



AgEcon SEARCH
RESEARCH IN AGRICULTURAL & APPLIED ECONOMICS

The World's Largest Open Access Agricultural & Applied Economics Digital Library

This document is discoverable and free to researchers across the globe due to the work of AgEcon Search.

Help ensure our sustainability.

Give to AgEcon Search

AgEcon Search
<http://ageconsearch.umn.edu>
aesearch@umn.edu

*Papers downloaded from **AgEcon Search** may be used for non-commercial purposes and personal study only. No other use, including posting to another Internet site, is permitted without permission from the copyright owner (not AgEcon Search), or as allowed under the provisions of Fair Use, U.S. Copyright Act, Title 17 U.S.C.*

**LEAST SQUARES AND ENTROPY AS
PENALTY FUNCTIONS**

by

Paul V. Preckel

Staff Paper # 98-16

August 18, 1998

**Dept. of Agricultural Economics
Purdue University**

Purdue University is committed to the policy that all persons shall have equal access to its programs and employment without regard to race, color, creed, religion, national origin, sex, age, marital status, disability, public assistance status, veteran status, or sexual orientation.

LEAST SQUARES AND ENTROPY AS PENALTY FUNCTIONS

by

Paul V. Preckel

Dept. of Agricultural Economics, Purdue University

West Lafayette, Indiana 47907-1145

preckel@agecon.purdue.edu

Staff Paper #98-16

August 18, 1998

Abstract

Mathematical measures of entropy as defined by Shannon (1948) and Kullback and Leibler (1951) are currently in vogue in the field of econometrics, primarily due to the comprehensive work by Golan, Judge, and Miller (1996). In this paper, an alternative interpretation of the entropy measure as a penalty function over deviations is presented.

Using this interpretation, a number of parallels are drawn with least squares estimators, and it is demonstrated that, with a minor modification of the traditional least squares estimator, both approaches may be applied to the general linear model. The advantages and disadvantages of each approach are discussed, and a philosophical approach to the selection of estimation technique is suggested.

Keywords: Entropy, econometrics, penalty functions, estimation

Helpful comments from Alice Harmon, John A.L. Cranfield, James Eales, and James Binkley are gratefully acknowledged. The author is responsible for any remaining errors.

Copyright © by Paul V. Preckel. All rights reserved. Readers may make verbatim copies of this document for non-commercial purposes by any means, provided that this copyright notice appears on all such copies.

LEAST SQUARES AND ENTROPY AS PENALTY FUNCTIONS

by
Paul V. Preckel

Introduction

Mathematical measures of entropy as defined by Shannon (1948) and Kullback and Leibler (1951) are currently in vogue in the field of econometrics, primarily due to the comprehensive work by Golan, Judge, and Miller (1996). Despite the thoroughness of this work, there remains confusion in the field of applied econometrics regarding when and how to use entropy. Much of this confusion seems to stem from the interpretation of the entropy measure, which was originally designed as a measure of the “distance” between probability distributions.

In this paper, an alternative interpretation of the entropy measure as a penalty function over deviations is presented. Throughout, the model of interest is the general linear model:

$$Y = X\mathbf{b} + \mathbf{e}$$

where Y is a vector of n observations of a single dependent variable, X is an n by m matrix of data where the rows correspond to observations, and the columns correspond to independent variables, \mathbf{b} is a vector of m coefficients to be estimated, and \mathbf{e} is a vector of n independent, identically distributed error terms. Thus, we are concerned with the single-equation case of an estimation problem with a simple error structure.

In the next section, the interpretation of the entropy function as a penalty function is presented, and parallels with the “least squares penalty function” over deviations are discussed. Examples are used to demonstrate that well-defined formulations for both entropy and least squares exist for this problem, and that the results (estimated values for \mathbf{b}) are quite similar. Further, it is demonstrated that the preceding statement is true in the underdetermined case. In the third section, the properties of Generalized Maximum Entropy (GME) estimators are discussed. In the fourth section, a discussion of instances where it may be advantageous to use entropy formulations is provided. The fifth section draws a few conclusions and makes suggestions regarding appropriate use of entropy estimators.

Entropy as a Penalty Function

Entropy of the sort used in econometrics (as opposed to the sort used in physics, which is loosely related as described, for example, by Georgescu-Roegen, 1987) has its origins in the field of information theory (Shannon, 1948). One interpretation is that the mathematical entropy measure indicates the *distance* (or, more accurately, discrepancy) between two distribution functions. In the case of Shannon’s measure of entropy, the measure reflects the degree to which the particular distribution being measured is different from a uniform distribution.

Shannon's measure of entropy for a distribution is given by the following expressions for the discrete and continuous cases:

$$-\sum_{i=1}^n p_i \ln(p_i) \quad \text{and} \quad -\int_{\Omega} f(x) \ln[f(x)] dx,$$

where p_i is the probability associated with the i -th support point in the discrete case, and $f(x)$ is the density at x and Ω is the support for the distribution in the continuous case. With Shannon's measure, the maximum possible value is bounded, and when the maximum is attained, the distribution being measured is precisely uniform. (That is, the distribution is uniform over the support points in the discrete case and over Ω in the continuous case).

This is not the usual concept of a distance. With vectors, we expect the distance between two vectors to be non-negative, and that the distance between them will be at its minimum value of zero if and only if they are identical. Fortunately, there is another entropy measure that behaves more like our intuitive concept of distance, namely the Kullback-Liebler measure of cross-entropy.

Cross-entropy measures the distance from one distribution to another. In the discrete and continuous cases, cross-entropy is stated mathematically as

$$\sum_{i=1}^n p_i \ln(p_i / q_i) \quad \text{and} \quad \int_{\Omega} f(x) \ln[f(x) / g(x)] dx,$$

where q_i is the probability associated with the i -th point in what we will call the "reference distribution," i.e., the distribution to which distance is measured, in the discrete case, and $g(x)$ is the density associated with x for the reference distribution in the continuous case. In Golan, Judge, and Miller (1996) – hereafter referred to as GJM – this distribution is often called the prior distribution. Here, this terminology is avoided to limit confusion with the more rigorous definition of a prior in the literature on Bayesian estimation.

It is well known (e.g., see Kapur and Kesavan, 1992, pp.163-166) that when the reference distribution is uniform, entropy and cross-entropy are equivalent in the sense that higher values for entropy correspond exactly to lower values for cross-entropy. For our comparisons with least squares approaches to parameter estimation, it is convenient to focus on minimizing cross-entropy rather than maximizing entropy. At this point, we merely note that, provided that the reference distribution is uniform, these will produce the same results.

Cross-entropy behaves more like distance. The measure is always non-negative and attains a value of zero when p_i equals q_i for all $i = 1, \dots, n$ in the discrete case, or $f(x)$ equals $g(x)$ for all $x \in \Omega$ in the continuous case. A detailed discussion of the properties of cross-entropy as a measure of discrepancy is presented in Kapur and Kesavan (1992, pp. 151-161). (Kapur and Kesavan call cross-entropy a measure of discrepancy rather than distance because there are some properties of standard distance measures that are not satisfied by cross-entropy.)

For the balance of this paper, we will focus on cross-entropy based on a two-point uniform distribution as the motivation for a penalty function over deviations. The use of the phrase penalty function is as a measure of undesirability of the sort that motivates the estimators for a least squares problem. That is, for an ordinary least squares problem, parameter values (\mathbf{b}) are chosen so as to minimize the sum of squares of the errors associated with the individual observations (\mathbf{e}_i). Thus, the least squares penalty for the fit of the estimated relationship for the i -th observation is $P_{ls}(\mathbf{e}) = \mathbf{e}_i^2$.

Following Golan, Judge, and Miller (1996), the contribution of the i -th observation to the objective to be minimized (using a cross-entropy framework with a uniform reference distribution) is

$$p_i \ln(p_i/0.5) + (1-p_i) \ln[(1-p_i)/0.5],$$

where the estimated value of the i -th error is $\mathbf{e}_i = p_i(-K) + (1-p_i)K$, and where the support for the two-point uniform reference distribution places the points equally K units on either side of zero. Solving the relationship for p_i as a function of \mathbf{e}_i and substituting it into the objective contribution yields

$$P_e(\mathbf{e}_i) = [1/2 - \mathbf{e}_i/(2K)] \ln(1 - \mathbf{e}_i/K) + [1/2 + \mathbf{e}_i/(2K)] \ln(1 + \mathbf{e}_i/K).$$

This entropy-based penalty over the error has a number of properties that are similar to the quadratic penalty. Both penalty functions attain their minimum at zero and are monotonically increasing in the absolute value of their argument. Both are symmetric about zero, i.e., $P_{ls}(\mathbf{e}) = P_{ls}(-\mathbf{e})$ and $P_e(\mathbf{e}) = P_e(-\mathbf{e})$. (Although these properties are presented here only for the case of a two-point reference distribution, they also apply to entropy measures based on more points provided that their support points are symmetrically located about zero and provided that the associated reference weights are also symmetric.)

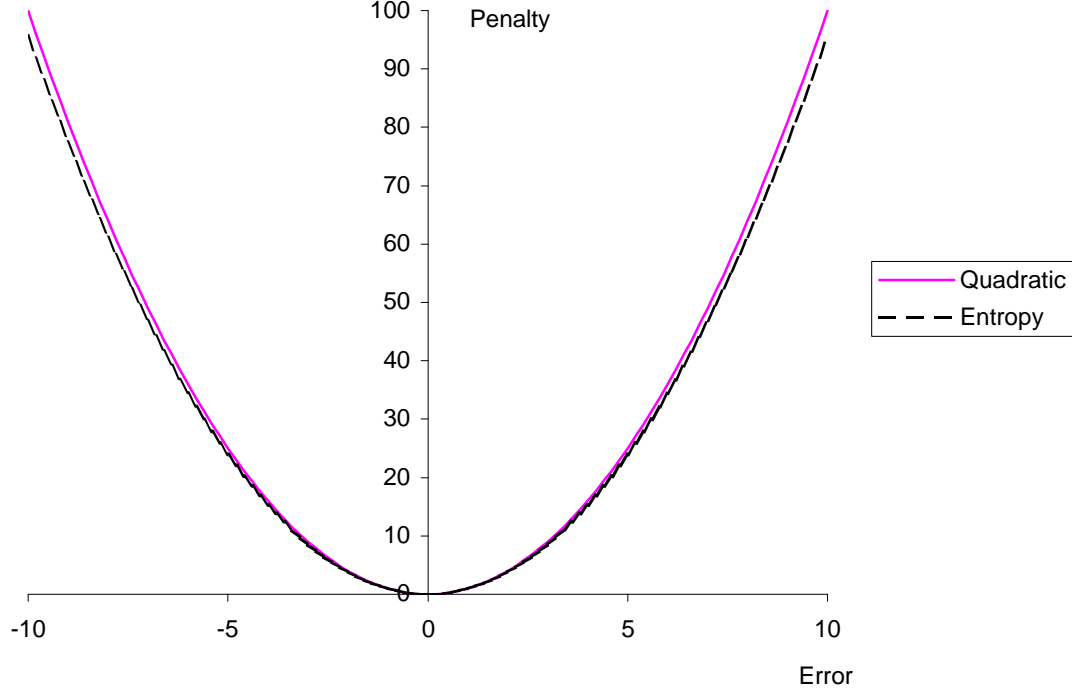
Further, if the entropy-based penalty is scaled by multiplying by a positive constant that makes the two penalties equal at $K/2$, then the graphs of the two penalties become very close with a maximum relative error of less than 4.5 percent over the range $[-K/2, K/2]$. Visually, these penalties are nearly indistinguishable on a graph over the $[-K/2, K/2]$ range. Figure 1 displays the entropy and quadratic penalty functions for a value of $K=100$ over the range $[-10, 10]$ for the case where the entropy penalty function has been scaled so that the two penalties are equal at $K/2$.

Hence, we have the following two similar problems. The standard least squares single-equation problem:

$$\text{Minimize } \sum_{i=1}^n \mathbf{e}_i^2$$

$$\text{Subject to: } Y_i - \sum_{j=1}^m X_{ij} \mathbf{b}_j = \mathbf{e}_i,$$

Figure 1. Quadratic and Entropy Penalty Functions Near Zero^{*}



^{*} The entropy penalty function is based on a reference distribution that puts equal weight at ± 100 .

and the GME problem where the errors receive an entropy treatment, but reference distributions are not specified for the parameters (hereafter denoted by GME._b):

$$\begin{aligned} &\text{Minimize} \quad \sum_{i=1}^n \left[\left(\frac{1}{2} - \frac{\mathbf{e}_i}{2K} \right) \ln \left(1 - \frac{\mathbf{e}_i}{K} \right) + \left(\frac{1}{2} + \frac{\mathbf{e}_i}{2K} \right) \ln \left(1 + \frac{\mathbf{e}_i}{K} \right) \right] \\ &\text{Subject to:} \quad Y_i - \sum_{j=1}^m X_{ij} \mathbf{b}_j = \mathbf{e}_i, \quad -K < \mathbf{e}_i < K. \end{aligned}$$

Based on our observations regarding the similarity between the least squares and entropy penalty functions over errors, we expect that the parameter estimates \mathbf{b} should be quite similar for these two problems.

Table 1. Parameter Estimates OLS versus GME. \mathbf{b}

	Intercept	Market Value of Firm	Value of Plant and Equipment
OLS	-9.9563	0.1517	0.0266
GME. \mathbf{b} , $K=75$	-10.5720	0.1503	0.0277
GME. \mathbf{b} , $K=300$	-9.9671	0.1516	0.0266
GME. \mathbf{b} , $K=300$	-9.9569	0.1517	0.0266
GME. \mathbf{b} , $K=1000$	-9.9563	0.1517	0.0266

To confirm this hypothesis, a well-known data set from Boot and DeWitt (1960) is used to compare the results of these problems. Using each of these formulations in turn, the level of gross investment is regressed on (a) market value of the firm at the end of the previous period, and (b) the value of the stock of plant and equipment at the end of the previous year. All values are for General Electric for the period 1935-1954.

Table 1 displays the results from regressions based on OLS and GME. \mathbf{b} using support values of $K=75$, $K=300$, $K=1200$ and $K=4800$. With $K=75$, the estimates, particularly for the intercept, are different from the OLS results. Increasing K by a factor of four to 300, the results are quite a bit closer to the OLS results. Increasing K by another factor of four to 1200 results in estimates that agree with the OLS results to four significant digits, and an additional factor of four increase in K results in estimates that agree with the OLS results to five significant digits. Hence, when the entropy measure is based on a broad support for the errors, the entropy results are very similar to the OLS results. (The rule of thumb suggested in GJM is to base K on an estimate of the standard deviation of the errors. They cite Pukelsheim's work as support for the use of a K value based on three standard deviations. This "three sigma" rule yields a K value of about 75 – the lowest value of K considered here.)

Of course GME was proposed by GJM not so much for the classic estimation problem considered above, but rather for the "ill-posed," or underdetermined case. One instance of the ill-posed case occurs if the number of observations is too low (or multicollinearity is too high), such that $X'X$ is only positive semi-definite (singular). In this case, OLS cannot proceed.

GJM manage to proceed by "re-parameterizing" \mathbf{b} to obtain a problem for which the \mathbf{b} values are uniquely determined. This is done by defining a support for \mathbf{b} and a set of weights which are often incorrectly called probabilities over the support. Rather than cloud the issue by referring to the support and weights of the reference distribution as a probability distribution,

here we will view them as parameters of our penalty function over deviations from the mean of our reference distribution. This penalty function, again using our cross-entropy perspective, has minimum value when the deviation is zero. We refer to the mean of the reference distribution as the target values for the parameters.

Following GJM, we formulate the GME minimum cross-entropy problem using the following assumptions: two-point supports with uniform weights for each of the parameters as well as for each error, and equal weighting of the cross-entropy penalty functions of the parameters and errors. If the target values for the coefficients, \mathbf{b} , are denoted by \mathbf{b}^0 , and the support points for the individual parameters are L above and below \mathbf{b}_i^0 , then the GME problem may be written as:

$$\begin{aligned} \text{Minimize } & \sum_{j=1}^m \left(\frac{1}{2} - \frac{\mathbf{b}_j - \mathbf{b}_j^0}{2L} \right) \ln \left(1 - \frac{\mathbf{b}_j - \mathbf{b}_j^0}{L} \right) + \left(\frac{1}{2} + \frac{\mathbf{b}_j - \mathbf{b}_j^0}{2L} \right) \ln \left(1 + \frac{\mathbf{b}_j - \mathbf{b}_j^0}{L} \right) + \\ & \sum_{i=1}^n \left(\frac{1}{2} - \frac{\mathbf{e}_i}{2K} \right) \ln \left(1 - \frac{\mathbf{e}_i}{K} \right) + \left(\frac{1}{2} + \frac{\mathbf{e}_i}{2K} \right) \ln \left(1 + \frac{\mathbf{e}_i}{K} \right) \\ \text{Subject to: } & Y_i - \sum_{j=1}^m X_{ij} \mathbf{b}_j = \mathbf{e}_i, \quad -K < \mathbf{e}_i < K, \quad -L < \mathbf{b}_j - \mathbf{b}_j^0 < L. \end{aligned}$$

Using our penalty function interpretation for the entropy- and least squares-based penalties on deviations, a similar problem can be formulated that could be viewed as a “generalized ordinary least squares” (GOLS, not to be confused with generalized least squares or GLS) problem. This problem is similar to the GME problem above because it includes penalties on deviations not only for the errors, but also for the parameters. Because of the parameter penalties, this estimation problem is also applicable to the ill-posed case. The formulation is as follows:

$$\begin{aligned} \text{Minimize } & \sum_{j=1}^m (\mathbf{b}_j - \mathbf{b}_j^0)^2 + \sum_{i=1}^n \mathbf{e}_i^2 \\ \text{Subject to: } & Y_i - \sum_{j=1}^m X_{ij} \mathbf{b}_j = \mathbf{e}_i. \end{aligned}$$

This formulation is not new. It is a special case of the mixed estimation formulation due to Theil and Goldberger (1961). This type of formulation also arises with stochastic restrictions as described in Kmenta (1986, pp. 497-500).

To demonstrate the similarity between these two problems, an ill-posed problem is specified as follows. To obtain target values for the coefficients, the first 18 observations (1935-1952) are used to obtain OLS estimates. These estimates are used as target values for both the

Table 2. OLS Results for the 1935-1952 Data Period, and GME versus GOLS Results for the 1953-1954 Period

	Intercept	Market Value of the Firm	Value of Plant and Equipment
OLS	-17.0335	0.1564	0.0296
GME	-17.0320	0.5198	-0.0925
GOLS	-17.0320	0.5198	-0.0925

GME and the GOLS problems, which are used to estimate the parameters based on only the last two observations (1953-1954). This might be the setup used if we believed that a structural change occurred in 1953 that we suspect made a fundamental change in the relationship between our dependent and independent variables. Thus, our estimation problems take into account prior knowledge about the former nature of the relationship (the means of the reference distributions for the coefficients) and the observations since the structural change.

Table 2 displays the target values for the coefficients, which are the OLS estimates based on 1935-1952 data, as well as the results for the GME and GOLS problems. For the GME problem, both K and L were set to 4800. This value is sufficiently large that doubling K and L has no effect on the parameter estimates. While the coefficient estimates for the two ill-posed problems are substantially different from the OLS target values, they are quite similar to each other. In fact, to the number of digits reported in the table, they are identical.

Hence, it appears that ill-posed problems are accessible not only from an entropy perspective, but also from a least squares perspective. The key in the ill-posed case is to apply the penalty function to both the deviations of the errors from zero and deviations of the parameters from their target values. In both the entropy and least squares cases, it is necessary to give explicit attention to the choice of weighting for the penalty over deviations of the errors from zero versus the penalty over deviations of the coefficients from target values. In the above examples the weightings were arbitrarily selected to be equal. The topic of weight selection is addressed for the entropy case in Golan, Judge, and Miller (1996), and for the least squares case in Theil and Goldberger (1961) and Theil (1963).

Properties of Entropy Estimators

Considerable effort has been directed towards developing the large and small sample properties of entropy estimators. (See e.g., Golan, Judge, and Miller, 1996; Mittlehammer, and Cardell, 1998; and Marsh, Mittlehammer, and Cardell, 1998.) When the entropy penalty functions over errors are based on symmetric reference distributions, it can be shown that the GME estimators are consistent and asymptotically normal. Given the unifying penalty function perspective presented in the previous section, one is led to speculate whether these asymptotic properties are satisfied for any penalty function that is symmetric about zero and is

monotonically increasing in its argument. However, this investigation is beyond the scope of the present paper.

In addition to these properties, it has been noted that by narrowing the support for the coefficients or the errors it is possible to obtain lower variance estimates than the analogous least squares estimates. It is important to inject a note of caution on this point. Lower variance with GME estimators, as with other shrinkage estimators, may come at a cost of bias (Judge, et al., 1985, pp. 82-90).

Consider an absurd case to make the point. One could define an infinitesimally narrow support for the coefficients and a wide support for the errors. Clearly, the variance of the GME estimator is small. However, the resulting coefficient estimates will have nothing to do with the data.

It is important to keep in mind the regularity conditions required to obtain the consistency results. (E.g., see Mittlehammer and Cardell, 1998.) Among other requirements, they include that the supports for the coefficients are sufficiently wide to contain their true values. In addition, the supports for the errors are required to be large enough to span the true support for the errors.

Any restriction of the support for coefficients or errors should be based on strong empirical evidence. Further, in reporting GME results it is absolutely critical to document the support selected and the motivation for the selection. Finally, it will probably be in the researcher's best interest to investigate the sensitivity of the coefficient values to the support values.

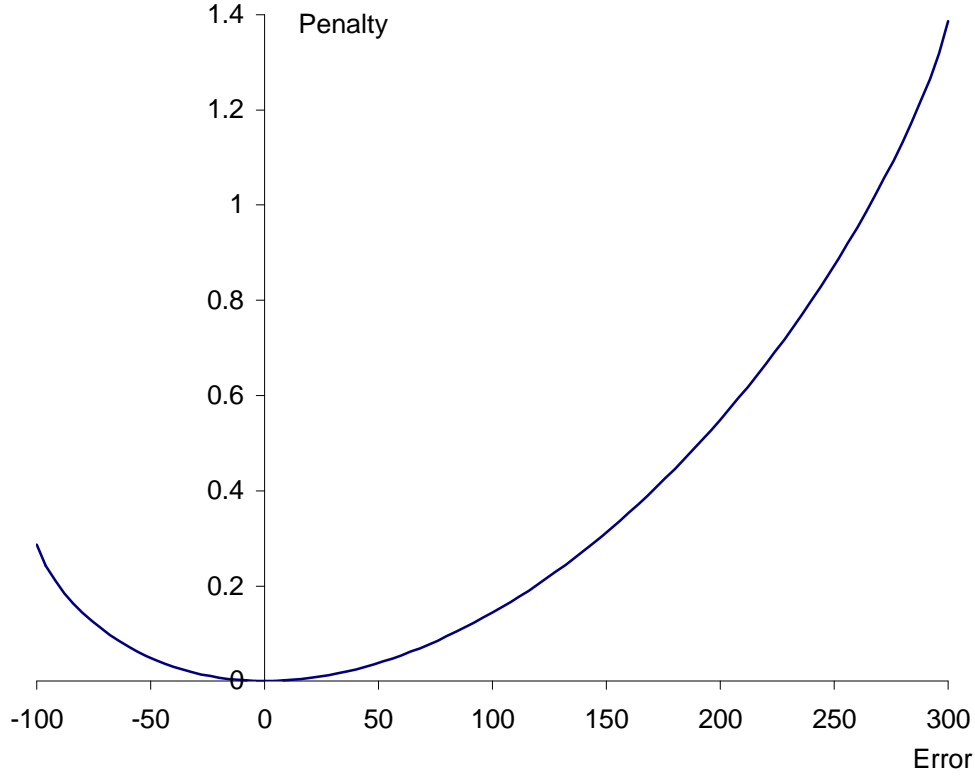
When Should We Use Entropy?

Let us begin with the premise that least squares has been, and will likely continue to be, a useful tool. Entropy approaches to econometrics do not make our lives easier. Unlike least squares, computing coefficient estimates using entropy requires more than just a bit of linear algebra. Furthermore, there are additional parameters to be specified – namely the reference distributions for errors and coefficients. Even for ill-posed problems, the GOLS approach suggested above eliminates the need to specify supports for the errors and coefficients. Hence, one is left wondering why we would want to use entropy estimation. There are three reasons.

First, there are problems where the estimation of a distribution is germane to the problem. For instance, Golan, Judge, and Perloff (1996) use entropy to estimate the size distribution of firms using government summary statistics. Similarly, Cranfield and Preckel (1997) use entropy to simultaneously estimate the income distribution and the parameters of a non-homothetic demand system.

The second obvious answer is in situations where it is desirable to have an asymmetric penalty function. For instance in the estimation of efficient frontier production functions (e.g., see Akridge, 1989), residual errors are either sign restricted, or negative and positive values are weighted differently. This is easily and elegantly accomplished in a cross-entropy framework by

Figure 2. Asymmetric Entropy Penalty Function (Reference Distribution Puts Weight of $\frac{3}{4}$ on -100 and $\frac{1}{4}$ on 300)



using an asymmetric reference distribution. For instance, by selecting a distribution that has a reference weight of $\frac{3}{4}$ on -100 , and $\frac{1}{4}$ on 300 , we obtain the penalty function

$$P_e(\mathbf{e}) = (3/4 - \epsilon/400) \ln(1 - \epsilon/300) + (1/4 + \epsilon/400) \ln(1 + \epsilon/100),$$

which puts much higher weight on deviations below zero than above. (Also note that this penalty function attains its minimum value of zero at $\epsilon=0$.) This cross-entropy penalty function is displayed in Figure 2. While the level of the penalty is higher at $+300$ than at -100 , the penalty is higher at -100 than at 100 . Thus, for deviations of equal magnitude, the penalty on the negative side is higher than on the positive side. By adding more points to the reference distribution and suitably selecting the weights, a wide class of penalty functions can be obtained.

The third answer is in situations where there exists a good, empirical basis for specifying narrow supports or highly peaked reference distributions for the parameters. It is important not to confuse this case with basing the supports on theoretical restrictions on the parameters. For instance, just because Cobb-Douglas exponents must be bounded between zero and one does not mean that a good reference distribution for the parameters of a two-input Cobb-Douglas puts equal weight at zero and one. (This substitutes a stochastic restriction for a deterministic

restriction.) The effect of such a support will be to bias the results towards values of one half for both exponents. There would also be a reduction in the variance of the estimators of the exponents, but the variance reduction would have nothing to do with the data! On the other hand, prior knowledge that the expenditure shares are in the ranges $[0.1, 0.2]$ and $[0.8, 0.9]$ might lead one to define an even more restrictive support that is appropriate because it incorporates prior knowledge into the estimation process.

Concluding Remarks

Rather than focusing on the distribution estimation problems that form the basis of entropy measures, perhaps we are better off looking at the entropy function as a penalty function over deviations. Re-examining the theory in the more general context of penalty functions may lead to useful, unifying results as well as help to identify other classes of “good” penalty functions.

For the general linear problem with symmetric penalty, there appears to be little or no gain to using GME over GOLS in the absence of strong prior knowledge (viz., justification for narrow supports for errors and/or coefficients). In the presence of such knowledge, efficiency gains will likely result. However, in the absence of strong prior knowledge, there is no strong justification for employing entropy for the general linear problem, for it requires specification of additional problem parameters (the reference distributions), thus opening the way for criticism for the choices made.

On the other hand, there exist instances where entropy provides a simple and elegant approach to problems that go beyond the general linear problem. For instance, problems wherein the estimation of a distribution is one of the fundamental goals are good candidates for an entropy approach. Likewise, cross-entropy is a mathematically elegant approach for formulating problems with asymmetric penalty functions. In this regard, there may be instances when it is attractive to “mix our penalty functions” by using, for instance, a least squares approach for the penalty function over errors combined with an entropy-based penalty function for deviations of parameters. Regardless of whether penalty functions are mixed or not, it is important to appropriately weight the penalties for the error deviations from zero versus the coefficient deviations from target whenever penalties are included for both.

As the applied economics profession proceeds forward with the application of entropy-based econometrics and investigates the properties of these estimators, we need to spend careful thought on developing the philosophy of this methodology. The goal of this paper is to provide focus and stimulation to the discussion. The problem addressed here – the general linear model – is narrow, but of sufficient importance to be worthy of analysis and discussion. Entropy provides a general and flexible approach to the construction of penalty functions, and as such, will be a useful addition to the applied econometrician’s toolbox.

References

- Akridge, J.T., 1989. "Measuring Productive Efficiency in Multiple Product Agribusiness Firms: A Dual Approach," *American Journal of Agricultural Economics*, **71**(1):116-125.
- Boot, J.C.G., and G.M. deWitt, 1960. "Investment Demand: An Empirical Contribution to the Aggregation Problem," *International Economic Review*, **1960**:3-30.
- Cranfield, J.A.L. and P.V. Preckel, 1997. "Maximum Entropy Estimation of a Demand System," *American Journal of Agricultural Economics*, **79**(5):1725.
- Georgescu-Roegen, N., 1987. "Entropy," pp.153-156 in *The New Palgrave: A Dictionary of Economics*, eds. J. Eatwell, M. Milgate, and P. Newman, Macmillan, London.
- Golan, A., G. Judge, and D. Miller, 1996. *Maximum Entropy Econometrics: Robust Estimation with Limited Data*, John Wiley & Sons, New York, New York.
- Golan, A., G. Judge, and J. Perloff, 1996. "Estimating the Size Distribution of Firms Using Government Summary Statistics," *Journal of Industrial Economics*, **44**(1):69-80.
- Judge, G.G., W.E. Griffiths, R.C. Hill, H. Lutkepohl, and T.-C. Lee, 1985. *The Theory and Practice of Econometrics*, Second Ed., John Wiley and Sons, New York, New York.
- Kmenta, J., 1986. *Elements of Econometrics*, 2nd Ed., Macmillan Publishing Company, New York.
- Kullback, S. and R.A. Leibler, 1951. "On Information and Sufficiency," *Annals of Mathematical Statistics*, **22**:79-86.
- Marsh, T.L., Mittlehammer, R.C., and N.S. Cardell, 1998. "A Structural-Equation GME Estimator," *American Journal of Agricultural Economics*, in press.
- Mittlehammer, R.C., and N.S. Cardell, 1998. "The Data-Constrained GME Estimator of the GLM: Asymptotic Theory and Inference," mimeo, Washington State University.
- Pukelsheim, F., 1994. "The Three Sigma Rule," *American Statistician*, **48**(4):88-91.
- Shannon, C.E., 1948. "A Mathematical Theory of Communication," *Bell System Technical Journal*, **27**:379-423,623-659.
- Theil, H., 1963. "On the Use of Incomplete Prior Information in Regression Analysis," *Journal of the American Statistical Association*, **58**:401-414.
- Theil, H., and A.S. Goldberger, 1961. "On Pure and Mixed Statistical Estimation in Economics," *International Economic Review*, **2**(1):65-78.