Information and Entropy Econometrics — A Review and Synthesis

Information and Entropy Econometrics — A Review and Synthesis

Amos Golan

Department of Economics
American University
4400 Massachusetts Avenue
NW Washington
DC 20016-8029
USA
agolan@american.edu

now
the essence of knowledge

Boston – Delft

Foundations and Trends® in Econometrics

Published, sold and distributed by:
now Publishers Inc.
PO Box 1024
Hanover, MA 02339
USA
Tel. +1-781-985-4510
www.nowpublishers.com
sales@nowpublishers.com

Outside North America:
now Publishers Inc.
PO Box 179
2600 AD Delft
The Netherlands
Tel. +31-6-51115274

The preferred citation for this publication is A. Golan, Information and Entropy Econometrics — A Review and Synthesis, Foundations and Trends® in Econometrics, vol 2, no 1–2, pp 1–145, 2006

ISBN: 978-1-60198-104-2

Foundations and Trends® in Econometrics

Volume 2 Issue 1–2, 2006

Editorial Scope

Foundations and Trends® in Econometrics will publish survey and tutorial articles in the following topics:

- Identification
- Model Choice and Specification Analysis
- Non-linear Regression Models
- Simultaneous Equation Models
- Estimation Frameworks
- Biased Estimation
- Computational Problems
- Microeconometrics
- Treatment Modeling
- Discrete Choice Modeling
- Models for Count Data
- Duration Models
- Limited Dependent Variables
- Panel Data
- Dynamic Specification
- Inference and Causality
- Continuous Time Stochastic Models
- Modeling Non-linear Time Series
- Unit Roots
- Cointegration
- Latent Variable Models
- Qualitative Response Models
- Hypothesis Testing
- Interactions-based Models
- Duration Models
- Financial Econometrics
- Measurement Error in Survey Data
- Productivity Measurement and Analysis
- Semiparametric and Nonparametric Estimation
- Bootstrap Methods
- Nonstationary Time Series
- Robust Estimation

Information for Librarians

Foundations and Trends® in Econometrics, 2006, Volume 2, 2 issues. ISSN paper version 1551-3076. ISSN online version 1551-3084. Also available as a combined paper and online subscription.

Foundations and Trends® in
Econometrics
Vol. 2, Nos. 1–2 (2006) 1–145

DOI: 10.1561/0800000004

Information and Entropy Econometrics — A Review and Synthesis*

Amos Golan

Department of Economics, American University, 4400 Massachusetts Avenue, NW Washington, DC 20016-8029, USA, agolan@american.edu

This work is dedicated to Marge and George Judge

Abstract

The overall objectives of this review and synthesis are to study the basics of information-theoretic methods in econometrics, to examine the connecting theme among these methods, and to provide a more detailed summary and synthesis of the sub-class of methods that treat the observed sample moments as stochastic. Within the above objectives, this review focuses on studying the inter-connection between information theory, estimation, and inference. To achieve these objectives, it provides a detailed survey of information-theoretic concepts and quantities used within econometrics. It also illustrates

* Part of this study was done during the senior fellowship at the Institute of Advanced Study and the faculty of statistics of the University of Bologna. The author thanks the Institute and the University for their Support. Detailed Comments from Douglas Miller and Mirko Degli Esposti and an anonymous reviewer on an earlier version of this manuscript, and comments from Essie Maasoumi on earlier sections of this survey, are greatly appreciated. The author also thanks Doug Miller for pointing some unknown references. Finally, the author thanks the Editor — Bill Greene — for all his helpful comments and suggestions throughout the process of composing this review.

the use of these concepts and quantities within the subfield of information and entropy econometrics while paying special attention to the interpretation of these quantities. The relationships between information-theoretic estimators and traditional estimators are discussed throughout the survey. This synthesis shows that in many cases information-theoretic concepts can be incorporated within the traditional likelihood approach and provide additional insights into the data processing and the resulting inference.

Keywords: Empirical likelihood; entropy, generalized entropy; information; information theoretic estimation methods; likelihood; maximum entropy; stochastic moments.

JEL codes: C13, C14, C49, C51

Preface

This review and synthesis is concerned with information and entropy econometrics (IEE). The overall objective is to summarize the basics of information-theoretic methods in econometrics and the connecting theme among these methods. The sub-class of methods that treat the observed sample moments as stochastic is discussed in greater detail.

Within the above objective, we restrict our attention to study the interconnection between information theory, estimation, and inference. We provide a detailed survey of information-theoretic concepts and quantities used within econometrics and then show how these quantities are used within IEE. We pay special attention for the interpretation of these quantities and for describing the relationships between information-theoretic estimators and traditional estimators.

In Section 1, an introductory statement and detailed objectives are provided. Section 2 provides a historical background of IEE. Section 3 surveys some of the basic quantities and concepts of information theory. This survey is restricted to those concepts that are employed within econometrics and that are used within that survey. As many of these concepts may not be familiar to many econometricians and economists, a large number of examples are provided. The concepts discussed include entropy, divergence measures, generalized entropy (known also as Cressie Read function), errors and entropy, asymptotic theory, and stochastic processes. However, it is emphasized that this is not a survey of information theory. A less formal discussion providing interpretation of information, uncertainty, entropy and ignorance, as viewed by scientists across disciplines, is provided at the beginning of that section. In Section 4, we discuss the classical maximum entropy (ME) principle (both the primal constrained model and the dual concentrated and unconstrained model) that is used for estimating under-determined, zero-moment problems. The basic quantities discussed in Section 3, are revisited again in connection with the ME principle. In Section 5, we discuss the motivation for information-theoretic (IT) estimators and then formulate the generic IT estimator as a constrained optimization problem. This generic estimator encompasses all the estimators within the class of IT estimators. The rest of this section describes the basics of specific members of the IT class of estimators. These members compose the sub-class of methods that incorporate the moment restrictions within the generic IT-estimator as (pure) zero moments' conditions, and include the empirical likelihood, the generalized empirical likelihood, the generalized method of moments and the Bayesian method of moments. The connection between each one of these methods, the basics of information theory and the maximum entropy principle is discussed. In Section 6, we provide a thorough discussion of the other sub-class of IT estimators: the one that views the sample moments as stochastic. This sub-class is also known as the generalized maximum entropy. The relevant statistics and information measures are summarized and connected to quantities studied earlier in the survey. We conclude with a simple simulated example. In Section 7, we provide a synthesis of likelihood, ME and other IT estimators, via an example. We study the interconnections among these estimators and show that though coming from different philosophies they are deeply rooted in each other, and understanding that interconnection allows us to understand our data better. In Section 8, we summarize related topics within IEE that are not discussed in this survey.

Readers of this survey need basic knowledge of econometrics, but do not need prior knowledge of information theory. Those who are familiar with the concepts of IT should skip Section 3, except Section 3.4 which is necessary for the next few sections. Those who are familiar with the ME principle can skip parts of Section 4, but may want to read the example in Section 4.7. The survey is self contained and interested readers can replicate all results and examples provided. No detailed proofs are provided, though the logic behind some less familiar arguments is provided. Whenever necessary the readers are referred to the relevant literature.

This survey may benefit researchers who wish to have a fast introduction to the basics of IEE and to acquire the basic tools necessary for using and understanding these methods. The survey will also benefit applied researchers who wish to learn improved new methods, and applications, for extracting information from noisy and limited data and for learning from these data.

Contents

1 Introductory Statement, Motivation, and Objective **1**

2 Historical Perspective **7**

3 Information and Entropy — Background, Definitions, and Examples **15**

3.1 Information and Entropy — The Basics 15
3.2 Discussion 26
3.3 Multiple Random Variables, Dependency, and Joint Entropy 28
3.4 Generalized Entropy 32
3.5 Axioms 34
3.6 Errors and Entropy 36
3.7 Asymptotics 39
3.8 Stochastic Process and Entropy Rate 45
3.9 Continuous Random Variables 46

4 The Classical Maximum Entropy Principle **49**

4.1 The Basic Problem and Solution 49
4.2 Duality — The Concentrated Model 51

4.3 The Entropy Concentration Theorem 55
4.4 Information Processing Rules and Efficiency 57
4.5 Entropy — Variance Relationship 58
4.6 Discussion 59
4.7 Example 61

5 Information-Theoretic Methods of Estimation — I 63

5.1 Background 63
5.2 The Generic IT Model 64
5.3 Empirical Likelihood 67
5.4 Generalized Method of Moments — A Brief Discussion 72
5.5 Bayesian Method of Moments — A Brief Discussion 81
5.6 Discussion 83

6 Information-Theoretic Methods of Estimation — II 85

6.1 Generalized Maximum Entropy — Basics 85
6.2 GME — Extensions: Adding Constraints 93
6.3 GME — Entropy Concentration Theorem and Large Deviations 94
6.4 GME — Inference and Diagnostics 96
6.5 GME — Further Interpretations and Motivation 99
6.6 Numerical Example 103
6.7 Summary 105

7 IT, Likelihood and Inference 107

7.1 The Basic Problem — Background 107
7.2 The IT-ME Solution 109
7.3 The Maximum Likelihood Solution 111
7.4 The Generalized Case — Stochastic Moments 112
7.5 Inference and Diagnostics 114

7.6 IT and Likelihood 117
7.7 Conditional, Time Series Markov Process 118

8 Concluding Remarks and Related Work Not Surveyed 131

References 137

1

Introductory Statement, Motivation, and Objective

All learning, information gathering and information processing, is based on limited knowledge, both *a priori* and data, from which a larger "truth" must be inferred. To learn about the true state of the world that generated the observed data, we use statistical models that represent these outcomes as functions of unobserved structural parameters, parameters of priors and other sampling distributions, as well as complete probability distributions. Since we will never know the true state of the world, we generally focus, in statistical sciences, on recovering information about the complete probability distribution, which represents the ultimate truth in our model. Therefore, all estimation and inference problems are translations of limited information about the probability density function (pdf) toward a greater knowledge of that pdf. However, if we knew all the details of the true mechanism then we would not need to resort to the use of probability distributions to capture the perceived uncertainty in outcomes that results from our ignorance of the true underlying mechanism that controls the event of interest.

Information theory quantities, concepts, and methods provide a unified set of tools for organizing this learning process. They provide a discipline that at once exposes more clearly what the existing methods do, and how we might better accomplish the main goal of scientific learning. This review first studies the basic quantities of information theory and their relationships to data analysis and information processing, and then uses these quantities to summarize (and understand the connection among) the improved methods of estimation and data processing that compose the class of entropy and information-theoretic methods. Within that class, the review concentrates on methods that use conditional and unconditional stochastic moments.

It seems natural to start by asking what is information, and what is the relationship between information and econometric, or statistical analysis. Consider, for example, Shakespeare's "Hamlet," Dostoevsky's "The Brothers Karamazov," your favorite poem, or the US Constitution. Now think of some economic data describing the demand for education, or survey data arising from pre-election polls. Now consider a certain speech pattern or communication among individuals. Now imagine you are looking at a photo or an image. The image can be sharp or blurry. The survey data may be easy to understand or extremely noisy. The US Constitution is still being studied and analyzed daily with many interpretations for the same text, and your favorite poem, as short as it may be, may speak a whole world to you, while disliked by others.

Each of these examples can be characterized by the amount of information it contains or by the way this information is conveyed or understood by the observer — the analyst, the reader. But what is information? What is the relationship between information and econometric analysis? How can we efficiently extract information from noisy and evolving complex observed economic data? How can we guarantee that only the relevant information is extracted? How can we assess that information? The study of these questions is the subject of this survey and synthesis.

This survey discusses the concept of *information* as it relates to econometric and statistical analyses of data. The meaning of "information" will be studied and related to the basics of Information Theory (IT) as is viewed by economists and researchers who are engaged in deciphering information from the (often complex and evolving) data, while taking into account what they know about the underlying process that generated these data, their beliefs about the (economic) system under investigation, and nothing else. In other words, the researcher wishes to extract the available information from the data, but wants to do it with minimal *a priori* assumptions. For example, consider the following problem taken from Jaynes's famous Brandeis lectures (1963). We know the empirical mean value (first moment) of, say one million tosses of a six-sided die. With that information the researcher wishes to predict the probability that in the next throw of the die we will observe the value 1, 2, 3, 4, 5 or 6. The researcher also knows that the probability is proper (sum of the probabilities is one). Thus, in that case, there are six values to predict (six unknown values) and two observed (known) values: the mean and the sum of the probabilities. As such, there are more unknown quantities than known quantities, meaning there are infinitely many probability distributions that sum up to one and satisfy the observed mean. In somewhat more general terms, consider the problem of estimating an unknown discrete probability distribution from a finite and possibly noisy set of observed (sample) moments. These moments (and the fact that the distribution is proper — summing up to one) are the only available information. Regardless of the level of noise in these observed moments, if the dimension of the unknown distribution is larger than the number of observed moments, there are infinitely many proper probability distributions satisfying this information (the moments). Such a problem is called an under-determined problem. Which one of the infinitely many solutions should one use? In all the IEE methods, the one solution chosen is based on an information criterion that is related to Shannon's information measure — entropy.

When analyzing a linear regression, a jointly determined system of equations, a first-order Markov model, a speech pattern, a blurry image, or even a certain text, if the researcher wants to understand the data but without imposing a certain structure that may be inconsistent with the (unknown) truth, the problem may become inherently under-determined. The criterion used to select the desired solution is an information criterion which connects statistical estimation and inference with the foundations of IT. This connection provides us with an

IT perspective of econometric analyses and reveals the deep connection among these "seemingly distinct" disciplines. This connection gives us the additional tools for a better understanding of our limited data, and for linking our theories with real observed data. In fact, information theory and data analyses are the major thread connecting most of the scientific studies trying to understand the true state of the world with the available, yet limited and often noisy, information.

Within the econometrics and statistical literature the family of IT estimators composes the heart of IEE. It includes the Empirical (and Generalized Empirical) Likelihood, the Generalized Method of Moments, the Bayesian Method of Moments and the Generalized Maximum Entropy among others. In all of these cases the objective is to extract the available information from the data with minimal assumptions on the data generating process and on the likelihood structure. The logic for using minimal assumptions in the IEE class of estimators is that the commonly observed data sets in the social sciences are often small, the data may be non-experimental noisy data, the data may be arising from a badly designed experiment, and the need to work with nonlinear (macro) economic models where the maximum likelihood estimator is unattractive as it is not robust to the underlying (unknown) distribution. Therefore, (i) such data may be ill-behaved leading to an ill-posed and/or ill-conditioned (not full rank) problem, or (ii) the underlying economic model does not specify the complete distribution of the data, but the economic model allows us to place restrictions on this (unknown) distribution in the form of population moment conditions that provide information on the parameters of the model. For these estimation problems and/or small and non-experimental data it seems logical to estimate the unknown parameters with minimum *a priori* assumptions on the data generation process, or with minimal assumptions on the likelihood function. Without a pre-specified likelihood, other non maximum likelihood methods must be used in order to extract the available information from the data. Many of these methods are members of the class of Information-Theoretic (IT) methods.

This survey concentrates on the relationship between econometric analyses, data and information with an emphasis on the philosophy leading to these methods. Though, a detailed exposition is provided here, the focus of this survey is on the sub-class of IT estimators that view the observed moments as stochastic. Therefore, the detailed formulations and properties of the other sub-class of estimators that view the observed moments as (pure) zero-moment conditions will be discussed here briefly as it falls outside the scope of that review and because there are numerous recent reviews and texts of these methods (e.g., Smith, 2000, 2005, 2007; Owen, 2001; Hall, 2005; Kitamura, 2006). However, the connection to IT and the ME principle, and the inter-relationships among the estimators, is discussed here as well.

References

Agmon, N., Y. Alhassid, and R. D. Levine (1979), 'An algorithm for finding the distribution of maximal entropy'. *Journal of Computational Physics* **30**, 250–259.

Antoine, B., H. Bonnal, and E. Renault (2007), 'On the efficient use of the informational content of estimating equations: Implied probabilities and Euclidean likelihood'. *Journal of Econometrics* **138**, 461–487.

Bayes, R. T. (1763), 'An essay toward solving a problem in the doctrine of chances'. *Philosophical Transactions of the Royal Soceity of London* **53**, 370–418.

Bera, A. K. and Y. Bilias (2002), 'The MM, ME, ML, EL, EF and GMM approaches to estimation: A synthesis'. *Journal of Econometrics* **107**, 51–86.

Bercher, J. F., G. L. Besnerais, and G. Demoment (1996), 'The maximum entropy on the mean method, noise and sensitivity'. In: J. Skiling and S. Sibisi (eds.): *Maximum Entropy and Bayesian Studies*. Kluwer.

Besnerais, L. G., J.-F. Bercher, and G. Demoment (1999), 'A new look at entropy for solving linear inverse problems'. *IEEE Transactions on Information Theory* **45**, 1565–1578.

Boltzmann, L. W. (1871). *Weiner Berichte* **63**, 397–418; 679–711; 712–732. Translated later.

Boole, G. (1854), *An Investigation of the Laws of Thought*. London: MacMillan. Reprinted by Dover Pub., Inc., New York, 1958.

Bryant, P. G. and O. I. Cordero-Brana (2000), 'Model selection using the minimum description length pronciple'. *The American Statistician* **54**, 257.

Burg, J. P. (1975), 'Maximum Entropy and Spectral Analysis'. Ph.D. thesis, Stanford University Press.

Chamberlain, G. (1987), 'Asymptotic efficiency in estimation with conditional moment restrictions'. *Journal of Econometrics* **34**, 305–334.

Chernoff, H. (1952), 'A measure of asymptotic efficiency for tests of hypothesis based on the sum of observations'. *Annals of Mathematical Statistics* **23**, 493–507.

Clarke, B. (2007), 'Information optimality and Bayesian modeling'. *Journal of Econometrics* **138**, 405–429.

Cover, T. M. and J. A. Thomas (1991), *Elements of Information Theory*. New York: John Wiley & Sons.

Cox, R. T. (1946), 'Probability, frequency and reasonable expectation'. *American Journal of Physics* **14**, 1–13.

Cressie, N. and T. R. C. Read (1984), 'Multinomial goodness-of-fit tests'. *Journal of Royal Statistical Soceity B* **46**, 440–464.

Csiszar, I. (1984), 'Sanov property, generalized I-projection and a conditional limit theorem'. *The Annals of Probability* **12**, 768–793.

Csiszar, I. (1991), 'Why least squares and maximum entropy? An axiomatic approach to inference for linear inverse problems'. *The Annals of Statistics* **19**, 2032–2066.

Csiszar, I. (1996). In: K. M. Hanson and R. N. Silver (eds.): *Maximum Entropy and Bayesian Studies*. Dordrecht: Kluwer.

Csiszar, I. and J. Korner (1981), *Information Theory: Coding Theorems for Discrete*. Budapest: Akademiai Kiado and New York: Academic Press.

Csiszar, I. and P. C. Shields (2004), 'Information theory and statistics: A tutorial'. *Foundations and Trends in Communications and Information Theory* **1**(4), 417–528.

Davis, H. T. (1941), *The Theory of Econometrics*. Indiana: The Principia Press.

Dembo, A. and O. Zeitouni (1998), *Large Deviations Techniques and Applications*. Second edition. Springer Verlag.

DiCiccio, T. J., P. Hall, and J. Romano (1991), 'Empirical likelihood is Bartlett-correctable'. *The Annals of Statistics* **19**, 1053–1991.

Donoho, D. L., I. M. Johnstone, J. C. Hoch, and A. S. Stern (1992), 'Maximum entropy and the nearly black object'. *Journal of the Royal Statistical Society Series B* **54**, 41–81.

Durbin, J. (1960), 'Estimation of parameters in time-series regression models'. *Journal of the Royal Statistical Society Series B* **22**, 139–153.

Ebrahimi, N., M. Habibullah, and E. S. Soofi (1992), 'Testing exponentiality based on Kullback-Leibler information'. *Journal of the Royal Statistical Society Series B* **54**(3), 739–748.

Econometric Reviews (2007). Special issue on IEE (Forthcoming).

Efron, B. and R. J. Tibshirani (1993), *An Introduction to the Bootstrap*. Chapman & Hall.

Ertl, K., W. von der Linden, V. Dose, and A. Weller (1996), 'Maximum entropy based reconstruction of soft X-ray emissivity profiles in W7-AS'. *Nuclear Fusion* **36**(11), 1477–1488.

Esscher, F. (1932), 'On the pProbability function in the collective theory of risk'. *Skandinavisk Aktuariedskrift* **15**, 175–195.

Fano, R. M. (1961), *Transmission of Information: A Statistical Theory of Communication*. New York: John Wiley.

Ferguson, T. S. (1958), 'A method of generating best asymptotically normal estimates with application to estimation of bacterial densities'. *Annals of Mathematical Statistics* **29**, 1046–1062.

Fisher, R. A. (1912), 'On an absolute criterion for fitting frequency curves'. *Messenger of Mathematics* **41**, 155–160.

Fisher, R. A. (1922), 'On the mathematical foundations of theoretical statistics'. *Philosophical Transactions of the Royal Society of London Series A* **222**, 309–368.

Gamboa, F. and E. Gassiat (1997), 'Bayesian methods and maximum entropy for ill-posed problems'. *The Annals of Statistics* **25**, 328–350.

Gerber, H. U. (1980), 'A characterization of certain families of distributions via essche transforms and independence'. *Journal of the American Statistical Association* **75**(372), 1015–1018.

Gibbs, J. W. (1902), *Elementary Principles in Statistical Mechanics*. New Haven, CT: Yale University Press.

Gokhale, D. and S. Kullback (1978), *The Information in Contingency Tables*. New York: Marcel Dekker.

Golan and D. Volker (2001), 'A generalized information theoretical approach to tomographic reconstruction'. *Journal of Physics A: Mathematical and General,* pp. 1271–1283.

Golan, A. (1988), *A discrete-stochastic model of economic production and a model of production fluctuations – theory and empirical evidence*. PhD Dissertation, UC Berkeley.

Golan, A. (2001), 'A simultaneous estimation and variable selection rule'. *Journal of Econometrics* **101**(1), 165–193.

Golan, A. (2002), 'Information and entropy econometrics – Editor's view'. *Journal of Econometrics* **107**(1–2), 1–15.

Golan, A. and H. Gzyl (2002), 'A generalized maxentropic inversion procedure for noisy data'. *Applied Mathematics and Computation* **127**, 249–260.

Golan, A. and H. Gzyl (2003), 'Priors and information-theoretic estimation'. *American Statistical Association Proceedings*.

Golan, A. and H. Gzyl (2006), 'Ill conditioned linear estimation problems: An information theoretic approach with priors'. Working Paper.

Golan, A. and G. Judge (1992), 'Recovering and processing information in the case of underdetermined economic inverse models'. Unpublished Manuscript, University of California, Berkeley.

Golan, A. and G. Judge (1993), *Recovering the parameters and forecasting in the case of ill-posed non-stationary inverse problems*. Unpublished manuscript, University of California, Berkeley.

Golan, A. and G. Judge (1996), 'A maximum entropy approach to empirical likelihood estimationand inference'. Unpublished Paper, University of California, Berkeley. (Presented at the 1997 Summer Econometric Society meetings).

Golan, A., G. Judge, and L. Karp (1996a), 'A maximum entropy approach to estimation and inference in dynamic models or counting fish in the sea with maximum entropy'. *Journal of Economic Dynamics and Control* **20**, 559–582.

Golan, A., G. Judge, and D. Miller (1996b), *Maximum Entropy Econometrics: Robust Estimation with Limited Data*. New York: John Wiley & Sons.

Golan, A., G. Judge, and J. Perloff (1996c), 'A generalized maximum entropy approach to recovering information from multinomial response data'. *Journal of the American Statistical Association* **91**, 841–853.

Golan, A., G. Judge, and J. Perloff (1997), 'Estimation and inference with censored and ordered multinomial response data'. *Journal of Econometrics* **79**, 23–51.

Golan, A., G. G. Judge, and S. Robinson (1994), 'Recovering information from incomplete or partial multisectoral economic data'. *Review of Economics and Statistics* **76**(3), 541–549.

Golan, A., L. Karp, and J. M. Perloff (2000), 'Estimating firms' mixed price and advertising strategies: Coke and Pepsi'. *Journal of Business Economic Statistics* **18**(4), 398–409.

Golan, A., J. Lane, and E. McEntarfe (2007), 'The dynamics of worker reallocation within and across industries'. *Economica* **74**, 1–20.

Golan, A. and J. Perloff (2002), 'Comparison of maximum entropy and higher-order entropy estimators'. *Journal of Econometrics* **107**(1–2), 195–211.

Golan, A., J. Perloff, and Z. Shen (2001), 'Estimating a demand system with non-negativity constraints: Mexican meat demand'. *Review of Economics and Statistics* **83**(3), 541–550.

Golan, A. and S. Vogel (2000), 'Estimation of non-stationary social accounting matrix coefficients with supply-side information'. *Economic Systems Research* **12**(4), 447–471.

Good, I. J. (1963), 'Maximum entropy for hypothesis formulation, especially for multidimensional contingency tables'. *Annals of Mathematical Statistics* **34**, 911–934.

Greene, W. H. (2008), *Econometric Analysis*. Sixth edition. Prentice Hall.

Gzyl, H. (1993), *The Maximum Entropy Method*. Singapore: Wirld Scientific Publishing.

Gzyl, H. (1998). In: G. Erickson (ed.): *Maximum Entropy and Bayesian Studies*. Dordrecht: Kluwer.

Hall, A. R. (2005), *Generalized Method of Moments*, Advanced Texts in Econometrics Series. Oxford University Press.

Hall, A. R., A. Inoue, K. Jana, and C. Shin (2007a), 'Information in eneralized method of moments estimation and entropy based moment selection'. *Journal of Econometrics* **138**, 488–512.

Hall, A. R., A. Inoue, and C. Shin (2007b), 'Entropy based moment selection in the presence of weak identification'. *Econometric Reviews*. (Forthcoming).

Hall, P. (1990), 'Pseudo-likelihood theory for empirical likelihood'. *The Annals of Statistics* **18**, 121–140.

Hansen, L. P. (1982), 'Large sample properties of generalized methods of moments estimators'. *Econometrica* **50**, 1029–1054.

Hansen, L. P., J. Heaton, and A. Yaron (1996), 'Finite-sample properties of some alternative GMM estimators'. *Journal of Business and Economic Statistics* **14**, 262–280.

Hanson, K. M. and R. N. Silver (1996), *Maximum Entropy and Bayesian Methods*. Dordrecht: Kluwer.

Hartley, R. V. L. (1928), 'Transmission of information'. *Bell System Technical Journal,* pp. 535–563.

Hellerstein, J. K. and G. W. Imbens (1999), 'Imposing moment restrictions from auxiliary data by weighting'. *The Review of Economics and Statistics* **81**, 1–14.

Holste, D., I. Grobe, and H. Herzel (1998), 'Bayes' estimators of generalized entropies'. *Journal of Physics A: Mathematical and General* **31**, 2551–2566.

Imbens, G. and R. Spady (2001), 'The performance of empirical likelihood and its generalizations'. Unpublished working paper, Department of Economics, University of California, Berkeley.

Imbens, G. W. (1997), 'One-step estimators for over-identified generalized method of moments models'. *Review of Economic Studies* **64**, 359–383.

Imbens, G. W. (2002), 'Generalized method of moments and empirical likelihood'. *Journal of Business & Economic Statistics* **20**, 493–507.

Imbens, G. W., P. Johnson, and R. H. Spady (1998), 'Information-theoretic approaches to inference in moment condition models'. *Econometrica* **66**, 333–357.

Jaynes, E. T. (1957a), 'Information theory and statistical mechanics'. *Physics Review* **106**, 620–630.

Jaynes, E. T. (1957b), 'Information theory and statistical mechanics II'. *Physics Review* **108**, 171–190.

Jaynes, E. T. (1963), 'Information theory and statistical mechanics'. In: K. W. Ford (ed.): *Statistical Physics*. New York: W. A. Benjamin, Inc., MIT Press, pp. 181–218.

Jaynes, E. T. (1978a), 'Information theory and statistical mechanics II'. In: K. W. Ford (ed.): *Statistical Physics*. New York: W.A. Benamin, Inc., pp. 181–218.

Jaynes, E. T. (1978b), 'Where do we stand on maximum entropy'. In: R. D. Levine and M. Tribus (eds.): *The Maximum Entropy Formalism*. MIT Press, pp. 15–118.

Jaynes, E. T. (1984), 'Prior information and ambiguity in inverse problems'. In: D. W. McLaughlin (ed.): *Inverse Problems, SIAM Proceedings*. Providence, RI: American Mathematical Society, pp. 151–166.

Jaynes, E. T. (2003), *Probability Theory: The Logic of Sciences*. Cambridge University Press.

Jeffreys, H. (1939), *Theory of Probability*. Oxford: Clarendon Press (later editions: 1948, 1961, 1967, 1988).

Journal of Econometrics (2002), 'Special issue on IEE'. **107**, 1–374.

Journal of Econometrics (2007), 'Special issue on IEE'. **138**, 379–585.

Judge, G. and R. C. Mittelhammer (2007), 'Estimation and inference in the case of competing sets of estimating equations'. *Journal of Econometrics* **138**, 513–531.

Kitamura, Y. (2006), 'Empirical likelihood methods in econometrics: Theory and practice'. Cowles Foundation Discussion Paper No. 1569.

Kitamura, Y. and T. Otsu (2005), 'Minimax estimation and testing for moment condition models via large deviations'. Manuscript, Department of Economics, Yale University.

Kitamura, Y. and M. Stutzer (1997), 'An information theoretic alternative to generalized method of moments estimation'. *Econometrica* **65**(4), 861–874.

Kitamura, Y. and M. Stutzer (2002), 'Corrections between entropic and linear projections in asset pricing estimation'. *Journal of Econometrics* **107**, 159–174.

Kitamura, Y., G. Tripathi, and H. Ahn (2004), 'Empirical likelihood based inference in conditional moment restriction models'. *Econometrica* **72**, 1667–1714.

Koehler, K. J. and K. Larntz (1980), 'An empirical investigation of goodness-of-fit statistics for sparse multinomial'. *Journal of the American Statistical Association* **75**, 336–344.

Kullback, S. (1954), 'Certain inequalities in information theory and the Cramer-Rao inequality'. *The Annals of Mathematical Statistics* **25**, 745–751.

Kullback, S. (1959), *Information Theory and Statistics*. New York: John Wiley & Sons.

Kullback, S. and R. A. Leibler (1951), 'On information and sufficiency'. *Annals of Mathematical Statistics* **22**, 79–86.

LaFrance, J. T. (1999), 'Inferring the nutrients with prior information'. *American Journal of Agricultural Economics* **81**, 728–734.

Laplace, P. S. (1774), 'Memoire sur la probabilite des causes par les evenemens'. *Mémoires de l'Académie Royale des Sciences* **6**, 621–656. Reprinted in Laplace (1878-1912), **8**, 27–65.

Levine, R. D. (1980), 'An information theoretical approach to inversion problems'. *Journal of Physics A* **13**, 91–108.

Levine, R. D. and M. Tribus (1979), *The Maximum Entropy Formalism*. Cambridge, MA: MIT Press.

Lindley, D. V. (1956), 'On a measure of the information provided by an experiment'. *Annals of Mathematics* **27**, 986–1005.

Maasoumi, E. (1993), 'A compendium to information theory in economics and econometrics'. *Econometric Reviews* **12**, 137–181.

Maasoumi, E. and J. Racine (2002), 'Entropy and predictability in the stock market returns'. *Journal of Econometrics* **107**, 291–312.

MacKay, D. J. C. (2003), *Information Theory, Inference, and Learning Algorithms*. Cambridge University Press.

Maxwell, J. C. (1859), 'Illustrations of the dynamical theory of gases'. In: W. D. Niven (ed.): *Collected Works*. London (1890), Vol. I: pp. 377–409.

Maxwell, J. C. (1876), 'On Boltzmann's theorem on the average distribution of energy in a system of material points'. In: W. D. Niven (ed.): *Collected Works*. London (1890), Vol. II: pp. 713–741.

McFadden, D. (1974), 'Conditional logit analysis of qualitative choice behavior'. In: P. Zarembka (ed.): *Frontiers of Econometrics*. New York: Academic Press, pp. 105–142.

Miller, D. J. (1994), Entropy and information recovery in linear economic models. PhD Dissertation, UC Berkeley.

Mittelhammer, R. and S. Cardell (1996), *On the Consistency and Asymptotic Normality of the Data Constrained GME Estimator of the GML* (mimeo). Pullman, WA: Washington State University.

Mittelhammer, R. C., G. G. Judge, and D. J. Miller (2000), *Econometric Foundations*. Cambridge: Cambridge University Press.

Navaza, J. (1986), 'The use of non-local constraints in maximum-entropy electron density reconstruction'. *Acta Crystallographica* **A42**, 212–223.

Newey, W. and D. McFadden (1994), 'Large sample estimation and hypothesis testing'. In: Engle and McFadden (eds.): *The Handbook of Econometrics* Vol. 4. New York: North-Holland.

Newey, W. K. and R. J. Smith (2004), 'Higher order properties of GMM and generalized empirical likelihood estimators'. *Econometrica* **72**, 219–255.

Neyman, J. and E. S. Pearson (1928a), 'On the use and interpretation of certain test criteria for purposes of statistical inference: Part I'. *Biomentrica* **20A**, 175–240.

Neyman, J. and E. S. Pearson (1928b), 'On the use and interpretation of certain test criteria for purposes of statistical inference: Part II'. *Biomentrica* **20A**, 263–294.

O'Sullivan, F. (1986), 'A statistical perspective on ill-posed inverse problems'. *Statistical Science* **1**, 502–527.

Otsu, T. (2006), 'Generalized empirical likelihood under weak identification'. *Econometric Theory* **22**, 513–527.

Owen, A. (1988), 'Empirical likelihood ratio confidence intervals for a single functional'. *Biometrika* **75**(2), 237–249.

Owen, A. (1990), 'Empirical likelihood ratio confidence regions'. *The Annals of Statistics* **18**(1), 90–120.

Owen, A. (1991), 'Empirical likelihood for linear models'. *The Annals of Statistics* **19**(4), 1725–1747.

Owen, A. (2001), *Empirical Likelihood*. Chapman & Hall/CRC.

Pearson, K. (1894), 'Contribution to the mathematical theory of evolution'. *Philosophical Transactions of the Royal Society of London Series A* **185**, 71–110.

Pearson, K. (1900), 'On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling'. *Philosophical Magazine Series* **50**, 157–175.

Perloff, J., L. Karp, and A. Golan (2007), *Estimating Market Power and Strategies*. Cambridge University Press.

Pukelsheim, F. (1994), 'The three sigma rule'. *American Statisticia* **48**, 88–91.

Qin, J. and J. Lawless (1994), 'Empirical likelihood and general estimating equations'. *The Annals of Statistics* **22**, 300–325.

Racine, J. and E. Maasoumi (2007), 'A versatile and robust metric entropy test of time-reversibility, and other hypotheses'. *Journal of Econometrics* **138**, 547–567.

Ramalho, J. J. S. and R. J. Smith (2002), 'Generalized empirical likelihood non-nested tests'. *Journal of Econometrics* **107**, 99–125.

Rényi, A. (1961), 'On measures of entropy and information'. In: *Proceedings of the 4th Berkeley Symposium on Mathematical Statistics and Probability,* Vol. I. Berkeley: University of California Press, pp. 547–561.

Rényi, A. (1970), *Probability Theory*. Amsterdam: North-Holland.

Sanov, I. N. (1961), 'On the probability of large deviations of random variables'. *Selected Translations in Mathematical Statistics and Probability* **I**, 213–244.

Sargan, J. D. (1958), 'The estimation of economic relationships using instrumental variables'. *Econometrica* **26**, 393–415.

Sargan, J. D. (1959), 'The estimation of relationships with autocorrelated residuals by the use of instrumental variables'. *Journal of the Royal Statistical Society Series B* **21**, 91–105.

Schennach, S. M. (2004), 'Exponentially tilted empirical likelihood'. Discussion Paper, University of Chicago.

Shannon, C. E. (1948), 'A mathematical theory of communication'. *Bell System Technical Journal* **27**, 379–423; 623–656.

Shore, J. E. and R. Johnson (1980), 'Axiomatic derivation of the principle of maximum entropy and the principle of minimum cross-entropy'. *IEEE Transactions on Information Theory* **IT-26**(1), 26–37.

Skilling, J. (1989a), 'The axioms of maximum entropy'. In: J. Skilling (ed.): *Maximum Entropy and Bayesian Methods in Science and Engineering*. Dordrecht: Kluwer Academic, pp. 173–187.

Skilling, J. (1989b), 'Classic maximum entropy'. In: J. Skilling (ed.): *Maximum Entropy and Bayesian Methods*. Dordrecht: Kluwer Academic Publishing, pp. 45–52.

Smith, R. J. (1997), 'Alternative semi-parametric likelihood approaches to generalized method of moments estimation'. *Economic Journal* **107**, 503–519.

Smith, R. J. (2000), 'Empirical likelihood estimation and inference'. In: M. Salmon and P. Marriott (eds.): *Applications of Differential Geometry to Econometrics*. Cambridge: Cambridge University Press, pp. 119–150.

Smith, R. J. (2004), *GEL criteria for moment condition models*. University of Warwick.

Smith, R. J. (2005), Local GEL methods for conditional moment restrictions. Working Paper, University of Cambridge.

Smith, R. J. (2007), 'Efficient information theoretic inference for conditional moment restrictions'. *Journal of Econometrics* **138**, 430–460.

Solomon, H. and M. A. Stephens (1978), 'Approximations to density functions using pearson curves'. *Journal of the American Statistical Association* **73**(361), 153–160.

Soofi, E. S. (1994), 'Capturing the intangible concept of information'. *Journal of the American Statistical Association* **89**(428), 1243–1254.

Soofi, E. S. (1996), 'Information theory and bayesian statistics'. In: D. A. Berry, K. M. Chaloner, and J. K. Geweke (eds.): *Bayesian Analysis in Statistics and Econometrics: Essays in Honor of Arnold Zellnge*. New York: Wiley, pp. 179–189.

Soofi, E. S. (2000), 'Principal information theoretic approaches'. *Journal of the American Statistical Association* **95**(452), 1349–1353.

Soofi, E. S. and J. J. Retzer (2002), 'Information indices: Unifications and applications'. *Journal of Econometrics* **107**, 17–40.

Stutzer, M. J. (2000), 'Simple entropic derivation of a generalized black-scholes model'. *Entropy* **2**, 70–77.

Stutzer, M. J. (2003a), 'Fund managers may cause their benchmarks to be priced 'Risks''. *Journal of Investment Management*. (Reprinted in Fong (ed.): *The World of Risk Management*, World Scientific Publishing, 2006).

Stutzer, M. J. (2003b), 'Optimal asset allocation for endowments: A large deviations approach'. In: Satchell and Scowcraft (eds.): *Advances in Portfolio Construction and Implementation*. Butterworth-Heineman.

Stutzer, M. J. (2003c), 'Portfolio choice with endogenous utility: A large deviations approach'. *Journal of Econometrics* **116**, 365–386.

Tikochinsky, Y., N. Z. Tishby, and R. D. Levine (1984), 'Consistent inference of probabilities for reproducible experiments'. *Physics Review Letters* **52**, 1357.

Tobias, J. and A. Zellner (2001), 'Further results on Bayesian method of moments analysis of the multiple regression model'. *International Economic Review* **42**(1), 121–139.

Tripathi, G. and Y. Kitamura (2003), 'Testing conditional moment restrictions'. *Annals of Statisics* **31**, 2059–2095.

Tsallis, C. (1988), 'Possible generalization of Boltzmann-Gibbs statistics'. *Journal of Statistical Physics* **52**, 479.

Ullah, A. (2002), 'Uses of entropy and divergence measures for evaluating econometric approximations and inference'. *Journal of Econometrics* **107**, 313–326.

Vasicek, O. (1976), 'A test for normality based on sample entropy'. *Journal of the Royal Statistical Society Series B* **38**(1), 54–59.

von Baeyer, H. C. (2004), *Information: The New Language of Science*. Harvard University Press.

Wu, X. and J. M. Perloff (2007), 'GMM estimation of a maximum entropy distribution with interval data'. *Journal of Econometrics* **138**, 532–546.

Zellner, A. (1988), 'Optimal information processing and bayes theorem'. *American Statistician* **42**, 278–284.

Zellner, A. (1991), 'Bayesian methods and entropy in economics and econometrics'. In: W. T. Grandy Jr. and L. H. Schick (eds.): *Maximum Entropy and Bayesian Methods*. Amsterdam: Kluwer, pp. 17–31.

Zellner, A. (1994), 'Bayesian method of moments/instrumental variable (BMOM/IV) analysis of mean and regression models'. *1994 Proceedings of the Section on Bayesian Statistical Science of the American Statistical Association*. 1995 (Paper presented at the August 1994 ASA meetings).

Zellner, A. (1996a), 'Bayesian method of moments/instrumental variable (BMOM/IV) analysis of mean and regression models'. In: J. C. Lee, W. C. Johnson, and A. Zellner (eds.): *Modeling and Prediction: Honoring Seymour Geisser*. Springer-Verlag, pp. 61–75.

Zellner, A. (1996b), 'Models, prior information, and Bayesian analysis'. *Journal of Econometrics* **75**, 51–68.

Zellner, A. (1997), 'Bayesian method of moments (BMOM): Theory and applications'. In: T. B. Fomby and R. C. Hill (eds.): *Advances in Econometrics: Applying Maximum Entropy to Econometric Problems*. Greenwich: JAI Press, pp. 85–105.

Zellner, A. (1999), 'New information-based econometric methods in agricultural economics: Discussion'. *American Journal of Agricultural Economics* **81**, 742–746.

Zellner, A. (2002), 'Information processing and Bayesian analysis'. *Journal of Econometrics* **107**, 41–50.

Zellner, A. (2007), 'Some aspects of the history of Bayesian information processing'. *Journal of Econometrics* **138**, 388–404.