No. 204 (83-27)

A VARIABLE SELECTION METHOD FOR TWO STAGE LEAST SQUARES

by

Haruo ONISHI

December 1983

A Variable Selection Method for Two Stage Least Squares

University of Tsukuba
Haruo ONISHI

ABSTRACT

The two stage least squares method is often used for the estimation of equations in a simultaneous equation model of economic activities. The variable classifications based on knowledge of economics and econometrics can generate not only economically meaningful but also just- or over-identifiable subsets from a given set of all possible included predetermined variable candidates, all possible right-hand side endogenous variable candidates, and all possible excluded predetermined variable candidates. If we can regard as the best subset the one which (i) is economically meaningful as well as econometrically estimable with the two stage least squares method, (ii) passes the criteria specified for the signs of coefficients, magnitudes of linear forms of coefficients, relative absolute error and turning point error tests for estimated values, and (iii) has the highest adjusted coefficient of determination, it is possible to let the computer automatically choose such a subset when it exists.

Keywords: two stage least squares, variable selection problem, variable classifications, meaningful subset

1. Introduction

It is difficult to perfectly solve the variable

selection problem for the ordinary least squares method (OLS), but it is possible to solve the best J subset problems for OLS by the package OEPP [4]. The best subset of the best subset problem (J=1) or the ultimately best subset among the best J subsets of the best J subset problems (J>1) can be regarded as a pragmatically best subset of the variable selection problem [3]. Furthermore, the best J subset problems for the linearly-constrained ordinary least squares method (COLS), the generalized least squares method (GLS), and the linearly-constrained generalized least squares method (CGLS) can be solved in one run of the computer by the OEPP [3]. In this paper, the author would like to give a clue to solve the variable selection problem for the two stage least squares method (2SLS) by H. Theil [5] and R.L. Basmann [1]. The author pointed out that the variable classifications for OLS, COLS, GLS, or CGLS are definitely necessary to find the best subset which is acceptable from the viewpoint of an applied field [2], [3]. These variable classifications can be easily extended to 2SLS. An additional variable classification is to distinguish among included predetermined candidates, RHS endogenous candidates and excluded predetermined candidates from the viewpoint of econometrics. It must be kept in mind that included predetermined variables are explanatory or predictive for the behavior or movement of a LHS endogenous variable but excluded predetermined variables are instrumental. When included predetermined variables, RHS endogenous variables, and excluded predetermined variables are not yet determined for a LHS endogenous variable and we try to find the best subset, we

would like to use words 'included predetermined candidates, RHS endogenous candidates and excluded predetermined candidates' from here on. The package OEPP for socio-economic analysis and forecasting can handle the variable selection method for 2SLS to be proposed.

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

Included predetermined, RHS endogenous and excluded predetermined candidates can be classified into 8 basic groups and 3 nested groups by knowledge of economics. Eight basic groups are (i) absolutely important (or forced or core), (ii) optionally important, (iii) gradually important, (iv) exclusively important, (v) gradually optional, (vi) exclusively optional, (vii) completely optional and (viii) fixed ones [2]. On the other hand, three cases of nested groups are (i) optionally important case, (ii) exclusively important case and (iii) exclusively optional case [3]. Let us postulate that variables or candidates are loaded by a functional form like $y=F(x_0,x_1,\ldots,x_K)$, where y stands for a LHS endogenous variable and x_k's stand for included predetermined, RHS endogenous, or excluded predetermined variables (or candidates). excluded predetermined variables (or candidates). Furthermore, let us enclose absolutely important candidates within /.../, optionally important candidates within <...>, gradually important candidates within <+...+>, exclusively important candidates within </.../>, gradually optional candidates within <-...->, exclusively optional candidates within <*...*>, fixed candidates within (...) and

completely optional candidates freely in a functional format. In the case of nested variable classifications, (i) an optionally important case can include </.../>, <+...+>, <*...*>, <-...->, (...) and completely optional candidates inside <...> like <</A,B/>C<*D,E,F*>>, (ii) an exclusively important case can include <...>, <+...+>, (...), and completely optional candidates inside </.../> like </<A,B,C><D,F><+G,H+>/>, and (iii) an exclusively optional case can include <...>, <-...->, (...), and completely optional candidates inside <*...*> like <*A,B<-C,D,E->(F,G)*>. These variable classifications are necessary to generate all meaningful subsets derivable from a given set of all possible included predetermined, RHS endogenous, and excluded predetermined candidates.

In addition to the above variable classifications, we need the variable classification done by knowledge of econometrics to generate all just- or over-identifiable subsets. We would like to postulate that (i) all possible included predetermined candidates and all possible RHS endogenous candidates are considered to explain the behavior or movement of a LHS endogenous variable, (ii) a group of all possible included predetermined candidates are entered first in a functional form, (iii) all possible RHS endogenous candidates follow all possible included predetermined candidates, (iv) all possible excluded predetermined candidates follow all possible RHS endogenous candidates, and (v) a group of all possible included predetermined candidates, a group of all possible RHS endogenous candidates and a group of all possible excluded predtermined candidates are separated with ":". As

explained later, non-uniquely included predetermined candidates which are not absolutely important are included in a group of excluded predetermined candidates in a functional format.

For instance, DB is a LHS endogenous variable, DB(-1) and Y are included predetermined variables, BPP, BPLP and BFP are RHS endogenous variables and Y(-1), DP(-1), DP(-2), DPL(-1), FP, FP(-1), BP(-1), PP and PLP(-1) are excluded predetermined variables, where, for instance, DB(-1) stands for DB with time lag number 1. Then, the following format can be used for 2SLS:

```
DB=F($C,DB(-1),Y:BPP,BPLP,BFP:Y(-1),DP(-1),DP(-2),DPL(-1),
    FP,FP(-1),BP(-1),PP,PLP(-1))                              (1)
```

where $C stands for the notation of a constant term. The estimated equation of format (1), which is over-identifiable, is expressed as follows:

$$DB=a_0+a_1DB(-1)+a_2Y+a_3BPP+a_4BPLP+a_5BFP \qquad (2)$$

where a_i's stand for coefficients. The excluded predetermined variables in format (1) do not appear in equation (2), although they are used for the calculation of coefficients. When the included predetermined, RHS endogenous, excluded predetermined variables in format (1) are candidates to find the best subset, equation (2) is one of all possible just- or over-identifiable subsets. If we assume that (i) all excluded predetermined candidates must always be used for the calculation of meaningful and just- or over-identifiable subsets, (ii) DB(-1) is completely optional, (iii) Y and BPP are absolutely important, and (iv) BPLP and BFP are exclusively important. Then, the following format can show the above assumptions:

```
DB=F($C,DB(-1)/Y:BPP/</BPLP,BFP/>:/Y(-1),DP(-1)
     DP(-2),DPL(-1),FP,FP(-1),BP(-1),PP,PLP(-1)/)   (3)
```

That all excluded predetermined candidates are always used implies that they must be treated as absolutely important candidates from the viewpoint of not economics but econometrics. Format (3) can generate and estimate the following 5 equations in addition to equation (2) (by a command):

$$DB=b_0+b_1BD(-1)+b_2Y+b_3BPP+b_4BPLP \quad (4)$$

$$DB=c_0+c_1DB(-1)+c_2Y+c_3BPP+c_4BFP \quad (5)$$

$$DB=d_0+d_1Y+d_2BPLP+d_3BPLP+d_4BFP \quad (6)$$

$$DB=e_0+e_1Y+e_2BPP+e_3BPLP \quad (7)$$

$$DB=f_0+f_1Y+f_2BPP+f_3BFP \quad (8)$$

where b_i, c_i, d_i, e_i and f_i stand for coefficients and all excluded predetermined candidates Y(-1), DP(-1), ..., PLP(-1) are used as instrumental candidates for each of equations (2) and (4) to (8) which are over-identified. Needless to say, the order of candidates in each group in format (3) is changeable. For instance, the following is equivalent to format (3):

```
DB=F($C/Y/DB(-1):</BPLP,BFP/>/BPP:Y(-1),DP(-1),
     DP(-2),DPL(-1),FP,FP(-1),BP(-1),PP,PLP(-1)/)   (9)
```

There is no guarantee that a multicollinearity problem does not occur among predetermined candidates. This implies that it may be risky to always treat all possible excluded predetermined candidates as absolutely important ones in application. For instance, suppose that there are three excluded predetermined candidates W1, W2 and W3 with respect to an equation at hand where W3=W1+W2. If all W1, W2 and W3 are always used for 2SLS, namely, if they are treated as absolutely important candidates, it is impossible to calculate coefficients by 2SLS, because of

the multicollinearity among W1, W2 and W3. In this case, </(W1,W2)W3/> can be entered in a group of all possible excluded predetermined candidates in a functional format. Therefore, pair W1 and W2 or a single candidate W3 is used for 2SLS and the multicollinearity problem can be avoided.

To generate all meaningful and just- or over-identifiable subsets, the computer has to recognize whether an included predetermined candidate is unique or non-unique to an equation at hand. Let us define a uniquely included predetermined candidate as the one which (i) could appear only in the equation at hand and (ii) never appears in any other equations and identities in a model. On the other hand, we define a non-uniquely included predetermined candidate as the one which (i) could appear in the equation at hand and (ii) appears in at least one of the other equations and/or identities in a model. Suppose that candidate ABC is a non-uniquely included predetermined candidate. If a meaningful subset does not have ABC as an included predetermined candidate, then ABC can be used as an excluded predetermined candidate for that subset. If ABC is a uniquely included predetermined candidate, a meaningful subset which does not include ABC must not use ABC as an excluded predetermined one. In OEPP, non-uniquely included predetermined candidates which are not absolutely important are included not only in a group of all possible included predetermined candidates but also in a group of all possible excluded predetermined candidates in a functional format. If a subset has at least one non-uniquely included predetermined candidate and has the same candidate as an excluded predetermined candidate, such a

subset must be ignored in the generation of all possible meaningful and just- or over-identifiable subsets. Of course, if a non-uniquely included predetermined candidate is absolutely important, such a non-uniquely included predetermined candidate must not be included in a group of excluded predetermined candidates.

Let us examine equations (6) to (8). Whether included predetermined candidate Y is unique or non-unique, Y must not be included in the group of excluded predetermined candidates in format (3), because Y is absolutely important. However, if DB(-1) is non-unique, equations (6) to (8) are estimated without excluded predetermined candidate DB(-1). To allow DB(-1) as an excluded predetermined candidate for equations (6) to (8), we add DB(-1) to the group of excluded predetermined candidates in format (3). The following format is suitable for this case:

```
DB=F($C,DB(-1)/Y:BPP/</BPLP,BFP/>:/Y(-1),DP(-1),DP(-2)
     DPL(-1),FP,FP(-1),BP(-1),PP,PLP(-1)/DB(-1))     (10)
```

Exactly the same equations as equations (2) and (4) to (8) derivable from format (3) can be estimated from format (10). However, equations (6) to (8) derivable from format (3) are estimated with excluded predetermined candidates Y(-1), DP(-1), ..., PLP(-1) but without DB(-1), whereas the corresponding equations derivable from format (10) are estimated with excluded predetermined candidates Y(-1), DP(-1), ..., PLP(-1), and DB(-1).

3. The j-th Best Subset Problem for Two Stage Least Squares

Here, instead of the variable selection problem for

2SLS, we formulate the j-th best subset problem from which exactly the same equation is obtained as in the case where an economist estimates and evaluates equations one at a time until he can find a satisfactory equation. We assume that there are N (cross-sectional) units (regions, industries, firms, households, classes, etc.) and T obervation times. Let us load a LHS endogenous variable (and its data vector) y, all possible included predetermined candidates (and their data vector) x_k^1's ($0 \leq k \leq K$), all possible RHS endogenous candidates (and their data vector) y_ℓ's ($1 \leq \ell \leq L$), and all possible excluded predetermined candidates (and their data vector) x_m^2's ($1 \leq m \leq M$) by $y=F(X^1:Y:X^{21})$ according to the rule of the variable classifications, where $X^1=(x_0^1,x_1^1,..,x_K^1)$, $Y=(y_1,y_2,...,y_L)$, $X^{21}=(x_1^2,x_2^2,...,x_M^2$, and a subset of x_k^1's which consists of non-uniquely included predetermined candidates which are not absolutely important). For instance, we can write $y=\{y_1(1),y_2(1),...,y_N(1),...,y_1(T),y_2(T),...,y_N(T)\}'$ where $y_n(t)$ stands for the datum of a LHS endogenous variable of unit n at time t for $1 \leq n \leq N$ and $1 \leq t \leq T$. Furthermore, let (X_i^1,Y_i,X_i^{21}) stand for the i-th subset derivable from (X^1,Y,X^{21}), respectively, A_i and B_i stand for the coefficient row vectors of X_i^1 and Y_i, respectively, and K_i, L_i, M_i stand for the numbers of candidates in X_i^1, Y_i, and X_i^{21}, respectively. We can express the i-th subset as follows:

$$y=X_i^1A_i+Y_iB_i+u \tag{12}$$

where u stands for a disturbance term and X_i^{21} are used in the first stage of the estimation by 2SLS. Now, we would like to formulate the j-th best subset problem for 2SLS as

follows:

find subset (X_i^1, Y_i, X_i^{21}) from a given set (X^1, Y, X^{21}) of all possible included predetermined, RHS endogenous, and excluded predetermined candidates specified for a LHS endogenous variable y and estimate coefficient vector $(\tilde{A}_i, \tilde{B}_i)$ such that

(I) subset (X_i^1, Y_i) is meaningful from the viewpoint of economics,

(II) subset (X_i^1, Y_i, X_i^{21}) is just- or over-identifiable,

(III) subset X_i^1 and subset X_i^{21} satisfies $X_i^1 X_i^{21} =$,

(IV) $(\tilde{A}_i, \tilde{B}_i)$ must be calculated by the following:

$$\begin{pmatrix} \tilde{A}_i' \\ \tilde{B}_i' \end{pmatrix} = \begin{pmatrix} X_i^{1\prime} X_i^1 & X_i^{1\prime} Y_i \\ Y_i' X_i^1 & Y_i' X_i (X_i' X_i)^{-1} X_i' Y_i \end{pmatrix}^{-1} \begin{pmatrix} X_i^{1\prime} y \\ Y_i' X_i (X_i' X_i)^{-1} X_i' y \end{pmatrix} \quad (13)$$

with the asymptotic variance-covariance matrix

$$\tilde{s}_i^2 \begin{pmatrix} X_i^{1\prime} X_i^1 & X_i^{1\prime} Y_i \\ Y_i' Y_i & Y_i' X_i (X_i' X_i)^{-1} X_i' Y_i \end{pmatrix}^{-1}$$

where

$$X_i = (X_i^1, X_i^2),\ \tilde{y}_i = X_i^1 \tilde{A}_i' + Y_i \tilde{B}_i',\ \tilde{s}_i^2 = (y - \tilde{y}_i)'(y - \tilde{y}_i)/(NT - K_i - L_i)$$

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

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

and/or $d_h^1 \leq D_{hi}^1 \tilde{C}_i' \pm |D_{hi}^2 \tilde{C}_i'| \pm |D_{hi}^3 \tilde{C}_i'| \leq d_h^2$ for $1 \leq h \leq H$ (14)

where D_{hi}^k for k=1,2,3 stands for a vector, d_h^1 and d_h^2 stand for a lower bound and an upper bound, respectively, $|D_{hi}^k \tilde{C}_i'|$ for k=2,3 stands for the absolute value of $D_{hi}^k \tilde{C}_i'$, "$\pm$" stands for "+" or "-", and "$\geq$" and "$\leq$" stand for ">" or "$\geqq$" and "<" or "$\leqq$", respectively,

(VI) $\tilde{y}_i$ with elements $\tilde{y}_{in}(t)$'s satisfies the following absolute relative error test, if necessary:

$$100\times|\{y_n(t)-\tilde{y}_{in}(t)\}/y_n(t)|\leqq w \text{ for all } 1\leqq n\leqq N \text{ and } 1\leqq t\leqq T \quad (15)$$

where w (%) is specified by a researcher,

(VII) $\tilde{y}_i$ satisfies the following turning point error test, if necessary:

if

$$\{y_n(t)-y_n(t-1)\}\{y_n(t+1)-y_n(t)\}<0 \quad (16)$$

and

$$100\times\mathrm{Min}[|\{y_n(t)-y_n(t-1)\}/y_n(t)|,|\{y_n(t)-y_n(t+1)\}/y_n(t)|]\geqq u \quad (17)$$

then

$$\{y_n(t)-y_n(t-1)\}\{\tilde{y}_{in}(t)-\tilde{y}_{in}(t-1)\}>0 \quad (18)$$

and

$$\{y_n(t+1)-y_n(t)\}\{\tilde{y}_{in}(t+1)-\tilde{y}_{in}(t)\}>0 \quad (19)$$

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

where u (%) is specified by a researcher,

(VIII) $(\tilde{A}_i,\tilde{B}_i)$ shows the j-th highest adjusted coefficient of determination calculated by RR_i:

$$RR_i=\mathrm{Max}\{0,1-(1-R_i)(NT-1)/(NT-K_i-L_i)\} \quad (20)$$

where

$$R_i=1-\tilde{e}_i'\tilde{e}_i/(y-\bar{y}E)'(y-\bar{y}E) \quad (21)$$

for $\tilde{e}_i=y-\tilde{y}_i$, $\bar{y}=\sum_{n=1}^{N}\sum_{t=1}^{T}y_n(t)/NT$, and

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

Let us explain the above problem briefly. Conditions (I) and (V) are required from knowledge of a research field, e.g., theories, field surveys on the behaviors of economic units (consumers, investors, the authorities of socio-economic policies, etc.), empirical studies, hypotheses in economics. Condition (II) implies a

necessary condition $L_i \leq M_i$ for estimation by 2SLS. Condition (III) guarantees that subsets X_i^1 and X_i^{21} do not have common predetermined candidates. Condition (IV) are formuli to calculate coefficients and their asymptotic variance-covariance matrix by 2SLS. Condition (VI) checks whether or not each observation $y_n(t)$ is estimated within the tolerance interval $[(1-w)y_n(t),(1+w)y_n(t)]$ by $\bar{y}_{in}(t)$. Condition (VII) is a useful criterion, when a model is used for the final test or forecasting in which the values of lagged endogenous variables are generated in a model. Condition (VIII) measures a goodness of fit. Conditions (I), (II), (V), (VI), and (VII) are called discrete criteria or pass-or-fail criteria. However, the last condition (VIII) is a continuous criterion. Thus, it is possible to rank by RR_i's the meaningful and just- or over-identifiable subsets which pass all discrete criteria applied. Accordingly, we define the j-th best subset as the one which (i) is meaningful and just- or over-identifiable, (ii) satisfies all discrete criteria applied from conditions (VI), (VII) and (VII), and (iii) has the j-th highest adjusted coefficient of determination. Unless a researcher has any other criterion to be used for the above problem, the solution to the best subset problem (J=1) can be regarded as an approximate solution to the variable selection problem for 2SLS. If he has a new criterion in addition to the criteria in the above problem, he can solve the best subset problem to the J-th best subset problem (say, J=10) in one run of the computer (by OEPP) and choose the ultimately best subset among the J best subsets through the new criterion if such a subset exists.

4. An Example

We would like to demonstrate the proposed variable selection method by estimating an agricultural production function of Cobb-Douglas type with the data of Japanese agriculture from 1965 to 1979. Let us introduce the following variable notations: LY=log(agricultural outputs), LL=log(labor), LKA=log(KA)=log(animal capital), LKP=log(KP) =log(plant capital), LK=log(agricultural capital)=log(KA+KP +machine capital), LKR=log(adjusted agricultural capital)= log(KA+KP+machine capital adjusted by a use rate), LQ= log(intermediate goods and services), LAX=log(A-X)=log (cultivated acreage minus abandoned and damaged acreage), LCAX=log((A-X) adjusted by a cropping index of rice), LRFI= log(real farm income), LRRPP=log(real producer price of rice), LRRWRPF=log(real wage rate of farming), LWIQ=log (wheat import quantity), DV.CS=dummy variable (normal or hot summer=0 and cold summer=1), T=time trend, and LT=log(T).

We assume that (i) a LHS endogenous variable is LY, (ii) DV.CS, T and LT are included predetermined ones, (iii) DV.CS and LT are uniquely included predetermined candidates, (iv) T is a non-uniquely included predetermined candidate, (v) T and LT are exclusively optional, (vi) LL, LK, LKR, LAX and LCAX are RHS endogenous candidates, (vii) LL is absolutely important, (viii) LK and LKR are exclusively important, (ix) LAX and LCAX are exclusively important, (x) LWIQ, LRFI, LRRPP, LRWRPF, LRFI(-1), LKA(-1) and LKP(-1) are excluded predetermined candidates and (xi) all excluded predetermined candidates are always used for the

estimation, implying that they are treated as absolutely important. The above assumptions concerned with candidates can be compactly expressed as follows:

```
LY=F($C<*T,LT*>DV.CS:/LL/</LAX,LCAX/></LK,LKR/>LQ:/LWIQ
    LRFI,LRRPP,LRWRPF,LRFI(-1),LKA(-1),LKP(-1)/T)     (22)
```

There are 80 meaningful and over-identifiable subsets in format (22). No meaningful and just-identifiable subsets exist. Since there are 17 non-constant candidates, the number of all possible subsets is $2^{17}-1=131,071$.

We would like to set the following sign conditions and magnitude conditions in order to check for and avoid unusual subsets: (i) a free sign (a positive or negative sign does not matter) for the constant term, positive signs for T, LT, LL, LAX, LCAX, LK, LKR and LQ, and a negative sign for DV.CS, (ii) 0.1<LL(0.5, (iii) 0.1<LAX+LCAX(0.6, (iv) 0.1<LK+LKR(0.5, (v) 0.1<LQ(0.3, (vi) 0.85(LL+LAX+LCAX +LK+LKR+LQ<1.15 and (vii) the minimum adjusted coefficient of determination is 0.8, where the variable notations in (ii) to (vi) imply their coefficients and "(" stands for "$\leq$". Although the sign conditions are redundant in this example, they are introduced to reduce the computer's checking time. The land factor is emphasized in (iii), but the intermediate goods and services factor is less emphasizied in (v). Unusually decreasing or increasing returns to scale in the agricultural production is checked and removed by (vi).

When the best two subsets are requested, the following equations are obtained in about 2 minutes and 38 seconds CPU time by the FACOM M-200 (about 13 MIPS):

```
LY=0.774713+0.01971*T+0.29748*LL+0.56407*LCAX+0.16598*LKR
  (1.61773)(0.00720) (0.13631)  (0.20161)     (0.05868)
```

R=0.9323, RR=0.9052, SD=0.01994, FA=-0.02247, DW=2.002 (23)

and

LY=0.740930+0.02197*T+0.36414*LL+0.50449*LAX+0.17548*LKR
(4.28449)(0.01183) (0.22128) (0.54837) (0.09515)

R=0.8606, RR=0.8049, SD=0.02860, FA=-0.1512, DW=2.251 (24)

where numbers in parentheses, R, RR, SD, FA and DW stand for standard deviations of asymptotic variances of coefficients, a coefficient of determination, an adjusted coeffcient of determination, standard deviation of a disturbance term, first-order autocorrelation coefficient, and Durbin-Watson statistic, respectively, and all excluded predetermined candidates are used for both equations. Since the adjusted coefficient of determination of equation (23) is much higher than that of equation (24) and we cannot find excellent features in equation (24) in comparison with equation (23), we would like to select equation (23) as the best subset.

5. Summary

The proposed variable selection method may be useful to find an approximate solution to the variable selection problem for the two stage least squares method. When the small sampling theory on the two stage least squares method is established, it is possible to let the computer automatically find a better solution to the variable selection problem.

References

1. Basmann, R.L., A Generalized Classical Method of Linear Estimation of Coefficients in a Structural Equation, Econometrica, 25, 77-83, (1957).

2. Onishi, H., A Variable Selection Procedure for Econometric Models, Computational Statistics and Data Analysis, 1, 2, 85-95, (1983).

3. Onishi, H. Variable Selecton in Regression Analysis and Package OEPP, Information Processing, Japan (pending), (1983).

4. Onishi, H., Computer Package OEPP for Socio-Economic Analysis and Forecasting, Institute of Socio-Economic Planning, University of Tsukuba, Japan (1983).

5. Theil, H., Economic Forecasts and Policy, 2nd ed., North-Holland, Amsterdam, (1958).