Image Reconstruction: An Information -Theoretic Approach

Avi Bhati[1], Bahattin Buyuksahin[2], Amos Golan[3]

(Partial Draft: June 2004)

Summary. The objective of this project is to develop an appropriate econometric (statistical) method for analyzing, with minimal assumptions, small and possibly incomplete and ill-behaved data. In this paper we develop an efficient and easy to apply image reconstruction (estimation) method for analyzing such data. The resulting method extends (and builds on the foundations of) information-theoretic methods by further relaxing some of the underlying assumptions, uses minimal distributional assumptions, performs well (relative to current methods of estimation and image reconstruction) and uses efficiently all the available information (hard and soft data). Further, this method is computationally efficient.

Key Words: Entropy, Image Reconstruction, Information

1 Urban Institute
2 American University
3 Department of Economics, American University, Roper 200, 4400 Massachusetts Ave., NW
Washington, DC 20016; Tel.: 202-885-3783; Fax: 202-885-3790; *E-mail address:* agolan@american.edu (A.Golan)

1. Introduction

1.1 Motivation and Basic Objectives

Most data sets (in economics and other social sciences) are non-experimental, very small, highly collinear, and are under-determined in the sense that there is not enough information for perfectly inverting to obtain a solution (ill-posed or ill-conditioned). Moreover, with these data, the underlying generation process, or likelihood, is unknown. The objective of this research is to develop an efficient and easy to apply image reconstruction (estimation) method for analyzing such data. The resulting method is an information-theoretic one that uses minimal distributional assumptions, performs well (relative to current methods of estimation), uses efficiently all the available information (hard and soft data) and is computationally efficient.

1.2 Background and Brief History

With the above in mind, and within the general objective of image reconstruction, estimation and inference for a large class of models, it seems that going back to the foundations of Information Theory and Maximum Entropy was inevitable and led to a whole class of information-theoretic methods. All of these information-theoretic methods could be viewed as approaches to solving ill-posed or under-determined problems in the sense that without a pre-specified likelihood or distribution, there are always more unknowns than knowns regardless of the amount of data. That is, since the observation matrix is irregular or ill-conditioned or since the number of unknowns exceeds the number of data points, the problem is ill-posed. To solve these problems, one has to (i) incorporate some prior knowledge, or constraints, on the solution, or (ii) specify a certain criterion to choose among the infinitely many solutions, or (iii) use both approaches. But what criterion and what constraints should one use?

It seems natural to employ an informational criterion together with variations of the observed moments. For example, Zellner (1997, p. 86) says, "The BMOM – Bayesian Method of Moments - approach is particularly useful when there is difficulty in formulating an appropriate likelihood function. Without a likelihood function, it is not possible to pursue traditional likelihood and Bayesian approaches to estimation and testing. Using a few simple assumptions, the BMOM approach permits calculation of post-data means, variances and other moments of parameters and future observations."

In the BMOM approach, one starts by maximizing the continuous entropy function subject to some side conditions (pure conservation laws) and normalization. This approach yields the average log-height of the density function, which is the least informative density given these side-conditions (e.g., Zellner, 1991; 1997).

Similarly, under the Empirical Likelihood (EL) objective, one starts by searching for the "natural" weight of each observation by maximizing the discrete likelihood objective subject to the exact moment restrictions and normalization. Under the Generalized Maximum Entropy (GME) approach, one maximizes the joint entropies of both signal and noise, but subject to noisy moment representation (noisy conservation laws). Other examples include some of the information-theoretic Generalized Method of Moments, GMM, (e.g., Kitamura and Stutzer, 1997; Imbens, Johnson and Spady, 1998) methods as well as the class of regularization methods, (e.g., Donoho et. al., 1992), or the class of models known as "quantified Maximum Entropy (ME)" (e.g., Skilling, 1989).[4] The solutions in all of these methods depend on the choice of regularization parameter, and the moments' representation, and all are derived as in the traditional ME approach.

In general, the common idea behind these approaches consists of transforming, with minimum a-priori assumptions, an ill-conditioned, linear problem, subject to non-linear but convex constraints, into a much simpler and much smaller problem in convex optimization. Even though these approaches may seem quite similar on the very abstract level, they differ in their treatment of the data (the constraints) as they are developed based on the view that the data one observes may be in terms of moments (pure or noisy) which takes us to the ME or MEM, or may be in terms of noisy observations which takes us to the GME. For a detailed discussion and comparison of these information-theoretic estimation rules see Golan (2002a).

In Section 2 we describe the basic Model and formulate the Information-Theoretic (IT) model. In Section 3, we present some examples and contrast (empirically) our method with others. We conclude in Section 4.

2. Model and Formulation

2.1 The Basic Model

The observed data of an image, D_{ij} , are binary 0-1. Thus, in each pixel we observe 0 or 1 (or black and white). Our objective is to reconstruct the blurry image to a "perfect" (clean) image as

4 A less well known class of ME-type methods that was developed for noisy data, known as the ME on the Mean, MEM, (and is related to the GME), is discussed in Gamboa and Gassiat (1997).

efficiently as possible. Since images have many pixels/cells, we need to formulate the model such that we keep the complexity level to a minimum. Further complication is that unlike the traditional discrete choice type models we do not have covariates (or independent variables) here. Instead, we try to take into account the possible nearest neighbors and the correlations among the pixels. To start, let

$$D_{ij} = p_{ij} \,. \tag{1}$$

where p_{ij} are the estimated probabilities (image). Taking into account the noise in the observed data, the correct noisy observed model is

$$D_{ij} = p_{ij} + \varepsilon_{ij} = \sum\nolimits_m z_m q_{ijm} + \sum\nolimits_l v_l w_{ijl} \tag{2}$$

where ε_{ij} represents the noise in the data. In this model, each ε_{ij} is naturally bounded in [-1, 1], $z \in [0,1]$ and $\sum_m q_{ijm} = \sum_l w_{ijl} = 1$. Thus, we converted all of our unknowns (signal and noise) into proper probability distributions. In this model, we allow correlations across neighbors.

To capture the relationship across the different pixels and to reduce the dimensionality, we transform model (2) into the following set of noisy moments:

$$\sum\nolimits_i D_{ij} = \sum\nolimits_i p_{ij} + \sum\nolimits_i \varepsilon_{ij} = \sum\nolimits_{i,m} z_m q_{ijm} + \sum\nolimits_{i,l} v_l w_{ijl} \qquad j=1, 2, \ldots, J \tag{3a}$$

$$\sum\nolimits_j D_{ij} = \sum\nolimits_j p_{ij} + \sum\nolimits_j \varepsilon_{ij} = \sum\nolimits_{j,m} z_m q_{ijm} + \sum\nolimits_{j,l} v_l w_{ijl} \qquad i=1, 2, \ldots, I \tag{3b}$$

The Basic (no correlation) Generalized Cross Entropy (GCE) model is

$$\begin{aligned} &\underset{\{p,w\}}{Min} \left\{ \sum_{ijm} q_{ijm} \log\left(q_{ijm} / q^0_{ijm}\right) + \sum_{ijl} w_{ijl} \log\left(w_{ijl} / w^0_{ijl}\right) \right\} \\ &s.t. \\ &(3a)-(3b) \\ &\sum_m q_{ijm} = 1; \sum_l w_{ijl} = 1 \end{aligned} \tag{4}$$

The GCE solution is

$$\hat{q}_{ijm} = \frac{q^0_{ijm} \exp\left[\left(\hat{\lambda}^i_j + \hat{\lambda}^j_i\right) z_m\right]}{\sum_m q^0_{ijm} \exp\left[\left(\hat{\lambda}^i_j + \hat{\lambda}^j_i\right) z_m\right]} \equiv \frac{q^0_{ijm} \exp\left[\left(\hat{\lambda}^i_j + \hat{\lambda}^j_i\right) z_m\right]}{\Omega_{ij}} \tag{5a}$$

and

$$\hat{w}_{ijl} = \frac{w^0_{ijl} \exp\left[\left(\hat{\lambda}^i_j + \hat{\lambda}^j_i\right) v_l\right]}{\sum_l w^0_{ijl} \exp\left[\left(\hat{\lambda}^i_j + \hat{\lambda}^j_i\right) v_l\right]} \equiv \frac{w^0_{ijl} \exp\left[\left(\hat{\lambda}^i_j + \hat{\lambda}^j_i\right) v_l\right]}{\Psi_{ij}} \tag{5b}$$

where $\hat{\lambda}^j_i$ and $\hat{\lambda}^i_j$ are the I+J estimated Lagrange multipliers associate with the data (Eqs. 3a-3b).

The estimated signal components are $\hat{p}_{ij} = \sum_m \hat{q}_{ijm} z_m$, and the estimated noise components are $\hat{\varepsilon}_{ij} = \sum_l \hat{w}_{ijl} v_l$.

2.2 The concentrated (dual) GCE

Instead of using the constrained optimization estimation model (4), the GCE can be formulated as an unconstrained, concentrated (or a generalized likelihood) model:

$$\ell(\lambda) = \sum_{i,j} \lambda^j_i D_{ij} + \sum_{i,j} \lambda^i_j D_{ij} - \sum_{i,j} \log \Omega_{ij}(\lambda) - \sum_{i,j} \log \Psi_{ij}(\lambda), \tag{6}$$

Maximizing (6) and solving for $\boldsymbol{\lambda}$, yields the estimated $\hat{\boldsymbol{\lambda}}$, which in turn yield the optimal probabilities $\hat{p}_{ij}$ and $\hat{w}_{ijl}$ via relationship (5).

2.3 Adding the Nearest Neighbor Correlation

Allowing for correlation of both the signal and the noise yields

$$D_{ij} = p_{ij} A_{ij} + \varepsilon_{ij} B_{ij} = \sum_m z_m q_{ijm} A_{ij} + \sum_l v_l w_{ijl} B_{ij} \tag{7}$$

where *A* and *B* are two correlation matrices associated with the signals and the errors correspondingly.

We now specialize that model, to a "nearest-neighbor" type model with exponential weights for the neighbors. Specifically,

$$\begin{aligned}\sum_i D_{ij} &= \sum_i \sum_{k=-\delta}^{+\delta} \sum_{h=-\delta}^{+\delta} p_{i+k,j+h} \exp\left(-\alpha c(k,h)\right) + \sum_i \sum_{k=-\delta}^{+\delta} \sum_{h=-\delta}^{+\delta} \varepsilon_{i+k,j+h} \exp\left(-\beta c(k,h)\right) \\ &= \sum_{i,m} \sum_{k=-\delta}^{+\delta} \sum_{h=-\delta}^{+\delta} z_m q_{i+k,j+h,m} \exp\left(-\alpha c(k,h)\right) + \sum_{i,l} \sum_{k=-\delta}^{+\delta} \sum_{h=-\delta}^{+\delta} v_l w_{i+k,j+h,l} \exp\left(-\beta c(k,h)\right)\end{aligned} \quad j=1,\ldots,J \quad \text{(8a)}$$

$$\begin{aligned}\sum_j D_{ij} &= \sum_j \sum_{k=-\delta}^{+\delta} \sum_{h=-\delta}^{+\delta} p_{i+k,j+h} \exp\left(-\alpha c(k,h)\right) + \sum_j \sum_{k=-\delta}^{+\delta} \sum_{h=-\delta}^{+\delta} \varepsilon_{i+k,j+h} \exp\left(-\beta c(k,h)\right) \\ &= \sum_{j,m} \sum_{k=-\delta}^{+\delta} \sum_{h=-\delta}^{+\delta} z_m q_{i+k,j+h,m} \exp\left(-\alpha c(k,h)\right) + \sum_{j,l} \sum_{k=-\delta}^{+\delta} \sum_{h=-\delta}^{+\delta} v_l w_{i+k,j+h,l} \exp\left(-\beta c(k,h)\right)\end{aligned} \quad i=1,\ldots,I \quad \text{(8a)}$$

where $\delta=0, 1, 2,\ldots$ represents the number of nearest neighbors, α reflects the rate of exponential decay for the signal, β reflects the decay rate for the noise, and $c(k,h) = Max\left(k^2, h^2\right)$. All of these parameters are determined a-priori. Together the pair ($c(k,h)$, α) represents the "weights" for the signal spatial correlations, and the pair ($c(k,h)$, β) represents the "weights" for the noise spatial correlations.

The GCE yields the solution

$$\hat{q}_{ijm} = \frac{q_{ijm}^{0} \exp\left[z_m \sum_{k=-\delta}^{+\delta} \sum_{h=-\delta}^{+\delta} \exp\left(-\alpha \mathrm{c(k,h)}\left(\hat{\lambda}_{j+h}^{i} + \hat{\lambda}_{i+k}^{j}\right)\right)\right]}{\sum_m q_{ijm}^{0} \exp\left[z_m \sum_{k=-\delta}^{+\delta} \sum_{h=-\delta}^{+\delta} \exp\left(-\alpha \mathrm{c(k,h)}\left(\hat{\lambda}_{j+h}^{i} + \hat{\lambda}_{i+k}^{j}\right)\right)\right]}$$

$$\equiv \frac{q_{ijm}^{0}\, \mathrm{e} \exp\left[z_m \sum_{k=-\delta}^{+\delta} \sum_{h=-\delta}^{+\delta} \exp\left(-\alpha \mathrm{c(k,h)}\left(\hat{\lambda}_{j+h}^{i} + \hat{\lambda}_{i+k}^{j}\right)\right)\right]}{\Omega_{ij}} \tag{9a}$$

and

$$\hat{w}_{ijl} = \frac{w_{ijl}^{0} \exp\left[v_l \sum_{k=-\delta}^{+\delta} \sum_{h=-\delta}^{+\delta} \exp\left(-\beta \mathrm{c(k,h)}\left(\hat{\lambda}_{j+h}^{i} + \hat{\lambda}_{i+k}^{j}\right)\right)\right]}{\sum_l w_{ijl}^{0} \exp\left[v_l \sum_{k=-\delta}^{+\delta} \sum_{h=-\delta}^{+\delta} \exp\left(-\beta \mathrm{c(k,h)}\left(\hat{\lambda}_{j+h}^{i} + \hat{\lambda}_{i+k}^{j}\right)\right)\right]}$$

$$\equiv \frac{w_{ijl}^{0} \exp\left[v_l \sum_{k=-\delta}^{+\delta} \sum_{h=-\delta}^{+\delta} \exp\left(-\beta \mathrm{c(k,h)}\left(\hat{\lambda}_{j+h}^{i} + \hat{\lambda}_{i+k}^{j}\right)\right)\right]}{\Psi_{ij}} \tag{9b}$$

Finally the concentrated GCE (with correlation) is exactly like model (6), but the normalizations (partition functions) of the two right hand side terms (Ω and Ψ) are replaced by the more general normalizations specified in equations (9a)-(9b) above.

3. Example

To present a simple example, consider the following data (discussion of the data to be added) presented in Figure 1. The first step image reconstruction is shown in Fig. 2.

Next step and Figures – to be added

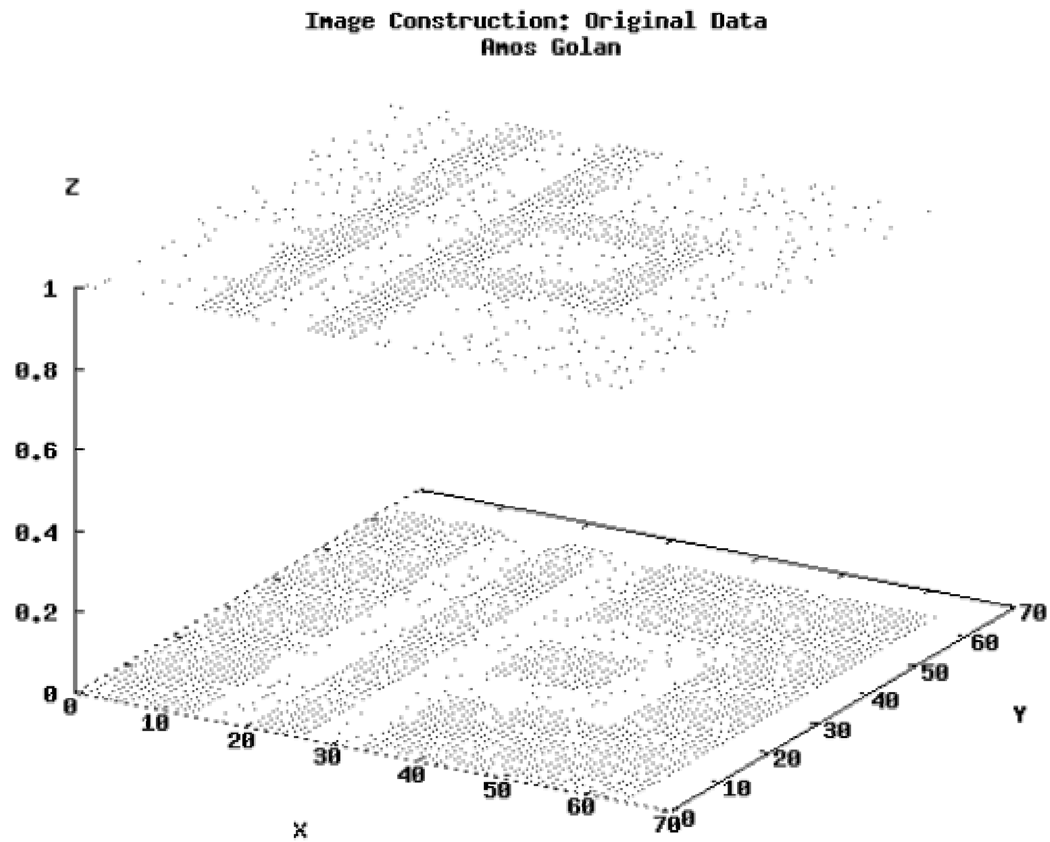
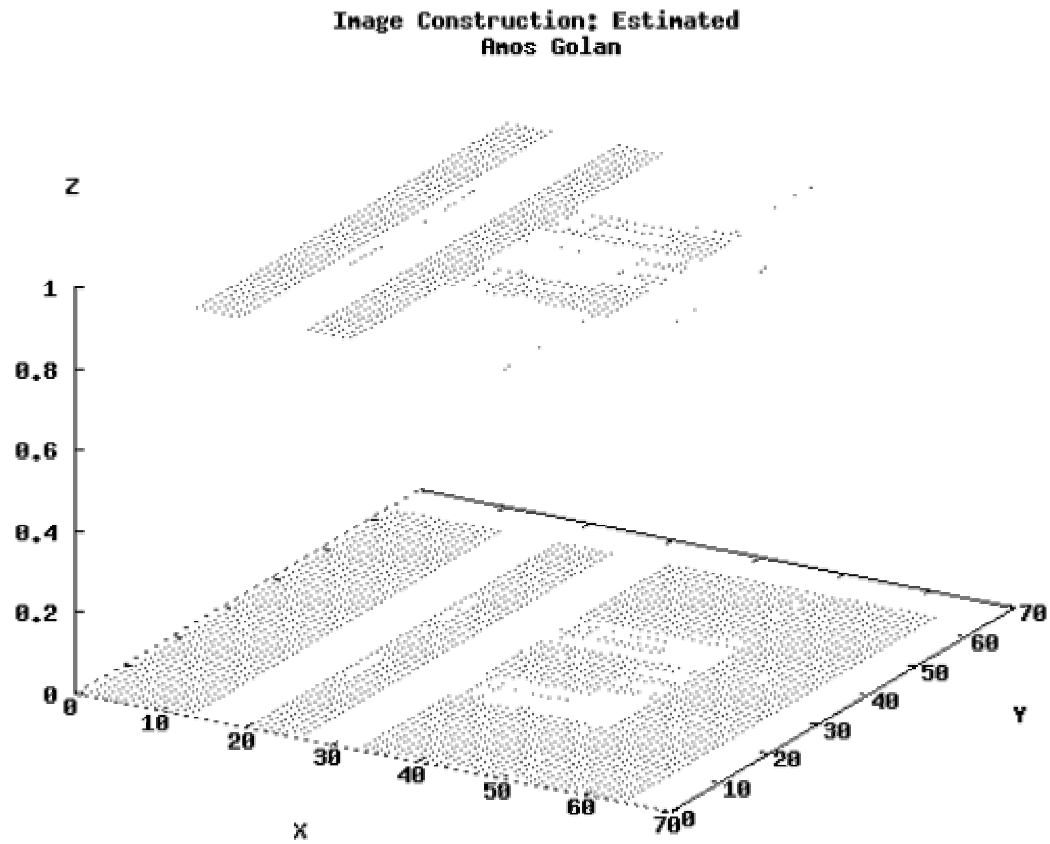


Image Construction: Original Data
Amos Golan
Z
1
0.8
0.6
0.4
0.2
0
0
10
20
30
40
50
60
70
X
0
10
20
30
40
50
60
70
Y

Figure 2: First Step Estimates



4. Conclusion

To be Added

References (not completed)

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

Cressie, N and T. R. C. Read, 1984. Multinomial goodness-of-fit tests. J. Royal Stat. Soc. B, 46, 440-464.

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, Ser. B, 54, 41-81.

Gamboa, F. and Gassiat, E. 1997. Bayesian methods and maximum entropy for ill-posed inverse problems. Annals of Statistics. 25(1) ,328-350

Golan, A. 2002a. A Simultaneous Estimation and Variable Selection Rule. J. of Econometrics 101:1, 165-193

Golan, A. 2002b. Information and Entropy Econometrics – Editor's View. J. of Econometrics, vol 107 (1-2) (2002), 1-15.

Golan, A. and Dose, V., 2001. A Generalized A Generalized Information Theoretical Approach to Tomographic Reconstruction, J. Physics A; Math. and General, 34, 1271-1283.

Golan, A. and G. Judge (1996), A Maximum Entropy Approach to Empirical Likelihood Estimation and Inference, (unpublished paper), University of California, Berkeley (presented at the 1997 Summer Econometric Society meetings).

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

Kitamura, Y., and M. Stutzer, 2001, Connection between entropic and linear projections in asset pricing estimation. J. of Econometrics (Forthcoming).

Imbens, G.W., Johnson, P. 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.

Kitamura, Y. and M. Stutzer, 1997. An information-theoretic alternative to generalized method of moment estimation. Econometrica 66 4, 861-874.

Kullback, S., 1959. Information Theory and Statistics. John Wiley & Sons, New York.

Kullback, S., and R. A. Leibler 1951. On information and sufficiency. The Annals of Math. Stat. 22, 79-86.

Mittelhammer, R., G. Judge, M van Akkeren and S. Cardell, 2001. Coordinate based empirical likelihood-like estimation in ill-conditioned inverse problems. Working paper.

Mittelhammer R., G. Judge amd D. Miller, "Econometric Foundations" (Cambridge, 2000).

Owen A. B., Empirical Likelihood for Linear Models, *Ann. Statistics*, (1991) 19, 1725-1747.

Qin J., and J. Lawless, Empirical Likelihood and General Estimating Equations, *The Ann. of Statistics*, (1994) 22, 300-325.

Shannon, C. E, 1948. A mathematical theory of communication. Bell System Technical Journal 27, 379-423.

Shore, J.E., and R.W. Johnson (1980), Axiomatic Derivation of the Principle of Maximum Entropy and the Principle of Minimum Cross-Entropy, *IEEE Trans. on Info. Theory*, IT-26(1), 26-37.

Skilling J., 1989. Classic Maximum Entropy. In: Skilling, J. (Ed.) Maximum Entropy and Bayesian Methods in Science and Engineering. Kluwer Academic, pp. 45-52.

Soofi, E. S., 1994. Capturing the intangible concept of information, J. of the American Statistical Association, 89, 1243-1254.

Zellner, A., 1991. Bayesian methods and entropy in economics and econometrics. In Grandy, W.T., Jr. and Schick, L. H. (Eds.), Maximum Entropy and Bayesian Methods, Kluwer, Amsterdam, 17-31.

Zellner, A., 1997. The Bayesian method of moments (BMOM): Theory and applications, in T. Fomby and R. Hill, eds., Advances in Econometrics 12, 85-105.