No. 206 (84-1)

A Variable Selection Procedure for the Limited Information Maximum Likelihood Method

by

Haruo ONISHI

January 1984

A Variable Selection Procedure for the Limited Information Maximum Likelihood Method

University of Tsukuba
Haruo ONISHI

Abstract: The j-th best subset problem is formulated for the variable selection problem of the limited information maximum likelihood estimation method (LIML) for a single equation. A procedure is proposed to solve the best subset problem (to the J-th best subset problem, e.g. J=10) in one run of a computer and regard the (ultimately) best subset (among the best J subsets) as an approximate solution to the variable selection problem. This procedure can find the same subset as in the case where a researcher estimates with LIML and evaluates all possible meaningful subsets one at a time until finding a satisfactory subset. The procedure is characterized by (i) the classifications of all possible included predetermined, explanatory endogenous, and excluded predetermined variable candidates specified for an explained endogenous variable with the information from a research field (for example, theories, field surveys or empirical studies in economics) and econometrics, (ii) the generation of all possible meaningful and just- or over-identifiable subsets, (iii) the evaluation of coefficients by the magnitude condition (including the sign condition) and of estimated values by the relative absolute error and turning point error tests, and (iv) the ranking of the meaningful and just- or over-identifiable subsets which satisfy (iii) by their adjusted coefficients of determination.

1. Introduction

In the case of the ordinary least squares method (OLS), the linearly-constrained ordinary least squares method (COLS), the generalized least squares method (GLS), or the linearly-constrained generalized least squares method (CGLS), the author proposed a variable selection method to solve the best subset problem to the J-th best subset problem and then to regard the (ultimately) best subset among the best J subsets as an approximate solution to the variable selection problem [7], [8]. Similarly, we can formulate the j-th best subset problem of LIML for a single equation in a simultaneous equation model (called 'model' from here on). To solve such a problem, we have to utilize the information from a research field in question and econometrics. Such information is concerned with (i) variable classifications for the generation of meaningful and just- or over-identifiable subsets and (ii) criteria for evaluation of estimated subsets. Econometrics (and statistics) cannot handle this kind of information which prevents a meaningless subset from being chosen as the best one. In solving the j-th best subset problem of LIML, all possible included predetermined, explanatory endogenous, and excluded predetermined variable candidates specified for an explained endogenous variable at hand must be classified through the information from the research field and econometrics ('variable candidates', 'explanatory endogenous' and 'explained endogenous' are called 'candidates', 'endogenous' and 'explained', respectively, from here on). Then, subsets each of which contains included predetermined and endogenous candidates necessary for a reasonable equation but does not contain any unnecessary included predetermined or endogenous candidates at all (implying 'meaningful') and can be estimated with LIML (implying 'just- or over-identifiable') can be generated. In most cases of application, estimated subsets are examined using information from research fields. When a researcher can find a subset which is reasonable from the viewpoints of his research field and econometrics and fits best to the observations of an explained variable, he adopts it. For instance, the sign conditions on estimated coefficients of a subset and/or magnitude conditions on values of linear functions of estimated coefficients need to be consistent with information from the research field.

The variable selection procedure to be proposed for LIML is available in the package OEPP developed for socio-economic analysis and forecasting by the author.

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

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 included predetermined candidates (and their data vectors), all possible endogenous candidates (and their data vectors) and all possible excluded predetermined candidates (and their data vectors), respectively, where x_0^1 stands for a constant term. We assume that there are N (cross-sectional) units and T observation times and the datum of y in unit n at time t is expressed as $y_n(t)$ for $1 \leq n \leq N$, $N \geq 1$, $1 \leq t \leq T$ and $T > 1$. Thus, we can express

$y = (y_1(1), y_2(1), \ldots, y_N(1), \ldots, y_n(t), \ldots, y_1(T), y_2(T), \ldots, y_N(T))'$. Here, we assume that x_k^1's and y_ℓ's are the included predetermined and endogenous candidates which are considered to explain the behavior or movement of y. On the other hand, x_m^2's are instrumental candidates used for the calculation of coefficients and their asymptotic variance-covariance matrix by LIML but never appear in an estimated equation.

First of all, it is of great importance to classify x_k^1's and y_ℓ's by means of information from the research field for the purpose of the generation of only meaningful subsets. The author pointed out that candidates are classified into 8 basic groups and 3 nested cases regardless of research fields. The eight basic groups are the (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 ones. The three nested cases are the (i) optionally important, (ii) exclusively important, and (iii) exclusively optional ones. Any kind of nested variable classification can be reduced into a combination of the basic variable classifications. A nested variable classification is compactly expressed, whereas the reduced basic variable classifications are long.

Second, the variable classification required from econometrics is to distinguish candidates among (i) uniquely included predetermined, (ii) non-uniquely included predetermined, (iii) endogenous, and (iv) excluded predetermined candidates. Let us explain the difference between a uniquely included predetermined candidate and a non-uniquely included predetermined candidate. A uniquely included predetermined candidate is defined as the one which (i) could appear in an equation at hand but (ii) never appears in any other equations and identities in a model. On the other hand, a non-uniquely included predetermined candidate is defined as the one which (i) could appear in an equation at hand and (ii) does appear in at least one of the other equations and/or identities in a model. If a uniquely included predetermined candidate is not selected in a subset, such a candidate must not be used as an excluded predetermined one for that subset. On the other hand, if a non-uniquely included predetermined candidate which is not absolutely important is not selected in a subset, such a candidate can be used as an excluded predetermined one for that subset. To handle this case, we need another variable classication. We postulate that (i) such non-uniquely included predetermined candidates are entered not only in a group of included predetermined candidates but also in a group of excluded predetermined candidates and (ii) they are enclosed within '...' when they are entered in a group of excluded predetermined candidates. If a non-uniquely included predetermined candidate is selected not only as an included predetermined one but also as an excluded predetermined one in a subset, such a subset must be ignored in the generation of all possible meaningful and just- or over-identifiable subsets.

A method is needed so that a researcher can easily load an explained variable and its classified candidates into a computer and have the computer generate, estimate with LIML and evaluate all possible meaningful and just- or over-identifiable subsets. Here, we would like to follow the variable-loading style (of the package OEPP) explained in [7].

Let us postulate that (i) y, x_k^1's, y_ℓ's and x_m^2's are loaded by a functional format where y is a function of x_k^1's, y_ℓ's and x_m^2's, (ii) x_k^1's are entered first and followed by y_ℓ's which are followed by x_m^2's and the non-uniquely included predetermined candidates among x_k^1's which are not absolutely important in a functional format, (iii) a group of x_k^1's, a group of y_ℓ's, and a mixed group of x_m^2's and the non-uniquely included predetermined candidates among x_k^1's which are not absolutely important are separated with ":". Thus, a functional format is expressed as $y=F(X^1:Y:X^{21})$ where X^1 stands for a set of x_k^1's, Y for a set of y_ℓ's, and X^{21} for a combined set of x_m^2's and the non-uniquely included predetermined candidates among x_k^1's which are not absolutely important. Of course, non-uniquely included predetermined candidates which are absolutely important must not be added to a group of excluded predetermined candidates. X^1, Y, X^{21} are classified through information from a research field in question and econometrics.

Let (X_i^1, Y_i, X_i^{21}) and (A_i, B_i) stand for the i-th meaningful and just- or over-identifiable subset and the coefficient vector for (X_i^1, Y_i), respectively, where X_i^1 and X_i^{21} do not possess any common candidate. Furthermore, I stands for an identity matrix, and X_i^1, Y_i and X_i^{21} have K_i, L_i and M_i candidates, respectively, where $K_i \leq K$, $1 \leq L_i \leq L$, $M_i \leq M$ and $L_i \leq M_i$. Then, we can express subset (X_i^1, Y_i, X_i^{21}) for y as follows:

$$y=X_i^1A_i'+Y_iB_i'+u \qquad (1)$$

where u stands for a disturbance term and X_i^{21} is used for the calculation of (A_i,B_i) and the asymptotic variance-covariance matrix.

Let us give a simple example. Suppose that (i) explained variable IMC is a function of all possible included predetermined candidates $C (constant term), SNC(-1) and Y(-1), all possible endogenous candidates DCB, IDCP and Y, and excluded predetermined candidates POP, RWR(-1), TAX, GASPRC and SR, (ii) SNC(-1) is a uniquely included predetermined and absolutely important candidate, (iii) Y(-1) is a non-uniquely included predetermined and completely optional candidate, (iv) DCB and IDCP are optionally important, (v) Y is absolutely important, (vi) all excluded predetermined candidates are always used, implying that they are treated as absolutely important candidates, where, for instance, Y(-1) stands for Y with time lag number 1. Needless to say, excluded predetermined candidates POP, RWR(-1), TAX, GASPRC, SR and Y(-1) appear in at least one of the other equations and/or identities in a model, whereas SNC(-1) never appears in any other equations and identities in a model. Then, the following functional format can be loaded:

IMC=F($C/SNC(-1)/Y(-1):<DCB,IDCP>/Y/:'Y(-1)'/POP
RWR(-1),TAX,GASPRC,SR/) (2)

It must be noticed that non-uniquely included predetermined candidate Y(-1) appears not only in a group of included predetermined candidates but also in a group of excluded predetermined candidates, because it is not absolutely important. From format (2), the following 6 meaningful and over-identifiable subsets (equations) are estimated:

$$IMC=a_0+a_1SNC(-1)+a_2Y(-1)+a_3DCB+a_4IDCP+a_5Y \quad (3)$$

$$IMC=b_0+b_1SNC(-1)+b_2Y(-1)+b_3IDCP+b_4Y \quad (4)$$

$$IMC=c_0+c_1SNC(-1)+c_2Y(-1)+c_3DCB+c_4Y \quad (5)$$

$$IMC=d_0+d_1SNC(-1)+d_2DCB+d_3IDCP+d_4Y \quad (6)$$

$$IMC=e_0+e_1SNC(-1)+e_2IDCP+e_3Y \quad (7)$$

$$IMC=f_0+f_1SNC(-1)+f_2DCB+f_3Y \quad (8)$$

where a_i, b_i, c_i, d_i, e_i and f_i stand for coefficients, all of POP, RWR(-1), TAX, GASPRC and SR are used for equations (3) to (5), and all of POP, RWR(-1), TAX, GASPRC, SR and Y(-1) are used for equations (6) to (8). Since there are 11 non-constant candidates in format (2), $2^{11}-1=2047$ possible subsets exist. However, the number of all possible meaningful and over-identifiable subsets is only 6. Equation (1) represents one of the equations (3) to (8). In this example, there is no just-identifiable subset, because the number of excluded predetermined candidates which are treated as absolutely important exceeds that of the maximum combination of endogenous candidates. The best subset must be found among equations (3) to (8), if it exists.

It should be kept in mind that if 'Y(-1)' in format (2) is replaced with Y(-1), three more equations in addition to equations (3) to (8) are estimated. The additional equations are expressed with equations (6) to (8) but excluded predetermined candidates POP, RWR(-1), ..., SR without Y(-1) are used for the estimation of these equations.

3. The j-th Best Subset Problem for LIML

At present, it is quite difficult to define the solution (i.e. the best subset) to the variable selection problem for LIML in such a way that the definition can be accepted by researchers in all fields where LIML is available. Accordingly, we would like to formulate the j-th best subset problem for LIML, define the solution to the problem and propose to regard the (ultimately) best subset as a pragmatically best subset for the variable selection problem for LIML.

We rearrange equation (1) as follows:

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

For equation (9), we would like to formulate the j-th best subset problem for LIML 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, endogenous, and excluded predetermined candidates specified for an explained variable y and estimate coefficient vector $(\bar{A}_i,\bar{b}_0,\bar{B}_i)$ 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^{21}) is just- or over-identifiable, namely, $1 \leqq L_i \leqq M_i$,

(III) subset X_i^1 and subset X_i^{21} do not have common candidates, namely, $X_i^1 \cap X_i^{21}=\phi$,

(IV) $(\bar{A}_i,\bar{b}_0,\bar{B}_i)$ must be calculated by the following [1]:

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

and

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

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

$$\bar{s}_i^2 \begin{bmatrix} G(X_i^1,X_i^1) & -G(X_i^1,Y_i) \\ -G(Y_i,X_i^1) & G(Y_i,Y_i) \end{bmatrix} \quad (12)$$

with respect to $\bar{r}_i$ that is obtained by minimizing the following ratio with respect to B_i:

$$r_i=B_iW_i^1B_i'/B_iW_iB_i' \quad (13)$$

where

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

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

$$\text{and } X_i=(X_i^1,X_i^{21}) \quad (15)$$

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

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

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

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

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

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

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

$$D_{hi}^1\bar{C}_i\underline{+}|D_{hi}^2\bar{C}_i|\underline{+}|D_{hi}^3\bar{C}_i|\underline{>}d_h^1,\ D_{hi}^1\bar{C}_i\underline{+}|D_{hi}^2\bar{C}_i|\underline{+}|D_{hi}^3\bar{C}_i|\underline{<}d_h^2,$$

$$\text{and/or } d_h^1\underline{<}D_{hi}^1\bar{C}_i\underline{+}|D_{hi}^2\bar{C}_i|\underline{+}|D_{hi}^3\bar{C}_i|\underline{<}d_h^2 \quad \text{for } 1\leqq h\leqq H \quad (22)$$

where D_{hi}^k for $1\leqq k\leqq 3$, d_h^1, and d_h^2 stand for a row vector, 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 $D_{hi}^k\bar{C}_i$, "$\underline{>}$" and "$\underline{<}$" indicate "<" or "$\leqq$" and ">" or "$\geqq$", respectively, and "$\underline{+}$" indicates "+" or "-",

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

$$100\times|\{y_n(t)-\bar{y}_{in}(t)\}/y_n(t)|\leq w \text{ for } 1\leq t\leq T \text{ and } 1\leq n\leq N \quad (23)$$

where w (%) is specified by a researcher,

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

if

$$\{y_n(t)-y_n(t-1)\}\{y_n(t+1)-y_n(t)\}<0 \quad (24)$$

and

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

then

$$\{y_n(t)-y_n(t-1)\}\{\bar{y}_{in}(t)-\bar{y}_{in}(t-1)\}>0 \quad (25)$$

and

$$\{y_n(t+1)-y_n(t)\}\{\bar{y}_{in}(t+1)-\bar{y}_{in}(t)\}>0 \quad (26)$$

for $2\leq t\leq T-1$, $T\geq 3$, $1\leq n\leq N$ and $N\geq 1$,

where u (%) is specified by a researcher,

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

$$RR_i=\mathrm{Max}\{0,1-(1-R_i)(NT-1)/(NT-K_i-L_i)\} \quad (27)$$

for

$$R_i=1-\tilde{e}_i'\tilde{e}_i/(y-\bar{y}E)'(y-\bar{y}E) \quad (28)$$

where

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

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

We would like to define the j-th best subset with respect to LIML as the one which (i) is derivable from a given set of all possible included predetermined, endogenous, and excluded predetermined candidates specified for an explained variable, (ii) is meaningful from the viewpoint of a research field in question, (iii) is just- or over-identifiable from the viewpoint of econometrics, (iv) is estimable with LIML, (v) satisfies conditions (V), (VI) and (VII), if applied, and (vi) shows the j-th highest adjusted coefficient of determination among the subsets which satisfy (i) to (v) above. Unless a researcher has any other new criterion, the best subset (j=1) is a pragmatically best solution to the variable selection problem for LIML. If there is a new criterion, he can take the following two steps: (i) solve the best subset problem to the J-th best subset problem (e.g. J=10) in one run of the computer and (ii) then select the ultimately best subset among the best J subsets himself through the new criterion. Thus, he can regard such a subset as a pragmatically best subset for the variable selection problem of LIML, when such a subset exists. Needless to say, if he cannot generate all possible meaningful and just- or over-identifiable subsets from a set of all possible included predetermined, endogenous, and excluded predetermined candidates, he should load separately several sets of possible included predetermined, endogenous, and excluded predetermined candidates and apply the j-th best subset problem for each of them.

4. An Example

We would like to demonstrate the proposed variable selection procedure by estimating the macro agricultural production function of Cobb-Douglas type using the data of Japanese agriculture from 1965 to 1979. Let us introduce the following variable notations: $C=constant term, LY=

log(agricultural outputs), 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, LQ=log(intermediate goods and services), LAX=log(A-X)=log(cultivated acreage minus abandoned and damaged acreage), LCAX= log((A-X)xMin(1,C)), C=a cropping index of rice, where C=1, C>1 and C<1 for average, rich and poor harvest, respectively, LRFI=log(real farm income), LRRPP=log(real producer price of rice determined by the government), LRWRPF=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 for technical progress, LT=log(T).

We assume that (i) LY is an explained variable, (ii) LT and DV.CS are uniquely included predetermined candidates, (iii) T is a non-uniquely included predetermined candidate, (iv) DV.CS 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, LKM(-1), LKA(-1) and LKP(-1) are excluded predetermined candidates, (xii) each meaningful and just- or over-identifiable subset must use all excluded predetermined candidates available to that subset as instrumental ones, implying that excluded predetermined candidates LWIQ to LKP(-1) must be treated as absolutely important and 'T' (not T) must be added to the group of excluded predetermined candidates in a functional format. Then, the following format can be loaded:

```
LY=F($C,DV.CS<*LT,T*>:/LL/</LK,LKR/></LAX,LCAX/>LQ:'T'
  /LWIQ,LRFI,LRRPP,LRWRPF,LKM(-1),LKA(-1),LKP(-1)/) (29)
```

Since there are 17 non-constant candidates in format (29), $2^{17}-1=131071$ possible subsets exist. However, the number of all possible meaningful and over-identifiable subsets derivable from format (29) is 48. No meaningful and just-identifiable subsets exist in format (29).

We would like to set the following sign 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, a negative sign for DV.CS and positive signs for LT, T, 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, and (vii) the minimum adjusted coefficient of determination is 0.7, where the variable notations in (i) to (vi) imply their coefficients and "(" stands for "$\leqq$". Although the sign conditions for LL, LK, LKR, LAX, LCAX and LQ are redundant, it is introcuded to reduce the computer's checking time. The land factor is emphasized in (iii), while the intermediate goods and services factor is less emphasized in (v). Condition (vi) regards as unsatisfactory those meaningful and over-identifiable subsets which show unusually increasing or decreasing returns to scale in the agricultural production.

When the best three subsets are requested under these conditions, only one equation, the following, was obtained in about 2 minutes 38 seconds CPU time by the FACOM M-200 (about 13 MIPS per CPU):

```
LY=1.748664+0.012987*T+0.151897*LL+0.152060*LK+0.582842*LAX
  (5.89655)(0.013675) (0.261810)  (0.110576)  (0.728765)
```

where numbers in parentheses, RR, SD, MEV, FA and DW stand for standard deviations of asymptotic variances of coefficients, adjusted coefficient of determination, standard deviation of asymptotic variance of a disturbance term, minimum eigen value, first-order autocorrelation coefficient and Durbin-Watson statistic [4], respectively. Excluded predetermined candidates LWIQ, LRFI, LRRPP, LRWRPF, LKM(-1), LKA(-1) and LKP(-1) were used for equation (30).

The adjusted coefficient of determination is rather small and the sum, 0.8868, of the elasticities of candidates LL, LK and LAX shows that decreasing returns to scale are slightly strong. Nonetheless, we may be able to regard equation (30) as a satifactory macro agricultural equation estimated by LIML.

5. Summary

The variable selection problem for LIML has not been discussed much in the literature. Since it is quite difficult to perfectly define the best subset for the variable selection problem for LIML as in the cases of the other estimation methods, the author would like to formulate the j-th best subset problem for LIML and solve only the best subset problem or the best subset problem to the J-th best subset problem in one run of the computer, depending on the use of a new criterion, and propose to regard the (ultimately) best subset as a pragmatically best subset of the variable selection problem for LIML. Then, the proposed procedure may make the work of estimation by LIML less time-consuming, labor-consuming and costly.

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 euqations, Annals of Mathematical Statistics 21 (1950) 570-582.

[2] A.R. Bergstrom, The covariance matrix of the limited information maximum likelihood estimator, Econometrica 40 (5) (1972) 899-900.

[3] H. Chernoff and H. Rubin, Asymptotic properties of limited-information estimates under generalized conditions, Studies in Econometric Methods, Cowles Foundation Monograph 14, ed. by W.C. Hood and T.C. Koopmanns, Yale Univ. Press, (1953) 200-212.

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

[5] F.M. Fisher and J.B. Kadane, The covariance matrix of the limited informaton estimator and the identification test: comments, Econometrica 40 (5) 1972) 901-903.

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

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

[8] H. Onishi, Variable selection in regression analysis and package OEPP, Information Processing, (forthcoming).

[9] H. Onishi, Computer Package OEPP for Socio-economic Analysis and Forecasting, Institute of Socio-Economic Planning, University of Tsukuba, 1983.