ENTROPY-BASED SEEMINGLY

UNRELATED REGRESSION

by

Alice Harmon, Paul V. Preckel, James Eales

Staff Paper #98-8

May 1998

Dept. of Agricultural Economics

Purdue University

ENTROPY-BASED SEEMINGLY UNRELATED REGRESSION

by
Alice Harmon
Paul V. Preckel
James Eales[1]
Dept. of Agricultural Economics, Purdue University
West Lafayette, Indiana 47907-1145
harmon@agecon.purdue.edu
preckel@agecon.purdue.edu
eales@agecon.purdue.edu
Staff Paper #98-8
May 1998

Abstract

We adapt the classical SUR procedure to a minimum cross entropy approach to estimate linear systems of equations where the errors across equations are correlated. We conclude that our entropy-based approach may provide a reasonable substitute for SUR in cases where classical methods may not be applied due to shortages of data.

Keywords: Entropy, Econometrics, Seemingly unrelated regression.

---

1 The authors are Graduate Research Assistant, Professor, and Assistant Professor, Department of Agricultural Economics, Purdue University, West Lafayette, Indiana.

ENTROPY-BASED SEEMINGLY UNRELATED REGRESSION

by

Alice Harmon, Paul V. Preckel, and James Eales
Dept. of Agricultural Economics, Purdue University
West Lafayette, Indiana 47907-1145
harmon@agecon.purdue.edu
preckel@agecon.purdue.edu
eales@agecon.purdue.edu

Staff Paper #98-8
May 1998

Entropy-based Seemingly Unrelated Regression

The standard least squares approach to regression problems is based on the model where y denotes our dependent variable, x denotes the independent variables, β denotes the parameters to be estimated, and ϵ denotes an error term. This standard approach assumes nothing about β. However,

$$y = x\beta + \epsilon \qquad (1)$$

it makes moderately strong assumptions regarding the error term, ϵ. The error term is assumed to be independent normal with mean zero. The error variance, σ^2, is unknown.

An alternative approach that has been gaining attention is based on the principal of maximum entropy. In contrast to the least squares approach, the entropy approach makes relatively strong assumptions regarding β. Assumptions are made on the bounds for the β's and "priors" regarding their values are assumed. However, fewer assumptions are made regarding the error terms. The error terms are assumed to have a bounded, discrete support and mean zero, but no other structure on the distribution of ϵ is assumed. While the advantages and disadvantages of these two approaches to estimating the relationships between variables are under evaluation, there is no clear consensus on the superiority of either approach.

Due to its long-standing use in estimation problems, least squares methods have been adapted for a number of problems related to (1). One adaptation is for the estimation of parameters in systems of equations. Due to the perpetual problem of omitted variables, an important feature of these system estimation problems is that the errors in (1) are often correlated, not across observations, but across the equations in the system. The Generalized Least Squares (GLS) approach, which encompasses many of the adaptations of (1), has been specialized to the estimation of systems where errors are correlated across equations, resulting in the method of Seemingly Unrelated Regressions (SUR).

Golan, Judge, and Miller (GJM, 1996) address GLS and specifically SUR problems in their recent book on using entropy for the general linear problem. However, the implementation of their approach is somewhat unclear, and their methods that recognize correlated errors are not easily operationalized. Their approach requires prior information on the correlations between the cross-equation errors in the system. This information is not usually known before estimation, but is a result of the estimation process.

The goal of this paper is to develop an easily implemented approach to estimation via the entropy method that closely parallels the SUR method, and then to compare the results of these two approaches to estimation. The well-known General Electric/Westinghouse data set (Boot and deWitt, 1960) will be used to develop the relationship between gross investment (the dependent variable) and the other variables. Estimates will be obtained for both the iterative cross-entropy approach to be developed as well as estimates for the more traditional methods of least squares, Aitken 2-step, and maximum likelihood using the general algebraic modeling system GAMS (Brooke, Kendrick, and Meeraus, 1992). Using the parameter estimates, the similarity of the results is tested by checking that the iterative cross-entropy estimates fall in a 95% confidence interval of the traditional estimates. Further comparison will be done by computing the summary statistics of the errors, including the sum, the variance, and the maximum and minimum values, for both methods and examining each step of the iterative process for both methods.

Background Theory

SUR is a general error covariance statistical model which can be viewed as a special case of GLS. In classical SUR, both the parameters of the conditional means and the error covariance matrix are estimated. The SUR label arises because a separate estimation of the relations ignores the possible correlation of the equation errors. Thus with SUR, the covariance matrix the potential for nonzero covariance between the errors in different equations, but not across observations.

The SUR estimation problem begins by estimating the system using ordinary least squares. Then the covariance matrix of the errors across equations is computed. At the next step, parameter estimates based on feasible GLS using the OLS estimate of the error covariance are computed. If iterations are halted at this point, the technique is often called Aitken 2-step. This estimation process results in new errors, and a new covariance matrix of the errors may be computed. The new matrix is used to weight the errors, and the GLS estimates may again be obtained. If this procedure is repeated until the estimates and covariance matrix converge, the result is the iterated SUR method and the estimates are maximum likelihood.

Adapting SUR to the Entropy Framework

GJM address the problem of seemingly unrelated regressions in their book. They suggest a Generalized Maximum Entropy (GME) approach combined with SUR. To reflect the non-zero covariances between errors across equations, they specify, within a maximum entropy format, an additional set of restrictions that force the error covariance to equal specified values. Because the errors are constrained to have the specified covariance structure, the covariance must be known *a priori*, and it is unclear how one would adapt the procedure to the normal case where the covariance matrix must also be estimated.

A Cross-Entropy Approach

Our entropy-based approach is iterative and parallels the classical SUR procedure closely. To focus attention on the treatment of correlations between the errors, the errors will be "reparameterized", but the coefficients β will not. Thus, the estimation problem will be changed from the least squares setting only via the treatment of the errors.

This approach is Bayesian in flavor and uses a prior distribution on the errors for computing the parameter estimates. As with the traditional SUR method, the initial assumption is that the errors are independent across the equations. The GJM entropy approach requires that we specify a discrete support for the errors. This support is chosen for each individual error to be -5σ, 0, 5σ with prior probabilities 1/50, 48/50, 1/50, respectively. The resulting distribution has mean zero, variance σ^2, no skewness and broad support.

These features are vaguely similar to the normal distribution. Because the errors across equations are correlated, they must be treated as jointly distributed. Thus to start the method, the prior distribution of the errors across equations is set to the n-fold Cartesian product of the priors for the individual errors for the n equation case. This means that for a two equation system, the prior joint distribution of the errors will contain nine points which will be organized as in Figure 1a. (The balance of the exposition will focus on this two equation case. The generalization to n equations is straightforward.)

For convenience in describing the parameterization of the errors, the equations are ordered as follows: (1,1), (2,1), (1,2), (2,2), (1,3), (2,3), ... where the first index corresponds to the equation number in the system, and the second index corresponds to the observation number. That is, rather than ordering the system equation by equation, it is ordered by observation. This simplifies the specification of the matrix **V** below. If the prior joint distribution's probabilities are denoted by q_j for $j = 1,2,...,9$, and the prior distribution points are denoted by a matrix **V**, where

$$\boldsymbol{V} = \begin{bmatrix} u & 0 & 0 & \cdots & 0 \\ 0 & u & 0 & \cdots & 0 \\ 0 & 0 & u & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \cdots & u \end{bmatrix} \quad , \quad u = \begin{bmatrix} -100 & -100 & -100 & 0 & 0 & 0 & 100 & 100 & 100 \\ -100 & 0 & 100 & -100 & 0 & 100 & -100 & 0 & 100 \end{bmatrix} ,$$

and the weights associated with the columns of **V** are $\mathbf{w} = [w_{11}, w_{12}, ..., w_{19}, w_{21}, ..., w_{n9}]$, then the minimum Cross Entropy problem that is analogous to the OLS problem that is used as the initial step of the traditional SUR approach is

$$\textbf{\textit{min}} \quad \sum_{i=1}^{n} \sum_{j=1}^{9} w_{ij} \ln\left(\frac{w_{ij}}{q_j} \right)$$

$$\textbf{\textit{subject to}} \quad y = X\beta + Vw \qquad (2)$$

$$\sum_{j=1}^{9} w_{ij} q_j = 1 .$$

This minimum Cross Entropy formulation chooses parameters β and weights **w** as "close" as possible to **q**, where **q** = $[q_1, q_2, ..., q_9, q_1, ..., q_9]$ is the prior distribution of probabilities (Kapur and Kesavan, 1992). The errors associated with each observation may be obtained as

$$\begin{bmatrix} \epsilon_{i1} \\ \epsilon_{i2} \end{bmatrix} = u \begin{bmatrix} w_{i1} q_1 \\ w_{i2} q_2 \\ \vdots \\ w_{i9} q_9 \end{bmatrix} .$$

The covariance matrix is computed based on these errors. The Cholesky decomposition of this covariance matrix is then used to linearly transform the points in the prior distribution, **u**, with the result that the covariance matrix of the prior is equal to the covariance matrix of the errors obtained from the previous problem. The model is re-estimated with the updated prior distribution.

If Σ is the covariance matrix of the errors, and D where $D^tD = \Sigma$ is obtained from the Cholesky factors of Σ, then the minimum Cross Entropy problem with the updated prior is identical to (2), but the matrix **V** is replaced by

$$V = \begin{bmatrix} D^t u & 0 & \cdots & 0 \\ 0 & D^t u & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & D^t u \end{bmatrix} .$$

After solving this problem, the estimates are analogous to an entropy version of Aitken 2-step. The covariance matrix of the errors is recomputed, a new Cholesky decomposition is computed, and the matrix V is transformed so that the covariance matrix of the prior again matches the covariance matrix of the errors from the previous estimation problem. As in traditional SUR, this process repeats until the β estimates converge.

Data

The General Electric/Westinghouse data set (given in Appendix A), reports gross investment, end-of-period capital stock, and end-of period value of outstanding shares. This data was used to compare our proposed method with the more traditional methods. We first employed the traditional methods of ordinary least squares, Aitken 2-step, and iterated SUR. Then our cross-entropy approach was applied to the same data set for comparison.

Results

The parameter estimates for the more traditional methods of least squares, Aitken 2-step, and maximum likelihood, as well as the parameter estimates for the corresponding Cross Entropy problems, are shown in the Table 1. The parameter estimates from these two approaches are quite similar, but not identical. The signs and magnitudes of the estimated coefficients agree across methods, and excluding the intercept, seven out of twelve of the estimates match the first significant digit. Furthermore, the Cross Entropy parameter estimates fall well within 95% confidence intervals of the traditional estimates. While the latter point does not imply that one method or the other is closer to "true" values, it does provide evidence that the methods produce similar results.

To extend the comparison, the summary statistics of the errors are examined. The sum, the maximum and minimum values, and the sum of the absolute errors were computed for both methods. The sum of the errors in the cross-entropy approach stand out, because they are not forced to sum to zero as they do automatically in the least squares case. However the maximum and the minimum values, compared between methods, are close. The sum of the absolute errors are also of similar magnitude but are uniformly lower for the entropy-based approach.

To give insight into the dynamics of the iteration process, Table 3 lists the covariance matrices for the errors for both methods. Comparing across the least squares and cross-entropy methods, these matrices are quite similar. All of the variances except one agree to one significant digit, and all of the covariances except one agree to two significant digits. In both cases, the disagreement is for the case where iterations continued to convergence (i.e., Max Likelihood for least squares, and CE Iterated for cross-entropy).

In the cross-entropy formulation, recall that the effect of the covariance matrix on the estimation problem is through rotation and scaling of the joint prior distribution for the errors across equations. The effect on the prior for this problem is illustrated in Figures 1a-1c. For both methods, the covariance matrix changes much more in going from the case of independent errors to the first GLS iteration than in all subsequent iterations combined. Hence, it is not surprising to see that the prior distribution changes far more in going from the independent case to the prior for the first rotation and scaling than for the subsequent rotations and scalings. That is, it is not surprising that Figures 1a and 1b are very different, while Figures 1b and 1c are qualitatively quite similar.

Table 1. Parameter Estimates for Least Squares and Entropy-Based SUR Methods & 95% Confidence Bounds for Least Squares Estimates

	General Electric			*Westinghouse*		
	Intercept	β_C	β_F	Intercept	β_C	β_F
Least Squares						
OLS	-9.956	0.152	0.027	-0.509	0.092	0.053
Aitken 2-Step	-27.719	0.139	0.038	-1.252	0.064	0.058
Maximum Likelihood	-30.748	0.136	0.041	-1.702	0.056	0.059
Cross Entropy						
CE Indep. Errors	-11.194	0.156	0.025	-0.559	0.106	0.051
CE 2-Step	-30.303	0.147	0.037	-0.374	0.075	0.054
CE Iterated	-32.928	0.145	0.038	-0.869	0.069	0.055
95% Confidence Intervals for Least Squares Estimates						
OLS	[-76.32,56.14]	[0.097,0.205]	[-0.006,0.059]	[-17.42,16.40]	[-0.026,0.211]	[0.020,0.086]
Aitken 2-Step	[-89.82,33.96]	[0.086,0.198]	[0.008,0.069]	[-17.20,14.64]	[-0.048,0.176]	[0.027,0.088]
Maximum Likelihood	[-93.60,31.62]	[0.082,0.190]	[0.010,0.071]	[-17.59,14.12]	[-0.056,0.167]	[0.029,0.090]

Table 2. Residual Analysis

	General Electric				*Westinghouse*			
	$\Sigma_i e_i$	***max*** e_i	***min*** e_i	$\Sigma_i \lvert e_i \rvert$	$\Sigma_i e_i$	***max*** e_i	***min*** e_i	$\Sigma_i \lvert e_i \rvert$
Least Squares								
OLS	0	58.737	-37.511	398.867	0	17.265	-13.505	163.674
Aiken 2-step	0	54.912	-34.511	420.555	0	16.493	-12.323	169.334
Maximum Likelihood	0	54.157	-36.013	428.741	0	16.562	-12.162	170.949
Cross-Entropy								
CE Indep. Errors	46.208	61.613	-36.589	390.966	5.030	17.903	-13.975	160.739
CE 2-Step	53.589	58.492	-30.390	408.956	16.449	17.347	-11.766	165.561
CE Iterated	53.957	57.559	-30.465	414.275	16.2758	17.234	-11.395	166.806

Table 3. Final Covariance Matrices of the Residual Errors

Least Squares		GE	W	Cross-Entropy		GE	W
OLS	GE	660.829	176.449	*CE Ind. Errors*	GE	667.045	176.351
	W	176.449	88.662		W	176.351	89.042
Aiken 2-step	GE	689.4188	190.636	*CE 2-Step*	GE	684.416	187.483
	W	190.636	90.065		W	187.483	90.160
Max Likelihood	GE	702.233	195.352	*CE Iterated*	GE	691.052	190.038
	W	195.352	90.953		W	190.038	90.510

Conclusions

We have developed an entropy-based analogue to the SUR method which is practical to apply. The method appears to function similarly to the traditional SUR method and produces similar estimates. Hence, it is possible to conclude that an entropy-based approach may provide a reasonable alternative to SUR. This approach is applicable in both the standard SUR case as well as in cases where there are cross equation restrictions such as in the case of demand systems. Given the sparse nature of demand data in many cases, the combined approach may be most useful for coping with correlated errors across equations in demand systems. While this paper does not advocate replacing SUR with the combined cross-entropy/SUR approach, this approach may be useful in cases where classical methods may not be applied due to shortages of data.

Figure 1a. The Initial Prior

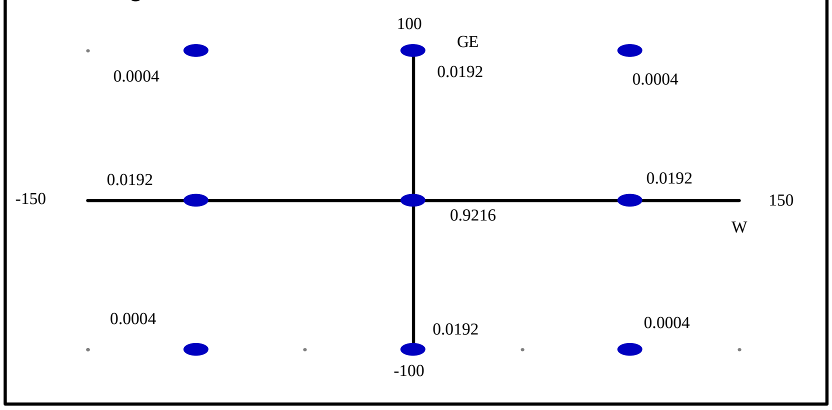


Figure 1b. The Prior After One Step

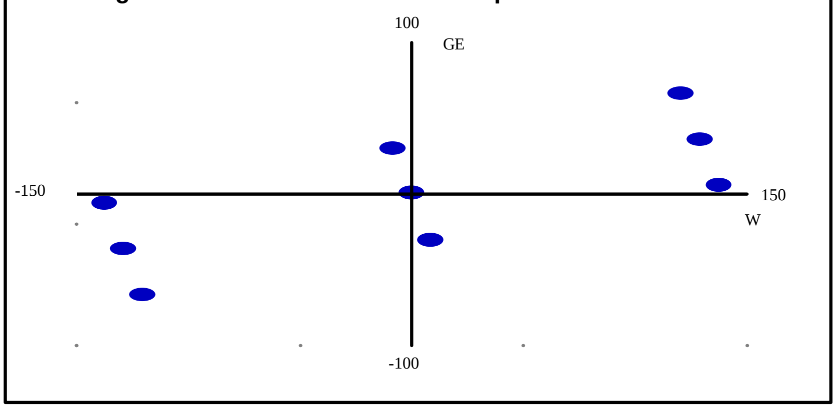
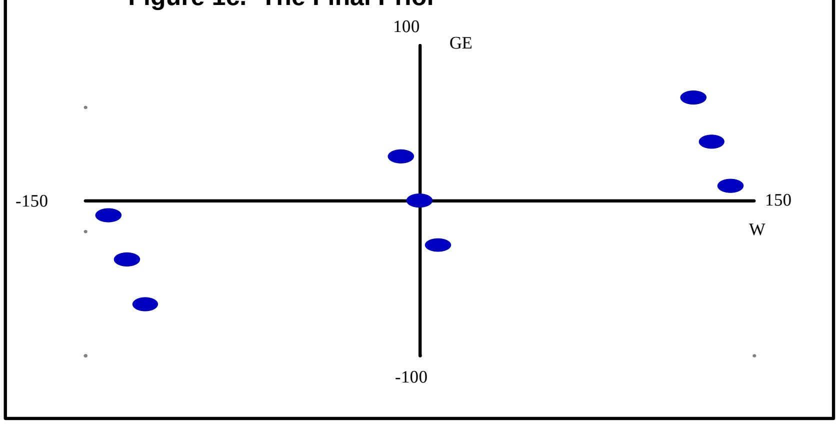


Figure 1c. The Final Prior



The axes labeled GE correspond to the errors for General Electric, and the axes labeled W correspond to the errors for Westinghouse. The decimal values shown near the points in Figure 1a are the corresponding probabilities. The same probabilities apply to the rotated and scaled points in Figures 1b and 1c.

References

Boot, J.C.G., and deWitt, G.M., 1960. "Investment Demand: An Empirical Contribution to the Aggregation Problem," *International Economic Review*, 1960:3-30.

Brooke, A., Kendrick, D., and Meeraus, A., 1992. *GAMS: A User's Guide*, Release 2.25, The Scientific Press, San Francisco, California.

Golan, A., Judge, G., Miller, D. 1996. *Maximum Entropy Econometrics: Robust Estimation with Limited Data*, John Wiley & Sons, Chichester.

Kapur, J. N., Kesavan, H. K.,1992. *Entropy Optimization Principles with Applications*, Academic Press, Boston.

Appendix A

	General Electric			Westinghouse		
	Gross Investment	End-of-period Capital Stock	End-of-period value of outstanding shares	Gross Investment	End-of-period Capital Stock	End-of-period value of outstanding shares
1935	33.10	1170.60	97.80	12.93	191.50	1.80
1936	45.00	2015.80	104.40	25.90	516.00	0.80
1937	77.20	2803.30	118.00	35.05	729.00	7.40
1938	44.60	2039.70	156.20	22.89	560.40	18.10
1939	48.10	2256.20	172.60	18.84	519.90	23.50
1940	74.40	2132.20	186.60	28.57	628.50	26.50
1941	113.00	1834.10	220.90	48.51	537.10	36.20
1942	91.90	1588.00	287.80	43.34	561.20	60.80
1943	61.30	1749.40	319.90	37.02	617.20	84.40
1944	56.80	1687.20	321.30	37.81	626.70	91.20
1945	93.60	2007.70	319.60	39.27	737.20	92.40
1946	159.90	2208.30	346.00	53.46	760.50	86.00
1947	147.20	1656.70	456.40	55.56	581.40	111.10
1948	146.30	1604.40	543.40	49.56	662.30	130.60
1949	98.30	1431.80	618.30	32.04	583.80	141.80
1950	93.50	1610.50	647.40	32.24	635.20	136.70
1951	135.20	1819.40	671.30	54.38	723.80	129.70
1952	157.30	2079.70	726.10	71.78	864.10	145.50
1953	179.50	2371.60	800.30	90.08	1193.50	174.80
1954	189.60	2759.90	888.90	68.60	1188.90	213.50

Table from data given in J.C.G. Boot and G.M. deWitt, "Investment Demand: An Empirical Contribution to the Aggregation Problem," *International Economic Review*, January 1, 1960, pp 3-30.