

Information Theoretic and Entropy Econometrics – Theory and Practice

Instructor: Amos Golan, American University

DATES: MAY 12-16, 2008

LOCATION: AMERICAN UNIVERSITY

Objectives and Scope

Background

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 toward a greater knowledge of that probability density function.

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. In this his course we will study the basic quantities of information theory and their relationships to data analysis and information processing, and then use these quantities to study and develop improved methods of estimation and data processing that compose the class of entropy and information-theoretic methods. Within the above class of Information-Theoretic models, this course concentrates on methods that use conditional and unconditional stochastic moments.

The Class

The focus of this course is on Information Theoretic (IT) methods of estimation in general, and on the econometric analysis of under-determined, noisy problems and econometric problems involving ill-conditioned or non-experimental data.

These IT methods are directly or indirectly build on the foundations of Information Theory (IT) and the principle of Maximum Entropy (ME). A common thread connecting all the IEE estimation methods is the objective of trying to better understand the data, while abstracting away from distributional assumptions or assumptions on the likelihood function. This class of methods includes the Bayesian Method of Moments (BMOM), the Empirical (or Generalized Empirical) Likelihood (EL/GEL), variations of the Generalized Method of Moments (GMM) and the Generalized ME (GME). All of these methods share the same basic objective of analyzing limited and noisy data using minimal assumptions.

The primary purpose of this class is to provide the background for understanding both (i) the theory, and (ii) to develop the necessary theoretical and empirical tools for practicing the theory in a wide range of economic/econometric estimation problems. Throughout the course, entropy estimators will be compared with their traditional counterparts and the computational aspects will be discussed and practiced with artificial and real data (your 'own' dataset).

The course will include lectures that will provide the basic background for understanding the theory, and daily tutorials in the computer laboratories. These tutorials will be done mainly in GAMS.

Preliminary time schedule

Monday May 12

09.00 - 12.30: Theory

1.30 – 5:00: Theory and Computer Lab

Tuesday May 17

09.00 - 12.30: Theory

2:00 -5:00: Computer Lab

Wednesday May 18

09.00 - 12.00: Theory

1.00 - 4.30: Computer Lab

Thursday May 19

09.00 - 12.00: Theory

1.00 - 4.30: Computer Lab

Friday May 20

09.00 - 12.00: Theory and Guests' Presentations

1.00 - 2:30: Computer Lab

2:30 - 4:30: Summary and Open Discussion

Content and Topics (Tentative)

Morning sessions: Theory

1. Background, motivation and philosophy

- Why IEE?
- What is Entropy?
- The axiomatic and combinatorial derivations of the entropy measure;
- The foundations of information theory and basic concepts
- The basic problem;
- The classical Maximum Entropy (ME) principle and formulation;
- The dual (concentrated) formulation;

- Basic diagnostics and test-statistics;
- Comparison with the standard ML and other estimation methods.

2. Derivation of the basic Generalized ME (GME) method - A simple example

Recovering the unknown coefficients of a matrix (matrix balancing) from aggregated data (e.g., estimating the coefficients of Input-Output Table, a Social Accounting Matrix, or first order Markov process, from limited aggregated data). A complete comparison with the traditional methods (e.g., ML) will be developed. Extensions that allow incorporating more variables (e.g., macro/policy) and accommodating for noisy data will be discussed in great detail.

3. The traditional linear regression model

- The basic set-up of the problem;
- Basic derivation of the GME;
- Primal vs. dual formulations;
- Diagnostics.
- Derivations and comparisons of other information-theoretic methods.

5. Extensions of the linear model to non-scalar identity covariance matrix (e.g., autocorrelation, heteroskedasticity).

6. Discussion of theoretical and empirical applications.

7. Special additional topics (to be presented, formulated and discussed if requested by the participants. We can discuss all/some of these topics as well as other topics of interest):

- Discrete choice models (ordered/unordered);
- Censored models;
- Model and variable selection;
- Linear and non-linear dynamic systems with control.
- Set of equations and simultaneous equations;
- Other

8. Summary and discussion of possible future directions

Afternoon sessions: Computer Lab and “Hands on Data”

Practical examples and open discussions will take place in the afternoons. In these sessions we:

1. Will use the generalized maximum entropy method to evaluate and estimate real world economic problems;
2. Exchange/develop relevant software; and
3. Discuss philosophical, practical, technical issues.

The main software package will be GAMS, but we will also use SAS and LIMDEP. In addition, some suggested computational problems (and solutions) will be provided.

Materials

- The main text is Golan, A., G.G. Judge, and D. Miller (1996), *Maximum Entropy Econometrics: Robust Estimation with Limited Data*, New York: John Wiley & Sons.
- And additional text is Golan, A., "[Information and Entropy Econometrics — A Review and Synthesis](#)", *Foundations and Trends® in Econometrics*: Vol. 2: No 1–2, pp 1-145.
- A Reader composed of the main reading for the class will be provided to each participant.
- A detailed reading list will be provided with the Final Syllabus. Handouts for SAS (and LIMDEP) will be provided as well.

Location

American University, Washington, D.C.

Registration

Download the necessary document/s from the [main page](#) and submit to the Economics department, or look at: <http://www.american.edu/american/registrar/>

About the instructor

[Amos Golan](#) is a professor in the Department of Economics at American University. He is an econometrician specialized in information and entropy econometrics with a special interest in processing and estimating incomplete, ill-behaved and/or non-experimental data. His work is both on the theoretical and applied level. He is the senior author of *Maximum Entropy Econometrics*, Guest Editor for the *Journal of Econometrics* and *Econometric Reviews* volumes on *Information and Entropy Econometrics* and he is currently on the editorial boards of *Econometric Reviews* and *Foundations and Trends in Econometrics*.