Journal of the Operations Research
Society of Japan
Vol. 30, No. 4, December 1987

A PRACTICAL VARIABLE SELECTION METHOD FOR GENERALIZED LEAST SQUARES AND ITS APPLICATION

Haruo Onishi
University of Tsukuba

(Received June 4, 1986; Revised March 4, 1987)

Abstract The j-th best subset problem for the generalized least squares is formulated in which statistical criteria as well as non-statistical conditions are introduced. Non-statistical conditions are based on a knowledge of the scientific field to which research is related, natural logic and common sense, while statistical criteria are t-test, Durbin-Watson serial correlation test, absolute relative error test, turning point test and fitting test, depending on the covariance matrix of a disturbance term and type of data. Various technical methods are devised to make a computer solve the first (j=1) to the J-th (e.g., J=10) best subset problems in one computer-run, depending on whether or not a researcher has a new criterion or appropriate values for the parameters used to evaluate meaningful subsets before estimation. Then, the ultimately best subset among the best J subsets is regarded as a practical solution to the variable selection problem for the generalized least squares. The System *OEPP* can handle the proposed variable selection method.

1. Introduction

A method to solve the variable selection problem for the generalized least squares, abbreviated as GLS from here on, introduced by Aitken [1] has not been proposed in the literature. The stepwise regression method, the forward selection method, the backward elimination method, the minimax principle method, the branch-and-bound method, the t-directed method, etc. ([3], [6], [7]) have been proposed for the ordinary least squares. However, these methods and the software systems which can handle them have been rarely used in actual research. All possible regressions are not helpful, especially if the number of explanatory variables is large. Usually, non-statistical conditions such as the scientific knowledge related to research at hand, natural logic and common sense are used with statistical criteria to find an

estimated equation which can be used for analysis and/or prediction. Since these methods ignore such non-statistical conditions, they often select an equation which is statistically satisfactory but unacceptable for research. As a result, instead of using these methods, researchers load, estimate and evaluate many equations one at a time by adding, removing or altering variables until they can find satisfactory equations. This present manner to find a satisfactory equation is quite time-consuming, laborious and costly.

Non-statistical conditions are concerned with (i) roles or meanings of variables, (ii) signs and/or magnitudes concerning regression coefficients and (iii) constraints and hypothetical relations on regression coefficients. The condition (i) is concerned with whether a subset is meaningful or meaningless for research at hand. A meaningful subset is defined as the one which includes variables necessary to explain the behavior of an explained variable as well as possible but does not include any unnecessary or redundant variables. The condition (ii) is widely applied to the magnitude tests for a single regression coefficient and the value of a linear function of regression coefficients and to the comparison of the absolute values of regression coefficients. The condition (iii) is seen, for example, in economics. Zero sum of all price elasticities and income elasticity in a Cobb-Doublas type demand function is used as a constraint to be imposed on regression coefficients or a hypothetical relation to test whether or not consumers are affected by money illusion.

Thus, we need to develop a method to solve systematically the variable selection problem for GLS in which statistical criteria as well as non-statistical conditions are used.

2. The j-th Best Subset Problem

We assume that T=number of observation times; $t=1,2,\ldots,T$; N=number of cross-sectional units (e.g., sectors and regions); $n=1,2,\ldots,N$; NT=sample size; $Y=(NT\times 1)$-vector of an explained variable; $y_n(t)$=observation of Y in unit n at time t so that $Y=\{y_1(1),y_2(1),\ldots,y_N(1),\ldots,y_1(T),y_2(T),\ldots,y_N(T)\}'$; $X=\{NT\times(K+1)\}$-matrix of all possible explanatory variables, including a constant term; $X_i=i$-th $\{NT\times(K_i+1)\}$-submatrix of X; $i=1,2,\ldots,2^K-1$; $u=(NT\times 1)$-vector of a disturbance term with $u \sim N(0,\sigma^2 V)$; σ^2=scale factor; V=positive definite $(NT\times NT)$-matrix in which all elements are known and the diagonal elements are normalized; A_i and $\tilde{A}_i=\{(K_i+1)\times 1\}$-vectors of true and estimated coefficients of X_i, respectively; $\tilde{Y}_i=(NT\times 1)$-vector of Y predicted by using

X_i; $\tilde{Y}_i=\{\tilde{y}_{i1}(1),\tilde{y}_{i2}(1),\ldots,\tilde{y}_{iN}(1),\ldots,\tilde{y}_{i1}(T),\tilde{y}_{i2}(T),\ldots,\tilde{y}_{iN}(T)\}'$; $\tilde{E}_i=Y-\tilde{Y}_i$; $\tilde{E}_i=\{\tilde{e}_{i1}(1),\tilde{e}_{i2}(1),\ldots,\tilde{e}_{iN}(1),\ldots,\tilde{e}_{i1}(T),\tilde{e}_{i2}(T),\ldots,\tilde{e}_{iN}(T)\}'$; $\tilde{s}_i^2=\sigma^2$ estimated by using $\tilde{E}_i$; and t_r^q=t-value of q% significance level and degrees of freedom r.

The i-th subset can be expressed as follows:

(1) $$Y=X_iA_i+u \quad \text{with } u \sim N(0,\sigma^2V)$$

$\tilde{Y}_i$ can be expressed as follows:

(2) $$\tilde{Y}_i=X_i\tilde{A}_i$$

The author would like to formulate the j-th best subset problem for GLS as follows:

Find subset X_i from a given set X of all possible explanatory variables specified for an explained variable Y and estimate $\tilde{A}_i$ and $\tilde{s}_i^2$ in one computer-run such that

[I] subset X_i is meaningful for the research at hand,

[II] $\tilde{A}_i$ and $\tilde{s}_i^2$ are calculated as follows:
in the case where no constraints are imposed on A_i,

(3) $$\tilde{A}_i=(X_i'V^{-1}X_i)^{-1}X_i'V^{-1}Y$$

or in the case where a set of constraints $B_iA_i=G_i$ is imposed on A_i,

(4) $$\tilde{A}_i=(X_i'V^{-1}X_i)^{-1}[X_i'V^{-1}Y+B_i'\{B_i(X_i'V^{-1}X_i)^{-1}B_i'\}^{-1}\{G_i-B_i(X_i'V^{-1}X_i)^{-1}X_i'V^{-1}Y\}]$$

(5) $$\tilde{s}_i^2=\tilde{E}_i'V^{-1}\tilde{E}_i/(NT-K_i-1+M_i)$$

where M_i is the rank of B_i but $M_i=0$ in (5), (7), (8), (9) and (25) unless constraints are imposed on A_i,

[III] $\tilde{A}_i$ satisfies the following magnitude conditions (including sign conditions) based on the scientific knowledge of the research, if necessary:

(6) $$f_h^1\le F_{hi}^1\tilde{A}_i\pm|F_{hi}^2\tilde{A}_i|\pm|F_{hi}^3\tilde{A}_i|,\ F_{hi}^1\tilde{A}_i\pm|F_{hi}^2\tilde{A}_i|\pm|F_{hi}^3\tilde{A}_i|\le f_h^2 \quad \text{and/or}$$
$$f_h^1\le F_{hi}^1\tilde{A}_i\pm|F_{hi}^2\tilde{A}_i|\pm|F_{hi}^3\tilde{A}_i|\le f_h^2 \quad \text{for } h=1,2,\ldots,H_i$$

where f_h^1, f_h^2 and F_{hi}^p for p=1,2,3 are a lower bound, an upper bound and a row coefficient vector of the h-th magnitude condition, respectively, $|F_{hi}^p\tilde{A}_i|$ for p=2,3 is the absolute value of $F_{hi}^p\tilde{A}_i$, and $\pm$ stands for + or - (of course, if F_{hi}^p is a zero vector, it must be ignored),

[IV] the following inequality must be satisfied to reject the null hypothesis H_0 and to accept the specified alternative hypothesis H_1, if necessary:

$$|C_{di}\tilde{A}_i - c_d| / \sqrt{\tilde{S}^2_{di}} > t^{q/2}_{NT-K_i-1+M_i} \quad \text{for } H_1 : C_{di}A_i \neq c_d \tag{7}$$

$$(C_{di}\tilde{A}_i - c_d) / \sqrt{\tilde{S}^2_{di}} > t^{q}_{NT-K_i-1+M_i} \quad \text{for } H_1 : C_{di}A_i > c_d \tag{8}$$

or

$$(c_d - C_{di}\tilde{A}_i) / \sqrt{\tilde{S}^2_{di}} > t^{q}_{NT-K_i-1+M_i} \quad \text{for } H_1 : C_{di}A_i < c_d \tag{9}$$

$$H_0 : C_{di}A_i = c_d \text{ for } d=1,2,\ldots,D_i \tag{10}$$

for

$$\tilde{S}^2_{di} = \tilde{s}^2_i C_{di}(X_i' V^{-1} X_i)^{-1} C_{di}' \tag{11}$$

where C_{di} and c_d are a row coefficient vector and a value of the d-th hypothetical relation, respectively, and q is a significance level (%) of a t-test specified by the researcher,

[V] the Durbin-Watson statistic DW_i defined below

$$DW_i = \tilde{E}_i' V^{-1/2} Q V^{-1/2} \tilde{E}_i / \tilde{E}_i' V^{-1} \tilde{E}_i \tag{12}$$

for

$$Q = \begin{pmatrix} 1 & -1 & 0 & 0 & & & \\ -1 & 2 & -1 & 0 & & & 0 \\ 0 & -1 & 2 & -1 & & & \\ & \cdots & \cdots & \cdots & \cdots & & \\ & & & & -1 & 2 & -1 & 0 \\ & 0 & & & 0 & -1 & 2 & -1 \\ & & & & 0 & 0 & -1 & 1 \end{pmatrix} \tag{13}$$

satisfies the following inequalities, if time series data for $N=1$, $T \geqq 6$ and $K_i - M_i \leqq 20$ are used ([4], [5], [9]) and V is diagonal:

$$DW_i \geqq d^W_{iU} \quad (\text{or } DW_i \geqq d^W_{iL}) \text{ if } DW_i \leqq 2 \tag{14}$$

or

$$4 - DW_i \geqq d^W_{iU} \quad (\text{or } 4 - DW_i \geqq d^W_{iL}) \text{ if } DW_i > 2 \tag{15}$$

and, if V is an identity matrix and (14) or (15) does not hold, calculate the following:

$$(16)\qquad V_i = \begin{pmatrix} 1 & r_i & \cdots & r_i^{T-1} \\ r_i & 1 & \cdots & r_i^{T-2} \\ & \cdots\cdots\cdots & & \\ r_i^{T-1} & r_i^{T-2} & \cdots & 1 \end{pmatrix}$$

for

$$(17)\qquad r_i = \sum_{t=2}^{T} \tilde{e}_{i1}(t)\tilde{e}_{i1}(t-1) \Big/ \sqrt{\sum_{t=2}^{T} \tilde{e}_{i1}(t)^2}\sqrt{\sum_{t=2}^{T} \tilde{e}_{i1}(t-1)^2}$$

and go back to condition [II] above after setting $V=V_i$, where d_{iU}^{W} and d_{iL}^{W} are the upper and lower limits of the Durbin-Watson serial correlation test of w (%) significance level specified by the researcher and the parentheses in (14) and (15) must hold when he would like to regard through his subjective judgement an inconclusive case as uncorrelated,

[VI] $\tilde{Y}_i$ satisfies the following absolute relative error test, if necessary: for all n and t

$$(18)\qquad 100\times\left|\{y_n(t)-\tilde{y}_{in}(t)\}/y_n(t)\right| \leqq v^1 \quad \text{if } y_n(t)\neq 0$$

and

$$(19)\qquad \left|\tilde{y}_{in}(t)\right| \leqq v^2 \quad \text{if } y_n(t)=0$$

where v^1 (%) and v^2 are specified by the researcher,

[VII] $\tilde{Y}_i$ satisfies the following turning point test, if necessary: if

$$(20)\qquad \{y_n(t)-y_n(t-t_i)\}\{y_n(t+t_i)-y_n(t)\}<0$$

and

$$(21)\qquad 100\times\mathrm{Min}\left[\left|\{y_n(t)-y_n(t-t_i)\}/y_n(t)\right|, \left|\{y_n(t+t_i)-y_n(t)\}/y_n(t)\right|\right] \geqq w^1 \quad \text{for } y_n(t)\neq 0$$

or

$$(22)\qquad \mathrm{Min}\left[\left|y_n(t-t_i)\right|, \left|y_n(t+t_i)\right|\right] \geqq w^2 \quad \text{for } y_n(t)=0$$

then

$$(23)\qquad \{y_n(t)-y_n(t-t_i)\}\{\tilde{y}_{in}(t)-\tilde{y}_{in}(t-t_i)\}>0$$

and

$$(24)\qquad \{y_n(t+t_i)-y_n(t)\}\{\tilde{y}_{in}(t+t_i)-\tilde{y}_{in}(t)\}>0$$

(i) for all n, $T\geqq 3$ and $2\leqq t\leqq T-1$ and $t_i=1$ in the case where no lagged explained variables are included in X_i,

(ii) for all n, $T\geqq 2T_i+1$, $T_i+1\leqq t\leqq T-T_i$ and $t_i=T_i$ in the case where only

one lagged explained variable, whose time lag number is T_i, is included in X_i, or

(iii) for all n, $T \geqq 2T_i+1$, $t_i+1 \leqq t \leqq T-t_i$ and $t_i=1,2,\ldots,T_i$ in the case where two or more lagged explained variables are included in X_i and T_i stands for the maximum among the time lag numbers of lagged explained variables,

where w^1 (%) and w^2 are specified by the researcher,

and

[VIII] $\tilde{A}_j$ yields the j-th highest Buse's coefficient of determination ([2]) adjusted by the number of explanatory variables among the subsets which have passed the above conditions [I] to [VII]:

$$RR_i=1-(1-R_i)(NT-1)/(NT-K_i-1+M_i) \tag{25}$$

where

$$R_i=1-\tilde{E}_i'V^{-1}\tilde{E}_i/(Y-\tilde{y}U)'V^{-1}(Y-\tilde{y}U) \quad \text{for } \tilde{y}=U'V^{-1}Y/U'V^{-1}U \tag{26}$$

and $U=(1,1,\ldots,1)'$ with dimension NT.

The author wants to propose the first ($j=1$) best subset as a practical solution to the variable selection problem for GLS, if the researcher knows appropriate values for the parameters (f_h^1, f_h^2, q, w, v^1, v^2, w^1, w^2, etc.) and does not have any new test to evaluate subsets. On the other hand, if he does not know appropriate values for the parameters before estimation or if he has a new test to evaluate subsets, he should solve the first to the J-th (e.g., $J=10$) best subset problems in one computer-run and find by himself the ultimately best subset among these J subsets by comparing them with each other or by applying the new test to these J subsets. Such an ultimately best subset can be regarded as a practical solution to the variable selection problem for GLS.

3. Derivation of All Meaningful Subsets from All Possible Explanatory Variables by the System *OEPP*

Let us show a method to make a computer derive all meaningful subsets from a given set of all possible variables. It must be noted that any kind of symbols and styles can be used instead of ours but the variable selection rules must be kept. We assume that an explained variable, Y, and a set, X, of its possible variables are loaded into a computer by a functional form like $Y=F(X)$. The X is classified into 3 kinds on the basis of a knowledge of applied research at hand. The first is for a group of P variables from which at least Q but at most R variables must be selected, where P, Q and R are integers satisfying $0 \leqq Q \leqq R \leqq P$ and $R \geqq 1$. $Q=0$ implies that an empty subset can be meaningful with respect to these variables. These meaningful subsets are

regarded as equivalent to each other concerning the selection of variables before estimation and evaluation. Let us enclose these P variables within $<Q<...>R>$ or $<R<...>Q>$. For example, $<0<WM,KS,LH,HG>1>$, where P=4, Q=0 and R=1, generates the following 5 meaningful subsets with respect to these variables: $\{\emptyset\}$, $\{WM\}$, $\{KS\}$, $\{LH\}$ and $\{HG\}$.

The second kind of classification is for a group of P variables whose degrees of importance are a priori known and among which the most important to the Q-th important variable must be always selected and the remaining variables are selected optionally, successively and additionally from the Q+1-st important to the least important variable. Lagged or powered variables belong to this classification. Let us assume that X_1 is more important than X_2 which is more important than X_3 ... X_{P-1} which is more important than X_P and enclose them within $<Q<...>>$ or $<<...>Q>$ in such a way that the most important variable X_1 to the least important variable X_P are entered from the $<Q<$ side to the $>>$ side or from the $>Q>$ side to the $<<$ side, respectively, where P and Q are integers satisfying $0 \leqq Q \leqq P$. Q=0 allows an empty subset as a meaningful subset with respect to these variables. Thus, we can express as $<Q<X_1,X_2,...,X_P>>$ or $<<X_P,X_{P-1},...,X_2,X_1>Q>$. For example, $<1<W,W(-1),W(-2)>>$ or $<<W(-2),W(-1),W>1>$, where $W(-t)$ is variable W with time lag number t, generates the following 3 meaningful subsets with respect to these variables: $\{W\}$, $\{W,W(-1)\}$ and $\{W,W(-1),W(-2)\}$. $<0<W,W(-1),W(-2)>>$ or $<<W(-2),W(-1),W>0>$ generates the following 4 meaningful subsets with respect to these variables: $\{\emptyset\}$, $\{W\}$, $\{W,W(-1)\}$ and $\{W,W(-1),W(-2)\}$.

The third classification is for grouped variables. Grouped variables are such that they cannot be selected separately but they must be selected as a whole just like a single variable in deriving all meaningful subsets. They are enclosed within (...) and appear in the above 2 classifications. For example, $<2<A,B(C1,C2)>3>$ derives the following 4 meaningful subsets with respect to these variables: $\{A,B\}$, $\{A,C1,C2\}$, $\{B,C1,C2\}$ and $\{A,B,C1,C2\}$.

Since a constant term has special meanings, it is better to treat it separately from non-constant variables. We give it a special symbol like @C and select it in all possible subsets, if it is needed.

Let us show that these 3 kinds of variable classifications can always derive all meaningful subsets. Suppose that the following equations correspond to all meaningful subsets:

$$Y=a_i^0+X_iA_i \quad \text{for } i=1,2,...,L \tag{27}$$

where a_i^0 and X_i are the constant term and the row vector of non-constant explanatory variables in the i-th equation and A_i is the column vector of the coefficients of X_i. (27) can be derived, whether nested or non-nested, from

(28) $Y=F(@C<1<(X_1)(X_2)\ldots(X_L)>1>)$

where all X_i's are treated as sets of grouped variables.

Let us show only how to derive the constraint suitable for each of all meaningful subsets from an aggregate constraint. An aggregate constraint is defined as a linear equation, like $bA=g$, of the regression-coefficients A of all possible explanatory variables X from which $b_iA_i=g$ for X_i is derived by selecting the regression-coefficients A_i and the equation-coefficients b_i corresponding to X_i, where b, b_i and g are the row vector of the known equation-coefficients of A, the row vector of the equation-coefficients of A_i and the value, respectively. In the System *OEPP* [8], $bA=g$, which is called an aggregate constraint, is loaded through $bX'=g$ by using explanatory variables instead of regression-coefficients. Since the equation-coefficients of some regression-coefficients are zeros, only $b^*X^{*\prime}=g$ is actually loaded, where b^* is the row vector of non-zero equation-coefficients and X^* is the row vector of the explanatory variables corresponding to b^*. This method can be easily applied to aggregate magnitude conditions and general linear hypothetical relations.

4. An Application to an Agricultural Production Function

We apply the proposed variable selection method for estimation of an agricultural production function of Cobb-Douglas type by using the data observed from 1965 to 1979 in Japan. As the Japanese economy has been growing, the ratio of the part-time farmers (second class), who are defined as farmers whose agricultural incomes are less than non-agricultural as a proportion of their total incomes, has become larger. Part-time farmers cannot take care of their crops and animals as much as full-time farmers (and part-time farmers of the first class), because the former are involved in their non-agricultural jobs more than the latter. As a result, it is considered that the agricultural production of the part-time farmers is easily affected by unfavorable weather so that it has larger variances than that of the full-time farmers. Hence, we assume that the variances of a disturbance term during the above estimation period are not constant but proportional to the ratio of the part-time farmers to all farmers and the covariances are zero.

We introduce the following notation; *LY* = log(products); *@C* = constant term; *DVCS* = dummy variable for cold summer; *LL* = log(labor); *LK* = log(capital) = log(KA+KP+KM) = log(animal+plant+machinery capital); *LKR* = log(capital adjusted by a use rate of machinery capital); *LAX* = log(*A*-*X*) = log(cultivated

acreage minus abandoned and damaged acreage); $LCAX$ = log(A-X adjusted by rice production index); LQ = log(intermediate inputs); and T = time trend for technical progress.

The following functional form is appropriate:

(29) $LY=F(@C<1<LL>1><0<DVCS,T,LQ>3><1<LK,LKR>1><1<LAX,LCAX>1>)$

which derives only $2^3\times2\times2=32$ meaningful subsets among $2^8-1=255$ possible subsets. Accordingly, the remaining 223 subsets are meaningless for this research.

To check for and avoid unrealistic equations, we introduce the following criteria: $DVCS<0$; $T>0$; $0.1\leqq LL<0.5$; $0.1\leqq LQ<0.3$; $0.1\leqq LK+LKR<0.5$; $0.1\leqq LAX+LCAX<0.6$; $0.85\leqq LL+LQ+LK+LKR+LAX+LCAX\leqq1.15$; 5% t-test; 5% Durbin-Watson serial correlation test (an inconclusive case is treated as acceptable by the researcher's subjective judgement); 1% absolute relative error test; and 0.5% slope turning point test, where the variables imply their regression coefficients.

When we tried to solve the first and second best subset problems in one computer-run of the System *OEPP*, we obtained only the following first best subset in less than 2 seconds CPU time by the FACOM M-380 (about 23 MIPS):

$$
\begin{array}{lllll}
LY & = 0.5837250 + 0.3158237*LL + 0.0208933*T + 0.1646213*LKR & & & \\
(S.E.) & (1.343301) & (0.1306969) & (0.0071497) & (0.0523064) \\
(T.R.) & (0.434545) & (2.416459) & (2.922250) & (3.147253)
\end{array}
\tag{30}
$$

$$
\begin{array}{l}
+ 0.5837928*LCAX \\
(0.1626294) \\
(3.589713)
\end{array}
$$

BR=0.9344, BRR=0.9081, SE=0.0200, FA=-0.0074, DW=2.032 where $S.E.$ and $T.R.$ in parentheses, BR, BRR, SE, FA and DW are standard errors and t-ratios of coefficients, Buse's coefficient of determination, Buse's adjusted coefficient of determination, scale factor of a disturbance term, first-order autocorrelation coefficient and Durbin-Watson statistic, respectively. The equation satisfies all tests applied and can be regarded as reasonable.

5. Concluding Remarks

The variable selection problem for GLS can be solved by using non-statistical conditions as well as statistical criteria. Since an overall evaluation function which reflects various non-statistical conditions and statistical criteria does not exist, the author formulated the j-th best subset problem, suggested to solve the first to the J-th (e.g., J=10) best subset problems in

one computer-run, depending on whether or not a researcher has appropriate values for the parameters or a new test, and proposed to regard the ultimately best subset among the best J subsets as a practical solution to the variable selection problem for GLS.

This method drastically saves time and reduces labor and cost, while it can improve the quality of applied research.

Acknowledgements

The author wishes to express his deep thanks to referees for helpful comments and suggestions on the paper.

References

[1] Aitken, A.C.: On Least-squares and Linear Combinations of Observations, *Proceedings of Royal Society*, Edinburgh, Vol. 55, 1934, 42-48.

[2] Buse, A.: Goodness of Fit in Generalized Least Squares Estimation, *American Statistician*, Vol. 27, No. 3, 1973, 106-108.

[3] Draper, N.R. and Smith, H.: *Applied Regression Analysis*, 2nd edition, Wiley, New York, 1981.

[4] Durbin, J. and Watson, G.S.: Testing for Serial Correlation in Least Squares Regression I, *Biometrika*, Vol. 37, 1950, 409-428.

[5] Durbin, J. and Watson, G.S.: Testing for Serial Correlation in Least Squares Regression II, *Biometrika*, Vol. 38, 1951, 159-178.

[6] Hocking, R.R.: The Analysis and Selection of Variables in Linear Regression, *Biometrica*, Vol. 32, 1976, 1-49.

[7] Judge, G., Griffiths, W.E., Hill, R. and Lee, T.: *The Theory and Practice of Econometrics*, Wiley, New York, 1980.

[8] Onishi, H.: *Professional Researcher System OEPP for Socio-Economic Analysis and Forecast*, Institute of Socio-Economic Planning, University of Tsukuba, Japan, 1987.

[9] Savin, N.E. and White, K.J.: The Durbin-Watson Test for Serial Correlation with Extreme Sample Sizes or Many Regressors, *Econometrica*, Vol. 45, No. 8, 1977, 1989-1996.

Haruo ONISHI: Institute of Socio-Economic Planning, University of Tsukuba, 1-1-1 Tennohdai, Tsukuba, Ibaraki, Japan 305