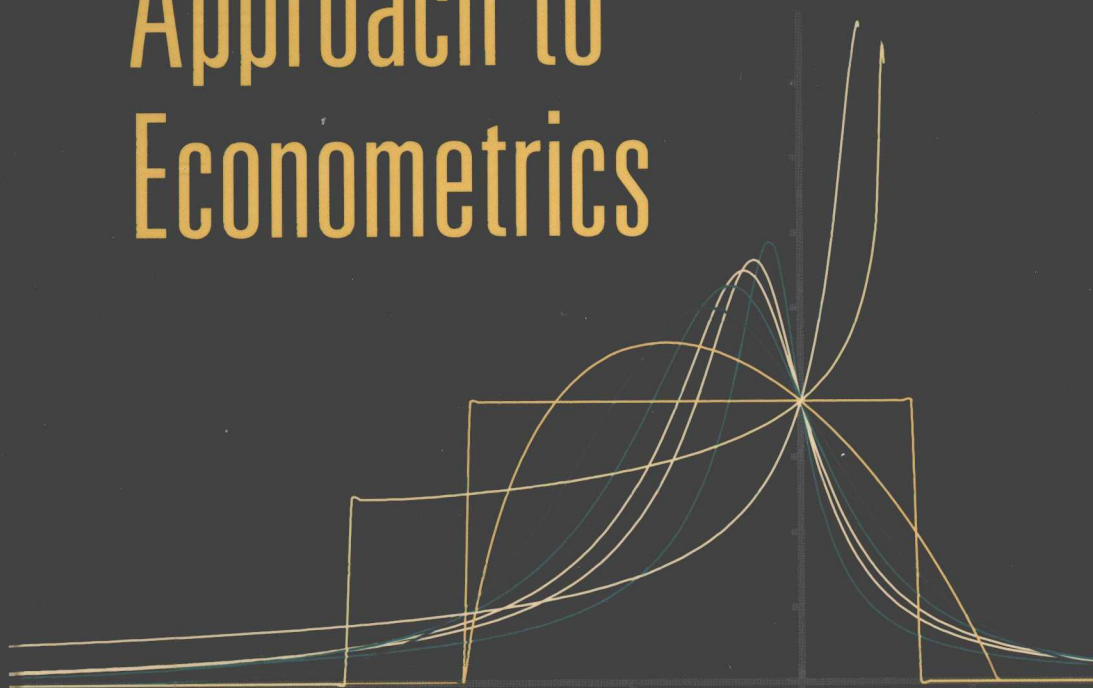


An Information Theoretic Approach to Econometrics



GEORGE G. JUDGE

RON C. MITTELHAMMER

0
CAMBRIDGE



30807611

An Information Theoretic Approach to Econometrics

GEORGE G. JUDGE

University of California, Berkeley

RON C. MITTELHAMMER

Washington State University



CAMBRIDGE
UNIVERSITY PRESS

CAMBRIDGE UNIVERSITY PRESS
Cambridge, New York, Melbourne, Madrid, Cape Town,
Singapore, São Paulo, Delhi, Tokyo, Mexico City

Cambridge University Press
32 Avenue of the Americas, New York, NY 10013-2473, USA

www.cambridge.org
Information on this title: www.cambridge.org/9780521689731

© George G. Judge and Ron C. Mittelhammer 2012

This publication 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 2012

Printed in the United States of America

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

Library of Congress Cataloging in Publication Data

Judge, George G.

An information theoretic approach to econometrics / George G. Judge, Ron C. Mittelhammer.
p. cm.

Includes bibliographical references and index.

ISBN 978-0-521-86959-1 (hardback) – ISBN 978-0-521-68973-1 (paperback)

1. Econometrics. I. Mittelhammer, Ron. II. Title.

HB139.J795 2011

330.01'5195–dc23 2011018358

ISBN 978-0-521-86959-1 Hardback

ISBN 978-0-521-68973-1 Paperback

Cambridge University Press has no responsibility for the persistence or accuracy of URLs for
external or third-party Internet Web sites referred to in this publication and does not guarantee
that any content on such Web sites is, or will remain, accurate or appropriate.

Preface

In one sense, the idea for this book started 15 years ago with a six-hour meeting of the two authors between planes at the O'Hare Hilton. At this meeting, maximum entropy and empirical likelihood principles were the major areas of discussion. Since that time the two of us have worked together and have only looked forward. As a result, *Econometric Foundations* appeared in 2000, and a range of related information theoretic articles emerged in the last decade. Pieces of some of these articles appear in this book.

This book was a pleasure to write. We hope the reader will feel our enthusiasm in entering the information theoretic world and leaving behind many conventional econometric methods that we spent a good part of our lives learning.

To write this book, we had the help of many colleagues. Several years of work with Amos Golan provided a base for dealing with pure and noisy inverse problems and the maximum entropy principle. Douglas Miller was involved in this early work and also worked with us on the *Econometric Foundations* book. Marian Grendar, a longtime colleague of one of the authors, worked with us concerning the theoretical underpinning of information theoretic methods and read and commented on many of the chapters in this book. Art Owen was always available to discuss issues relating to the empirical likelihood approach to estimation and inference. Wendy Tam Cho was a creative partner in solving a range of important pure inverse applied problems. To a large group of colleagues, too numerous to mention, we express our warm thanks and appreciation and hope they have been appropriately acknowledged in references throughout the book.

For a book to reach the publication stage, two persons are necessary in addition to the author. One involves a person that takes incoherent words and other symbols and converts them to a working copy. In this context, Danielle Engelhardt, with intelligence, good humor, and word-processing

skills, worked with us to get each chapter just right and to turn it into a beautiful copy. A second person is an editor who understands the subject matter and shares your goals. In Scott Parris, we found such a person and a full partner every step of the way. For the two authors, this has been a joint venture; the order of the names has only alphabetical significance.

George G. Judge
Ron C. Mittelhammer

Contents

<i>Preface</i>	<i>page xv</i>
1 Econometric Information Recovery	1
1.1 Book Objectives and Problem Format	1
1.2 Organization of the Book	3
1.3 Selected References	4
PART I TRADITIONAL PARAMETRIC AND SEMIPARAMETRIC ECONOMETRIC MODELS: ESTIMATION AND INFERENCE	
2 Formulation and Analysis of Parametric and Semiparametric Linear Models	7
2.1 Data Sampling Processes (DSPs) and Notation	7
2.2 A Parametric General Linear Model	10
2.2.1 The Parametric Model and Maximum Likelihood (ML) Estimation of β and σ^2	12
2.2.2 The Parametric Model and Inference	15
2.3 A Semiparametric General Linear Model	16
2.3.1 The Squared Error Metric and the Least Squares (LS) Principle	17
2.3.2 The LS Estimator	18
2.3.3 Finite Sample Statistical Properties of the LS Estimator	19
2.3.4 Consistency and Asymptotic Normality of the LS Estimator	19
2.3.5 Linear Semiparametric Model Inference	20
2.3.6 Inferential Asymptotics	21
2.3.7 Hypothesis Testing: Linear Equality Restrictions on β	22

2.4	General Linear Model with Stochastic X	24
2.4.1	Linear Model Assumptions	25
2.4.2	LS Estimator Properties: Finite Samples	26
2.4.3	LS Estimator Properties: Asymptotics	26
2.4.4	ML Estimation of β and σ^2 under Conditional Normality	27
2.4.5	Hypothesis Testing and Confidence Region Estimation	28
2.4.5a	Semiparametric Case	28
2.4.5b	Parametric Case	28
2.4.6	Summary: Statistical Implications of Stochastic X	30
2.5	Extremum (E) Estimation and Inference	30
2.5.1	ML and LS Estimators Expressed in E Estimator Form	31
2.5.2	Asymptotic Properties of E Estimators	32
2.5.3	Inference Based on E Estimation	33
2.5.4	Summary and Forward: E Estimators	34
2.6	Selected References	35
3	Method of Moments, Generalized Method of Moments, and Estimating Equations	36
3.1	Introduction	36
3.1.1	A Just-Determined Moment System with Random Sampling of Scalars	37
3.2	Just-Determined Moment Systems, Random Sampling, and Method of Moments (MOM)	39
3.2.1	General Asymptotic Properties	40
3.2.2	Linear Model Semiparametric Estimation through Moment Equations	41
3.2.3	MOM Conclusions	42
3.3	Generalized Method of Moments (GMM)	43
3.3.1	GMM Framework	43
3.3.2	GMM Linear Model Estimation	44
3.3.2a	Optimal GMM Weight Matrix	45
3.3.2b	Sampling Properties of Estimated Optimal GMM (EOGMM) Estimator	46
3.3.2c	Hypothesis Testing and Confidence Regions	47
3.3.2d	Additional Properties of the GMM Approach	48
3.3.2e	Summary and Forward: The GMM Approach	49

3.4	Estimating Equations	50
3.4.1	Duality between Estimating Equations (EEs) and E Estimators	51
3.4.2	Linear Estimating Functions (EFs)	52
3.4.3	Optimal Unbiased EFs	54
3.4.3a	Unbiasedness	54
3.4.3b	Optimal Estimating Functions (OptEFs): The Scalar Case	56
3.4.3c	OptEFs: The Multivariate Case	57
3.4.4	Inference in the Context of EE Estimation	59
3.4.4a	Wald (W) and Z Tests and Confidence Regions	59
3.4.4b	Generalized Score (Lagrange Multiplier-Type) Tests and Confidence Regions	60
3.4.4c	Pseudo-Likelihood Ratio Tests and Confidence Regions	61
3.5	E Estimation with Instrumental Variables	62
3.6	Summary and Forward	63
3.7	Selected References	64

PART II FORMULATION AND SOLUTION OF STOCHASTIC INVERSE PROBLEMS

4	A Stochastic-Empirical Likelihood Inverse Problem: Formulation and Estimation	69
4.1	Introduction	69
4.2	A Stochastic Linear Inverse Problem	71
4.2.1	Addressing the Indeterminacy of Unknowns	73
4.3	Nonparametric ML Solutions to Inverse Problems	74
4.3.1	Nonparametric ML	74
4.3.2	Empirical Likelihood (EL) Function for θ	76
4.3.3	Comparing the Use of Estimating Functions in EE and EL Contexts	78
4.3.4	The Functional Form of the EL Function	80
4.3.5	Summary of the EL Concept	81
4.3.6	Maximum Empirical Likelihood (MEL) Estimation of a Population Mean	82
4.3.7	MEL Linear Model Estimation for Stochastic X	85
4.4	Epilogue	86
4.5	Selected References	87
	Appendix 4. A Numerical Example: Computing MEL Estimates	87

5	A Stochastic Empirical Likelihood Inverse Problem: Estimation and Inference	90
5.1	Introduction	90
5.2	MEL Inference: <i>iid</i> Case	90
5.2.1	MEL Efficiency Property	91
5.3	Empirical Example of MEL Estimation Based on Two Moments	94
5.4	Hypothesis Tests and Confidence Regions: <i>iid</i> Case	95
5.4.1	Empirical Likelihood Ratio Tests and Confidence Regions for $c(\theta)$	95
5.4.2	Wald Tests and Confidence Regions for $c(\theta)$	96
5.4.3	Lagrange Multiplier Tests and Confidence Regions for $c(\theta)$	97
5.4.4	Z-Test of Inequality Hypotheses for the Value of $c(\theta)$	97
5.4.5	Testing the Validity of Moment Equations	98
5.4.6	MEL Testing and Confidence Intervals for Population Mean	99
5.4.7	Illustrative MEL Confidence Interval Example	100
5.5	Concluding Comments	101
5.6	Selected References	103
6	Kullback-Leibler Information and the Maximum Empirical Exponential Likelihood	104
6.1	Introduction	104
6.1.1	Solutions to Systems of Estimating Equations and Kullback-Leibler Information	104
6.2	Kullback-Leibler Information Criterion (KLIC)	106
6.2.1	Relationship between Maximum Empirical Exponential Likelihood (MEEL) and KL Information	108
6.2.1a	Objective of MEEL	109
6.3	The General MEEL Alternative Empirical Likelihood Formulation	111
6.3.1	The MEEL Estimator and Alternative Empirical Likelihood	111
6.3.2	MEEL Asymptotics	112
6.3.3	MEEL Inference	114
6.3.3a	Testing $H_0 : c(\theta) = r$	114
6.3.3b	Testing $H_0 : c(\theta) \leq r$ or $H_0 : c(\theta) \geq r$	115
6.3.3c	Testing the Validity of Moment Equations	115
6.3.3d	Confidence Regions	116

6.3.4	Contrasting the Use of Estimating Functions in EE and MEEL Contexts	116
6.4	Combining Estimation Equations under Kullback-Leibler Loss	117
6.4.1	Combating Model Uncertainty: General Combining Formulations	117
6.4.2	Example: A Combined Estimator	119
6.4.2a	Finite Sample Performance	120
6.4.2b	Implications	121
6.5	An Informative Reference Distribution	121
6.6	Concluding Remarks	123
6.7	Reader Idea Checklist	124
6.8	Selected References	125
	Appendix 6.A Relationship between the Maximum Empirical Likelihood (MEL) Objective and KL Information	126
	Appendix 6.B Numerical Illustration of MEEL and MEL Estimation of a Probability Distribution	128
	Appendix 6.C Shannon's Entropy – Some Historical Perspective	130
PART III A FAMILY OF MINIMUM DISCREPANCY ESTIMATORS		
7	The Cressie-Read Family of Divergence Measures and Empirical Maximum Likelihood Functions	135
7.1	Introduction	135
7.1.1	Family of Likelihood Functions	136
7.2	The Cressie-Read (CR) Power Divergence Family	137
7.3	Three Main Variants of $I(\mathbf{p}, \mathbf{q}, \gamma)$	139
7.4	Minimum Power Divergence and Empirical Maximum Likelihood (EML) Estimation	140
7.5	Inference	142
7.5.1	Test Statistics	142
7.5.1a	Moment Validity Tests	143
7.5.1b	Tests of Parameter Restrictions	144
7.6	Concluding Remarks	145
7.7	Selected References	146
	Appendix 7.A Propositions, Proofs, and Definitions	148
	Appendix 7.B Entropy Families	152
8	Cressie-Read-MPD-Type Estimators in Practice: Monte Carlo Evidence of Estimation and Inference Sampling Performance	153
8.1	Introduction	153
8.2	Design of Sampling Experiments	154

8.3	Sampling Results	156
8.3.1	Estimator MSE Performance	156
8.3.2	Bias and Variance	157
8.3.3	Prediction MSE	159
8.3.4	Size of Moment Validity Tests	159
8.3.5	Confidence Interval Coverage and Expected Length	160
8.3.6	Test Power	161
8.4	Summary Comments	161
8.5	Selected References	163
	Appendix 8.A Computational Issues and Numerical Approach	164
PART IV BINARY-DISCRETE CHOICE MINIMUM POWER DIVERGENCE (MPD) MEASURES		
9	Family of MPD Distribution Functions for the Binary Response-Choice Model	169
9.1	Introduction	169
9.2	The Statistical Model Base	171
9.2.1	Parametric	171
9.2.2	Nonparametric Stochastic Inverse Problem Representation	172
9.3	Minimum Power Divergence Class of CDFs for the Binary Response Model	173
9.3.1	Applying MPD to Conditional Bernoulli Probabilities	174
9.3.2	Conditional Reference Probabilities	175
9.3.3	The Class of CDFs Underlying p	177
9.3.4	Properties of the MPD Class of Probability Distribution Functions	179
9.3.4a	Moments	181
9.3.4b	Concavity of $\ln(F(w; q, \gamma))$ and $\ln(1 - F(w; q, \gamma))$ in w	182
9.4	Summary and Extensions	182
9.5	Selected References	183
	Appendix 9.A Additional Properties of MPD Distributions	184
10	Estimation and Inference for the Binary Response Model Based on the MPD Family of Distributions	187
10.1	Introduction	187
10.2	MPD Solutions for p and λ as Estimators in Binary Response Models	188

10.2.1	Interpreting λ as an Estimator of β	188
10.2.2	Estimating the Marginal Probability Effects of Changes in Response Variables	190
10.3	Asymptotic Estimator Properties	191
10.3.1	Asymptotic Inference	192
10.4	Estimation Alternatives	193
10.4.1	EML Estimation	193
10.4.2	NLS-MPD Estimation	195
10.5	Sampling Performance	195
10.5.1	Sampling Design	196
10.5.2	Sampling Results	198
10.6	Summary and Extensions	200
10.7	Selected References	201
	Appendix 10.A Asymptotic Properties of $\widehat{\text{MPD}}(q, \gamma)$ – Applicability of Assumptions	202
	Appendix 10.B Consistency	203
	Appendix 10.C Asymptotic Normality	204
 PART V OPTIMAL CONVEX DIVERGENCE		
11	Choosing the Optimal Divergence under Quadratic Loss	207
11.1	Introduction	207
11.2	Econometric Model and the Cressie-Read (CR) Family	208
11.3	Choosing a Minimum Loss Estimation Rule	209
11.3.1	Distance–Divergence Measures	210
11.3.2	A Minimum Quadratic Risk Estimation Rule	211
11.3.3	The Case of Two CR Alternatives	212
11.3.4	Empirical Calculation of α	213
11.4	Finite Sample Implication	214
11.5	Estimator Choice, $\gamma = (1, 0, -1)$	216
11.6	Sampling Performance	217
11.7	Concluding Remarks	218
11.8	Selected References	219
	Appendix 11.A A $\gamma = (0, -1)$ Special Case Convex Estimation Rule	220
12	Epilogue	221
	<i>Abbreviations</i>	223
	<i>Index</i>	227

ONE

Econometric Information Recovery

1.1 Book Objectives and Problem Format

The objectives of this book are to

- i) develop a plausible basis for reasoning in situations involving incomplete-partial econometric model information,
- ii) develop principles and procedures for learning or recovering information from a sample of indirect noisy data, and
- iii) provide the reader with a firm conceptual and empirical understanding of basic information theoretic econometrics models and methods.

What makes the econometric information recovery process interesting is that

- economic-behavioral systems, such as physical and biological systems, are statistical in nature;
- the conceptual econometric model contains parameters and noise components that are unknown and unobserved and, indeed, not subject to direct observation or measurement;
- the recovery of information on the unknown parameters or components requires, for analysis purposes, the use of indirect noisy measurements based on observable data and the solution of an inverse problem that maps the indirect noisy observations, into information on the unknown model and its unobservable components;
- the models may be ill-posed or, in the context of traditional procedures, may be undetermined and the solution not amenable to conventional rules of logic or to being written in closed form.

These problems, taken either individually or in some combination, represent the intellectual challenge of modern econometric analysis and research.

Building on the productive efforts of our precursors in the areas of theoretical economics and inferential statistics, we hope, in this book, to provide an operational understanding of a rich set of information theoretic methods that may be used in theoretical and applied econometrics.

Econometrics is a work in progress. Anyone who doubts this should review a sampling of econometric books starting in the mid-1930s and map the development of econometrics over time. Advances in econometric methodology have been substantial in both content and number, and they continue at a geometric rate.

Information theoretic methods, which have a base in statistical mechanics in physics, have developed in econometrics over the last two decades. In this book we provide a conceptual and empirical understanding of information theoretic methods in some of the major areas of econometrics. Because in econometrics, and in other subject-matter areas, we must work with indirect noisy observations and ill-posed econometric models, traditional econometric methods may not be applicable in answering many of the quantitative questions we wish to ask.

To be a bit more specific, as noted previously, in econometric analyses the unknown and uncontrolled components of the econometric model cannot generally be observed directly. Thus, the analyst must use indirect noisy observations based on observable data to recover information relative to these unknown and unobservable components. This situation is associated with a concept in systems and information theory called the *inverse problem*, which is the problem of recovering information about unknown and uncontrolled components of a model from indirect noisy observations. The adjective *indirect* refers to the fact that although the observed data are considered to be directly influenced by the values of model components, the observations are not themselves the direct values of these components but only indirectly reflect the influence of the components. Thus, the relationship characterizing the effect of unobservable components on the observed data must be somehow inverted to recover information about the unobservable model components from the data-observations. Because econometric relations generally contain a systematic and a noise component, the problem of recovering information about unknowns and unobservables (θ , ϵ) from sample observations (y , x) within the context of an econometric model $Y = \eta(X, \theta, \epsilon)$ is referred to as an *inverse problem with noise* or as a *stochastic inverse problem*. A solution to this inverse problem is of the general form $(y, x) \Rightarrow (\theta, \epsilon)$. Because most econometric analyses are of this form, it would seem natural that solution methods should be used that are consistent with the underlying information recovery problem.

1.2 Organization of the Book

To establish notation and connect with reader knowledge, the book starts with the specification and analysis of the simplest parametric and semiparametric probability models. The book is organized in five parts.

Part I is concerned with estimation and inference procedures for parametric and semiparametric linear probability models. In Chapter 2, a notational basis for the econometric models ahead is specified. Estimation and inference is considered for both parametric and semiparametric models and the idea of extremum estimation is introduced. In Chapter 3, estimation and inference methods are introduced for obtaining information on parameters that are functionally related to moments of the data sampling distribution. The focus is on just and overdetermined cases and the chapter starts with an examination of both the method of moments and the generalized method of moments. Finally, the concept of estimating equations, which subsumes the moment equation approaches, is introduced. With an eye to the chapters ahead, we note how extremum estimation relates to the aforementioned methods.

In Part II, we leave the traditional econometric world and focus on econometric models in the form of ill-posed stochastic inverse problems, where information recovery is based on indirect noisy observations and asymptotic considerations come into play. In Chapter 4, a nonparametric stochastic linear inverse problem (econometric model) is defined and a solution (estimation method) is proposed. In Chapter 5, we consider the problem of inference as it is related to the evolving stochastic empirical likelihood inverse problem. In Chapter 6, we introduce the Kullback-Leibler information criterion and the Shannon-Jaynes maximum entropy principle, and provide an estimation and inference basis for the evolving maximum empirical exponential likelihood estimator.

In Part III, we introduce the Cressie-Read family goodness of fit-power divergence measures and provide, in Chapter 7, a framework for estimation and inference. In Chapter 8, sampling experiments are used to illustrate the finite sampling properties of this family of estimators. Recognizing there may be uncertainty regarding the specification of the estimating equations, a choice rule under quadratic loss is proposed.

In Part IV, we consider an information theoretic approach to the binary-discrete choice stochastic inverse problem. In Chapter 9, a minimum power divergence (MPD) family of distribution functions for the binary response discrete choice model is proposed, an estimation framework is specified, and the corresponding statistical properties are identified. In Chapter 10,

an estimation and inference basis for the binary MPD family is demonstrated and sampling experiments are used to illustrate finite sample results.

In Part V, we recognize that a basic limitation of traditional likelihood-divergence approaches is the impossibility of describing or identifying estimators-distributions of an arbitrary form. In Chapter 11, we address this estimation and inference problem by suggesting a loss-function-based way of choosing an optimum member from the Cressie-Read family.

Finally, in Chapter 12, we look back over the preceding chapters with a critical eye and make predictions about the econometric information theoretic road ahead.

1.3 Selected References

- Cavalier, L., 2008. "Nonparametric Statistic Inverse Problems," *Inverse Problems*, 24:1–19.
- Mittelhammer, R., M. Judge, and D. Miller, 2000. *Econometric Foundations*, Cambridge University Press, New York.

PART I

TRADITIONAL PARAMETRIC AND SEMIPARAMETRIC ECONOMETRIC MODELS: ESTIMATION AND INFERENCE

In Part I, we use a familiar data sampling process to focus on parametric and semiparametric econometric models. This grouping nicely reflects the information, real or imagined, that the analyst uses in terms of the economic and data sampling process that is being modeled. In contrast to fully defined parametric models, semiparametric models cannot be fully defined in terms of the values of a finite number of parameters. In particular, there is no assertion made that a particular parametric family of probability distribution is known that fully defines the probability distribution underlying the data sampling process. Building on this base, in Part I we move in the direction of extremum formulations for analyzing models of this type. For a more complete discussion of the material and relevant proofs in Chapters 2 and 3, see Mittelhammer, Judge, and Miller (2000) and Mittelhammer (1996).