No.191 (83-14)

Variable Selection in Regression Analysis and Package OEPP

by

Haruo ONISHI

October 1983

Variable Selection in Regression Analysis and Package OEPP

University of Tsukuba

Haruo ONISHI

A variable selection procedure is proposed which can be commonly used for the ordinary least squares, the linearly-constrained ordinary least squares, the generalized least squares, and the linearly-constrained generalized least squares. The procedure starts from the generation of all meaningful subsets and ends up with finding the j-th best subset. All possible subsets are divided into meaningful and meaningless subsets by using knowledge of a research field in question. A subset which makes sense as a reasonable equation from the viewpoint of the research field is called a meaningful subset. Only meaningful subsets are generated by variable classifications, estimated by one of the above four methods, and evaluated not only by statistical criteria but also by criteria unique to the research field in one run of the computer. Computer package OEPP can handle this procedure.

1. Introduction

Regression analysis is one of the most popular methods to analyze complicated relationships among variables representing activities, elements, powers, or the like in research and is used in many fields covering natural sciences as well as social sciences. At present, applied researchers who make regression analyses, especially in social sciences, usually estimate and evaluate equations one at a time until they can find the most satisfactory equation. A variable selection problem is that of how to obtain the most satisfactory equation in the most efficient way with respect to time, labor, and cost. Various procedures and methods have been proposed. They include the stepwise regression procedure, the forward selection method, the backward elimination method, the stagewise regression procedure, and the min-max regret principle method [5], [8], [16], [18], [26]. BMDP, SAS, SPSS, MINITAB, and some other computer program packages can handle these procedures [7], [17], [22], [25]. However, unfortunately these procedures are seldom used in application. The reason is that there is a gap between statisticians' interests and applied researchers' needs.

A procedure by all possible regressions means to efficiently estimate in one run of the computer all possible subsets derived from a set of all possible predictor variable candidates (called 'candidates' from here on) which are considered to affect a dependent variable. This approach can eventually lead a researcher to find the most satisfactory subset, if such a subset exists. For instance, if there are 20 different candidates for a dependent variable, $2^{20}-1= 1,048,575$ possible subsets exist. The computer output by all possible regressions becomes huge in volume and expensive. It is quite difficult for a researcher to select by himself the most satisfactory subset by checking 1,048,575 subsets by statistical criteria as well as by non-statistical criteria unique to the field. If some of the 1,048,575 subsets include candidates necessary for reasonable equations, it is quite troublesome to choose those equations from 1,048,575 equations. Subsequently, a researcher loads, estimates, and evaluates by statistical and non-statistical criteria some number of possible equations for a dependent variable one at a time until he can find a rather good one. During this process, an applied researcher always utilizes knowledge of the research field, say economic theories, field surveys, and/or empirical studies, when he specifies candidates necessary for a dependent variable. Therefore, these possible equations must make sense with respect to the specification of candidates. This process is similar to all possible regressions but avoids meaningless regressions which all possible regressions may include. This approach is still quite time-consuming, labor-consuming, and costly.

Thus, a variable selection procedure has been universally requested to automatically find the most satisfactory subset derived from a set of all possible candidates most efficiently with respect to time, labor, and cost. This paper proposes a procedure which automatically finds the pragmatically best subset, when the ordinary least squares (OLS), constrained ordinary least squares (COLS), generalized least squares (GLS), or constrained generalized least squares (CGLS) method is utilized for estimation. This procedure is embodied in package OEPP designed for socio-economic analysis and forecasting.

2. Factors and Candidates in Regression

An estimation problem is concerned with the determination of what function, what factors, what candidates (or sets of data), and what estimation method are suitable for a regression analysis in question. Whether a function is linear or nonlinear is an important matter. What kind of nonlinear form is appropriate is a complicated problem. There are cases in which some candidates which have different data are considered to alternatively or complementarily explain one phase of the movement (or behavior) of a dependent variable. We would like to call the common characteristics of these candidates a factor. The simplest case is that there is one candidate for one factor. In most natural sciences in which experiments or obervations can be repeated in a rather short time, the case in which one factor is fully represented by one candidate is usual. However, especially in social sciences like economics, politics and sociology, experiments for the purpose of obtaining data are quite difficult. Even if possible, the observations of experiments are quite time-consuming and expensive. This reality in social sciences leads to the following problems: more than one candidate possibly represents a factor, and a researcher cannot determine which candidate or what combination of candidates represents most appropriately that factor before estimation and evaluation. For instance, the rich can buy a commodity, say beef, more than the poor, if all other things are equal. It is reasonable to consider that the factor wealth is important to explain the purchase of the commodity. What set of data represents most appropriately the factor wealth ? The current real income, a combination of the current and the last period's real income, the average of the current and the last period's real income, the current real income plus the real liquid assets (savings in banks and post offices, bonds, securities, stocks, etc.) are possible candidates for the factor wealth. Many estimation methods have been developed. Sometimes it is difficult to know which estimation method is most suitable for an equation in question. Thus, an estimation problem is quite complicated. Here, we put emphasis on practicability. We would like to assume that a linear form with respect to coefficients is appropriate and an appropriate estimation method among OLS, COLS, GLS, and CGLS is known. Then, we can focus on a variable selection problem in which factors and candidates are determined under the assumptions of linearity of coefficients and a predetermined estimation method.

3. Basic Variable Classification Based on Information Unique to an Applied Field

When subsets are derived from a set of all possible candidates considered to affect a dependent variable, it is quite important to recognize whether or not a subset is meaningful from the viewpoint of knowledge of an applied field before estimation. Statistics cannot cover such knowledge. For instance, in the field of economics, economists are interested in economically meaningful subsets but ignore economically meaningless ones. Thus, economists have to distinguish economically meaningful subsets from economically meaningless ones by using economic knowledge. A meaningful subset includes necessary candidates for a reasonable equation for a dependent variable and should be estimated and evaluated. Usually, knowledge is obtained through theories including hypotheses to be tested, field surveys (experiments in natural sciences), and/or empirical studies. In general, possible candidates for a dependent variable are basically classified into eight groups, regardless of applied fields, by knowledge of the research field [23]. They are absolutely important (or forced or core), optionally important, exclusively important, gradually important, exclusively optional, gradually optional, completely optional, and fixed groups. Let us explain briefly basic classification. Suppose that Y=dependent variable, V_k=the k-th candidate with constant term V_0, a_k=coefficient of candidate V_k, $\mathbb{X}$=set of all possible candidates V_k's, $\mathbb{X}_i$=the i-th meaningful subset, A_i=coefficient vector of $\mathbb{X}_i$, and $\mathbb{X}_*$=set of all meaningful subsets $\mathbb{X}_i$'s. Let a dependent variable and all possible candidates be loaded, under the assumption of linearity of coefficients, in a functional format as in package OEPP, where a variable is expressed with a maximum of 8 alphanumeric characters starting from an alphabetic character, and a lagged variable is expressed with the notation of that non-lagged variable followed and attached by a minus lag number in parentheses, like ABCD(-3) of variable ABCD. For instance, unless all possible regressions are requested, the following functional format, which is actually loaded,

$$Y=F(V_0,V_1,V_2,V_3,V_4,V_5) \qquad (1)$$

leads to the estimation of

$$Y=a_0+a_1V_1+a_2V_2+a_3V_3+a_4V_4+a_5V_5 \qquad (2)$$

If all possible regressions are requested, (1) leads to $2^5-1=31$ possible subsets, where V_0 is always selected and V_1 to V_5 are selected in the combinatorial manner.

Absolutely important candidates, denoted by enclosing them within /.../ in a functional format, are such that they must be included in all meaningful subsets. Entry /V_1,V_2,V_3/ in a functional format, just like Y=F(.../V_1,V_2, V_3/...) implies that candidates V_1, V_2, and V_3 all must be included in a meaningful subset. Optionally important candidates, denoted by enclosing them within <...>, are such that at least one of them must be included in a meaningful subset. Entry <V_4,V_5,V_6> like in Y=F(...<V_4,V_5, V_6>...) requires at least one of V_4, V_5, and V_6 be included in a meaningful subset. Exclusively important candidates, denoted by enclosing them within </.../>, are such that only one of them must be included in a meaningful subset. Entry </V_7,V_8,V_9/> like in Y=F(...</V_7,V_8,V_9/>...) requires only one of V_7, V_8, and V_9 be included in a meaningful subset. Gradually important candidates, denoted by entering them within <+...+>, are such that they are gradually selected, beginning from the most important to the least important among them. It is postulated that the most important candidate among them is entered in the left-most position in <+...+> and the least important in the right-most position. Entry <+V_{10},V_{11},V_{12}+> like in Y=F(...<+V_{10},V_{11},V_{12}+>...) permits one of the following three cases to be included in a meaningful subset: (a) only V_{10}, (b) V_{10} and V_{11}, or (c) V_{10}, V_{11}, and V_{12}. The information used in this entry is that candidate V_{10} is more important than candidate V_{11} which is more important than candidate V_{12}. Exclusively optional candidates, denoted by entering them within <*...*>, are such that at most one of them must be included in a meaningful subset. Entry <*V_{13},V_{14},V_{15}*> like in Y=F(...<*V_{13},V_{14},V_{15}*>...) permits (a) none of candidates V_{13}, V_{14}, and V_{15}, (b) only V_{13}, (c) only V_{14}, or (d) only V_{15} to be included in a meaningful subset. Gradually optional candidates, denoted by entering them within <-...->, are such that a meaningful subset does not necessarily include any of them but once the subset includes at least one of them, it must include them in such a way that gradually important candidates are included in a meaningful subset. Entry <-V_{16},V_{17},V_{18}-> like in Y=F(...<-V_{16},V_{17},V_{18}->...) permits one of the following four cases in a meaningful subset: (a) none of them is included, (b) only V_{16}, (c) V_{16} and V_{17}, or (d) V_{16}, V_{17}, and V_{18}. Completely optional candidates, denoted by entering them freely in a functional format, are such that a subset cannot be considered to be meaningful or meaningless by whether or not they are included in a subset before estimation and their importance is determined only through evaluation. Entry V_{19},V_{20} like in Y=F(...V_{19},V_{20},...) permits (a) none of candidates V_{19} and V_{20}, (b) only V_{19}, (c) only V_{20}, or (d) V_{19} and V_{20} to be included in a meaningful subset. Finally, fixed candidates, denoted by entering them within (...) in a functional format, are such that they cannot be separately included in a meaningful subset and once they are selected in a meaningful subset, all of them as a group must be included. Fixed candidates are also optionally important, exclusively important, gradually important, exclusively optional, gradually optional, or completely optional. Entry </V_{21},V_{22}(V_{23}, V_{24})/> like in Y=F(...</V_{21},V_{22}(V_{23},V_{24})/>...) leads one of the following cases to be included in a meaningful subset: (a) only V_{21}, (b) only V_{22}, or (c) V_{23} and V_{24}. It is postulated that candidates are separated with " ", ",", "<", ">", "/", "<", ">", "</", "/>", "<+", "+>", "<*", "*>", "<-", "->", "(", or ")". Let us give a simple example. If all meaningful subsets are requested to be estimated by a command, the next functional format, which is loaded by a user,

$$Y=F(V_0/V_1,V_2/<+V_3,V_4+>V_5) \qquad (3)$$

leads to the estimation of the following four equations:

$$Y=a_0+a_1V_1+a_2V_2+a_3V_3+a_4V_4+a_5V_5 \qquad (4)$$

$$Y=a_0+a_1V_1+a_2V_2+a_3V_3+a_5V_5 \qquad (5)$$

$$Y=a_0+a_1V_1+a_2V_2+a_3V_3+a_4V_4 \qquad (6)$$

$$Y=a_0+a_1V_1+a_2V_2+a_3V_3 \qquad (7)$$

Needless to say, $\mathbb{X}=\{V_0,V_1,V_2,V_3,V_4,V_5\}$, $\mathbb{X}_1=\{V_0,V_1,V_2,V_3,V_4,V_5\}$, $\mathbb{X}_2=\{V_0,V_1,V_2,V_3,V_5\}$, $\mathbb{X}_3=\{V_0,V_1,V_2,V_3,V_4\}$, $\mathbb{X}_4=\{V_0,V_1,V_2,V_3\}$, and $\mathbb{X}_*=\{\mathbb{X}_1,\mathbb{X}_2,\mathbb{X}_3,\mathbb{X}_4\}$. Let us see the candidates in detail. Constant term V_0 is always included in a meaningful subset, if it appears in a functional format. Candidates V_1 and V_2 are absolutely important, so that V_1 and V_2 must be always included in a meaningful subset. Candidates V_3 and V_4 are gradually important. Then, only V_3 or pair V_3 and V_4 must be included in a meaningful subset. Candidate V_5 is completely optional, so that some meaningful subsets, i.e., (4) and (5) can include V_5 and the others, i.e., (6) and (7), do not include V_5. All meaningful subsets derivable from (3) are (4) to (7). Any other possible subsets among $2^5-1=31$ possible subsets are regarded as meaningless.

4. Nested Variable Classification

It is possible to derive from a given set of all possible candidates only meaningful subsets for a research field in question through the basic variable classification. In some cases, it becomes quite troublesome to load all possible candidates in accordance with the rule of basic variable classifications. However, if we permit nested variable classifications, we can load the candidates compactly in a functional format. Here we would like to confine ourselves to the nested variable classifications of optionally important candidates (entered inside <...>), exclusively important candidates (entered inside </.../>, and exclusively optional candidates (entered inside <*...*>) which are sometimes used in application. Although the nested variable classifications of gradually important or gradually optional candidates are mathematically or statistically interesting, we do not refer to them here, because the author cannot find good examples for such cases in application. It must be kept in mind that nested variable classifications can be reduced to the basic ones.

3.1 Nested Variable Classification of Optionally Important Candidates

Optionally important candidates are entered into <...> in the basic variable classification. Now we permit:

(i) </.../>, <+...+>, <*...*>, <-...->, (...), and a single candidate can be entered inside <...>.

(ii) A group of candidates in </.../>, <+...+>, <*...*>, <-...->, or (...) are treated like a single candidate when outer <...> is expanded.

(iii) </.../> and <+...+> must be always selected but <*...*>, <-...->, (...), and a single candidate are selected like a completely optional candidate when outer <...> is expanded.

(iv) At least one candidate must be selected when outer <...> is expanded.

(v) Once </.../>, <+...+>, <*...*>, <-...->, or (...) is selected like a single candidate in the expansion of nested <...>, the candidates in </.../>, <+...+>, <*...*>, <-...->, or (...) are subordinated to the rule of the basic variable classifications.

(vi) Redundant subsets must be ignored, if they occur.

Let us give an example.

$$<</V_1,V_2/><*V_3,V_4*>> \qquad (8)$$

Let us put $P=</V_1,V_2/>$ and $R=<*V_3,V_4*>$. Then, (8) becomes <P,R>. Since P must be always selected, we have two cases: (a) P or (b) P,R. We have (a-1) V_1 or (a-2) V_2 from (a) and (b-1) V_1, (b-2) V_2, (b-3) V_1,V_3, (b-4) V_1,V_4, (b-5) V_2,V_3, or (b-6) V_2,V_4 from (b). (a-1) is the same as (b-1), while (a-2) is the same as (b-2). By rule (vi), (b-1) and (b-2) are redundant. Accordingly, the following six cases are possible: (a-1), (a-2), and (b-3) to (b-6). It is easily recognized that $<</V_1,V_2/>\ <*V_3,V_4*>>$ is equivalent to $</V_1,V_2(V_1,V_3)(V_1,V_4)\ (V_2,V_3)(V_2,V_4)/>$.

3.2 Nested Variable Classification of Exclusively Important Candidates

(i) <...>, <+...+>, (...), and a single candidate are permitted to be entered inside </.../>.

(ii) A group of candidates in <...>, <+...+>, or (...) is treated like a single candidate when outer </.../> is expanded.

(iii) Only one candidate must be selected in the expansion of outer </.../>.

(iv) Once <...>, <+...+>, or (...) is selected in the expansion of outer </.../>, the candidates in <...>, <+...+>, or (...) must follow the rule of the basic variable classifications.

(v) Redundant subsets must be ignored, if they occur.

Let us give an example.

</<V_1,V_2,V_3>V_4/> (9)

Let us put P=<V_1,V_2,V_3> so that (9) becomes </P,V_4/>. We have (a) P or (b) V_4. (a) leads to (a-1) V_1, (a-2) V_2, (a-3) V_3, (a-4) V_1,V_2, (a-5) V_1,V_3, (a-6) V_2,V_3, or (a-7) V_1,V_2,V_3. Finally, we have the following eight cases: (a-1) to (a-7), or (b). Needless to say, </<V_1,V_2,V_3> V_4/> is equivalent to </V_1,V_2,V_3,V_4(V_1,V_2)(V_1,V_3)(V_2,V_3) (V_1,V_2,V_3)/>.

3.3 Nested Variable Classification of Exclusively Optional Candidates

(i) <...>, <-...->, (...), and a single candidate are permitted to be entered inside <*...*>.

(ii) A group of candidates in <...>, <-...->, or (...) is treated like a single candidate when outer <*...*> is expanded.

(iii) At most one candidate must be selected in the expansion of outer <*...*>.

(iv) Redundant subsets must be ignored, if they occur. Let us give an example.

$$<*<-V_1,V_2,V_3->(V_4,V_5)*> \qquad (10)$$

Let us put $P=<-V_1,V_2,V_3->$ and $R=(V_4,V_5)$. Then we have <*P,R*> which leads to (a) P, (b) R, or (c) empty. (a) generates (a-1) V_1, (a-2) V_1,V_2, (a-3) V_1,V_2,V_3, or (a-4) empty. (b) is V_4,V_5. Hence, we have (a-1) to (a-3), (b), or (c). Thus, (10) is equivalent to $<*V_1(V_1,V_2),(V_1,V_2,V_3)(V_4,V_5)*>$.

Let us give an example of nested variable classification in economics. We assume that a demand quantity of beef depends on candidates representing the factor wealth, the factor relative price, and the factor inertia effect on eating habit. For simplicity, we assume that the factor wealth is represented by (real) income, the factor relative price is represented by prices of beef relative to those of pork, poultry, mutton, ham, sausage, and fish, and the factor inertia effect is represented by at most the last two periods' demand quantities. Since there may be a lag of income spent on beef, we introduce a new candidate of average income for two periods in addition to income itself. The demand for beef is believed to be affected not only by the current income but also by the last two periods' incomes. Then, we can load the following functional format:

```
BD=F($C</<+Y,Y(-1),Y(-2)+><+YY,YY(-1),YY(-2)+>/></<</B1P1,
     B1P2/>B1PL,B1M,B1H,B1S,B1F><</B2P1,B2P2/>B2PL,B2M,
     B2H,B2S,B2F>/><-BD(-1),BD(-2)->)                      (11)
```

where BD=beef demand quantity, $C=constant term, Y=real income, Y(-i)= i-periods-ago Y, YY={Y+Y(-1)}/2, YY(-i)= i-periods-ago YY, Bj=(consumer) price of beef of quality j (for instance, sirloin for j=1 and chopped meat for j=2), Pj=price of pork of quality j, PL=price of poultry, M=price of mutton, H=price of ham, S=price of sausage, F=price of fish, BiPj=Bi/Pj for i,j=1,2 and i≠j, BiPL=Bi/PL for i=1,2, BiM=Bi/M for i=1,2, BiH=Bi/H for i=1,2, BiS=Bi/S for i=1,2, BiF=Bi/F for i=1,2, and BD(-i)= i-periods-ago beef demand quantity. Because (i) <+Y,Y(-1),Y(-2)+> and <+YY,YY(-1), YY(-2)+> are exclusive of each, (ii) either of them must be selected for the factor wealth, and (iii) the current income may be more important than the previous period's income which may be more important than the two-periods-ago income, we introduced </<+Y,Y(-1),Y(-2)+><+YY,YY(-1), YY(-2)+>/>. It should be noted that the selection of, say, Y and Y(-2) before estimation and evaluation is avoided, because it is rather difficult to affirm before statistical tests that the previous period's income Y(-1) does not affect the current demand for beef but the two-periods-ago income Y(-2) does. The number of all possible subsets is $2^{22}-1=4,194,303$, because there are 22 non-constant candidates. However, as calculated by (16) in the next chapter, the number of all meaningful subsets is $(2x3)x\{2x(2x2^5)\}x(2+1)=2,304$. The remaining 4,191,999 subsets are meaningless from the viewpoint of economics. The feasible set for estimation consists of 2,304 subsets in the domain of 4,194,303 subsets. Functional format

(11) is equivalent to the following four functional formats:

```
BD=F($C<+Y,Y(-1),Y(-2)+><</B1P1,B1P2/>B1PL,B1M,B1H,
     B1S,B1F><-BD(-1),BD(-2)->)                            (12)

BD=F($C<+Y,Y(-1),Y(-2)+><</B2P1,B2P2/>B2PL,B2M,B2H,
     B2S,B2F><-BD(-1),BD(-2)->)                            (13)

BD=F($C<+YY,YY(-1),YY(-2)+><</B1P1,B1P2/>B1PL,B1M,
     B1H,B1S,B1F><-BD(-1),BD(-2)->)                        (14)

BD=F($C<+YY,YY(-1),YY(-2)+><</B2P1,B2P2/>B2PL,B2M,
     B2H,B2S,B2F><-BD(-1),BD(-2)->)                        (15)
```

5. Number of Meaningful Subsets

We assume that nested variable classifications are reduced to basic ones and a group of fixed candidates is counted as a single candidate in the calculation of the number of all meaningful subsets. Since the computer cost depends on the number of all meaningful subsets, it is convenient to know how many meaningful subsets are generated, estimated, and evaluated. The number of all meaningful subsets is calculated by the following formula:

$$NMS=\prod_{i=1}^{I}(2^{i_P}-1)\times\prod_{j=1}^{J}j_Q\times\prod_{k=1}^{K}k_R\times\prod_{\ell=1}^{L}(\ell_S+1)\times\prod_{m=1}^{M}(m_T+1)\times 2^N-g \qquad (16)$$

where I groups each of which has i_P optionally important candidates for $1\leq i\leq I$, J groups each of which has j_Q exclusively important candidates for $1\leq j\leq J$, K groups each of which has k_R gradually important candidates for $1\leq k\leq K$, L groups each of which has ℓ_S gradually optional candidates for $1\leq \ell\leq L$, M groups each of which has m_T exclusively optional candidates for $1\leq m\leq M$, and N completely optional candidates appear in a functional format, g=0 if at least one of /.../, </.../>, and <+...+> is used, and g=1 if none of /.../, </.../>, and <+...+> is used. If some groups of candidates do not appear in a functional format, the corresponding terms in (16) must be ignored. The number of absolutely important candidates and use of a constant term do not change the number of all meaningful subsets, because they are always included in each meaningful subset. For instance, A=F($C/B,C/<D,E,F></G,H/><*I,J,K*>(L,M)(N,O,P)R) has $NMS=(2^3-1)x2x(3+1)x2^3=448$ meaningful subsets among $2^{16}-1=65{,}535$ possible subsets, where I=1, $1_P=3$, J=1, $1_Q=2$, K=0, L=0, M=1, $1_S=3$, N=3, and g=1. The feasible set for estimation consists of 448 subsets in the domain of 65,535 subsets.

6. The j-th Best Subset Problem

A variable selection problem must be solved not only by statistical knowledge but also by knowledge unique to a research field in question. Since regression analysis is made in many fields, it is quite difficult to formulate a variable selection problem in a way common to all research fields. Accordingly, we would like to formulate the j-th best subset problem. If a researcher does not have any criterion other than the criteria in the j-th best subset problem defined below, solving the best subset problem (j=1) means solving the variable selection problem. If he has a unique criterion, he can solve the best subset problem to the J-th best subset problem, which are called here the first J best subset problems, and may be able to find the ultimately best subset among the best J subsets.

In OEPP, the specifications of appropriate criteria and parameter BEST=J can solve the first J best subset problems, and the best to the J-th best subset (called subequation in OEPP) are obtained in this order with/without the confidence intervals, tables, and graphs in one run of the computer.

To treat time series data, cross-sectional data, and pooled data (time series as well as cross-sectional data), we would like to assume that there are N cross-sectional units (firms, households, local governments, prefectures, plots, classes, etc.) and T observation times. Case $N=1$ and $T>1$ indicates time series data, case $N>1$ and $T=1$ indicates cross-sectional data, and case $N>1$ and $T>1$ indicates pooled data. Let Y=vector of a dependent variable's data $Y_n(t)$ for $1 \leq n \leq N$ and $1 \leq t \leq T$, $\mathbb{Y}$=a dependent variable loaded, $\bar{Y}_i$=Y estimated by the i-th meaningful subset X_i with element $\bar{Y}_{in}(t)$, X=matrix of all possible candidates' data, $\mathbb{X}$=set of all possible candidates, X_i=matrix of the data of the i-th meaningful subset including K_i candidates, $\mathbb{X}_i$=the i-th meaningful subset of $\mathbb{X}$, $\mathbb{X}_*$=set of all $\mathbb{X}_i$'s, A=coefficient vector of $\mathbb{X}$, A_i=coefficient vector of $\mathbb{X}_i$. When $\mathbb{Y}=F(\mathbb{X})$ is loaded, Y, X, and X_i are generated and used for calculation.

The j-th Best Subset Problem

Find meaningful subset $\mathbb{X}_i \in \mathbb{X}_*$ derivable from set $\mathbb{X}$ of all possible candidates considered to affect a dependent variable $\mathbb{Y}$ and estimate $\bar{A}_i$ such that

(I) $\bar{A}_i$ of relation $Y=X_i\bar{A}_i+U$ where $U \sim N(0,\sigma^2 Q)$ for known positive definite matrix Q minimizes the following quadratic form:

$$(Y-X_i\bar{A}_i)'Q^{-1}(Y-X_i\bar{A}_i) \quad (17)$$

(II) $\bar{A}_i$ satisfies the following constraints, if imposed on $\bar{A}_i$:

$$B_{gi}\bar{A}_i=b_g \text{ for } 1\leqq g\leqq G \quad (18)$$

where B_{gi} and b_g stand for a row vector and a scalar of the g-th constraint imposed on $\bar{A}_i$, respectively,

(III) $\bar{A}_i$ satisfies the following magnitude conditions, if necessary:

$$C^1_{hi}\bar{A}_i\underline{+}|C^2_{hi}\bar{A}_i|\underline{+}|C^3_{hi}\bar{A}_i|\underline{>}c^1_h,\ C^1_{hi}\bar{A}_i\underline{+}|C^2_{hi}\bar{A}_i|\underline{+}|C^3_{hi}\bar{A}_i|\underline{<}c^2_h,$$

$$\text{or } c^1_h\underline{<}C^1_{hi}\bar{A}_i\underline{+}|C^2_{hi}\bar{A}_i|\underline{+}|C^3_{hi}\bar{A}_i|\underline{<}c^2_h \text{ for } 1\leqq h\leqq H \quad (19)$$

where C^k_{hi}, c^1_h, and c^2_h stand for a row vector, a lower bound, and an upper bound of the h-th magnitude condition applied for $\bar{A}_i$ for $1\leqq k\leqq 3$, respectively, $|C^k_{hi}\bar{A}_i|$ is the absolute value of $C^k_{hi}\bar{A}_i$ for k=2 and 3, "$\underline{>}$" and "$\underline{<}$" indicate ">" or "$\geqq$" and "<" or "$\leqq$", respectively, and "$\underline{+}$" indicates "+" or "-",

(IV) the following hypotheses about $\bar{A}_i$ are accepted at a specified significance level of a t-test or a specified t-value:

$$D_{pi}\bar{A}_i\neq d_p,\ D_{pi}\bar{A}_i>d_p,\ D_{pi}\bar{A}_i<d_p, \text{ or}$$

$$D_{pi}\bar{A}_i=d_p \text{ for } 1\leqq p\leqq P \quad (20)$$

where D_{pi} and d_p stand for a row vector and a scalar of the relationship of the p-th hypothesis performed for $\bar{A}_i$,

(V) the following estimated variance or scale factor $\bar{s}^2_i$ is significant at a specified significance level of a χ^2 test:

$$\bar{s}_i^2=\bar{E}_i'Q^{-1}\bar{E}_i/(NT-K_i+G) \qquad (21)$$

where $\bar{E}_i=Y-\bar{Y}_i=Y-X_i\bar{A}_i$,

(VI) the following Durbin-Watson statistic DW_i is significant at a specified significance level, when $N=1$, $T\geq2$, and $Q=I$ (identity matrix):

$$DW_i=\sum_{t=2}^{T}\{\bar{E}_{i1}(t)-\bar{E}_{i1}(t-1)\}^2/\sum_{t=1}^{T}\bar{E}_{i1}(t)^2 \qquad (22)$$

where $\bar{E}_{i1}(t)=Y_1(t)-\bar{Y}_{i1}(t)$ for $1\leq t\leq T$ [9], [10], [11],

(VII) $\bar{Y}_i$ satisfies the following absolute relative error test for specified w (%), if necessary:

$$100\times|\{Y_n(t)-\bar{Y}_n(t)\}/Y_n(t)|\leq w \text{ for } 1\leq t\leq T \text{ and } 1\leq n\leq N \qquad (23)$$

(VIII) $\bar{Y}_i$ satisfies the following turning point error test for specified u (%), if necessary:

If

$$\{Y_n(t)-Y_n(t-1)\}\{Y_n(t+1)-Y_n(t)\}<0 \qquad (24)$$

and

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

then (25)

$$\{Y_n(t)-Y_n(t-1)\}\{\bar{Y}_{in}(t)-\bar{Y}_{in}(t-1)\}>0 \qquad (26)$$

and

$$\{Y_n(t+1)-Y_n(t)\}\{\bar{Y}_{in}(t+1)-\bar{Y}_{in}(t)\}>0 \qquad (27)$$

for $2\leq t\leq T-1$, $T\geq3$, $1\leq n\leq N$, and $N\geq1$,

(IX) $\bar{A}_i$ shows the j-th best degree of goodness of fit.

If $Q=I$ and $G=0$, the above problem is for OLS. If $Q=I$ and $G\geq1$, the above problem is for COLS. If $Q\neq I$ and $G=0$, the above problem is for GLS. Finally, if $Q\neq I$ and $G\geq1$, the above problem is for CGLS.

7. Goodness of Fit

Various measures for goodness of fit have been proposed. They include a coefficient of determination adjusted for the degree of freedom, a Buse's coefficient of determination adjusted for the degree of freedom, the Akaike's information criterion, and the Schwarz's information criterion [4], [1], [29]. Each of the above measures has advantages and disadvantages. We would like to confine ourselves to these four measures, although there are other proposed ones. The adoption of the above measures depends on the nature of a problem. Accordingly, we would like to rewrite (IX) in the j-th Best Subset Problem with the introduction of each of the four measures.

In the case where an adjusted coefficient of determination is adopted as a measure of goodness of fit, (IX) can be rewritten as follows:

(IX) $\bar{A}_i$ shows the j-th highest adjusted coefficient of determination, RR_i, for OLS and COLS:

$$RR_i=\mathrm{Max}\{0,(1-R_i)(NT-1)/(NT-K_i+G)\} \quad (28)$$

where

$$R_i=1-\bar{E}_i'\bar{E}_i'/(Y-\bar{y}M)'(Y-\bar{y}M) \quad (29)$$

$\bar{y}=Y'M/NT$ and $M=(1,1,1,\ldots,1)'$ with NT dimension and G=0 unless constraints are imposed on $\bar{A}_i$.

In the case where the Buse's adjusted coefficient of determination is adopted as a measure, (IX) can be rewritten as follows:

(IX) $\bar{A}_i$ shows the j-th highest Buse's adjusted coefficient of determination, BRR_i, for GLS and CGLS:

$$BRR_i=\mathrm{Max}\{0,(1-BR_i)(NT-1)/(NT-K_i+G)\} \quad (30)$$

where

$$BR_i = 1 - \bar{E}_i' Q^{-1} \bar{E}_i / (Y - \bar{\bar{y}}M)' Q^{-1} (Y - \bar{\bar{y}}M) \qquad (31)$$

$$\bar{\bar{y}} = Y' Q^{-1} M / M' Q^{-1} M$$

and G=0 unless constraints are imposed on $\bar{A}_i$.

In the case where the Akaike's information criterion is adopted as a measure, (IX) can be rewritten as follows:

(IX) $\bar{A}_i$ shows the j-th lowest Akaike's information criterion value, AIC_i, for OLS, COLS, GLS, and CGLS:

$$AIC_i = NT\log 2 + NT\log \bar{E}_i' Q^{-1} \bar{E}_i / NT + NT + 2(K_i + 1 - G) \qquad (32)$$

where σ^2 is counted as an unknown in (32) and G=0 unless constraints are imposed on $\bar{A}_i$.

In the case where the Schwarz's information criterion is adopted as a measure, (IX) can be rewritten as follows:

(IX) $\bar{A}_i$ shows the j-th highest Schwarz's information criterion value, SIC_i, for OLS, COLS, GLS, and CGLS:

$$SIC_i = -\{NT\log 2 + NT\log \bar{E}_i' Q^{-1} \bar{E}_i / NT + NT + (K_i + 1 - G)\log K\}/2 \qquad (33)$$

where

K stands for the number of different candidates in $\bar{X}$, i.e., a fixed candidate is counted as one if it appears more than once in $\bar{X}$, and σ^2 is counted as an unknown in (33).

In package OEPP, the user can specify one of "RR", "BRR", "AIC", and "SIC" and, furthermore, a criterion value which is used as a tolerance level to accept a well-fitted equation. For instance, if he specifies "RR" and "0.85", a meaningful subset which satisfies the criteria adopted from (I) to (VIII) and has an adjusted coefficient of determination equal to or greater than 0.85 can be a solution for the j-th best subset problem. On the other hand, a meaningful subset which passes the criteria adopted from (I) to (VIII) but has an adjusted coefficient of determination less than 0.85 is regarded as unacceptable, because the goodness of fit is considered to be insufficient.

8. Derivation of Linear Relationships of Constraints, Magnitude Conditions, or Hypotheses

The elements of B_{gi} in (II), C^k_{hi} in (III), and D_{pi} in (IV) of the j-th Best Subset Problem depend on $\bar{A}_i$. Since it is quite troublesome to load all $B_{gi}\bar{A}_i$'s, $C^k_{hi}\bar{A}_i$'s, and $D_{pi}\bar{A}_i$'s, they must be derived from respective appropriate formats B_gA, C^k_hA, and D_pA. It is convenient to use $\bar{\bar{X}}$ instead of coefficients A, when $B_gA=b_g$, $C^1_hA\pm|C^2_h\bar{A}|\pm|C^3_h\bar{A}|\geq c^1_h$, etc., $D_pA\neq d_p$, etc. are loaded. Then, $B_g\bar{\bar{X}}=b_g$, $C^1_h\bar{\bar{X}}\pm ABS(C^2_h\bar{\bar{X}})$ $\pm ABS(C^3_h\bar{\bar{X}})\geq c^1_h$, and $D_p\bar{\bar{X}}\neq d_p$ are actually loaded as linear functions and generate $B_gA=b_g$, $C^1_hA\pm|C^2_hA|\pm|C^3_hA|\geq c^1_h$, and $D_pA\neq d_p$, respectively, where ABS(...) indicates an absolute value. In OEPP, only non-zero elements of B_g, C^k_h, and D_p and the corresponding candidates of $B_g\bar{\bar{X}}=b_g$, $C^1_h\bar{\bar{X}}\pm ABS(C^2_h\bar{\bar{X}})$ $\pm ABS(C^3_h\bar{\bar{X}})\geq c^1_h$, and $D_p\bar{\bar{X}}\neq d_h$ are loaded in linear forms, where "*" is needed between an element (number) and a candidate when the element is not either 1 or -1. It must be noted that $B_gA=b_g$, $C^1_hA\pm|C^2_hA|\pm|C^3_hA|\geq c^1_h$, and $D_pA\neq d_p$ may not make sense as a constraint, a magnitude condition, or a hypothesis, but $B_{gi}A_i=b_g$, $C^1_{hi}A_i\pm|C^2_{hi}\bar{A}_i|\pm|C^3_{hi}\bar{A}_i|\geq c^1_h$, and $D_{pi}A_i\neq d_p$ derived from $B_gA=b_g$, $C^1_hA\pm|C^2_hA|\pm|C^3_hA|\geq c^1_h$, and $D_pA\neq d_p$ must make sense as a constraint imposed on, a magnitude condition applied for, or a hypothesis performed for the coefficients $\bar{A}_i$ of the i-th meaningful subset $\bar{\bar{X}}_i$, respectively. In OEPP, scalars, b_g's, c_h^k's, and d_p's, of linear relationships in (II), (III), and (IV) are entered on the right-hand sides of the respective relationships to notify a computer of the ends of the relationships on (continued) cards.

Let us give a simple example of a magnitude condition. Suppose that the following conditions are required due to the stability condition of a dynamic model and a priori information that the inertia effect (whether the effect is positive or negative is not known a priori) of an eating habit tapers off, as time passes: (a) the coefficient, say, a_3 of candidate C(-1) must be smaller than 1 but greater than -1, (b) the coefficient, say, a_4 of candidate C(-2) must be smaller than a_3 of C(-1) in the absolute value, and (c) once C(-2) is used as a candidate, C(-1) must be included together with C(-2) in a meaningful subset, where candidate C stands for the current consumption and candidate C(-k) indicates candidate C with time lag number k for k=1 and 2. Candidates C(-1) and C(-2) can be regarded as gradually important. Let Y=C, V_3=C(-1), and V_4=C(-2) in functional format (3), where other V_k's are appropriate candidates.

The following two magnitude conditions (H=2) are required:

$$|\bar{a}_3|<1 \tag{34}$$

and

$$|\bar{a}_3|>|\bar{a}_4| \tag{35}$$

which is transposed as follows:

$$|\bar{a}_3|-|\bar{a}_4|>0 \text{ (or } -|\bar{a}_4|+|\bar{a}_3|>0\text{) or}$$

$$|\bar{a}_4|-|\bar{a}_3|<0 \text{ (or } -|\bar{a}_3|+|\bar{a}_4|<0\text{)} \tag{36}$$

(34) is loaded through

$$ABS(C(-1))<1, \quad -1<C(-1)<1, \text{ or } 1>C(-1)>-1 \qquad (37)$$

On the other hand, (36) is loaded through

$$ABS(C(-1))-ABS(C(-2))>0 \text{ (or } -ABS(C(-2))+ABS(C(-1))>0\text{) or}$$

$$ABS(C(-2))-ABS(C(-1))<0 \text{ (or } -ABS(C(-1))+ABS(C(-2))<0\text{)} \qquad (38)$$

If $ABS(C(-1))<1$ is loaded, then $C_1^1=(0,0,0,0,0,0)$, $C_1^2=(0,0,0,1,0,0)$, $C_1^3=(0,0,0,0,0,0)$, and $c_1^1=1$ for $\tilde{A}=(\tilde{a}_0,\tilde{a}_1,\tilde{a}_2,\tilde{a}_3,\tilde{a}_4,\tilde{a}_5)$ are recognized together with "<" by a computer. If $-1<C(-1)<1$ or $1>C(-1)>-1$ is loaded, then $C_1^1=(0,0,0,1,0,0)$, $C_1^2=(0,0,0,0, 0,0)$, $C_1^3=(0,0,0,0,0,0)$, $c_1^1=-1$, and $c_1^2=1$ for $\tilde{A}$ are recognized together with "<", "+" of "$+ABC(C_1^2\tilde{X})$", and "-" of "$-ABS(C_1^3\tilde{X})$" by a computer. If $ABS(C(-1))-ABS(C(-2))>0$ is loaded, then $C_2^1=(0,0,0,0,0,0)$, $C_2^2=(0,0,0,1,0,0)$, $C_2^3=(0,0,0,0,1,0)$, and $c_2^1=0$ for $\tilde{A}$ are recognized together with ">", "+" of "$+ABS(C_2^2\tilde{X})$", and "-" of "$-ABS(C_2^3\tilde{X})$" by a computer. It is assumed that $\tilde{a}_3=0$ and $\tilde{a}_4=0$ are refused by a t-test. However, if a t-test is not performed,

$$0<|\tilde{a}_3|<1 \text{ and } |\tilde{a}_3|>|\tilde{a}_4|>0 \qquad (39)$$

are loaded through

$$0<ABS(C(-1))<1 \text{ or } 1>ABS(C(-1))>0 \qquad (40)$$

and

$$ABS(C(-2))>0 \qquad (41)$$

together with (38) where H=3. When a meaningful subset clears the magnitude conditions (34) and (35) with a t-test or (39) with/without a t-test, such a subset is regarded as satisfactory with respect to this magnitude condition test.

Next, we would like to give a simple example of hypothesis testing. Suppose that the following hypothesis (P=1) is loaded when functional format (3) is loaded:

$$0.95*V_1+V_2-1.5*V_3-1.5*V_4\#1 \quad (42)$$

where "#" is used for "≠" in OEPP.

Let H_0^i and H_1^i stand for a null and an alternative hypothesis performed for the $\bar{A}_i$, respectively. Then, D_1 and d_1 corresponding to X are as follows:

$$D_1=(0,0.95,1,-1.5,-1.5,0) \quad \text{and} \quad d_1=1$$

The alternative hypothesis for the coefficients of (5) is as follows:

$$H_1^2: D_{12}\bar{A}_2=(0,0.95,1,-1.5,0)(\bar{a}_0,\bar{a}_1,\bar{a}_2,\bar{a}_3,\bar{a}_5)'\neq 1=d_1$$

or H_1^2: $0.95\bar{a}_1+\bar{a}_2-1.5\bar{a}_3\neq 1$. The null hypothesis H_0^2 is considered to be H_0^2: $0.95\bar{a}_1+\bar{a}_2-1.5\bar{a}_3=1$. A two-tailed t-test is automatically performed for the above hypotheses H_0^2 and H_1^2 at a significance level specified by the user. The alternative hypothesis H_1^3 of the third meaningful subset (6) is H_1^3: $0.95\bar{a}_1+\bar{a}_2-1.5\bar{a}_3-1.5\bar{a}_4\neq 1$, while the null hypothesis H_0^3 becomes H_0^3: $0.95\bar{a}_1+\bar{a}_2-1.5\bar{a}_3-1.5\bar{a}_4=1$.

If (43) is loaded instead of (42),

$$0.95*V_1+V_2-1.5*V_3-1.5*V_4>1 \quad (43)$$

the alternative hypothesis H_1^1 for the coefficients of, say, (4) is regarded as

$$H_1^1: D_{11}\bar{A}_1=(0,0.95,1,-1.5,-1.5,0)(\bar{a}_0,\bar{a}_1,\bar{a}_2,\bar{a}_3,\bar{a}_4,\bar{a}_5)'>1=d_1$$

or H_1^1: $0.95\bar{a}_1+\bar{a}_2-1.5\bar{a}_3-1.5\bar{a}_4>1$. The null hypothesis H_0^1 becomes H_0^1: $0.95\bar{a}_1+\bar{a}_2-1.5\bar{a}_3-1.5\bar{a}_4\leq 1$. A one-tailed t-test is performed for H_0^1 and H_1^1 at a specified significance level.

Thus, a meaningful subset which accepts all hypotheses derived for that subset is regarded as satisfactory with respective to hypothesis testing.

9. An Example

Let us demonstrate the proposed procedure for OLS and GLS methods by estimating a Cobb-Douglas production function of Japan's agriculture with data from 1965 to 1979. Three main factors of agricultural production are labor, capital, and land. Minor factors are intermediate goods (and services) and technical progress. Factor climate is taken into consideration through use of a dummy candidate of cold summer. We introduce the following candidates: LY=log(Y)=log(total outputs), LL=log(L)=log(labor), LKA=log(KA)=log(animal capital), LKP=log(KP)=log(plant capital), LKM=log(KM)=log(machinery capital), LK=log(KA+KP+KM) =log(total agricultural capital), LKMR=log(KMxR)=log(machinery capital adjusted by an estimated use rate), LKR=log(KA +KP+KMxR)=log(total capital used for agricultural production), LAX=log(A-X)=log(planted acreage minus abandoned and damaged acreage), LCAX=log((A-X)xMin(C,1))=log((A-X) adjusted by rice cropping index C, where average harvest C(t)=1, favorable harvest C(t)>1, and poor harvest C(t)<1), LT=log(T)=log(time trend for technical progress), LQ=log(Q)=log(intermedidate goods), and DV.CS=dummy variable for cold summer, where DV.CS=1 if summer is cold and DV.CS=0 otherwise.

The following functional format is appropriate for OLS method:

```
LY=F($C/LL/</LK,LKR(LKA,LKP,LKM)(LKA,LKP,LKMR)/>
     </LAX,LCAX/><*T,LT*>LQ,DV.CS)                          (44)
```

The criteria are as follows:

(i-1) sign of coefficient of constant term is free or undetermined, (i-2) sign of coefficient of dummy variable candidate DV.CS is negative, (i-3) signs of coefficients of all other candidates must be positive, (ii-1) coefficient(s) of a candidate (or candidates) representing the factor labor, capital, land, or intermediate goods must be greater than 0.1 but less than 0.5, (ii-2) the sum of coefficients of candidates representing the factors labor, capital, land, and intermedidate goods must be greater than 0.9 but less than 1.1 (avoiding unusually increasing or decreasing returns to scale of inputs), (iii) 5 % t-test for coefficients of all non-constant candidates, (iv) 5 % Durbin-Watson statistic test, (v) 5 % absolute relative error test, (vi) 1 % turning point error test, and (vii) minimum requirement of an adjusted coefficient of determination is 0.8.

The number of all meaningful subsets is $4 \times 2 \times (2+1) \times 2^2 = 96$, while that of all possible subsets is $2^{13}-1=8191$, because there are 13 different non-constant candidates in (44). These 96 meaningful subsets compose the feasible set in the domain of 8,191 possible subsets for estimation. If the (statistically as well as economically) best subset exists, it must be in this feasible set. It is quite difficult for a researcher to choose 96 meaningful subsets from 8,191 possible subsets and then find the best subset which satisfies the above criteria by himself, if all possible regressions are used. How can the stepwise regression procedure and similar procedures find the best subset in the domain of 8,191 subsets ?

We do not have any other criterion to evaluate all meaningful subsets, so that we specify parameter BEST=1.

Then, the following equation was obtained as the statistically as well as economically best in less than 22 seconds CPU time by FACOM M-200:

```
LY=0.5241875+0.3184192*LL+0.1568185*LKR+0.5828521*LCAX
  (0.380375) (2.530431)   (2.850482)    (3.570088)

  +0.0214367*T                                          (45)
   (3.205218)
```

R=0.9330, RR=0.9062, SD=0.019836, FA=-0.007405, DW=1.975

where numbers in parentheses, R, RR, SD, FA, and DW stand for t-ratios, a coefficient of determination, an adjusted coefficient of determination, standard deviation of a disturbance term, a first-order autocorrelation coefficient, and a Durbin-Watson statistic, respectively.

To demonstrate the consistency between the proposed procedures for OLS and GLS methods and the accuracy of package OEPP, we divide the dependent variable and all candidates used for OLS method by candidate LL and load the following functional format for GLS method:

```
LYLL=F($C/ILL/</LKLL,LKRLL(LKALL,LKPLL,LKMLL)(LKALL,LKPLL,
   LKMRLL)/></LAXLL,LCAXLL/><*TLL,LTLL*>LQLL,DV.CSLL)  (46)
```

where LYLL=LY/LL, ILL=1/LL, LKLL=LK/LL, LKRLL=LKR/LL, LKALL=LKA/LL, LKPLL=LKP/LL, LKMLL=LKM/LL, LKMRLL=LKMR/LL, LAXLL=LAX/LL, LCAXLL=LCAX/LL, TLL=T/LL, LTLL=LT/LL, LQLL=LQ/LL, and DV.CSLL=DV.CS/LL.

The diagonal elements of Q^{-1} are the reciprocals of squared LL(t) for all t. It must be noticed that the coefficients of $C and LL in (44) correspond to those of ILL and $C in (46). The same criteria as the case of OLS method are loaded for GLS method. Then, we can have the same equation as the best one obtained under OLS method in about 22 seconds CPU time except for rounding errors. Since goodness of fit is measured not by RR but by BRR, the

BRR for GLS method differs from the RR for OLS method.

$$LYLL=0.3184195+0.5241877*ILL+0.1568184*LKRLL$$
$$(2.530434)\quad(0.3803761)\quad(2.850482)$$

$$+0.5828519*LCAXLL+0.02143676*T \qquad (47)$$
$$(3.570102)\quad(3.205231)$$

BR=0.9773, BBR=0.9682, SD=0.019836

If $RR_i > RR_j$ for the i-th and j-th meaningful subsets in the above example, $BRR_i > BRR_j$ for the corresponding i and j. Since the procedures used for OLS and GLS methods yield exactly the same equation and confidence intervals (not shown in this paper), it is considered that the package OEPP has accuracy and convenience enough to be used for socio-economic analysis.

10. Conclusion

To solve a variable selection problem, especially for social sciences, the author formulated the j-th best subset problem and proposed a variable selection procedure to solve the best to the J-th best subset problem in one run of the computer. The proposed procedure puts emphasis on practical use and is available in computer package OEPP. This procedure saves much research time, labor, and other resources like paper and electricity, and improves the quality of research.

Acknowledgements

The author wishes to thank Professor K. Nakayama, Head of the Science Information Processing Center, University of Tsukuba, for offering me a terminal.

References

1. AKaike, H., Information Theory and an Extension of the Maximum Likelihood Principle, The Second International Symposium on Information Theory, Budapest (1973).

2. Allen, D.M., Mean Square Error of Prediction as a Criterion for Selecting Variables, Technometrics, 13, 469-475 (1971).

3. Beale, E.M.L. et al, The Discarding of Variables in Multivariate Analysis, Biometrika, 54, 357-365 (1967).

4. Buse, A., Goodness of Fit in Generalized Least Squares Estimation, The American Statistician, 27, 3, 106-108 (1973).

5. Daniel, C. and Wood, F.S., Fitting Equations to Data Wiley, New York (1971).

6. Dhrymes, P.J. et al, Criteria for Evaluation of Econometric Model, Annals of Economic and Social Measurements, 291-324 (1972).

7. Dixon, W.J. et al, BMDP, Health Sciences and Computing Faculty, Department of Biomathematics, School of Medicine, University of California (1981).

8. Draper, N.R. and Smith, H., Applied Regression Analysis Wiley, New York, 2nd ed. (1981)

9. Durbin, J. and Watson, G.S., Testing for Serial Correlation in Least Squares Regression I, Biometrika, 37, 409-428 (1950).

10. Durbin, J. and Watson, G.S., Testing for Correlation in Least Squares Regression II, Biometrika, 38, 159-178 (1951).

11. Durbin, J. and Watson, G.S., Testing for Serial Correlation in Least Squares Regression III, Biometrika, 58, 1-19 (1971).

12. Furnival, G.M., All Possible Regressions with Less Computation, Technometrics, 13, 403-408 (1971).

13. Garside, M.J., The Best Subset in Multiple Regression Analysis, Applied Statistics, 14, 196-200 (1965).

14. Gorman, J.W. and Toman, R.J., Selection of Variables for Fitting Equations, Technometrics, 8, 27-51 (1966).

15. Hocking, R.R. and Leslie, R.N., Selection of the Best Subset in Regression Analysis, Technometrics, 9, 531-540 (1967).

16. Hocking, R.R., The Analysis and Selection of Variables in Linear Regression, Biometrics, 32, 1-49 (1976).

17. Holwig, J.T. et al, SAS, SAS Institute, Raleigh, NC (1980).

18. Judge, G.G. et al, The Theory and Practice of Econometrics, Wiley, New York, (1980).

19. La Motte , L.R. and Hocking, R.R., Computational Efficiency in the Selection of Regression Variables, Techometrics, 12, 83-93 (1970).

20. Mallows, C.L., Some Comments on C_p, Technometrics, 15, 661-675 (1973).

21. Mantel, N., Why Stepdown Procedures in Variable Selection, Technometrics, 12, 621-625 (1970).

22. Nie, N.H. et al, SPSS, National Opinion Research Center, University of Chicago (1980).

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

24. Onishi, H., OEPP, Institute of Socio-Economic Planning, University of Tsukuba, Ibaraki, Japan (1982).

25. Ryan, T.A.J. et al, MINITAB-80 Statistics Department, Pennsylvania State University (1980).

26. Sawa, T., Selection of Variables in Regression Analysis, The Economic Studies Quarterly, 53-63 (1968).

27. Seber, G.A.F., Linear Regression Analysis, Wiley, New York (1977).

28. Schatzoff, M. et al, Efficient Calculation of All Possible Regressions, Technometrics, 10, 769-779 (1968).

29. Schwarz, G., Estimating the Dimension of a Model, The Annals of Statistics, 6, 2, 461-465 (1978).

30. Weisberg, S., Applied Linear Regression, Wiley, New York (1980).

31. Wherry, R.J., A New Formula for Predicting the Shrinkage of the Coefficient of Multiple Correlation, Occupational Counseling Techniques 440-451 (1940).