COWLES COMMISSION DISCUSSION PAPER: STATISTICS NO. 385

Some Recent Work of H. Theil on Estimation in Systems of Simultaneous Equations

by R. Radner and F. Bobkoski

May 28, 1954

1. Introduction

Dr. Hans Theil of the Netherlands School of Economics, Rotterdam, has recently proposed a new method of estimating the coefficients of a single equation in a system of simultaneous linear equations, and has shown how this method is a member of a general class of estimators, of which the least squares and limited information estimators are also members [1]. Since Dr. Theil's work is not likely to become generally available for some time, it seems worthwhile to summarize it for those who are currently interested in the field of estimation of simultaneous linear equations. At the same time the notation has been changed to correspond fairly closely to that of Cowles Commission Monograph 14 [2].

In what follows, sections 2-5 summarize the main theoretical points of Dr. Theil's paper. It should be noted, however, that many points in his paper are not covered, and that the selection and presentation here are our own. Section 6 gives our comparison of the computational aspects of the limited information method and a method proposed by Dr. Theil.

2. Two Rounds Estimation: A Heuristic Approach

We shall be considering the system of simultaneous linear equations

(1) $$B\, y(t) + \Gamma z(t) + u(t) = 0 \quad , t = 1, \ldots, T$$

where $y(t)$ and $z(t)$ are the vectors of jointly dependent and predetermined variables, respectively, and B and Γ are matrices of unknown coefficients. We wish to estimate the coefficients in one of these equations, say the first.* Suppose that in the whole system there are G jointly dependent and H predetermined variables, whereas only g of the jointly dependent and h of the predetermined variables appear in the first equation. Suppressing the subscript 1 we may write the first equation of (1) as:

(2) $$\sum_{i=1}^{g} \beta_i\, y_i(t) + \sum_{j=1}^{h} \gamma_j\, z_j(t) + u(t) = 0$$

The "normal equations" of least squares estimation suggest that we form the sample moments of equation (2) with all the predetermined variables. This gives us

(3) $$\sum_{i=1}^{g} \beta_i\, m(y_i z_n) + \sum_{j=1}^{h} \gamma_j\, m(z_j, z_n) + m(u, z_n) = 0$$

$$n = 1, \ldots, H.$$

where $m(y_i, z_n) = \frac{1}{T} \sum_{t=1}^{T} y_i(t)\, z_n(t)$, etc. Since the expected value of $m(u, z_n)$ is zero, we might think of estimating the β_i and γ_j by numbers b_i and c_j determined by

---

* Even if we want to estimate all the equations we still may prefer doing them "one at a time", as apposed to using the full maximum liklihood method.

$$(4) \qquad \sum_{i=1}^{g} b_i \, m(y_i, z_n) + \sum_{j=1}^{h} c_j \, m(z_j, z_n) = 0 \quad ,$$

$$n = 1, \ldots, H$$

As a matter of fact, if equation (2) is <u>just identified</u> in the system (1), then (4) will, in general, have a solution (up to an arbitrary constant of proportionality) and the estimates so derived will be the limited information estimates.

If (2) is <u>over-identified</u> (in general, if $H > g + h - 1$) then there will be more equations than unknowns, and (4) will have no solution. Suppose we normalize (3) by setting $\beta_1 = -1$.

If we let

$$Y_i(n) = m(y_i, z_n)$$
$$Z_j(n) = m(z_j, z_n)$$
$$v(n) = m(u, z_n)$$

then (3) becomes

$$(3') \qquad Y_1(n) = \sum_{i=2}^{g} \beta_i Y_i(n) + \sum_{j=1}^{h} \gamma_j \, Z_j(n) + v(n)$$

$$n = 1, \ldots, H.$$

Equation (3') looks like a regression of Y_1 on $Y_i (i = 2, \ldots g)$ and $Z_j (j = 1, \ldots, h)$, with H observations and disturbances $v(n)$. If the original disturbances $u(t)$ were uncorrelated in time, then we can determine the covariance matrix of the $v(n)$ up to multiplication by a constant. Of course, the $v(n)$ are not independent of $Y_2(n), \ldots, Y_g(n)$. However, we can calculate those estimates $\hat{b}_i$ and $\hat{c}_j$ which would be minimum variance linear unbiassed estimates of the β_i and γ_j <u>if the $v(n)$ were independent of $Y_2(n), \ldots, Y_g(n)$.</u> (See [3], for example).* These will be called the <u>second-round</u> estimates of the β_i and γ_j. They are, under certain conditions, consistent but not unbiassed. <u>They are closely related to limited</u>

* It should be noted that, under the assumptions of section 4, the covariances between $v(m)$ and the $Y_i(n)$ are of the order of $1/T$.

information estimates, and this relationship will be brought out in the next section.

3. Two-Rounds, Limited Information and Least Squares Estimators in the General Class of R(k) Estimators.

In order to discuss the various estimators, we will need some notation to describe the second order sample moments, of which the estimates are functions. Let the symbol Δ refer to the jointly dependent variables which appear in the first equation of (1), and let $*$ refer to those predetermined variables which appear in the first equation. Finally, when we normalize equation (2) by setting $\beta_1 = -1$, the symbol $\bar{\Delta}$ will refer to the remaining variables $y_2, \ldots, y_g$ which make up y_Δ. As for the sample moment matrices, $M_{\Delta,\Delta}$ will denote the matrix of second order moments (about 0) of y_Δ with itself, $M_{\Delta,*}$ the moments of y_Δ with z_*, $M_{\bar{\Delta},1}$ the moments of $y_{\bar{\Delta}}$ with y_1, etc.

For any non-negative number k we now define estimates b_i of $\beta_i (i=2,\ldots,g)$ and c_j of γ_j $(j = 1, \ldots, h)$ by the following equation

$$(5) \qquad r(k) = R(k) \begin{bmatrix} c \\ b \end{bmatrix}$$

$$R(k) = \begin{bmatrix} M_{\bar{\Delta},\bar{\Delta}} - k\, W_{\bar{\Delta},\bar{\Delta}} & M_{\bar{\Delta},*} \\ M_{*,\bar{\Delta}} & M_{*,*} \end{bmatrix} \text{ (matrix)}$$

$$r(k) = \begin{bmatrix} M_{\bar{\Delta},1} - k\, W_{\bar{\Delta},1} \\ M_{*,1} \end{bmatrix} \text{ (column vector)}$$

$$b' = (b_2, \ldots, b_g)$$

$$c' = (c_1, \ldots, c_h)$$

Here $W_{\Delta\Delta}$ is the sample moment matrix of residuals in the least squares regressions (reduced form) of the variables $y_1, \ldots, y_g$ on $z_1, \ldots\ldots, z_H$; $W_{\bar{\Delta}\bar{\Delta}}$ is $W_{\Delta\Delta}$ with the first row and column deleted; $W_{\bar{\Delta}1}$ is the first column of $W_{\Delta\Delta}$ with the first element deleted.

The estimates determined by (5) will be called R(k)-estimates. We can now state the first main fact about these estimates.

Theorem 1. (i) For $k = 1$, the R(k) estimates are the second round estimates discussed in section 2. (ii) Setting $k = 1 + v$ gives the limited information estimates, where v is the smallest root of the determinental equation

$$| W^*_{\Delta\Delta} - (1 + v) W_{\Delta,\Delta}| = 0$$

and $W^*_{\Delta\Delta}$ is the sample moment matrix of residuals in the least squares regressions of $y_1, \ldots, y_g$ on $z_1, \ldots, z_h$. (iii) Setting $k = 0$ gives the least squares estimates when y_1 is taken as the dependent variable and $y_2, \ldots, y_g, z_1, \ldots, z_h$ as the independent variables.

4. Asymptotic Properties of R(k) Estimates.

We first list the assumptions made about the system (1). It must be clear that all probability statements made are conditional statements given the values of the predetermined variables $z(1), \ldots, z(T)$. Since we shall be interested in asymptotic properties we must assume we are given an infinite sequence of predetermined (vector) variables $z(1)$, $z(2)$, etc.

The assumptions are

I. B is non-singular

II. For every T, $M_{\Delta,\Delta}$ is non-singular

III. For every T, the matrix $\begin{bmatrix} M_{\Delta,\Delta} & M_{\Delta,*} \\ M_{*,\Delta} & M_{*,*} \end{bmatrix}$ has rank $(g + h - 1)$,

and as T increases, this matrix converges in probability to a matrix with rank $(g + h - 1)$.

IV. The $u(t)$ form a sequence of independent, identically distributed (vector) variables, and the conditional distribution of $u(t)$ given $z(t)$ is independent of $z(t)$.

Theorem 2.

Under assumptions I-IV, the $R(k)$ estimates $b_2, \ldots, b_g, c_1, \ldots, c_h$ defined by (5) have the following properties (with k possibly a function of the sample):

(i) if $\underset{T \to \infty}{\text{plim}} (k - 1) = 0$, then they are consistent*

(ii) if $\underset{T \to \infty}{\text{plim}} (k - 1) \sqrt{T} = 0$, then the estimates multiplied by $\sqrt{T}$ have the asymptotic covariance matrix

$$\sigma^2 \underset{T \to \infty}{\text{plim}} \left[R(1) \right]^{-1} \tag{6}$$

where σ^2 is the variance of $u_1(t)$.

Corollary 1. Limited information and second round estimates have the same asymptotic covariance matrix (6).

This follows from the fact that v, as defined in theorem 1, (ii), is of order $\frac{1}{T}$, in probability.

Corollary 2. If in addition to assumptions I-IV, the disturbances $u(t)$ are normally distributed, then all $R(k)$ estimates for which $\underset{T \to \infty}{\text{plim}} (k-1) \sqrt{T} = 0$ are asymptotically efficient.

This follows from the fact that the corollary is true for limited information estimates.

5. A Coefficient of Simultaneous Correlation

To measure the closeness of fit of the entire system of equations, Theil suggests a coefficient of simultaneous correlation S, where $(1 - S^2)$ is the ratio of the generalized variance of all the disturbances in the reduced

* "plim" means limit in probability.

form to the generalized variance of all jointly dependent variables. The quantity $(1 - S^2)$ may be estimate by (for example) $\frac{|W_{\Delta\Delta}|}{|M_{\Delta\Delta}|}$

6. Comparison of Computations Necessary for Theil's Method and the Limited Information Method.

For both Theil's method and the Limited Information Single Equation Method (L.I.S.E.) we compute the moment matrices $M_{\Delta z}$, the moment matrix of the jointly dependent variables in the equation to be estimated on all of the predetermined variables, and M_{zz}, the moment matrix of all predetermined variables. L.I.S.E. requires in addition $M_{\Delta\Delta}$, the moment matrix of the jointly dependent variables in the equation. For both methods we compute $M_{\Delta z} M_{zz}^{-1} M_{z\Delta}$.

In Theil's method we solve $r(1) = R(1)\binom{c}{b}$. The matrix $R(1)$ is inverted in order to obtain the covariance matrix of the estimate of $\binom{c}{b}$. The computation $r(1)\,R(1)^{-1}$ completes the estimation of $\binom{c}{b}$. The sum of squares of the residuals divided by some appropriate number of degrees of freedom estimates the variance of the disturbances. This estimate multiplied by $R(1)^{-1}$ completes the computation using Theil's method.

L.I.S.E. requires computation of $W_{\Delta\Delta} = M_{\Delta\Delta} - M_{\Delta z} M_{zz}^{-1} M_{z\Delta}$, $P' = M_{z^*z^*}^{-1} M_{z^*\Delta}$, $B = M_{\Delta z} M_{zz}^{-1} M_{z\Delta} - M_{\Delta z^*} P'$, and $A = B^{-1} W_{\Delta\Delta}$, where z^* refers to the predetermined variables in the equation to be estimated. The largest characteristic root of A is obtained and its associated vector λ properly normalized, gives the estimate of b, then $c = -P'b$. In order to estimate the covariance matrix we compute the following quantities:

$$x \quad c = (1 + \frac{1}{\lambda})\, \beta\, W_{y^*y^*}\beta, \quad F_{\Delta b} = \left[{}_{11}\left(B_{y^*y^*} - \frac{1}{\lambda\, b\, W_{y^*y^*} b}, W_{y^*y^*}\, b'\, b\, W_{y^*y^*}\right) \right]^{-1}$$

$F'_{bc} = {}_{01}[P^{*'}]\, F_{bb}\, F_{cc} = {}_{01}[P^{*'}]\, F_{\beta\gamma} + M_{z}^{-1}{}_{*z^*}$. The covariance matrix of the coefficients is $\frac{C}{C-H} \begin{pmatrix} F_{bb} & -F_{bc} \\ -F_{cb} & F_{bc} \end{pmatrix}$

Making a time comparison is general is impossible since the relative times will depend on the number of dependent variables in the equation, the number of predetermined variables in the equation and the number of predetermined variables in the system of equations. We are going to be specific and make a comparison in two computations to be described below. In making this comparison we will assume that the residuals will be computed in the L.I.S.E. computation.

The first is an equation involving three dependent variables and three predetermined variables in the equation, and six predetermined variables not in the equation. There were thirty observations on each variable and logs of the data to six decimal places were used in the computation. Table 1 gives our estimates of the time necessary to complete both computations.

Table 1

L.I.S.E.		Two Rounds (Theil)
Moment matrix	25 hrs.	24 hrs.
$M_{\Delta z}\, M_{zz}^{-1}\, M_{z\Delta}$	12	12
estimate of $\binom{c}{b}$	3	2
residuals and covariance matrix	4	2
Total	44	40

The variance of these time estimates is, we suspect, rather large, but we do not believe that it is large enough to make the comparison worthless.

The time saving is not very large; however, we note that the computation of the moment matrix and $M_{\Delta z}M_{zz}^{-1}N_{z\Delta}$ takes up most of the computation time and that the matrix involved in the iteration and $M_{\Delta\Delta}$ are three by three matrices. This example then minimizes the time saving one might expect by using Theil's method.

We have also estimated the time it would take to complete the computations by both methods in the Girshick and Haavelmo paper on the analysis of demand for food [4]. The time estimates for all equations are combined and appear in Table 2.

Table 2

	L.I.S.E.	Two Rounds
Moment matrix	$7\frac{1}{2}$ hrs.	$5\frac{1}{2}$ hrs.
$M_{yz}M_{zz}^{-1}M_{zy}$	2	2
coefficient $\binom{c}{b}$	$8\frac{1}{2}$	3
residuals and covariance matrices	10	6
Total	28	$16\frac{1}{2}$

The relative time saving here using Theil's method is substantial since it is necessary to compute 30 moments as compared to 45 in L.I.S.E. and the computations were carried out on the original data instead of on the logs. Secondly the order of $R(1)$ in each of these equations is small, so that computing $\binom{c}{b}$ by Theil's method is short. Even so, the full advantages of Theil's method are not apparent because only two of the equations require an iteration in L.I.S.E.

References

[1] H. Theil, "Estimation and Simultaneous Correlation in Complete Equation Systems", Centraal Plan-Bureau, The Hague, June 23, 1953 (mimeographed). See also a paper to appear in the proceedings of the 28th International Statistical Institute in Rome.

[2] W. C. Hood and T. C. Koopmans (ed.), "Studies in Econometric Method, (especially Ch. VI), Wiley, New York, 1953.

[3] A. C. Aitken, "On Least Squares and Linear Combination of Observations", Proc. Roy. Soc. of Edinburgh, 55, (1934-5), p. 42.

[4] M. A. Girshick and T. Haavelmo, "Statistical Analysis of the Demand for Food", Econometrica, Vol. 15, (1947), pp. 79-109.