of the ordering $\succcurlyeq$.

Two important properties of $\hat{F}$ should be noted, and will be proved below:

(1) If the set of admissible mappings F is subject to some mild regularity conditions, then the optimal F is precisely the conditional expectation of y given x, $\hat{F}(x) = E(y \mid x)$. (Theorem 1.6.1.)
(2) If the joint distribution of x and y is normal, then the optimal F, which is $E(y \mid x)$ as above, is an affine function $\hat{F}(x) = \alpha + x\mathrm{B}$. (Theorem 1.4.1.)

As a result of these two properties, Doob (1953, p. 77) introduced the concept of "wide-sense conditional expectation of y given x", defined as that *affine* mapping $F(x) = \alpha + x\mathrm{B}$ which minimizes the mean-square error (1.1.5) among all such affine mappings, and denoted $\hat{F}(x) = \hat{E}(y \mid x)$. The same concept was introduced by Cramér (1946, p. 272), and called the "minimum-mean-square regression of y on x."

It is convenient to introduce a similar concept in which the set of affine mappings $F(x) = \alpha + x\mathrm{B}$ is replaced by the set of linear(-homogeneous) ones $F(x) = x\mathrm{B}$. The resulting optimal F may be called the *best homogeneous linear predictor of y given x*, denoted $\hat{\mathcal{P}}(y \mid x)$.

In order to relate the above concepts to the standard linear regression model used in econometrics, it will be useful first to derive the expression for the best homogeneous linear predictor of y given x.

THEOREM 1.1.2 *If the random variables x and y of* (1.1.2) *have a joint distribution satisfying* (1.1.3), *then among all linear mappings $F(x) = x\mathrm{B}$, the one that* *minimizes the matrix mean-square error, or "risk"*

$$R(\mathrm{B}) = E\{(y - x\mathrm{B})'(y - x\mathrm{B})\}, \tag{1.1.6}$$

is given by

$$\hat{\mathcal{P}}(y \mid x) = x\hat{\mathrm{B}}, \tag{1.1.7}$$

where $\hat{\mathrm{B}}$ *is any solution of the equation*

$$\mathrm{M}_{xx}\hat{\mathrm{B}} = \mathrm{M}_{xy} \tag{1.1.8}$$

(*which always exists*). *The value* (1.1.7) *is unique with probability* 1, *and the minimum risk is*

$$R(\hat{\mathrm{B}}) = \mathrm{M}_{yy} - \mathrm{M}_{yx}\hat{\mathrm{B}}. \tag{1.1.9}$$

In order to prove this theorem it will be necessary first to introduce some definitions and then to prove two lemmas.

DEFINITION 1.1.2 For any $n \times k$ matrix X, a (weak) generalized inverse of X, denoted X^-, is defined as any $k \times n$ matrix satisfying $XX^-X = X$.

The matrix X^- is called a g-inverse by Rao, C. R. (1966).[2] It will be shown in the next chapter (Theorem 2.3.1) that any matrix possesses a weak generalized inverse.

For the following standard concepts and results see, e.g., Thrall and Tornheim (1957, pp. 73–78) and Afriat (1957, p. 801).

DEFINITION 1.1.3 The row space of an $n \times k$ matrix X is the set of $1 \times k$ row vectors

$$\mathcal{R}(X) = \{\xi \in \Re^k : \xi = y'X \text{ for some } n \times 1 \text{ column vector } y \in \Re^n\},$$

where $\Re^m$ denotes m-dimensional Euclidean space. The column space of X is defined as the set of $n \times 1$ column vectors

$$\mathcal{C}(X) = \{\mu \in \Re^n : \mu = X\beta \text{ for some } k \times 1 \text{ column vector } \beta \in \Re^k\}.$$

The *column null space* (or *column kernel*) of X is the set of $k \times 1$ vectors

$$\mathcal{K}(X) = \{\beta \in \Re^k : X\beta = 0\}.$$

The dimensions of $\mathcal{C}(X)$ and $\mathcal{K}(X)$ are respectively the rank, $\rho(X)$, and column nullity, $\nu(X)$, of X, where $\rho + \nu = k$.

LEMMA 1.1.1 *Let A, B' be $m \times n$ matrices. Then* $\text{tr}(AB) = \text{tr}(BA)$.

Proof: Putting $A = [a_{ij}]$, $B = [b_{ij}]$, we have

$$\text{tr}(AB) = \sum_{i=1}^{m} \sum_{j=1}^{n} a_{ij} b_{ji} = \sum_{j=1}^{n} \sum_{i=1}^{m} b_{ji} a_{ij} = \text{tr}(BA). \qquad \square$$

LEMMA 1.1.2 *Let z be a row vector with moment matrix $E\{z'z\} = M$. Then z belongs to the row space of M with probability* 1.

Proof:[3] Let M^- be any symmetric weak generalized inverse of M. Then we may write

$$z = zM^-M + z(I - M^-M). \tag{1.1.10}$$

Defining

$$e = z(I - M^-M), \tag{1.1.11}$$

we see from (1.1.11) and Definition 1.1.2 that the moment matrix of e is

$$E\{e'e\} = E\{(I - MM^-)z'z(I - M^-M)\} = (I - MM^-)M(I - M^-M) = 0.$$

From Lemma 1.1.1 and the above equation, $E\{ee'\} = \text{tr}(E\{e'e\}) = 0$. Now a theorem of Billingsley (1995, p. 203, Theorem 15.2(ii)) implies that if the integral of a positive-valued function is zero, then the set on which this function is positive has measure zero. Applying this to the positive-valued random variable ee' we conclude that[4]

$$E\{ee'\} = 0 \Rightarrow \Pr\{ee' = 0\} = 1.$$

Since ee' is a sum of squares, this implies in turn that $\Pr\{e = 0\} = 1$. Going back now to (1.1.10) we see that since the second term on the right (which is (1.1.11)) is zero with probability 1, therefore

$$\Pr\{z = zM^-M \equiv wM\} = 1 \quad (\text{where } w \equiv zM^-).$$

This states that z belongs to the row space of M with probability 1. $\square$