No. 214 (84-10)

On a Variable Selection Procedure for Fuller's Modified Limited Information Maximum Likelihood Method

by

Haruo ONISHI

February 1984

On a Variable Selection Procedure for Fuller's Modified Limited Information Maximum Likelihood Method

University of Tsukuba
Haruo ONISHI

Abstract: W.A. Fuller [5] proposed a modification of the limited information maximum likehood (LIML) method developed by W.T. Anderson and H. Rubin [1]. The Fuller's modified LIML (F-LIML) method possesses exact moments which the LIML method does not. Also, the F-LIML method is asymptotically unbiased to $O(T^{-1})$ when $\alpha=1$. Thus, the F-LIML method is appropriate to be used for the estimation of equations in a simultaneous equation model. The j-th best subset problem is formulated for the variable selection problem of the F-LIML method. The (ultimately) best subset obtained by solving the first (to the J-th, e.g. J=10) best subset problem(s) in one computer-run is proposed as an approximate solution, which can be a pragmatically best subset, to the variable selection problem for the F-LIML method. Such a subset has the following characteristics in addition to the above statistical properties: (i) meaningful from the viewpoint of a research field in question, (ii) just- or over-identifiable, (iii) satisfying the magnitude conditions on (values of linear functions of) estimated coefficients (including the sign conditions), the Durbin-Watson test ([2], [3], [4]), the relative absolute error test and the turning point test [8], if applied, and (vi) showing the highest (or sufficiently high) adjusted coefficient of determination.
Thus, the laborious work of estimation by the F-LIML method can be drastically reduced.

1. Introduction

Since the work of estimation requires various information to be used not only from a research field in question but also from econometrics, it is quite laborious. Econometricians and statisticians have made efforts to develop new methodologies by which estimation can be efficiently made. Inspite of their efforts, no variable selection procedures for estimation methods for a single equation in a simultaneous equation model (called a model from here on) have been proposed so far in the literature. Thus, much brain labor and other resources like paper and electricity are wasted in the world every year on research which needs estimation. However, if we examine the work of estimation in detail, we can find a gap between theoretical econometricians' interests and applied econometricians' needs which must be filled to make the work of estimation less time-consuming, labor-consuming and costly. Theoretical econometricians pay much attention to properties of estimators, tests, how to measure a goodness of fit, etc. On the other hand, applied econometricians are more sensitive to selection of correct variables, signs of estimated coefficients and magnitudes of (values of linear functions of) estimated coefficients, etc. in addition to statistical properties. Suppose, for instance, that a statistically best equation for NO_2 in the atmosphere shows a negative coefficient of the explanatory variable which represents the number of traffic vehicles. If the number of traffic vehicles is increased, other things being equal, then the equation states that the amount of NO_2 will be reduced, even though traffic vehicles emit but never remove NO_2 gas. Applied researchers cannot accept such an equation in analysis and prediction. As a result, they repeatedly estimate and evaluate many equations through appropriate criteria by trial and error until obtaining a satisfactory equation.

This paper proposes a variable selection procedure for the F-LIML method. The proposed procedure is available in the computer package OEPP, designed for socio-economic analysis and forecasting, which has been developed by the author [9].

2. Derivation of All Possible Meaningful and Just- or Over-identifiable Subsets

When N non-constant explanatory variable candidates are specified for an explained variable, we can derive 2^N-1 possible subsets ('variable candidates' are called 'candidates' from here on). It is rare in application that all 2^N-1 possible subsets are meaningful from the viewpoint of common sense or the information (theories, surveys, experiments, empirical studies and so on) obtained from a research field in question, especially when N is large. Usually, all possible subsets are partitioned into a group of meaningful subsets and a group of meaningless subsets. A meaningful subset is defined as the one which (i) has candidates necessary for an equation which is reasonable for the research but (ii) does not have any other unnecessary candidates. Unless a subset satisfies (i) or (ii) of the above conditions, it is meaningless from the viewpoint of the research. Since the F-LIML method is an estimation method for a single equation in a model, a subset must be just- or over-identifiable (called identifiable from here on) from the viewpoint of econometrics. Thus, the difficulty in solving the variable selection problem for the F-LIML method is how to derive all possible meaningful and identifiable subsets from a given set of N explanatory and instrumental candidates specified for an explained endogenous variable (called an explained variable from here on), regardless of research fields.

The author [8] pointed out that when the ordinary least squares (OLS) method is utilized, variable classifications based on the information from a research field in question can lead to the derivation of all possible meaningful subsets from a given set of explanatory candidates specified for an explained variable. In general, candidates are classified on the basis of information from a research field into the following eight groups: (i) absolutely important (or forced or core), (ii) optionally important, (iii) exclusively important, (iv) gradually important, (v) exclusively optional, (vi) gradually optional, (vii) completely optional and (viii) fixed. The variable classifications for the OLS method can be used for the F-LIML method, too. Table 1 summarizes the meanings and characteristics of the 8 categories and selection of classified candidates.

The F-LIML method requires that candidates specified for an explained endogenous variable be divided into (i) a set of included predetermined candidates (called included candidates from here on), (ii) a set of explanatory endogenous candidates (called endogenous candidates from

Table 1. Variable Classifications for the F-LIML Method by Functional Format $y=F(X^1:Y:X^2)$ in Case of Three Candidates A, B and C

Names	Classifications	Selection of Candidates
Absolutely Important	/A,B,C/	(1) A,B,C
Optionally Important	<A,B,C>	(1) A,B,C; (2) A,B; (3) A,C; (4) B,C; (5) A; (6) B; or or (7) C
Exclusively Important	</A,B,C/>	(1) A; (2) B; or (3) C
Gradually Important	<+A,B,C+>	(1) A,B,C; (2) A,B; or (3) A
Exclusively Optional	<*A,B,C*>	(1) A; (2) B; (3) C; or (4) Empty (no selection)
Gradually Optional	<-A,B,C->	(1) A,B,C; (2) A,B; (3) A; or (4) Empty
Completely Optional	A,B,C	(1) A,B,C; (2) A,B; (3) A,C; (4) B,C; (5) A; (6) B; (7) C; or (8) Empty
Fixed	(D,E)	For instance, if B=(D,E) above, all B's must be replaced with D,E.
NAI & NU Included Predetermined	'A,B,C'	If all or some of A, B and C are not selected as included candidates, they must be selected as excluded ones.
Constant Term	$C	Always selected, if any.

Footnotes: (1) NAI & NU stands for non-absolutely-important and non-uniquely. (2) y=explained variable, X^1=set of included candidates, Y=set of endogenous candidates, X^2=set of excluded or NAI & NU included candidates. (3) X^1, Y and X^2 are classified by the above rules. (4) Candidates A, B and C can be expressed with at most 8 alphanumeric symbols (and minus time lag numbers in parentheses like HLWK(-2)). (5) Candidates in a functional format are separated from each other by a blank, a comma (,), /, <, >, </, />, <+, +>, <*, *>, <-, ->, (,), : or '. (6) For example, AA=F($C/BC/D:<+E E(-1)+>FG</I(I1,I2)/>:<J,K>'D'/L,M,N,PQ/). (7) OEPP can handle nested variable classifications.

here on) and (iii) a set of excluded predetermined candidates (called excluded candidates from here on) [6]. Furthermore, we need to classify and divide included candidates into (i) uniquely included ones and (ii) non-uniquely included ones. Special attention is paid only to non-uniquely included candidates which are not absolutely important. Non-uniquely included candidates which are absolutely important can be treated just like uniquely included ones. Let us explain the difference between a uniquely included candidate and a non-uniquely included candidate. A uniquely included candidate is defined as the one which (i) could appear in the equation at hand but (ii) never appears in any other equations and identities in a model. On the other hand, a non-uniquely included candidate is defined as the one which (i) could appear in the equation at hand and (ii) does appear in at least one of the other equations and/or identities in a model. Suppose that a non-uniquely included candidate x is not absolutely important. If candidate x is not selected as an included candidate in a meaningful subset, x can be used as an excluded candidate for that subset. Even if such a subset is under-identified because the number of the endogenous candidates exceeds by one that of the excluded candidates selected in the subset, it can be just-identified by selecting x as an excluded candidate. However, if candidate x is treated as absolutely important, x is always selected as an included candidate in all meaningful and identifiable subsets. In this case, candidate x must not be selected as an excluded candidate for all meaningful and identifiable subsets. To allow for these cases, we need to distinguish non-uniquely-included and non-absolutely-important candidates from other included candidates.

Let y, x_k^1's ($0 \leq k \leq K$), y_ℓ's ($1 \leq \ell \leq L$), and x_m^2's ($1 \leq m \leq M$) stand for an explained variable (and its data vector), all possible (uniquely or non-uniquely) included candidates (and their data vectors), all possible endogenous candidates (and their data vectors), and all possible excluded or non-uniquely-included and non-absolutely-important candidates (and their data vectors), respectively, where x_0^1 stands for a constant term. Furthermore, the relationship between y, x_k^1's, y_ℓ's and x_m^2's is loaded into a computer through a functional format like $y=F(X^1:Y:X^2)$, where X^1, Y and X^2 stand for sets (and data matrices) of x_k^1's, y_ℓ's and x_m^2's, respectively, and X^1, Y and X^2 are classified on the basis of information on all possible candidates. Thus, non-uniquely-included and non-absolutely-important candidates are entered into both X^1 and X^2 in a functional format. It must be noted that X^1, Y and X^2 are separated from each other with a colon (:) in a functional format and, furthermore, X^1 is entered first on the right-hand side of a functional format, followed by Y which is followed by X^2. In deriving only meaningful and identifiable subsets from $y=F(X^1:Y:X^2)$, subsets must be ignored which (i) contain non-uniquely-included and non-absolutely-important candidates as included as well as excluded ones or (ii) do not have at least one of those candidates.

To allow for time series, cross-sectional and pooled data, we assume that there are N (cross-sectional) units (e.g. sectors, divisions, areas, plots, households) and T observation times and the datum of an explained variable y in unit n at observation time t is expressed as $y_n(t)$ for $1 \leq n \leq N$ and $1 \leq t \leq T$. In the case of pooled data, y is expressed as $y = (y_1(1), y_2(1), \ldots, y_N(1), \ldots, y_n(t), \ldots, y_1(T), y_2(T), \ldots, y_N(T))'$. It is assumed that x_k^1's and y 's are considered to explain the behavior or movement of an explained variable y and x_m^2's are instrumental candidates for the F-LIML method.

It is essential to classify x_k^1's, y_ℓ's and x_m^2's by using the information from a research field in question and econometrics in order to derive only meaningful and identifiable subsets from a given set of all possible included, endogenous and excluded candidates specified for an explained variable.

Let us give an example. Suppose that a researcher loads the following entry into a computer:

```
KU=F($C/OK,NSD/HRNK:<*(ANDSN,RBN)CD*>/TNGW/:'HRNK'
     /YKW,TMNG,FKI/)                                  (1)
```

where $C stands for a constant term. Then, he can have the following 6 equations:

$$KU = a_0 + a_1 OK + a_2 NSD + a_3 HRNK + a_4 ANDSN + a_5 RBN + a_6 TNGW \quad (2)$$

$$KU = b_0 + b_1 OK + b_2 NSD + b_3 HRNK + b_4 CD + b_5 TNGW \quad (3)$$

$$KU = c_0 + c_1 OK + c_2 NSD + c_3 HRNK + c_4 TNGW \quad (4)$$

$$KU = d_0 + d_1 OK + d_2 NSD + d_3 ANDSN + d_4 RBN + d_5 TNGW \quad (5)$$

$$KU = e_0 + e_1 OK + e_2 NSD + e_3 CD + e_4 TNGW \quad (6)$$

$$KU = f_0 + f_1 OK + f_2 NSD + f_3 TNGW \quad (7)$$

where a_i's, b_i's, c_i's, d_i's, e_i's and f_i's stand for coefficients. 'HRNK' in format (1) implies that HRNK is a non-uniquely included candidate. Whether OK and NSD are uniquely or non-uniquely included candidates does not matter, because they are treated as absolutely important. Equations (2) to (4) are estimated by using excluded candidates YKW, TMNG and FKI, so that equation (2) is just-identified but equations (3) and (4) are over-identified. Equations (5) to (7) possess excluded candidates HRNK, YKW, TMNK and FKI, so that they are over-identified.

If 'HRNK' is replaced with HRNK in format (1), 3 more equations are obtained in addition to the above 6 equations. Since HRNK can be treated as a completely optional excluded candidate in cases where it is not selected as an included candidate, equations (5) to (7) which do not have candidate HRNK are estimated with two different sets of excluded candidates. The additional equations have exactly the same forms as equations (5) to (7), but their excluded candidates are YKW, TMNG and FKI. Of course, equations (5) to (7) estimated with different sets of excluded candidates usually show different coefficients.

3. The j-th Best Subset Problem for Fuller's Modified Limited Information Maximum Likelihood Method

Let (X_i^1, Y_i, X_i^2) stand for the i-th meaningful and identifiable subset derivable from $y=F(X^1{:}Y{:}X^2)$ and satisfying $X_i^1 \cap X_i^2 = \phi$ and (A_i, B_i) stand for a row coefficient vector of (X_i^1, Y_i). Furthermore, K_i, L_i and M_i stand for the numbers of candidates in X_i^1, Y_i and X_i^2, respectively. Then, the i-th meaningful and identifiable subset can be expressed as follows:

$$y = X_i^1 A_i' + Y_i B_i' + u \tag{8}$$

where $u \sim N(0, \sigma^2 I)$ stands for a disturbance term and X_i^2 is used for the calculation of coefficients and their variance-covariance matrix.

We can rewrite equation (8) as follows:

$$(y,Y_i)(b_0,-B_i)'-X_i^1A_i'=u \quad \text{for } b_0=1 \tag{9}$$

For the above equation, we would like to formulate the j-th best subset problem for the F-LIML method as follows:

Find subset (X_i^1,Y_i,X_i^2) derivable from a given set $(X^1;Y,X^2)$ of all possible included, endogenous, and excluded candidates specified for an explained variable y and estimate coefficient vector $(\bar{A}_i,\bar{B}_i)$ such that

(I) subset (X_i^1,Y_i) is meaningful from the viewpoint of a research field in question,

(II) subset (X_i^1,Y_i,X_i^2) is just- or over-identifiable, namely, $1 \leq L_i \leq M_i$,

(III) subsets X_i^1 and X_i^2 have no common candidates, namely, $X_i^1 \cap X_i^{21}=\phi$,

(IV) $(\bar{A}_i,\bar{b}_0,\bar{B}_i)$ and its asymptotic variance-covariance matrix must be calculated as follows:

$$(W_i^1-\tilde{r}_iW_i)(-\bar{b}_0,\bar{B}_i)'=0 \text{ with } \bar{b}_0=1 \tag{10}$$

and

$$\bar{A}_i'=(X_i^{1\prime}X_i^1)^{-1}X_i^{1\prime}(y,Y_i)(-\bar{b}_0,\bar{B}_i)' \tag{11}$$

with the estimated asymptotic variance-covariance matrix [6] of $(\bar{A}_i,\bar{B}_i)$

$$\bar{s}_i^2 \begin{vmatrix} G(X_i^1,X_i^1) & -G(X_i^1,Y_i) \\ -G(Y_i,X_i^1) & G(Y_i,Y_i) \end{vmatrix}$$

$$\tilde{r}_i=\bar{q}_i-1/(NT-K_i-M_i) \tag{12}$$

for $\bar{q}_i$ that is obtained by minimizing the following ratio with respect to $\bar{B}_i$:

$$q_i=\bar{B}_iW_i^1\bar{B}_i'/\bar{B}_iW_i\bar{B}_i' \tag{13}$$

where

$$W_i^1=(y,Y_i)'Z_i^1(y,Y_i) \text{ for } Z_i^1=I-X_i^1(X_i^{1\prime}X_i^1)^{-1}X_i^{1\prime} \quad (14)$$

$$W_i=(y,Y_i)'Z_i(y,Y_i) \text{ for } Z_i=I-X_i(X_i'X_i)^{-1}X_i' \quad (15)$$

$$X_i=(X_i^1,X_i^2)$$

$$G(Y_i,Y_i)=\{Y_i'(\tilde{r}_iZ_i-Z_i^1)Y_i\}^{-1} \quad (16)$$

$$G(Y_i,X_i^1)=G(Y_i,Y_i)Y_i'X_i^1(X_i^{1\prime}X_i^1)^{-1} \quad (17)$$

$$G(X_i^1,Y_i)=(X_i^{1\prime}X_i^1)^{-1}X_i^{1\prime}Y_iG(Y_i,Y_i) \quad (18)$$

$$G(X_i^1,X_i^1)=(X_i^{1\prime}X_i^1)^{-1}\{I+X_i^{1\prime}Y_iG(Y_i,Y_i)Y_i'X_i^1(X_i^{1\prime}X_i^1)^{-1}\} \quad (19)$$

$$\bar{y}_i=X_i^1\bar{A}_i'+Y_i\bar{B}_i' \quad (20)$$

$$\bar{s}_i^2=(y-\bar{y}_i)'(y-\bar{y}_i)/(NT-K_i-L_i) \quad (21)$$

(V) $\bar{C}_i=(\bar{A}_i,\bar{B}_i)$ satisfies the following magnitude condition (including the sign condition), if necessary:

$$D_{hi}^1\bar{C}_i \pm |D_{hi}^2\bar{C}_i| \pm |D_{hi}^3\bar{C}_i| \geq d_h^1,\quad D_{hi}^1\bar{C}_i \pm |D_{hi}^2\bar{C}_i| \pm |D_{hi}^3\bar{C}_i| \leq d_h^2,$$

and/or $d_h^1 \leq D_{hi}^1\bar{C}_i \pm |D_{hi}^2\bar{C}_i| \pm |D_{hi}^3\bar{C}_i| \leq d_h^2$ for $1 \leqq h \leqq H$ (22)

where D_{hi}^k for $k=1,2,3$, d_h^1 and D_h^2 stand for a row vector of coefficients, a lower bound and an upper bound of the h-th magnitude condition, respectively, $|D_{hi}^k\bar{C}_i|$ for $k=2,3$ stands for the absolute value of $D_{hi}^k\bar{C}_i$, "$\geq$" and "$\leq$" indicate "$>$" or "$\geqq$" and "$<$" or "$\leqq$", respectively, and "$\pm$" indicates "+" or "-",

(VI) the Durbin-Watson statistic DW_i defined below is significant at a level specified by a researcher, when $N=1$, $T \geqq 15$ and $K_i+L_i \leqq 5$:

$$DW_i=\sum_{t=2}^{T}\{\bar{e}_{i1}(t)-\bar{e}_{i1}(t-1)\}^2 / \sum_{t=1}^{T}\bar{e}_{i1}(t)^2 \quad (23)$$

for $\bar{e}_{i1}(t)=y_1(t)-\bar{y}_{i1}(t)$ for $1 \leqq t \leqq T$,

(VII) $\bar{y}_i$ whose elements are denoted by $\bar{y}_{in}(t)$'s satisfies the following relative absolute error test, if necessary:

$$100 \times |\{y_n(t)-\bar{y}_{in}(t)\}/y_n(t)| \leqq w \text{ for } 1 \leqq n \leqq N \text{ and } 1 \leqq t \leqq T \quad (24)$$

where w (%) is specified by a researcher,

(VIII) $\bar{y}_i$ satisfies the turning point test, if necessary:

if

$$\{y_n(t)-y_n(t-1)\}\{y_n(t+1)-y_n(t)\}<0 \quad (25)$$

and

$$100\mathrm{x}\mathrm{Min}[|\{y_n(t)-y(t-1)\}/y_n(t)|,|\{y_n(t+1)-y_n(t)\}/y_n(t)|]\geqq v \quad (26)$$

then

$$\{y_n(t)-y_n(t-1)\}\{\tilde{y}_{in}(t)-\tilde{y}_{in}(t-1)\}>0 \quad (27)$$

and

$$\{y_n(t+1)-y_n(t)\}\{\tilde{y}_{in}(t+1)-\tilde{y}_{in}(t)\}>0 \quad (28)$$

for $1\leqq n\leqq N$, $2\leqq t\leqq T-1$, and $T\geqq 3$

where v (%) is specified by a researcher,

and

(IX) $(\bar{A}_i,\bar{B}_i)$ shows the j-th highest adjusted coefficient of determination measured by RR_i:

$$RR_i=\mathrm{Max}\{0,1-(1-R_i)(NT-1)/(NT-K_i-L_i)\} \quad (29)$$

where

$$R_i=1-(y-\tilde{y}_i)'(y-\tilde{y}_i)/(y-\bar{y}E)'(y-\bar{y}E)$$

$$\bar{y}=\sum_{n=1}^{N}\sum_{t=1}^{T}y_n(t)/NT \text{ and}$$

$E=(1,1,1,\ldots,1)'$ with dimension (NTx1).

Let us briefly explain the above problem. Conditions (I) and (V) are necessary from the viewpoint of information from a research field in question. Conditions (II), (III), (IV) and (VI) are related to statistics or econometrics. Conditions (VII), (VIII) and (IX) are important criteria to evaluate meaningful and identifiable subsets through the comparison of estimated observations with actual ones. Condition (VIII) is quite useful, when an estimated model which contains lagged endogenous candidates is utilized for the final test and/or forecasting. Condition (IX) measures a goodness of fit. Although various measures have been proposed, we adopted an adjusted coefficient of determination to measure a goodness of fit, because it has been used so often in the literature of applied research and can assume a value between 1 (for the perfect fitting) and 0 (for the worst fitting).

If a researcher does not need to use a new criterion, the best subset is defined as the one which (i) is meaningful, identifiable and estimable with the F-LIML method, (ii) satisfies all discrete (or pass-or-fail) criteria applied from (V), (VI), (VII) and (VIII), and (iii) has the highest adjusted coefficient of determination. However, if a new criterion needs to be satisfied, the ultimately best subset is defined as the one which (i) is one of the J best subsets obtained by solving the first to the J-th best subset problems in one computer-run and (ii) satisfies the new criterion. Of course, the ultimately best subset must be found among the J best subsets through the new criterion by a researcher. Finally, we regard the (ultimately) best subset as an approximate solution, which can be a pragmatically best subset, to the variable selection problem for the F-LIML method.

4. An Example

We would like to demonstrate the proposed variable selection procedure by estimating a macro agricultural production function of Cobb-Douglas type with the data of Japanese agriculture from 1965 to 1979. Let us introduce the following notations: LY=log(agricultural products), LL=log(labor), LKA=log(KA)=log(animal capital), LKP=log(KP)

=log(plant capital), LKM=log(KM)=log(machine capital), LK= log(KA+KP+KM), LKR=log(KA+KP+KM*R), R=an estimated use rate of machine capital, where $0 \leqq R \leqq 1$, LAX=log(A-X)=(cultivated acreage minus damaged and abandoned acreage), LCAX=log ((A-X)xMIn(C,1)), C=a cropping index of rice, where C=1, C>1 and C<1 for average, rich and poor harvest, respectively, LQ=log(intermediate goods and services), LWIQ=log(wheat import quantity), LRFI=log(real farm income), LRRPP= log(real producer price of rice determined by the government), LRWRPF=log(real wage rate of farming), DVCS=dummy variable which assumes 1 for cold summer and 0 for normal or hot summer, T=time trend, LT=log(T) and $C= constant term.

We assume that (i) LY is an explained variable, (ii) LT and DVCS are uniquely included candidates, (iii) T is a non-uniquely included candidate, (iv) DVCS is completely optional, (v) LT and T are exclusively optional, (vi) LL, LK, LKR, LAX, LCAX and LQ are endogenous candidates, (vii) LL is absolutely important, (viii) LK and LKR are exclusively important, (ix) LAX and LCAX are exclusively important, (x) LQ is completely optional, (xi) LWIQ, LRFI, LRRPP, LRWRPF, LRFI(-1), LKA(-1) and LKP(-1) are excluded candidates, where, for example, LRFI(-1) stands for LRFI with time lag number 1, and (xii) each meaningful and identifiable subset must use all excluded candidates available to that subset as instrumental ones. The last assumption (xii) implies that excluded candidates must be treated as absolutely important and 'T' (not T) must be added to the group of excluded candidates in a functional format. Then, the following format can be loaded:

LY=F($C<*LT,T*>DVCS:/LL/</LAX,LCAX/></LK,LKR/>LQ:'T'
/LWIQ,LRFI,LRRPP,LRWRPF,LRFI(-1),LKA(-1),LKP(-1)/) (30)

where y, X^1, Y and X^2 in the j-th best subset problem are as follows:

y={LY}, X^1={$C<*LT,T*>DVCS},

Y={/LL/</LAX,LCAX/></LK,LKR/>LQ}, and

X^2={'T'/LWIQ,LRFI,LRRPP,LRWRP,LRFI(-1),LKA(-1),LKP(-1)/}

17 candidates in format (30) are non-constant, so that $2^{17}-1=131071$ possible subsets can be derived from format (30). However, the number of all possible meaningful as well as identifiable subsets derivable from format (30) is only 48. Accordingly, the remaining 131,023 subsets are meaningless or cannot be estimated with F-LIML method. All possible meaningful and identifiable subsets are over-identified. Meaningful and just-identifiable subsets cannot be derived from format (30).

The positions of candidates in a (non-econometrical) group can be switched with each other and, furthermore, those of groups in X^1, Y or X^2 can be switched with each other. For instance, the following functional format is equivalent to format (30):

LY=F($C<*LT,T*>DVCS:</LCAX,LAX/>LQ</LK,LKR/>/LL/:'T'
/LWIQ,LRFI(-1),LRFI,LKA(-1),LKP(-1),LRRPP,LRWRPF/) (31)

It is possible to derive all meaningful and identifiable subsets from format (30). However, we need more information to find the best subset (j=1) or the best J subsets (J>1). Let us introduce the following information for production which is consistent with both common sense and economics: (i) a free sign (a sign is not determined before estimation) for $C, a negative sign for DVCS and positive signs for LL, LK LKR, LAX, LCAX and LQ, (ii) 0.1<LL(0.5,

(iii) 0.1<LK+LKR(0.5, (iv) 0.1<LAX+LCAX(0.6, (v) 0.1<LQ(0.3, (vi) 0.85<LL+LK+LKR+LAX+LCAX+LQ(1.15, (vii) 5 % Durbin-Watson statistic test, (viii) 5 % relative absolute error test, (ix) 10 % turning point test and (x) the minimum adjusted coefficient of determination is 0.7, where the variable notations in (i) and (vi) imply their coefficients and "(" stands for "≦". The land factor is emphasized more than the others by (iv). The intermediate goods and services factor is emphasized less than the others by (v). Condition (vi) judges as unsatisfactory those meaningful and identifiable subsets which show unusually increasing or decreasing returns to scale in the agricultural production.

When we solved the first best subset problem in one computer-run, we obtained the following equation in about 2 minutes 36 seconds CPU time by the FACOM M-200 (about 13 MIPS/CPU):

```
LY=1.04457+0.017302*T+0.238956*LL+0.596084*LCAX+0.148097*LK
  (2.3492)(0.008884) (0.170867)  (0.321405)    (0.059907)
RR=0.9015, SD=0.02032, MEV=2.192, K=2.025, FA=0.01233,
DW=1.933                                                (32)
```

where numbers in parentheses, RR, SD, MEV, K, FA and DW stand for standard deviations of asymptotic variances of coefficients, adjusted coefficient of determination, minimum eigen value, k of k-class estimator, first-order autocorrelation coefficient and Durbin-Watson statistic, respectively. The excluded candidates used for equation (32) are LWIQ, LRFI, LRRPP, LRWRPF, LRFI(-1), LKA(-1) and LKP(-1), respectively. Since candidate T is selected as an included candidate in equation (32), candidate T was not selected as an excluded candidate in the equation, as assumed above.

The sum of the coefficients of candidates LL, LCAX and LK in equation (32) is 0.983137 so that slightly decreasing returns to scale prevails in the agricultural sector. We can accept equation (32) as a macro agricultural production function of Cobb-Douglas type.

5. Summary

Fuller's modified limited information maximum likelihood method is useful for the estimation of equations in a simultaneous equation model. The author formulated the j-th best subset problem for this estimation method and proposed to regard the (ultimately) best subset of the first (to the J-th, e.g. J=10) best subset problem(s) as a pragmatically best subset in the variable selection problem. The proposed procedure may be able to reduce drastically the laborious trial-and-error work of estimation by this method.

References

[1] T.W. Anderson and H. Rubin, The asymptotic properties of estimators of the parameters of a single equation in a complete system of stochastic equations, Annals of Mathematical Statistics 21 (1950) 570-582.

[2] J. Durbin and G.S. Watson, Testing for serial correlation in least squares regression I, Biometrica 37 (1950) 409-428.

[3] J. Durbin and G.S. Watson, Testing for serial correlation in least squares regression II, Biometrica 38 (1951) 159-178.

[4] J. Durbin and G.S. Watson, Testing for serial correlation in least squares regression III, Biometrica 58 (1971) 1-19.

[5] W.A. Fuller, Some properties of a modification of the limited information estimator, Econometrica 45 (4) (1977) 939-953.

[6] A.S. Goldberger, Econometric Theory (3rd ed., Wiley, NY, 1966).

[7] T.C. Liu and J.B. Breen, The covariance matrix of the limited information estimator and the identification test, Econometrica 37 (2) (1969) 222-227.

[8] H. Onishi, A variable selection procedure for econometric models, Computational Statistics and Data Analysis 1 (2) (1983) 85-95.

[9] H. Onishi, Computer Package OEPP for Socio-Economic Analysis and Forecasting (Institute of Socio-Economic Planning, University of Tsukuba, Japan).