No. 963

An Informatic and Computational Variable Selection Method for Almon Distributed Lag Regression

by

Haruo Onishi

December 2001

An Informatic and Computational Variable Selection Method for Almon Distributed Lag Regression

Haruo ONISHI*

Institute of Policy and Planning Sciences
University of Tsukuba, Tsukuba, Ibaraki, JAPAN

Abstract: A regression equation may include many consecutively lagged variables with/without other explanatory variables. The explanatory variables may cause multicollieaircity or the number of the explanatory variables may exceed the sample size. In this case, the Almon distributed lag regressionmay be useful. Although all possible regressions, the forward selection, backward elimination, stepwise regression, mini-max regret principle methods and others have been studied in the literature, a variable selection method for the Almon distributed lag regression has not been studied much. As a variable selection problem for the Almon distributed lag regression (ADLR), the j-th ADLR-best subset problem is proposed and how to solve it is shown. The Intellectual Statistical System **OEPP** is developed to solve the first j ADLR-best subset problems in a run of a computer.

Keywords: Almon distributed lag regression; Scientific variable classification; Grouped variable; Combinatorial variable; Sequential variable; Meaningful subset; Variable selection problem; J-th best subset problem; Practically best regression equation; Intellectual Statistical System OEPP.

1 Introduction

Now that the technology, hard and soft, concerned with a computer has matured, statistical science can make not superficial but positive contributions to societies and the world through substantial sciences. The contributions are possible only when

*Correspondent: Professor Dr. Haruo ONISHI, Institute of Policy and Planning Sciences, University of Tsukuba, Tsukuba, JAPAN 305-8573. E-mail: onishi@sk.tsukuba.ac.jp, Tel: 0298-53-5008, Fax: 0298-55-3849. 23:20, December 2 (Sunday), 2001. File olspaper.tex.

statistical theories are well interwoven with informatic and computational methods to handle the professional knowledge related to research in question.

In some economic research, many consecutively lagged variables appear in a regression equation. They may cause multicollinearity or the number of them and other variables, if any, may exceed the sample size. In this circumstance, it is impossible to apply regression analysis. S. Almon (1965) proposed a method to reduce the number of regression coefficients to be estimated. Part of the changes of a dependent variable over time is explained by a set of current and consecutively lagged data of a factor and the remaining is explained by the other factors. She put a decisive assumption that the regression coefficients of these variables may be approximated by a polynomial of some appropriate degree on the basis of the Weierstrass's theorem, which states that a continuous function in a closed interval may be approximated by a polynomial of an appropriate degree, and successfully reduced the number of regression coefficients to be estimated. It is better to obtain some statistical results than to do nothing. If a user, i.e., an applied researcher, knows accurately all variables which statistically explain the movement or behavior of a dependent variable and has a reliable data set, then he can easily estimate the regression equation with the Almon distributed lag regression, abbreviated as ADLR hereafter, and proceed to analysis, prediction and/or policy simulation (scenario) by using the estimated regression equation. However, in most actual regression analyses, a user does not know a priori such a specification and may not have a precise data set. It is quite difficult to know a priori a functional form between a dependent variable and explanatory variables, define the correct data of some explanatory variable and determine the most appropriate proxy variable if the data of a variable are unavailable. These occurrences force a user to search for the scientifically reasonable and statistically best regression equation somewhat through trial and error.

Some efficient algorithms to calculate all possible regressions in a run of a computer have been developed in the literature. If the number of variables is, for instance, 20, then 1,048,575 regression equations are estimated and printed in the computer output, because $2^{20} - 1$=1048575 where 1 is for an empty subset. A user has to obtain a heavy and 10-meter high printout and select the best among 1,048,575 regression equations. It is impossible or quite difficult. All possible regressions are not so useful as expected.

The forward selection, backward elimination, stepwise regression, mini-max regret principle methods and others have been proposed to search for the best regression equation in a run of a computer. Unfortunately, these conventional variable selection methods often select different best regression equations, especially if the number of explanatory variables exceeds 5. It is quite strange that the best regression equation for a dependent variable depends on a variable selection method employed. Also, they are not useful for ADLR.

As a result, a user adds a new variable to or removes another from the initial set of explanatory variables by paying close attention to each new set of explanatory variables in such a way that no contradiction occurs, loads a new set of explanatory variables with the dependent variable into a computer and then evaluates the estimated regression equation through scientific, statistical and data-analytic criteria. This process is repeated at least until he obtains a scientifically reasonable and statistically satisfactory regression equation. This primitive one-at-a-time procedure is quite laborious, time-consuming, costly and resource-wasting and does not take advantage of the rapid improvement of cost performance of a computer in contrast with the increasing price of output paper. Unless all intended regression equation candidates are estimated and evaluated through scientific, statistical and data-analytic criteria, the quality of regression analysis may be lowered.

A crucial defect common to these conventional variable selection methods is that they cannot take advantage of the professional knowledge needed to solve a variable selection problem in the research at hand. Statistics does not surpass all other sciences but is independent of them and just as important. A variable selection problem can be solved by statistics on the basis of the professional knowledge related to research at hand. Unless professional knowledge needed can be generally handled and processed by a computer, a variable selection problem cannot be solved by the computer. Fortunately, it is possible to make a computer recognize and process professional knowledge, whatever the research, loaded by a user. If a software system possesses not only a knowledge database of professional knowledge but also (a database of) statistical and data-analytic criteria, it is called a statistical expert system. So far, professional knowledge which can be used for any kind of research cannot be stored in a software system except where research is narrowly limited. It is better to incorporate user's professional knowledge regarding the research at hand into a software system and utilize it together with statistical and data-analytic tests. Such a software system is not an expert system but can be called an intellectual system. The author developed the *Intellectual Statistical System* **OEPP**[1] in which a user-knowledge-based variable selection method for ADLR is available.

2 Almon Distributed Lag Regression

To simplify the explanation about ADLR, we here omit a constant term and non-ADLR explanatory variables and deal with only one kind of ADLR explanatory variables (e.g., current and lagged interest rates for capital formation).

Let X=current variable and X_{-k} = k-time-lagged variable of X. Suppose that current and consecutively lagged variables, X, X_{-1}, X_{-2}, $\cdots$, X_{-K}, are explanatory variables for the dependent variable Y and the regression equation is expressed as

[1] The Intellectual Statistical System **OEPP** consists of the main program and about 550 subroutines written in FORTRAN 77 and leading to about 90,000 lines in total and is at present available for FACOM and IBM-compatible main frame machines, workstations and personal computers. An 8-hour-a-day work of estimation by a primitive procedure can be done in less than one second of CPU time at a cost of half a US dollar by a notebook-type PC.

follows:

$$Y = a_0X + a_1X_{-1} + a_2X_{-2} + \cdots + a_KX_{-K} + U = \boldsymbol{X}A + U \tag{2.1}$$

where Y=dependent variable or $(T \times 1)$-vector of its data; T=sample size of time series data; X=current explanatory variable; K=far-end time lag number with $K > 0$; $X_{-k} = k$-time-lagged variable of X or $(T \times 1)$-vector of its time series data for $k = 0, 1, 2, \cdots, K$ with $X \equiv X_{-0}$; a_k=true regression coefficient of X_{-k}; U=disturbance term or $(T \times 1)$-vector of its disturbances with $\mathcal{E}(U) = \mathbf{0}_T$ and $\mathcal{V}(U) = \sigma^2\mathbf{I}_T$; $\mathbf{0}_T = (T \times 1)$-zero vector; $\mathbf{I}_T = (T \times T)$-unit-matrix; $\boldsymbol{X} = (K+1)$-variate set $\{X, X_{-1}, X_{-2}, \cdots, X_{-K}\}$ or $\{T \times (K+1)\}$-matrix of their data; and $A = (a_0, a_1, a_2, \cdots, a_K)'$.

Suppose that (2.1) cannot be directly estimated by OLS (ordinary least squares), because $|\boldsymbol{X}'\boldsymbol{X}|$=0 or $K + 1 > T$. We assume that regression coefficient a_k is expressed as the following polynomial of degree P with respect to k:

$$a_k = b_0 + b_1k + b_2k^2 + \cdots + b_Pk^P \quad \text{for all } k = 0, 1, 2, \cdots, K \tag{2.2}$$

where P=integer, satisfying $1 \leq P < \min\{K, T\} - 1$ and b_p =parameter to be estimated of a polynomial for $p = 0, 1, 2, \cdots, P$.

The explanatory variables in the regression equation (2.1) satisfying (2.2) are here called Almon variables. We can express (2.2) in a matrix form as

$$a_k = W_kB \quad \text{for all } k \tag{2.3}$$

where $W_k = (1, k, k^2, \cdots, k^P)$ and $B = (b_0, b_1, b_2, \cdots, b_P)'$. Then we have

$$\begin{pmatrix} a_0 \\ a_1 \\ a_2 \\ a_3 \\ \vdots \\ a_K \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & \cdots & 0 \\ 1 & 1 & 1 & 1 & 1 & \cdots & 1 \\ 1 & 2 & 2^2 & 2^3 & 2^4 & \cdots & 2^P \\ 1 & 3 & 3^2 & 3^3 & 3^4 & \cdots & 3^P \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & K & K^2 & K^3 & K^4 & \cdots & K^P \end{pmatrix} \begin{pmatrix} b_0 \\ b_1 \\ b_2 \\ b_3 \\ \vdots \\ b_P \end{pmatrix} \tag{2.4}$$

and express (2.4) as

$$A = \boldsymbol{W}B \tag{2.5}$$

where $\boldsymbol{W} = (W_0', W_1', W_2', \cdots, W_K')' = \{(K+1) \times (P+1)\}$-matrix of fixed elements and $B = (b_0, b_1, b_2, \cdots, b_P)'$. Substituting (2.5) into (2.1), we can rewrite (2.1) as

$$Y = ZB + U \tag{2.6}$$

where

$$\boldsymbol{Z} = \boldsymbol{X}\boldsymbol{W}. \tag{2.7}$$

If $|\boldsymbol{Z}'\boldsymbol{Z}| \neq 0$, B may be estimated as

$$\hat{B} = (Z'Z)^{-1}Z'Y \tag{2.8}$$

which leads to $\mathcal{E}(\hat{B}) = B$ and $\mathcal{V}(\hat{B}) = \sigma^2(\boldsymbol{Z}'\boldsymbol{Z})^{-1}$. The regression coefficient vector, A, may be estimated as

$$\hat{A} = \boldsymbol{W}\hat{B}. \tag{2.9}$$

$\hat{A}$ has the following statistical properties: $\mathcal{E}(\hat{A}) = \boldsymbol{W}B = A$ and $\mathcal{V}(\hat{A}) = \sigma^2\boldsymbol{W}(\boldsymbol{Z}'\boldsymbol{Z})^{-1}\boldsymbol{W}'$.

An assumption that variable $X_{-(K+1)}$ is not included in (2.1) may require $a_{K+1}=0$. If we regard it as a constraint, we have the following constraint imposed on B by substituting $K+1$ into k in (2.2):

$$b_0 + (K+1)b_1 + (K+1)^2 b_2 + \cdots + (K+1)^P b_P = 0. \tag{2.10}$$

On the other hand, constraint $\sum_{k=0}^{K} a_k=1$, implying constant returns to scale over time, yields the following constraint imposed on B:

$$(K+1)b_0 + (\sum_{k=1}^{K} k)b_1 + (\sum_{k=1}^{K} k^2)b_2 + \cdots + (\sum_{k=1}^{K} k^P)b_P = 1. \tag{2.11}$$

Let X not be included in (2.1) so that $Y = a_1X_{-1}+a_2X_{-2}+a_3X_{-3}+\cdots+a_KX_{-K}+U$ is estimated. We have $a_0 = 0$ which lead to $b_0 = 0$. Constraints like (2.10) and (2.11) can be rewritten as a constraint set $\boldsymbol{C}B = \boldsymbol{c}$.

In actual research, one or more sets of Almon variables with/without non-Almon variables appear in a regression equation. Let X^ℓ be the ℓ-th Almon current variable for all ℓ. A general form of ADLR may be expressed as follows:

$$Y = a_0 + \sum_{k=1}^{K^0} a_k^0 X_k^0 + \sum_{\ell=1}^{L}\sum_{k_\ell=0}^{K^\ell} a_{k_\ell}^\ell X_{-k_\ell}^\ell + U \tag{2.12}$$

where L=number of kinds (or groups) of Almon variables; X_k^0,a_k^0 = k-th non-Almon variable and its true regression coefficient for $k = 1, 2, \cdots, K^0$, respectively, if $K^0 > 0$; and $X_{-k_\ell}^\ell$,$a_{k_\ell}^\ell$ = k_ℓ-time-lagged variable of the current X^ℓ in the ℓ-th Almon variable group and its true regression coefficient for $k_\ell = 0, 1, 2, \cdots, K^\ell$ with $X_{-0}^\ell \equiv X^\ell$, respectively. For instance, if Y=capital, then ℓ=1 may be "an interest rate", ℓ=2 "economic growth rate", ℓ=3 "retained earnings", and so on.

3 A Variable Selection Problem for ADLR

We focus on the selection of the practically best regression equation when all possible Almon and non-Almon variables are given for a dependent variable. Such a regression equation has been obtained through a trial and error process in most research. A user may not necessarily have perfect professional knowledge and precise data needed to find such a regression equation but may possess fairly sufficient professional knowledge and considerably reliable data. Making allowance for this situation, we formulate the j-th ADLR-best subset problem as a variable selection problem for ADLR in which a user can specify an appropriate positive integer for

j which depends on the degree of the certainty and confidence in his professional, statistical and data-analytic knowledge about the research. If he has sufficient professional knowledge, precise data and rich experience in model building and knows appropriate significance levels for statistical tests and data-analytic criteria, then j=1 can be specified and a solution to the (first) ADLR-best subset problem becomes the practically best regression equation. He should otherwise specify 4 or 5 for j, continuously solve the first 4 or 5 ADLR-best subset problems in a run of a computer and then select the ultimately best among the solutions by comparing them with each other or using his own new criterion.

3.1 Notation

We introduce the following notation:

T ≡sample size of time series data where $t = 1, 2, \cdots, T$ for $T > 1$;
Y ≡dependent variable or $(T \times 1)$-vector of its data y_t's, i.e., $Y = (y_1, y_2, \cdots, y_T)'$;
y ≡average of Y, i.e., $y = \sum_{t=1}^{T} y_t/T$;
L ≡number of sets (or kinds or groups) of Almon variables;
ℓ ≡set number of Almon variables, where $\ell = 1, 2, \cdots, L$ for $L \geq 1$;
ℓ_* ≡number assigned to a set of all possible non-Almon variables, like a time trend variable or dummy variables for seasons (quarterly data), oil crises, cold summer and liberalization of a foreign exchange rate, as $\ell_* = 0$ or sets of all possible Almon variables as all ℓ_* like $\ell_* = 1$ (or $\ell = 1$) for a set of the Almon variables for the current and lagged interest rates, $\ell_* = 2$ (or $\ell = 2$) for a set of the Almon variables for the current and lagged economic growth rates and $\ell_* = 3$ (or $\ell = 3$) for a set of the Almon variables for the current and lagged retained earnings in capital formation where $\ell_* = 0, 1, 2, \cdots, L$;
K^0 ≡number of all possible non-Almon variables not including a constant term for $K^0 \geq 0$ where $K^0 = 0$, unless any non-constant non-Almon variables are used;
$K^\ell, \underline{K}^\ell$ ≡possible farthest and nearest end-points of the time lag of current Almon variable X^ℓ, respectively, where K^ℓ and $\underline{K}^\ell$ are specified by a user and an optimal far-end point is determined among $\kappa_\ell = \underline{K}^\ell, \underline{K}^\ell + 1, \underline{K}^\ell + 2, \cdots, K^\ell$ for $1 < \underline{K}^\ell \leq K^\ell$;
$K \equiv K^0 + \sum_{\ell=1}^{L}(K^\ell + 1)$ =number of all possible explanatory nonconstant variables so that $2^K - 1$ possible nonempty subsets exist;
$i \equiv 1, 2, 3, \cdots, 2^K - 1$ =number assigned to each of all possible subsets or submatrices of a set or a data matrix $\boldsymbol{X}$ of all possible explanatory variables;
X_0 ≡constant term or $(T \times 1)$-vector of its data 1's, i.e., $X_0 = (1, 1, \cdots, 1)'$;
$X_k^0 \equiv k$-th possible non-Almon variable or $X_k^0 \equiv (x_{1k}^0, x_{2k}^0, \cdots, x_{Tk}^0)' = (T \times 1)$-vector of its data x_{tk}^0's for $k = 1, 2, \cdots, K^0$ if $K^0 \geq 1$ and X_k^0 must be ignored, if $K^0 = 0$;
X^ℓ ≡current Almon variable of the ℓ-th set of Almon variables or $X^\ell \equiv (x_1^\ell, x_2^\ell, \cdots, x_T^\ell)'$

$= (T \times 1)$-vector of its data x_t^ℓ's for all ℓ;

$X_{-k^\ell}^\ell$ ≡possible k^ℓ-time-lagged Almon variable of X^ℓ or $X_{-k^\ell}^\ell \equiv (x_{1-k^\ell}^\ell, x_{2-k^\ell}^\ell, \cdots, x_{-1}^\ell, x_0^\ell, x_1^\ell, \cdots, x_{T-k^\ell}^\ell)' = (T \times 1)$-vector of its data $x_{t-k^\ell}^\ell$'s for $k^\ell = 1, 2, \cdots, K^\ell$ for all ℓ;

$\boldsymbol{X}^0$ ≡set $\{X_0, X_1^0, X_2^0, \cdots, X_{K^0}^0\}$ or data matrix $(X_0, X_1^0, X_2^0, \cdots, X_{K^0}^0)$ of all possible non-Almon variables, including a constant term;

K_i^0 ≡number of non-constant non-Almon variables in $\boldsymbol{X}_i^0$ for all i where $K_i^0 \geq 0$ for all i;

$M_i^0 \equiv K_i^0 + 1$ for some i;

$\boldsymbol{X}^\ell \equiv \ell$-th Almon variable set $\{X^\ell, X_{-1}^\ell, X_{-2}^\ell, \cdots, X_{-\underline{K}^\ell}^\ell, \cdots, X_{-K^\ell}^\ell\}$ or data matrix $(X^\ell, X_{-1}^\ell, X_{-2}^\ell, \cdots, X_{-\underline{K}^\ell}^\ell, \cdots, X_{-K^\ell}^\ell)$ for all ℓ;

U ≡disturbance term or $(T \times 1)$-vector of disturbances u_t's, i.e., $U = (u_1, u_2, \cdots, u_T)'$;

$\boldsymbol{X}$ ≡set $\{\boldsymbol{X}^0, \boldsymbol{X}^1, \boldsymbol{X}^2, \cdots, \boldsymbol{X}^L\}$ or data matrix $(\boldsymbol{X}^0, \boldsymbol{X}^1, \boldsymbol{X}^2, \cdots, \boldsymbol{X}^L)$ of all possible explanatory variables, including a constant term, for the dependent variable Y;

$\boldsymbol{X}_i \equiv i$-th subset of $\boldsymbol{X}$ or i-submatrix of $\boldsymbol{X}$ for all i, where if $\boldsymbol{X}_i^\ell \neq \emptyset$ for all ℓ and some i, then $\boldsymbol{X}_i = \{\boldsymbol{X}_i^0, \boldsymbol{X}_i^1, \boldsymbol{X}_i^2, \cdots, \boldsymbol{X}_i^L\}$ or $\boldsymbol{X}_i = (\boldsymbol{X}_i^0, \boldsymbol{X}_i^1, \boldsymbol{X}_i^2, \cdots, \boldsymbol{X}_i^L)$;

K_i^ℓ ≡far-end point of $\boldsymbol{X}_i^\ell$ for all ℓ and all i for $\underline{K}^\ell \leq K_i^\ell \leq K^\ell$;

$M_i^\ell \equiv K_i^\ell + 1$ for all ℓ and some i when all $\boldsymbol{X}_i^\ell$'s for all ℓ are meaningful subsets where $M_i^{\ell_*} = K_i^{\ell_*} + 1$ for all ℓ_*;

$\boldsymbol{X}_i^\ell$ =meaningful and ALDR-applicable M_i^ℓ-subset of $\boldsymbol{X}^\ell$ for all ℓ and some i, then $\boldsymbol{X}_i^\ell = \{X^\ell, X_{-1}^\ell, X_{-2}^\ell, \cdots, X_{-K_i^\ell}^\ell\} = (T \times M_i^\ell)$-subset of $\boldsymbol{X}^\ell$ or $\boldsymbol{X}_i^\ell = (X^\ell, X_{-1}^\ell, X_{-2}^\ell, \cdots, X_{-K_i^\ell}^\ell) = (T \times M_i^\ell)$-submatrix of $\boldsymbol{X}^\ell$ for $\underline{K}^\ell \leq K_i^\ell \leq K^\ell$;

$M_i \equiv \sum_{\ell_*=0}^L M_i^{\ell_*}$ =number of all variables in $\boldsymbol{X}_i$, including a constant term;

$A_i^{\ell_*} \equiv (a_i^{\ell_* 0}, a_i^{\ell_* 1}, a_i^{\ell_* 2}, \cdots, a_i^{\ell_* K_i^{\ell_*}})' = (M_i^{\ell_*} \times 1)$-vector of true regression coefficients of $\boldsymbol{X}_i^0$ or meaningful $\boldsymbol{X}_i^\ell$ for all ℓ and some i;

$A_i \equiv (A_i^{0\prime}, A_i^{1\prime}, A_i^{2\prime}, \cdots, A_i^{L\prime})'$ of $\boldsymbol{X}_i$ when $\boldsymbol{X}_i^\ell$'s for all ℓ are meaningful for some i;

P_ℓ, Q_ℓ ≡possible smallest and largest numbers of degrees of a polynomial used for the variables in the ℓ-th Almon variable set, respectively, and specified by a user where $1 \leq P_\ell \leq Q_\ell < \underline{K}^\ell$;

r_ℓ ≡number assigned to a polynomial of degrees p_ℓ applied to the variables in the ℓ-th Almon variable set where $r_\ell = 1, 2, \cdots, Q_\ell - P_\ell + 1$ for all ℓ;

$R \equiv \Pi_{\ell=1}^L (Q_\ell - P_\ell + 1)$ =number of all possible permutations of the degrees of L polynomials applied to L Almon-variable sets;

p_ℓ ≡possible degree of the r_ℓ-th polynomial candidate for meaningful $\boldsymbol{X}_i^\ell$ where $p_\ell = P_\ell, P_\ell + 1, \cdots, Q_\ell$ for all ℓ so that $p_\ell = P_\ell + r_\ell - 1$;

$W_{k_i^\ell r_\ell}^{\ell p_\ell} \equiv (1, k_i^\ell, (k_i^\ell)^2, (k_i^\ell)^3, \cdots, (k_i^\ell)^{p_\ell}) = \{1 \times (p_\ell + 1)\}$-vector for $k_i^\ell = 0, 1, 2, \cdots, K_i^\ell$, all ℓ and some i so that $W_0^{\ell p_\ell} = (1, 0, 0, \cdots, 0)$, $W_1^{\ell p_\ell} = (1, 1, 1, \cdots, 1)$, $W_2^{\ell p_\ell} = (1, 2, 4, \cdots, 2^{p_\ell})$, $W_3^{\ell p_\ell} = (1, 3, 9, \cdots, 3^{p_\ell})$, etc. for all ℓ and some i;

$\boldsymbol{W}_{ir_\ell}^\ell \equiv (W_{0r_\ell}^{\ell p_\ell \prime}, W_{1r_\ell}^{\ell p_\ell \prime}, W_{2r_\ell}^{\ell p_\ell \prime}, \cdots, W_{K_i^\ell r_\ell}^{\ell p_\ell \prime})' = \{M_i^\ell \times (p_\ell + 1)\}$-conversion matrix for meaningful $\boldsymbol{X}_i^\ell$ for all r_ℓ, all ℓ and some i;

$\boldsymbol{Z}_{ir_\ell}^\ell \equiv \{Z_{ir_\ell}^{\ell 0}, Z_{ir_\ell}^{\ell 1}, Z_{ir_\ell}^{\ell 2}, \cdots, Z_{ir_\ell}^{\ell p_\ell}\} = (p_\ell + 1)$-set of new converted variables or $\boldsymbol{Z}_{ir_\ell}^\ell \equiv (Z_{ir_\ell}^{\ell 0}, Z_{ir_\ell}^{\ell 1}, Z_{ir_\ell}^{\ell 2}, \cdots, Z_{ir_\ell}^{\ell p_\ell}) = \{T \times (p_\ell + 1)\}$-matrix converted from meaningful $\boldsymbol{X}_i^\ell$ through the r_ℓ-th polynomial of degrees p_ℓ, i.e., $\boldsymbol{Z}_{ir_\ell}^\ell = \boldsymbol{X}_i^\ell \boldsymbol{W}_{ir_\ell}^\ell$ for all r_ℓ, all ℓ and some i where $\boldsymbol{Z}_{ir_\ell}^\ell$ is here called the converted Almon variables from the ℓ-th Almon variable
group;

$r \equiv$number assigned to the polynomial-index set $\{r_1, r_2, \cdots, r_L\}$ from which the corresponding set of degrees of polynomials is derived as $\{p_1, p_2, \cdots, p_L\}$, i.e., $\{r_1 + P_1 - 1, r_2 + P_2 - 1, \cdots, r_L + P_L - 1\}$ where $r = 1, 2, \cdots, R$ so that (i) $r = 1$ corresponds to the polynomial-index set $\{1, 1, \cdots, 1\}$ and the polynomial-degree set $\{P_1, P_2, \cdots, P_L\}$; (ii) $r = 2$ corresponds to the polynomial-index set $\{2, 1, \cdots, 1\}$ and the polynomial-degree set $\{P_1 + 1, P_2, \cdots, P_L\}$, $\cdots$; (iii) $r = Q_1$ corresponds to the polynomial-index set $\{R_1, 1, \cdots, 1\}$ and the polynomial-degree set $\{Q_1, P_2, \cdots, P_L\}$; (iv) $r = Q_1 + 1$ corresponds to the polynomial-index set $\{1, 2, \cdots, 1\}$ and the polynomial-degree set $\{P_1, P_2 + 1, \cdots, P_L\}$, $\cdots\cdots$; (v) $r = Q_1 + 2$ corresponds to the polynomial-index set $\{2, 2, \cdots, 1\}$ and the polynomial-degree set $\{P_1 + 1, P_2 + 1, \cdots, P_L\}$, $\cdots\cdots$; (vi) $r = R$ corresponds to the polynomial-index set $\{R_1, R_2, \cdots, R_L\}$ and the polynomial-degree set $\{Q_1, Q_2, \cdots, Q_L\}$;

$K_{ir} \equiv M_i^0 + \sum_{\ell=1}^L (p_\ell + 1)$ =number of all non-Almon and converted Almon variables related to $\boldsymbol{X}_i$ for all r and some i where all $\boldsymbol{X}_i^\ell$'s for all ℓ are meaningful;

$\boldsymbol{W}_{ir} \equiv (M_i \times K_{ir})$-matrix which consists of block-diagonal matrices $\mathbf{I}_{M_i^0}$, $\boldsymbol{W}_{ir_1}$, $\boldsymbol{W}_{ir_2}$, $\cdots$, $\boldsymbol{W}_{ir_L}$ and all zero off-diagonal matrices where it is assumed that r corresponds to not only the polynomial-degree set $\{p_1, p_2, \cdots, p_L\}$ but also the polynomial-index set $\{r_1, r_2, \cdots, r_L\}$ for all r and some i where all $\boldsymbol{X}_i^\ell$'s for all ℓ are meaningful (see Subsection 3.2);

$\boldsymbol{Z}_{ir} = K_{ir}$ new converted variables or $\boldsymbol{Z}_{ir} \equiv r$-th $(T \times K_{ir})$-matrix converted from $\boldsymbol{X}_i$ through the r-th polynomial-degree set $\{p_1, p_2, \cdots, p_L\}$, i.e., $\boldsymbol{Z}_{ir} = \boldsymbol{X}_i \boldsymbol{W}_{ir} = (\boldsymbol{X}_i^0, \boldsymbol{Z}_{ir_1}^1, \boldsymbol{Z}_{ir_2}^2, \cdots, \boldsymbol{Z}_{ir_L}^L)$ for all r and some i when all $\boldsymbol{X}_i^\ell$'s for all ℓ are meaningful;

$A_{ir}^0 \equiv (M_i^0 \times 1)$-vector of regression coefficients of $\boldsymbol{X}_i^0$ of $\boldsymbol{Z}_{ir}$ for all r and all i;

$A_{ir_\ell}^\ell \equiv \boldsymbol{W}_{ir_\ell}^\ell B_{ir_\ell}^\ell = (M_i^\ell \times 1)$-vector of regression coefficients of $\boldsymbol{X}_i^\ell$ of $\boldsymbol{Z}_{ir}$ for all r and all i;

$B_{ir_\ell}^\ell \equiv (b_{ir_\ell}^{\ell 0}, b_{ir_\ell}^{\ell 1}, b_{ir_\ell}^{\ell 2}, \cdots, b_{ir_\ell}^{\ell p_\ell})' = (p_\ell + 1)$-vector of regression coefficients of $\boldsymbol{Z}_{ir_\ell}^\ell$ of $\boldsymbol{Z}_{ir}$ for all r_ℓ, all ℓ and some i where $A_{ir_\ell}^\ell = \boldsymbol{W}_{ir_\ell}^\ell B_{ir_\ell}^\ell$;

$B_{ir} \equiv (A_{ir}^{0\prime}, B_{ir_1}^{1\prime}, B_{ir_2}^{2\prime}, \cdots, B_{ir_L}^{L\prime})' = (K_{ir} \times 1)$-vector of true regression coefficients of $\boldsymbol{Z}_{ir}$ for all r and some i;

$N_i^0 \equiv$rank$(\boldsymbol{C}_i^0)$ of constraint set $\boldsymbol{C}_i^0 A_i^0 = \boldsymbol{c}_i^0$ where it is assumed that N_i^0 =0 if no constraint is imposed on A_i^0 for all i;

$N_i^\ell \equiv$rank$(\boldsymbol{C}_{ir_\ell}^\ell)$ of constraint set $\boldsymbol{C}_{ir_\ell}^\ell B_{ir_\ell}^\ell = \boldsymbol{c}_i^\ell$ where it is assumed that $N_{ir_\ell}^\ell = 0$ if no constraint is imposed on $B_{ir_\ell}^\ell$ for all i;

$N_i \equiv$number of all constraints imposed on A_i, i.e., $N_i = \sum_{\ell_*=0}^L N_i^{\ell_*}$ for all i;

$\boldsymbol{C}_{ir}$ ≡matrix whose diagonal submatrices are $\boldsymbol{C}_i^0$, $\boldsymbol{C}_{ir_1}^1$, $\boldsymbol{C}_{ir_2}^2$, $\cdots$, $\boldsymbol{C}_{ir_L}^L$ and all off-diagonal submatrices are zero matrices where $\boldsymbol{C}_i^0$ and/or $\boldsymbol{C}_{ir_\ell}^\ell$ must be eliminated with the corresponding rows if the constraints are not imposed on A_i^0 and/or $B_{ir_\ell}^\ell$ (see Subsection 3.2);

$c_i^{\ell_*} \equiv (N_i^{\ell_*} \times 1)$-vector of scalars (constraint values) for all ℓ_* and i;

$\hat{A}_{ir}^0 \equiv (\hat{a}_{ir}^{00}, \hat{a}_{ir}^{01}, \hat{a}_{ir}^{02}, \cdots, \hat{a}_{ir}^{0K_i^{\ell_*}})'$ =estimate of A_i^0 based on $\boldsymbol{Z}_{ir}$ for all i;

$\hat{A}_{ir_\ell}^\ell \equiv (\hat{a}_{ir_\ell}^{\ell 0}, \hat{a}_{ir_\ell}^{\ell 1}, \hat{a}_{ir_\ell}^{\ell 2}, \cdots, \hat{a}_{ir_\ell}^{\ell K_i^\ell})'$ =estimate of A_i^ℓ based on $\boldsymbol{Z}_{ir}$ for all ℓ and i;

$\hat{A}_{ir} \equiv (\hat{A}_{ir_\ell}^{0\prime}, \hat{A}_{ir_\ell}^{1\prime}, \hat{A}_{ir_\ell}^{2\prime}, \cdots, \hat{A}_{ir_\ell}^{L\prime})'$ =estimate of A_i based on $\boldsymbol{Z}_{ir}$ for all r and i;

$\hat{s}_{ir}^{\ell_* m_i^\ell}$ ≡standard deviation (or error) of $\hat{a}_{ir}^{\ell_* m_i^\ell}$ for all ℓ_* and m_i^ℓ;

$m_i^\ell = 0, 1, 2, \cdots, M_i - 1$;

$\hat{t}_{ir}^{\ell_* m_i^\ell} \equiv t$-ratio of $\hat{a}_{ir}^{\ell_* m_i^\ell}$ for all ℓ_* and m_i^ℓ

$\hat{y}_t^{ir}$ ≡(partial-test) estimate of y_t based on $\boldsymbol{Z}_{ir}$ for all r and i;

$\hat{Y}_{ir} \equiv (\hat{y}_1^{ir}, \hat{y}_2^{ir}, \cdots, \hat{y}_T^{ir})'$ =(partial-test) estimate of Y by $\boldsymbol{Z}_{ir}$ for all r and i;

$\hat{e}_t^{ir} \equiv y_t - \hat{y}_t^{ir}$ =(partial-test) residual from y_t based on $\boldsymbol{Z}_{ir}$ for all r and i;

$\hat{E}_{ir} \equiv (\hat{e}_1^{ir}, \hat{e}_2^{ir}, \cdots, \hat{e}_T^{ir})' = (T \times 1)$-vector of (partial-test) residuals $\hat{e}_t^{ir}$'s;

$T_{ir}^* \equiv T - K_{ir} + N_i$ =degrees of freedom for all r and i;

$\mathbf{I}_n \equiv (n \times n)$-identity-matrix;

$\mathbf{0}_n \equiv (n \times 1)$-zero-vector;

$\mathbf{O}_{mn} \equiv (m \times n)$-zero-matrix;

σ^2, σ ≡unknown variance and standard deviation of the disturbance u_t, respectively;

$\hat{\sigma}_{ir}^2, \hat{\sigma}_{ir}$ ≐estimates of σ^2 and σ based on $\boldsymbol{Z}_{ir}$ for all r and i, respectively;

α_h^1 ≡a priori known lower bound of the h-th sign or magnitude condition about regression coefficients;

α_h^2 ≡a priori known upper bound of the h-th sign or magnitude condition about regression coefficients;

β ≡percentage of a significance level for a one-tailed or two-tailed t-test ($0 < \beta < 0.25$);

γ ≡percentage of a significance level for the Durbin-Watson test (γ =0, 0.01 or 0.05 at present);

ν ≡percentage of a significance level for a two-tailed t-test for detecting a residual outlier ($0 < \nu < 0.25$);

η ≡percentage of a significance level for the Jarque-Bera normality test ($0 < \eta < 0.25$);

ε ≡criterion value of a standardized residual test;

ζ_1 ≡definition value (%) of a turning point for $y_t \neq 0$;

ζ_2 ≡definition value of a turning point for $y_t = 0$;

θ ≡minimum requirement of an adjusted coefficient of determination.

3.2 Assumptions

(1) A user must have fairly sufficient knowledge about his research.

(2) If there are L kinds of Almon variable groups, a subset becomes meaningless unless any variables are selected from each of L groups. If $\boldsymbol{X}_i^\ell$ is $\{X^\ell, X_{-1}^\ell, X_{-2}^\ell, \cdots, X_{-K_i^\ell}^\ell\}$, it is called meaningful and ADLR-applicable when $\underline{K}^\ell \leq K_i^\ell \leq \overline{K}^\ell$.

(3) If regression coefficient $a_i^{\ell k_i^\ell}$ of A_i^ℓ is estimated with $\boldsymbol{Z}_{ir}$, it is represented by a polynomial of degree p_ℓ in the r-th polynomial-degree set $\{p_1, p_2, \cdots, p_\ell, \cdots, p_L\}$ as follows:

$$\begin{aligned} a_{ir_\ell}^{\ell k_i^\ell} &= b_{ir_\ell}^{\ell 0} + b_{ir_\ell}^{\ell 1} k_i^\ell + b_{ir_\ell}^{\ell 2} (k_i^\ell)^2 + \cdots + b_{ir_\ell}^{\ell p_\ell} (k_i^\ell)^{p_\ell} \\ &= W_{k_i^\ell r_\ell}^{\ell p_\ell} B_{k_i^\ell r_\ell}^{\ell p_\ell} \quad \text{for } k_i^\ell = 0, 1, 2, \cdots, K_i^\ell, \text{ all } \ell \text{ and } i \end{aligned} \tag{3.13}$$

(4) OLS cannot be directly used, because
$|\boldsymbol{X}_i'\boldsymbol{X}_i| = 0$ or $M_i - N_i > T$ for all i
but the following holds:
$|\boldsymbol{X}_i^{0\prime}\boldsymbol{X}_i^0| \neq 0$ or $T > M_i^0 + \sum_{\ell=1}^{L}(Q_\ell + 1)$ for at least some i.

(5) The disturbance term U is normally distributed as $U \sim N(0_T, \sigma^2 \mathrm{I}_T)$.

(6) $|\boldsymbol{Z}_{ir}'\boldsymbol{Z}_{ir}| \neq 0$ for at least some i.

and

(7) $\boldsymbol{X}_i$ is nonstochastic or independent of U if stochastic for all i.

If the i-th subset $\boldsymbol{X}_i$ is a meaningful subset[2] for the research at hand, it can be expressed as the following regression equation:

$$Y = \boldsymbol{X}_i A_i + U = \boldsymbol{X}_i^0 A_i^0 + \boldsymbol{X}_i^1 A_i^1 + \boldsymbol{X}_i^2 A_i^2 + \cdots + \boldsymbol{X}_i^L A_i^L + U \tag{3.14}$$

which is estimated with R possible polynomial-degree sets. Thus, R regression equation candidates are estimated for (3.14). Concretely writing $\boldsymbol{W}_{ir}$, $\boldsymbol{C}_{ir}$ and $\boldsymbol{c}_i$, we have with the polynomial-index set $\{r_1, r_2, \cdots, r_L\}$ and the polynomial-degree set $\{p_1, p_2, \cdots, p_L\}$ which correspond to r with $r_\ell = p_\ell - P_\ell + 1$ for $p_\ell = P_\ell, P_\ell + 1, \cdots, Q_\ell$

[2] A meaningful subset was originally defined in H. Onishi (1983). A meaningful subset is defined as a subset which includes all necessary explanatory variables but does not include any unnecessary, redundant or contradiction-causing explanatory variables so that it deserves to be estimated and evaluated as a candidate for the best regression equation. Thus, it corresponds to a regression equation in the primitive one-regression-equation-at-a-time procedure where a user repeatedly loads into a computer a set of variables by carefully adding a new variable(s) to or removing an already-entered variable(s) from the set. On the other hand, a subset which does not include at least one of the necessary explanatory variables or includes at least one unnecessary, redundant or contradiction-causing explanatory variable is called a meaningless subset. In the primitive procedure, a wise user never loads any meaningless sets into a computer and estimate them.

$$\boldsymbol{W}_{ir}=\begin{pmatrix} \mathbf{I}_{M_i^0} & \mathbf{O}_{M_i^0,p_1+1} & \mathbf{O}_{M_i^0,p_2+1} & \cdots & \mathbf{O}_{M_i^0,p_L+1} \\ \mathbf{O}_{M_i^1,M_i^0} & \boldsymbol{W}_{ir_1} & \mathbf{O}_{M_i^1,p_2+1} & \cdots & \mathbf{O}_{M_i^1,p_L+1} \\ \mathbf{O}_{M_i^2,M_i^0} & \mathbf{O}_{M_i^2,p_1+1} & \boldsymbol{W}_{ir_2} & \cdots & \mathbf{O}_{M_i^2,p_L+1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{O}_{M_i^L,M_i^0} & \mathbf{O}_{M_i^L,p_1+1} & \mathbf{O}_{M_i^L,p_2+1} & \cdots & \boldsymbol{W}_{ir_L} \end{pmatrix}; \quad B_{ir}=\begin{pmatrix} A_{ir}^0 \\ A_{ir}^1 \\ A_{ir}^2 \\ \vdots \\ A_{ir}^L \end{pmatrix};$$

$$\boldsymbol{C}_{ir}=\begin{pmatrix} \boldsymbol{C}_i^0 & \mathbf{O}_{N_i^0,p_1+1} & \mathbf{O}_{N_i^0,p_2+1} & \cdots & \mathbf{O}_{N_i^0,p_L+1} \\ \mathbf{O}_{N_i^1L_i^0} & \boldsymbol{C}_{ir_1}^1 & \mathbf{O}_{N_i^1,p_2+1} & \cdots & \mathbf{O}_{N_i^1,p_L+1} \\ \mathbf{O}_{N_i^2M_i^0} & \mathbf{O}_{N_i^2,p_1+1} & \boldsymbol{C}_{ir_2}^2 & \cdots & \mathbf{O}_{N_i^2,p_L+1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{O}_{N_i^LM_i^0} & \mathbf{O}_{N_i^L,p_1+1} & \mathbf{O}_{N_i^L,p_2+1} & \cdots & \boldsymbol{C}_{ir_L}^L \end{pmatrix}; \quad \boldsymbol{c}_i=\begin{pmatrix} \boldsymbol{c}_i^0 \\ \boldsymbol{c}_i^1 \\ \boldsymbol{c}_i^2 \\ \vdots \\ \boldsymbol{c}_i^L \end{pmatrix};$$

where if constraint sets $\boldsymbol{C}_i^0 A_i^0 = \boldsymbol{c}_i^0$ and/or $\boldsymbol{C}_{ir_\ell}^\ell B_{ir_\ell}^\ell = \boldsymbol{c}_i^\ell$ are not imposed on A_i^0 and/or $B_{ir_\ell}^\ell$, respectively, the corresponding rows in $\boldsymbol{C}_{ir}$ and vectors in $\boldsymbol{c}_i$ must be eliminated.

If (3.14) is estimated by ADLR with the r-th polynomial-degree set $\{p_1, p_2, \cdots, p_L\}$ which corresponds to the polynomial-index set $\{r_1, r_2, \cdots, r_L\}$, it is converted into

$$\begin{aligned} Y &= \boldsymbol{X}_i^0 A_{ir}^0 + \boldsymbol{Z}_{ir_1}^1 B_{ir_1}^1 + \boldsymbol{Z}_{ir_2}^2 B_{ir_2}^2 + \cdots + \boldsymbol{Z}_{ir_L}^L B_{ir_L}^L + U \\ &= \boldsymbol{Z}_{ir} B_{ir} + U \end{aligned} \tag{3.15}$$

which is actually estimated.

3.3 The J-th ADLR-Best Subset Problem

We concretely formulate the j-th ADLR-best subset problem[3] as a variable selection problem for ADLR as follows:

The J-th ADLR-Best Subset Problem

Search for subset $\boldsymbol{X}_i$, i.e., $\{\boldsymbol{X}_i^0, \boldsymbol{X}_i^1, \boldsymbol{X}_i^2, \cdots, \boldsymbol{X}_i^L\}$, from set $\boldsymbol{X}$ of all possible non-Almon and Almon variables specified for the dependent variable Y, determine the optimal far-end points $\{K_{ir}^1, K_{ir}^2, K_{ir}^3, \cdots, K_{ir}^L\}$ of $\boldsymbol{X}_i$ and the optimal polynomial-degrees $\{p_1, p_2, p_3, \cdots, p_L\}$ which corresponds to r, and estimate $\hat{A}_{ir}$ with their standard errors $\hat{s}_{ir}^{\ell_* k_i^{\ell_*}}$'s and t-ratios $\hat{t}_{ir}^{\ell_* k_i^{\ell_*}}$'s, variance $\hat{\sigma}_{ir}^2$ and standard error $\hat{\sigma}_{ir}$ of the disturbance term U and some other important statistics under the criterion set $\{\alpha_h^1\text{'s}, \alpha_h^2\text{'s}, \beta, \gamma, \eta, \varepsilon, \zeta_1, \zeta_2, \theta\}$ specified by a user in a run of a computer such that

[3] For simplicity, a case of lagged dependent variables is not referred to here. Hence, the Schur stability condition, the Durbin h-test for serial correlation and final test are not incorporated in the j-th ADLR-best subset problem, although the System **OEPP** can handle them.

[I] X_i^0 is a meaningful subset for the research at hand and X_i^ℓ is meaningful and ADLR-applicable for the research, i.e., $X_i^\ell = \{X^\ell, X_{-1}^\ell, X_{-2}^\ell, \cdots, X_{-K_i^\ell}^\ell\}$ for all ℓ;

[II] $\hat{A}_{ir}$, $\hat{Y}_{ir}$, $\hat{E}_{ir}$, $\hat{V}_{ir}$, $\hat{\sigma}_{ir}^2$, $\hat{\sigma}_{ir}$, $\hat{s}_{ir}^{\ell_* k_i^{\ell_*}}$'s, $\hat{t}_{ir}^{\ell_* k_i^{\ell_*}}$'s are estimated as follows:

$$\hat{Y}_{ir} = X_i \hat{A}_{ir}, \quad \hat{E}_{ir} = Y - \hat{Y}_{ir}, \quad \hat{\sigma}_{ir}^2 = \frac{\hat{E}_{ir}' \hat{E}_{ir}}{T_{ir}^*}, \quad \hat{\sigma}_{ir} = \sqrt{\hat{\sigma}_{ir}^2}, \quad \hat{s}_{ir}^{\ell_* k_i^{\ell_*}} = \sqrt{\hat{V}_{ir}^{k_* k_*}}$$

$$\text{and } \hat{t}_{ir}^{\ell_* k_i^{\ell_*}} = \frac{\hat{a}_{ir}^{\ell_* k_i^{\ell_*}}}{\hat{s}_{ir}^{\ell_* k_i^{\ell_*}}} \text{ for } k_i^{\ell_*} = 0, 1, 2, \cdots, K_i^{\ell_*} \text{ and } k_* = \sum_{n=1}^{\ell-1} L_i^n + k_i^{\ell_*} + 1$$

$$\text{for } \ell_* = 0, 1, 2, \cdots, L \tag{3.16}$$

(i) in an unconstrained case ($N_i = 0$)

$$\hat{A}_{ir} = W_{ir}(Z_{ir}' Z_{ir})^{-1} Z_{ir}' Y \tag{3.17}$$

with the estimated covariance matrix

$$\hat{V}_{ir} = \hat{\sigma}_{ir}^2 W_{ir}(Z_{ir}' Z_{ir})^{-1} W_{ir}' \tag{3.18}$$

(ii) in a constrained case of $C_{ir} A_i = c_i$ ($N_i > 0$)

$$\begin{aligned} \hat{A}_{ir} &= W_{ir}(Z_{ir}' Z_{ir})^{-1}[Z_{ir}' Y + C_{ir}'\{C_{ir}(Z_{ir}' Z_{ir})^{-1} C_{ir}'\}^{-1} \\ &\quad \times \{c_i - C_{ir}(Z_{ir}' Z_{ir})^{-1} Z_{ir}' Y\}] \end{aligned} \tag{3.19}$$

with the estimated covariance matrix

$$\begin{aligned} \hat{V}_{ir} &= \hat{\sigma}_{ir}^2 W_{ir}[(Z_{ir}' Z_{ir})^{-1} - (Z_{ir}' Z_{ir})^{-1} C_{ir}'\{C_{ir}(Z_{ir}' Z_{ir})^{-1} C_{ir}'\}^{-1} \\ &\quad \times C_{ir}(Z_{ir}' Z_{ir})^{-1}] W_{ir}' \end{aligned} \tag{3.20}$$

where $\hat{V}_{ir}^{k_* k_*}$ = (k_*, k_*)-diagonal element of $\hat{V}_{ir}$;

[III] $\hat{A}_{ir}$ must satisfy the following sign and/or magnitude conditions if necessary from the viewpoints of the professional knowledge related to the research in question: for $h_i = 1, 2, \cdots, H_i$

$$f_{h_i}(\hat{A}_{ir}) \geq \alpha_{h_i}^1, \quad f_{h_i}(\hat{A}_{ir}) \leq \alpha_{h_i}^2 \quad \text{or} \quad \alpha_{h_i}^1 \leq f_{h_i}(\hat{A}_{ir}) \leq \alpha_{h_i}^2 \tag{3.21}$$

where $f_{h_i}(\hat{A}_{ir})$ =function of $\hat{A}_{ir}$ which is linear in most cases but has absolute values and square roots in some cases;

[**IV**] the following Jarque-Bera normality test statistic $\widehat{\mathcal{JB}}_{ir}$ must hold to maintain the null hypothesis $\mathrm{H}_0 : U$ is normally distributed with the expectation $\mathcal{E}(U) = \mathbf{0}_T$ at a significance level 100η %[4] :

$$\widehat{\mathcal{JB}}_{ir} = T\Big(\frac{\widehat{\mathcal{S}}^2_{ir}}{6} + \frac{(\widehat{\mathcal{K}}_{ir} - 3)^2}{24}\Big) \leq \chi^2_2(\eta) \tag{3.22}$$

for

$$\widehat{\mathcal{S}}^2_{ir} = \frac{\Big\{\sum_{t=1}^{T}(\hat{e}^{ir}_t)^3/T\Big\}^2}{\Big\{\sum_{t=1}^{T}(\hat{e}^{ir}_t)^2/T\Big\}^3} \quad \text{and} \quad \widehat{\mathcal{K}}_{ir} = \frac{\sum_{t=1}^{T}(\hat{e}^{ir}_t)^4/T}{\Big\{\sum_{t=1}^{T}(\hat{e}^{ir}_t)^2/T\Big\}^2}$$

where $\chi^2_2(\eta) = \eta$ percentile of a χ^2-distributon with 2 degrees of freedom,

[**V**] the following inequality must be satisfied to adopt the specified alternative hypothesis H_1 or maintain the specified null hypothesis H_0, depending on the purpose of the research[5] , if necessary: for $d_i = 1, 2, \cdots, D_i$

(i) if H_1 must be adopted for $\mathrm{H}_0 : G_{id_i}A_i = g_{d_i}$ against $\mathrm{H}_1 : G_{id_i}A_i \neq g_{d_i}$,

$$\frac{|G_{id_i}\hat{A}_{ir} - g_{d_i}|}{\hat{S}^{d_i}_{ir}} > t_{T^*_{ir}}(\beta/2) \tag{3.23}$$

(ii) if H_1 must be adopted for $\mathrm{H}_0 : G_{id_i}A_i = g_{d_i}$ against $\mathrm{H}_1 : G_{id_i}A_i > g_{d_i}$,

$$\frac{G_{id_i}\hat{A}_{ir} - g_{d_i}}{\hat{S}^{d_i}_{ir}} > t_{T^*_{ir}}(\beta) \tag{3.24}$$

(iii) if H_1 must be adopted for $\mathrm{H}_0 : G_{id_i}A_i = g_{d_i}$ against $\mathrm{H}_1 : G_{id_i}A_i < g_{d_i}$,

$$\frac{g_{id_i} - G_{id_i}\hat{A}_{ir}}{\hat{S}^{d_i}_{ir}} > t_{T^*_{ir}}(\beta) \tag{3.25}$$

or

(vi) if H_0 must be maintained for $\mathrm{H}_0 : G_{id_i}A_i = g_{d_i}$ against $\mathrm{H}_1 : G_{id_i}A_i \neq g_{d_i}$,

$$\frac{|G_{id_i}\hat{A}_{ir} - g_{d_i}|}{\hat{S}^{d_i}_{ir}} \leq t_{T^*_{ir}}(\beta/2) \tag{3.26}$$

---

[4] The Shapiro-Wilk normality test is alternatively available.

[5] It is possible to specify a different significance level (% or a critical value) for each hypothesis test in the System **OEPP**. If an alternative hypothesis ($\neq$ is represented as #) is loaded into a computer, it is required that it be adopted. On the other hand, if a null hypothesis is loaded into a computer, it is required that it be maintained. Concretely speaking, $G_{id_i}A_i \neq g_{d_i}$ is derived from aggregate linear hypothetical relation $G_dA \neq g_d$ which is loaded through $G_d\boldsymbol{X}\#g_d$ into a computer and converted into $G_{id_i}\boldsymbol{W}_{ir}B_{ir}$ leading to $G_{id_i}\boldsymbol{W}_{ir}\hat{B}_{ir}$ which is equal to $G_{id_i}\hat{A}_{ir}$.

for

$$(\hat{S}_{ir}^{d_i})^2 = G_{id_i}\hat{V}_{ir}G'_{id_i} \text{ and } \hat{S}_{ir}^{d_i} = \sqrt{(\hat{S}_{ir}^{d_i})^2} \tag{3.27}$$

where $G_{id_i} = (1 \times \sum_{\ell=0}^{L} M_i^{\ell})$-vector of known coefficients of A_i in the d_i-th hypothetical relation for the i-th subset; g_{d_i} =known value of the d_i-th hypothetical relation; and $t_{T_{ir}^*}(\beta/2)$ and $t_{T_{ir}^*}(\beta)$ denote the critical points for 100β % two-tailed and one-tailed t-tests with T_{ir}^*, i.e., $T - K_{ir} + N_i$ degrees of freedom, respectively;

[VI] (1) the Durbin-Watson serial correlation statistic, $\widehat{DW}_{ir}$, defined below must satisfy the following inequality, if annual time series data (m =1) or quarterly time series data (m =4) are used and no lagged dependent variables are in $\boldsymbol{X}_i$ (Durbin and Watson (1950,1951)):

$$\widehat{DW}_{ir} > d_{ir}(\gamma) \quad \text{if } \widehat{DW}_{ir} \leq 2 \quad \text{or} \quad 4-\widehat{DW}_{ir} > d_{ir}(\gamma) \quad \text{if } \widehat{DW}_{ir} > 2 \tag{3.28}$$

for

$$\widehat{DW}_{ir} = \frac{\sum_{t=1+m}^{T}(\hat{e}_t^{ir} - \hat{e}_{t-m}^{ir})^2}{\sum_{t=1}^{T}(\hat{e}_t^{ir})^2} \tag{3.29}$$

where $d_{ir}(\gamma) = d_{T,K_{ir}-N_i}^{u}(\gamma)$ if he subjectively treats an inconclusive case as unacceptable or $d_{ir}(\gamma) = d_{T,K_{ir}-N_i}^{\ell}(\gamma)$ if he subjectively treats an inconclusive case as acceptable, $d_{T,K_{ir}-N_i}^{u}(\gamma)$ =upper limit of a 100γ % significance level of the Durbin-Watson serial correlation test with $(T, K_{ir}-N_i)$ degrees of freedom, $d_{T,K_{ir}-N_i}^{\ell}(\gamma)$ = lower limit;

[VII] $\hat{E}_{ir}$ must satisfy the following outlier test[2] [3] (Sawa, 1979), if necessary:

$$\max_t \widehat{OT}_t^{ir} \leq t_{T_{ir}^*-1}(\nu/2T) \quad \text{for some } t = 1, 2, \cdots, T \tag{3.30}$$

for

$$\widehat{OT}_t^{ir} = \frac{|\hat{e}_t^{ir}|/\sqrt{1-\hat{q}_t^{ir}}}{\sqrt{\hat{E}_{ir}'\hat{E}_{ir} - (\hat{e}_t^{ir})^2/(1-\hat{q}_t^{ir})}/\sqrt{T_{ir}^* - 1}}$$

[2] $\hat{E}_{ir} \sim \mathcal{N}(\mathbf{0}_T, \sigma^2\{\mathrm{I}_T - Z_{ir}(Z_{ir}'Z_{ir})^{-1}Z_{ir}'\})$ in an unconstrained case or $\hat{E}_{ir} \sim \mathcal{N}(\mathbf{0}_T, \sigma^2 [\mathrm{I}_T - Z_{ir}(Z_{ir}'Z_{ir})^{-1}Z_{ir}' + Z_{ir}(Z_{ir}'Z_{ir})^{-1}C_{ir}'\{C_{ir}(Z_{ir}'Z_{ir})^{-1}C_{ir}'\}^{-1}C_{ir}(Z_{ir}' \times Z_{ir})^{-1}Z_{ir}'])$ in a constrained case. It turns out that $\hat{\hat{e}}_t^{ir} \equiv \hat{e}_t^{ir}/(\sigma\sqrt{1-\hat{q}_t^{ir}}) \sim \mathcal{N}(0,1)$. As $T \to +\infty$, $\nu/2T \to 0$. Hence, the larger the sample size, the weaker the outlier test. It may be possible to set a criterion value ξ to define badlly-fitted estimates of Y as $|y_t - \hat{y}_t^{ir}| > \xi$'s and use the number T^* of badlly-fitted estimates instead of T and the critical point $t_{T^*-K_{ir}+N_i+1}(\nu/2T^*)$. For large T, one may use a standardized residual test.

[3] The following standardized residual test for $\hat{\hat{e}}_t^{ir}$ can be used, if necessary, by defining $\hat{\hat{e}}_t^{ir} \equiv \hat{e}_t^{ir}/(\hat{\sigma}_{ir}\sqrt{1-\hat{q}_t^{ir}})$: $\max_t |\hat{\hat{e}}_t^{ir}| \leq \varepsilon$ for $t = 1, 2, \cdots, T$ with respect to the criterion value ε specified by a researcher where $\hat{\hat{e}}_t^{ir} \sim \mathcal{N}(0,1)$; $\Pr\{|\hat{\hat{e}}_t^{ir}| \leq 1\} \doteq 0.6827$; $\Pr\{|\hat{\hat{e}}_t^{ir}| \leq 2\} \doteq 0.9545$; and $\Pr\{|\hat{\hat{e}}_t^{ir}| \leq 3\} \doteq 0.9983$.

and/or

(*ii*) all standardized residuals $\hat{\hat{e}}^i_{st}$'s defined below must not exceed the user-specified criterion value ε[6] , if necessary[7] :

$$\max_{s,t} |\hat{\hat{e}}^i_{st}| \leq \varepsilon \quad \text{for some } s = 1, 2, \cdots, S \text{ and } t = 1, 2, \cdots, T$$

for

$$\hat{\hat{e}}^i_{st} \equiv \frac{\hat{e}^i_{st}}{\hat{\sigma}_i \sqrt{1 - \hat{q}^i_{pp}}}$$

where $t_{T^*_{ir}-1}(\nu/2T)$=$50\nu/T$ percentile of a t-test with $T^*_{ir}-1$, i.e., $T-K_{ir}+N_i-1$ degrees of freedom and $\hat{q}^{ir}_t$=t-th diagonal element of $\boldsymbol{Z}_{ir}(\boldsymbol{Z}'_{ir}\boldsymbol{Z}_{ir})^{-1}\boldsymbol{Z}'_{ir}$ in an unconstrained case or $\boldsymbol{Z}_{ir}(\boldsymbol{Z}'_{ir}\boldsymbol{Z}_{ir})^{-1}\boldsymbol{Z}'_{ir} - \boldsymbol{Z}_{ir}(\boldsymbol{Z}'_{ir}\boldsymbol{Z}_{ir})^{-1}\boldsymbol{C}'_{ir}\{\boldsymbol{C}_{ir}(\boldsymbol{Z}'_{ir}\boldsymbol{Z}_{ir})^{-1} \times\boldsymbol{C}'_{ir}\}^{-1}\boldsymbol{C}_{ir}(\boldsymbol{Z}'_{ir}\boldsymbol{Z}_{ir})^{-1}\boldsymbol{Z}'_{ir}$ in a constrained case;

[VIII] $\hat{Y}_{ir}$ must satisfy the following turning point test, when time series data are used and at least one lagged dependent variable is included in $\boldsymbol{X}_i$ and if it is necessary: for $t = 1+n, 2+n, \cdots, T-n$ and $n = 1, 2, \cdots, t^*_i$,
if

$$(y_{t+n} - y_t)(y_t - y_{t-n}) < 0 \tag{3.31}$$

and

$$100 \times \min\Big\{\Big|1 - \frac{y_{t-n}}{y_t}\Big|, \Big|1 - \frac{y_{t+n}}{y_t}\Big|\Big\} \geq \zeta_1 \quad \text{for } y_t \neq 0 \tag{3.32}$$

or

$$\min\{|y_{t-n}|, |y_{t+n}|\} \geq \zeta_2 \quad \text{for } y_t = 0, \tag{3.33}$$

then

$$(y_t - y_{t-n})(\hat{y}^{ir}_t - \hat{y}^{ir}_{t-n}) > 0 \tag{3.34}$$

and

$$(y_{t+n} - y_t)(\hat{y}^{ir}_{t+n} - \hat{y}^{ir}_t) > 0 \tag{3.35}$$

where t^*_i =maximum time lag number among the time lag numbers of all lagged dependent variables in $\boldsymbol{X}_i$;

and

[IX] the following fitting criterion must be met:

the adjusted coefficient of determination $(\widehat{\mathcal{R}}^r_i)^2$ defined below is greater than

[6] $\hat{\hat{e}}^i_{st} \sim \mathcal{N}(0,1)$. $\Pr\{|\hat{\hat{e}}^i_{st}| \leq 1\} \doteq 0.6827$; $\Pr\{|\hat{\hat{e}}^i_{st}| \leq 1.6449\} \doteq 0.9000$; $\Pr\{|\hat{\hat{e}}^i_{st}| \leq 1.9600\} \doteq 0.9500$; $\Pr\{|\hat{\hat{e}}^i_{st}| \leq 2\} \doteq 0.9545$; and $\Pr\{|\hat{\hat{e}}^i_{st}| \leq 3\} \doteq 0.9983$.

[7] In regression analysis, an outlier test to identify an outlier in Y before estimation is less important than an outlier test through residuals. If an outlier-like y_{st} is well tracked with outlier-like $x^{k_i i}_{s,t-\tau_{k_i}}$'s, it is not regarded as an outlier.

or equal to θ and the j-th highest among the adjusted coefficients of determination of the subsets which satisfy the above conditions [**I**] to [**VIII**] and are greater than or equal to θ^6 :

$$(\widehat{\mathcal{R}}_{ir})^2 = \max\left[1 - \frac{\{1-(\hat{R}_{ir})^2\}(T-1)}{T^*_{ir}}, 0\right] \geq \theta \tag{3.36}$$

for

$$(\hat{R}_{ir})^2 = 1 - \frac{\widehat{E}'_{ir}\widehat{E}_{ir}}{(Y-yX_0)'(Y-yX_0)}.\square$$

A criterion set $\{\alpha_h^1$'s,α_h^2's$\}$ depends on the professional knowledge related to the research in question. There are no optimal criteria for $\{\beta,\gamma,\eta,\varepsilon,\nu,\theta\}$. Conventionally used criteria are usually set for them. For example, β =0.05 (5 %) or β =0.1 (10 %), γ =0.05 (5 %), η =0.1 (10 %), ε =1.65, ν =0.1 (10 %), θ =0.6. A data-analytic criterion set $\{\zeta_1,\zeta_2\}$ depends on the data of a dependent variable. If the data of a dependent variable are expressed as ratios, rates or percentages, the above criterion for θ may be reduced, since empirical studies indicate that it is not easy to obtain high $\hat{\mathcal{R}}^2$. No solution may exist if criteria are very severe. It may be impossible to reverse a priori known sign condition or alter lower and upper bounds of a magnitude condition except where a new fact is found and correct the established professional knowledge. If a user is uncertain about a sign or magnitude condition, he should not use it. If a sign condition (positive or negative) is applied, a one-tailed t-test is automatically applied. A two-tailed t-test is applied, otherwise. The System **OEPP** can evaluate all regression equations and optionally print with respect to each of all regression equations all reasons or only the first reason if a regression equation is unsatisfactory. If no solution existed, a diagnosis is printed whether all regression equations failed to pass all sign and/or magnitude conditions or some regression equations passed them but failed to pass statistical and/or data-analytic criteria. In the latter case, it will be printed that a solution exists if weak statistical and data-analytic criteria are reset and the job is rerun. It should be kept in mind that a solution to the j-th ADLR-best subset problem is the practically j-th best regression equation in the sense that (i) various scientific, statistical and data-analytic tests must be applied, (ii) there exists no overall evaluation function which incorporates all scientific, statistical and data-analytic tests, (iii) an inconclusive case often occurs in the Durbin-Watson serial correlation test, (vi) no optimal criterion set exists at present and (v) the solution depends on a criterion set specified by a user.

[6] $\widehat{AIC}_{ir} = T\{\ln 2\pi + 1 + \ln(\widehat{E}'_{ir}\widehat{E}_{ir}/T)\} + 2(K_{ir} - N_i + 1)$ can be used instead of $(\widehat{\mathcal{R}}_{ir})^2$ and the phrase "the j-th highest" must be replaced with "the j-th lowest". $T(\ln 2\pi + 1) + 4$ is not essential, because it is constant where X_0 is taken into consideration and already counted in K_{ir}. The 1 of $2(K_{ir} - N_i + 1)$ implies that σ^2 is counted as an unknown.

4 Scientific Variable Classifications and Meaningful Subsets

The author has developed the System **OEPP** for a user-knowledge-based variable selection method to solve the first j ADLR-best subset problems in a run of a computer, which is available for any kind of research. The main characteristics are (i) the automatic generation and estimation of only all meaningful subsets through scientific variable classifications rendered by a user and (ii) the automatic evaluation of the regression equation candidates by scientific, statistical and data-analytic criteria loaded by him.

Let us call a priori known sign and magnitude conditions about regression coefficients **scientific conditions**, a regression equation of a meaningful (sub)set which satisfies all scientific conditions a **scientifically reasonable regression equation**, a regression equation which satisfies the user-specified criteria for all statistical hypothesis and data-analytic tests and, furthermore, fits the j-th best to the data of a dependent variable a **statistically j-th best regression equation**, and a scientifically reasonable and statistically j-th best regression equation a **practically j-th best regression equation**. A user tries to search for the practically (first) best regression equation (j =1) or one among the first j practically best regression equation candidates.

It is quite amazing that all possible explanatory variables for ADLR can be trebly classified by the professional knowledge regarding research at hand in order to generate only meaningful subsets. They are (i) **single** or **grouped variables**, (ii) **combinatorial** or **sequential variables**. Furthermore, sequential variables are classified into (iii) **non-Almon** or **Almon (sequential) variables**. These variable classifications are called **scientific variable classifications**, distinguished from **econometric variable classifications**[8] . Three kinds of variable classifications are not only necessary but also sufficient for the sake of generation of only meaningful subsets for any kind of research.

4.1 Single Variable VS Grouped Variables

In variable selection by a run of a computer, a user has to notify a computer of the meanings or roles of explanatory variables in regression analysis. First of all, all possible explanatory variables must be classified according to whether an explanatory variable has its own meanings or role by itself or as a group. If an explanatory variable can have some meanings or role by itself and independently of any other explanatory variables, it is called a single variable. On the other hand, if an explanatory variable can have some meanings or role only when it is used together with the other(s), it is called a grouped variable. For instance, in population

[8] Included predetermined variables, right-hand side endogenous variables and excluded predetermined variables in the estimation of a simultaneous equation system are called econometric variable classifications which play different roles in variable selection.

movement from a region to another or job change from a company to another not because of unemployment but because of his desire, a pair of a pushing-out factor and receiving-in factor may be expected to appear in a regression equation. Computer hardware and software are complementary goods, regarded as a set of two grouped variables. Hardware without software is just a plastic box. Software without hardware cannot output anything. The information about the signs of some regression coefficients is often available from the professional knowledge related to research at hand. A user may want to test hypotheses about the signs of some regression coefficients. In this situation, it is convenient to attach the a priori known or to-be-hypothetically-tested signs, + or −, to the fronts of such explanatory variables in loading a dependent variable together with all possible explanatory variables into a computer. Let $\diamondsuit$ denote +, − or nothing and braces { and } stand for a set or subset of variables. Therefore, $\diamondsuit X$ implies $+X$, $-X$ or X. $\{+X_1, -X_2, X_3\}$ implies a (sub)set of variables X_1, X_2 and X_3 the a priori known signs of whose regression coefficients are positive, negative and unavailable, respectively. We postulate that (i) a set of grouped explanatory variables $\diamondsuit X_1, \diamondsuit X_2, \cdots, \diamondsuit X_K$ is enclosed within (and) like

$$(\diamondsuit X_1, \diamondsuit X_2, \cdots, \diamondsuit X_K) \quad \text{for } K = 2, 3, \cdots \tag{4.37}$$

and (ii) treated just like a single variable in generating only meaningful subsets.

Let $\mathcal{X}_k$ = k-th single variable $\diamondsuit X_k$ or k-th set of M_k grouped variables, $\diamondsuit X_{k1}$, $\diamondsuit X_{k2}, \cdots, \diamondsuit X_{kM_k}$, which is counted just like a single variable and enclosed within (and) like $(\diamondsuit X_{k1}, \diamondsuit X_{k2}, \cdots, \diamondsuit X_{kM_k})$ in the following combinatorial and sequential variable classifications but not in a generated subset denoted by $\{\cdots\}$ like $\{\cdots, \diamondsuit X_{k1}, \diamondsuit X_{k2}, \cdots, \diamondsuit X_{kM_k}, \cdots\}$. For simplicity, we call $\mathcal{X}_k$ a condensed variable hereafter.

Example 1: If we regard $X_1, +X_2, (X_3, -X_4), +X_5, (-X_6, +X_7, X_8)$ as $\mathcal{X}_1, \mathcal{X}_2, \mathcal{X}_3, \mathcal{X}_4, \mathcal{X}_5$ like

$$X_1, +X_2, (X_3, -X_4), +X_5, (-X_6, +X_7, X_8) = \mathcal{X}_1, \mathcal{X}_2, \mathcal{X}_3, \mathcal{X}_4, \mathcal{X}_5$$

then $\mathcal{X}_1 = X_1$; $\mathcal{X}_2 = +X_2$; $\mathcal{X}_3 = (X_3, -X_4)$; $\mathcal{X}_4 = +X_5$; and $\mathcal{X}_5 = (-X_6, +X_7, X_8)$ so that X_1, X_2 and X_5 are single variables, X_3 and X_4 are grouped variables in a set, and X_6, X_7 and X_8 are grouped variables in another set. The scientifically reasonable regression coefficients of X_2, X_5 and X_7 are positive from the viewpoints of the professional knowledge at hand, whereas those of X_4 and X_6 are negative. However, those of X_1, X_3 and X_8 are not a priori known, implying that the regression coefficients of X_1, X_3 and X_8 can be estimated as either positive or negative. The number of all possible subsets of variables $X_1, X_2, \cdots, X_8$ is 2^8 =256, whereas that of condensed variables $\mathcal{X}_1, \mathcal{X}_2, \cdots, \mathcal{X}_5$ is 2^5 =32. The use of information or professional knowledge about sets of grouped variables drastically reduces the number of subsets to be generated.

4.2 Basic Combinatorial Variables

A set of combinatorial variables is defined as one which consists of explanatory variables which can be selected in a combinatorial way to generate meaningful subsets. Let K =number of the condensed variables of $\mathcal{X}_1$, $\mathcal{X}_2$, $\cdots$, $\mathcal{X}_K$; M, N =nonnegative integers such that $0 \leq M, N \leq K$ and $0 < M + N$, $M^* = \min\{M, N\}$ and $N^* = \max\{M, N\}$. Nonnegative integers M and N and variables with/without the signs of their regression coefficients of $\mathcal{X}_k$'s reflect part of user's professional knowledge and specified by a user. Suppose that $\mathcal{X}_1, \mathcal{X}_2, \cdots, \mathcal{X}_K$ be a set of combinatorial condensed variables at least M^* but at most N^* of which should be selected. We postulate that (i) they are enclosed within $< M <$ and $> N >$ or within $< N <$ and $> M >$ as follows: for $K = 1, 2, 3, \cdots$

$$< M < \mathcal{X}_1, \mathcal{X}_2, \cdots, \mathcal{X}_K > N > \quad \text{or} \quad < N < \mathcal{X}_1, \mathcal{X}_2, \cdots, \mathcal{X}_K > M >, \tag{4.38}$$

(ii) the subsets of $\{\mathcal{X}_1, \mathcal{X}_2, \cdots, \mathcal{X}_K\}$ which include at least M^* but at most N^* condensed variables of $\mathcal{X}_1$, $\mathcal{X}_2$, $\cdots$, $\mathcal{X}_K$ are generated, and (iii) the signs of the estimated regression coefficients of variables are examined whether or not they coincide with the a priori known signs of the corresponding variables. The subsets generated become meaningful with respect to these variables. A user can specify the appropriate integers M and N through the professional knowledge of his research. Needless to say, $\sum_{m=M^*}^{N^*} \binom{K}{m}$ meaningful subsets are generated with respect to $\mathcal{X}_1$, $\mathcal{X}_2$, $\cdots$, $\mathcal{X}_K$. It must be noted that the positions of combinatorial variables within $< M <$ and $> N >$ or within $< N <$ and $> M >$ do not matter. $M = 0$ or $N = 0$ allows an empty subset. It is easily understood that $< 0 < \mathcal{X}_1, \mathcal{X}_2, \cdots, \mathcal{X}_K > K >$ or $< K < \mathcal{X}_1, \mathcal{X}_2, \cdots, \mathcal{X}_K > 0 >$ generates 2^K meaningful subsets including an empty subset but $< 1 < \mathcal{X}_1, \mathcal{X}_2, \cdots, \mathcal{X}_K > K >$ or $< K < \mathcal{X}_1, \mathcal{X}_2, \cdots, \mathcal{X}_K > 1 >$ generates $2^K - 1$ meaningful subsets excluding an empty subset. Let us give examples.

Example 2:

$$< 1 < +X_1, -X_2, (X_3, +X_4) > 2 > \quad \text{or} \quad < 2 < (X_3, +X_4), +X_1, -X_2 > 1 >$$

which is regarded as $< 1 < \mathcal{X}_1, \mathcal{X}_2, \mathcal{X}_3 > 2 >$ or $< 2 < \mathcal{X}_3, \mathcal{X}_1, \mathcal{X}_2 > 1 >$, respectively, with $\mathcal{X}_1 = +X_1$, $\mathcal{X}_2 = -X_2$ and $\mathcal{X}_3 = (X_3, +X_4)$, generates $\{\mathcal{X}_1\}$; $\{\mathcal{X}_2\}$; $\{\mathcal{X}_3\}$; $\{\mathcal{X}_1, \mathcal{X}_2\}$; $\{\mathcal{X}_1, \mathcal{X}_3\}$; and $\{\mathcal{X}_2, \mathcal{X}_3\}$, i.e., $\{+X_1\}$; $\{-X_2\}$; $\{X_3, +X_4\}$; $\{+X_1, -X_2\}$; $\{+X_1, X_3, +X_4\}$; and $\{-X_2, X_3, +X_4\}$ but treats $\emptyset$; $\{X_3\}$; $\{+X_4\}$; $\{+X_1, X_3\}$; $\{+X_1, +X_4\}$; $\{-X_2, X_3\}$; $\{-X_2, +X_4\}$; $\{+X_1, -X_2, X_3\}$; $\{+X_1, -X_2, +X_4\}$; and $\{+X_1, -X_2, X_3, +X_4\}$ as meaningless and does not generate them where X_1 and X_2 are single and combinatorial variables while X_3 and X_4 are grouped and combinatorial variables.

Example 3:

$$< 1 < +X_1 > 1 > \; < 0 < X_2, -X_3 > 2 > \quad \text{or} \quad < 2 < -X_3, X_2 > 0 > \; < 1 < +X_1 > 1 >$$

generates $\{+X_1\}$; $\{+X_1, X_2\}$; $\{+X_1, -X_3\}$; and $\{+X_1, X_2, -X_3\}$, which are led by the combinations of [1] the only one partially meaningful subset $\{+X_1\}$ generated by $< 1 < +X_1 > 1 >$ and [2] four partially meaningful subsets $\emptyset$, $\{X_2\}$; $\{-X_3\}$; $\{X_2, -X_3\}$ generated by $< 0 < X_2, -X_3 > 2 >$ or $< 2 < -X_3, X_2 > 0 >$. Meaningless subsets $\emptyset$; $\{X_2\}$; $\{-X_3\}$; and $\{X_2, -X_3\}$ are never generated.

4.3 Basic Sequential Variables

The third variable classification is for sequential variables. The Almon distributed lag regression is available only for a special case of time series data. At first, we introduce the entry format for non-Almon sequential variables and then modify it in a way that a computer can recognize (i) each set of Almon variables, (ii) whether the far-end point of time lag is fixed (known) or varying (to be determined), and (iii) whether a degree of a polynomial is fixed or varying (to be determined).

4.3.1 Non-Almon Sequential Variables

Sometimes, a priori information is available that explanatory variables in a set must be sequentially selected. Typical examples are consecutively lagged variables or power variables in a polynomial. Suppose that X_k denotes the k-th power of variable X, i.e., $X_k = (X)^k$ obtained through a variable transformation for $k = 2, 3, \cdots, K$, the K is the maximum degree of a polynomial, all variables with/without the a priori known signs of their regression coefficients from X to the K-th power of X, namely, $\Diamond X, \Diamond X_2, \Diamond X_3, \cdots, \Diamond X_K$, must appear in a polynomial regression equation, and the K varies, for instance, from 3 to 6. Then the relevant subsets are expressed as $\{\Diamond X, \Diamond X_2, \Diamond X_3\}$, $\{\Diamond X, \Diamond X_2, \Diamond X_3, \Diamond X_4\}$, $\{\Diamond X, \Diamond X_2, \Diamond X_3, \Diamond X_4, \Diamond X_5\}$ and $\{\Diamond X, \Diamond X_2, \Diamond X_3, \Diamond X_4, \Diamond X_5, \Diamond X_6\}$ which are the third-, fourth-, fifth- and sixth-degree polynomials of X, respectively.

We postulate that (i) a set of non-Almon sequential condensed variables $\mathcal{X}_1, \mathcal{X}_2, \cdots, \mathcal{X}_K$ is enclosed within $< M <$ and $>>$ or $<<$ and $> M >$ as follows: for $0 \le M \le K$ and $K = 1, 2, 3, \cdots$

$$< M < \mathcal{X}_1, \mathcal{X}_2, \cdots, \mathcal{X}_M, \cdots, \mathcal{X}_K >> \quad \text{or} \quad << \mathcal{X}_K, \cdots, \mathcal{X}_M, \cdots, \mathcal{X}_2, \mathcal{X}_1 > M > \tag{4.39}$$

and (ii) the subsets of $\{\mathcal{X}_1, \mathcal{X}_2, \cdots, \mathcal{X}_K\}$ which include the first $\mathcal{X}_1, \mathcal{X}_2, \cdots, \mathcal{X}_M$ and then one by one additionally from $\mathcal{X}_{M+1}$ to $\mathcal{X}_K$ are generated. $M = 0$ allows an empty subset as meaningful. The $(K - M + 1)$ generated subsets become meaningful with respect to these condensed variables. It must be noted that the positions of sequential variables within $< M <$ and $>>$ or within $<<$ and $> M >$ play decisively important roles in generating correct meaningful subsets.

Example 4:

$$< 0 < -X_1, X_2, +X_3, X_4 >> \quad \text{or} \quad << X_4, +X_3, X_2, -X_1 > 0 >$$

generates $\emptyset$; $\{-X_1\}$; $\{-X_1, X_2\}$; $\{-X_1, X_2, +X_3\}$; and $\{-X_1, X_2, +X_3, X_4\}$.

4.3.2 Almon Sequential Variables

Since Almon variables are sequential as well as single, we follow the format of non-Almon sequential variables (4.39) mentioned above but attach the degree or degree range (the degree is changed within the range) of a polynomial with a slash to it. Accordingly, a computer can distinguish non-Almon sequential variables from Almon sequential variables and recognize the degree or degree range of a polynomial to be used. Non-Almon sequential variables are used whether or not data are times series. To emphasize that Almon sequential variables are available only for time series data, we introduce $\Diamond X$ as an Almon current variable with/without the a priori known sign and $\Diamond X_{-k}$ as the k-time lagged variable with/without the a priori known sign of X for $k = 1, 2, \cdots, K$. Thus, $\mathcal{X}_1 = \Diamond X$ and $\mathcal{X}_{k+1} = \Diamond X_{-k}$ for $k = 1, 2, \cdots, K$.

(I) Fixed Degree of a Polynomial

Suppose that $(K+1)$ Almon current and lagged variables $X, X_{-1}, X_{-2}, \cdots, X_{-K}$ (in a group) follow the postulation of non-Almon sequential variables in (4.39) and the regression coefficients of the Almon variables in each of $(K - M + 2)$ meaningful subsets are represented by a polynomial of degree P where $0 < M \leq K + 1$ and $1 \leq P < \min\{M, T\} - 1$. Then we postulate that (i) Almon sequential variables $\Diamond X$, $\Diamond X_{-1}$, $\Diamond X_{-2}$, $\cdots$, $\Diamond X_{-K}$ are enclosed within $< M/P <$ and $>>$ from the $< M/P <$ side or within $<<$ and $> P/M >$ from the $> P/M >$ side like

$$< M/P < \Diamond X, \Diamond X_{-1}, \Diamond X_{-2}, \cdots, \Diamond X_{-(M-1)}, \Diamond X_{-M}, \cdots, \Diamond X_{-K} >> \quad (4.40)$$ [9]

or

$$<< \Diamond X_{-K}, \cdots, \Diamond X_{-M}, \Diamond X_{-(M-1)}, \cdots, \Diamond X_{-2}, \Diamond X_{-1}, \Diamond X > P/M >, \quad (4.41)$$

[9] Suppose that "ABC" is the user's notation of an Almon current variable so that $X = ABC$. In the System **OEPP**, X_{-k} can be directly treated by "$ABC(-k)$" where k =time lag number. If the current variable, ABC, is a transformed variable, it must be obtained through variable transformations prior to introducing $ABC(-k)$. For instance, if K =8 and P =2, then the following entry is loaded into a computer: $< 9/2 < +ABC, -ABC(-1), ABC(-2), \cdots, ABC(-7), ABC(-8) >>$ or $<< ABC(-8), ABC(-7), \cdots, -ABC(-1), +ABC > 2/9 >$, implying that the first 9 variables counted from the $< 9/2 <$ or $> 2/9 >$ side become the first meaningful subset of Almon variables whose regression coefficients are expressed as a polynomial of degree 2, i.e., $a_{10}ABC + a_{11}ABC(-1) + a_{12}ABC(-2) + \cdots + a_{18}ABC(-8)$ and $a_{1k} = b_{10} + b_{11}k + b_{12}k^2$ for $k = 0, 1, 2, \cdots, 8$, where $\hat{a}_{10} > 0$ and $\hat{a}_{11} < 0$ are scientific conditions. Thus, if $ABC(-9)$ is included as in $< 9/2 < +ABC, \cdots, ABC(-8), ABC(-9) >>$ or $<< ABC(-9), ABC(-8), \cdots, +ABC > 2/9 >$, the first 10 variables counted from the $< 9/2 <$ or $> 2/9 >$ side become the second meaningful subset of the Almon variables whose regression coefficients are expressed as a polynomial of degree 2, i.e., $a_{20}ABC + a_{21}ABC(-1) + a_{22}ABC(-2) + \cdots + a_{28}ABC(-8) + a_{29}ABC(-9)$ and $a_{2k} = b_{20} + b_{21}k + b_{22}k^2$ for $k = 0, 1, 2, \cdots, 9$, where $\hat{a}_{20} > 0$ and $\hat{a}_{21} < 0$ are scientific conditions. Two Almon distributed lag regressions are automatically conducted by loading the latter entry into a computer.

(ii) the $(K-M+2)$ meaningful subsets

$$\boldsymbol{X}_1 = \{\Diamond X, \Diamond X_{-1}, \Diamond X_{-2}, \cdots, \Diamond X_{-(M-1)}\};$$

$$\boldsymbol{X}_2 = \{\Diamond X, \Diamond X_{-1}, \Diamond X_{-2}, \cdots, \Diamond X_{-(M-1)}, \Diamond X_{-M}\}; \quad \cdots\cdots\cdots;$$

$$\boldsymbol{X}_{K-M+2} = \{\Diamond X, \Diamond X_{-1}, \Diamond X_{-2}, \cdots, \Diamond X_{-(K-1)}, \Diamond X_{-K}\} \tag{4.42}$$

are generated and then (iii) the corresponding converted sets

$$\boldsymbol{Z}_i = \{Z_i^0, Z_i^1, Z_i^2, \cdots, Z_i^P\} \quad \text{for } \boldsymbol{X}_i \; i = 1, 2, \cdots, K-M+2$$

are derived whose data matrices are calculated as $\boldsymbol{Z}_i = \boldsymbol{X}_i \boldsymbol{W}_i$
where

$$\boldsymbol{W}_1 = (W_1^{0\prime}, W_1^{1\prime}, W_1^{2\prime}, \cdots, W_1^{M-1\prime})'; \quad \boldsymbol{W}_2 = (W_2^{0\prime}, W_2^{1\prime}, W_2^{2\prime}, \cdots, W_2^{M\prime})';$$

$$\cdots\cdots; \text{ and } \boldsymbol{W}_{K-M+2} = (W_{K-M+2}^{0\prime}, W_{K-M+2}^{1\prime}, W_{K-M+2}^{2\prime}, \cdots, W_{K-M+2}^{K\prime})' \tag{4.43}$$

and $W_i^{k_i} = (1, k_i, (k_i)^2, \cdots, (k_i)^P)$ for $k_i = 0, 1, 2, \cdots, M+i-2$ and $i = 1, 2, \cdots, K-M+2$. $\boldsymbol{Z}_i$ is used as regressors. It must be noticed that $\boldsymbol{X}_i$ and $\boldsymbol{Z}_i$ are used not only as variable sets but also data matrices. $L=1$ and $R=1$ so that $\ell = 1$ and $r = 1$ are omitted.

If $0 < M < K+1$ in (4.40) or (4.41), then the optimal far-end point must be determined from $M-1$ to K inclusive. If the far-end point K of Almon variables X, X_{-1}, X_{-2}, $\cdots$, X_{-K} is known, M must be equal to $K+1$ in (4.40) or (4.41). In this case, only one meaningful $(K+1)$-variate subset $\{\Diamond X, \Diamond X_{-1}, \Diamond X_{-2}, \cdots, \Diamond X_{-K}\}$ is generated and then converted into a $(P+1)$-variate set $\{Z_1^0, Z_1^1, Z_1^2, \cdots, Z_1^P\}$.

If the starting point of the time lag is J such that $0 < J \leq K$, the following format is suitable:
$< M/P < \Diamond X_{-J}, \Diamond X_{-(J+1)}, \cdots, \Diamond X_{-(J+M-1)}, \Diamond X_{-(J+M)}, \cdots, \Diamond X_{-K} >>$ or $<< \Diamond X_{-K}, \cdots, \Diamond X_{-(J+M)}, \Diamond X_{-(J+M-1)}, \cdots, \Diamond X_{-(J+1)}, \Diamond X_{-J} > P/M >$. $B_{ir_\ell}^{\ell p_\ell}$ must be estimated subject to $b_{ir_\ell}^{\ell 0} + b_{ir_\ell}^{\ell 1} k_i^\ell + b_{ir_\ell}^{\ell 2} (k_i^\ell)^2 + \cdots + b_{ir_\ell}^{\ell p_\ell} (k_i^\ell)^P = 0$ for $k_i^\ell = 0, 1, 2, \cdots, J-1$, $i = 1, 2, \cdots, K-M-J+2$, $r_\ell = 1$ and $\ell = 1$.

Example 5: Case of $L=1$, $K=9$, $M=10=K+1$ and $P=2$ in which the far-end point is known and the degree of a polynomial is fixed at 2.

$$< 10/2 < +X, +X_{-1}, +X_{-2}, +X_{-3}, X_{-4}, X_{-5}, X_{-6}, X_{-7}, X_{-8}, X_{-9} >>$$

which is equivalent to

$$<< X_{-9}, X_{-8}, X_{-7}, X_{-6}, X_{-5}, X_{-4}, +X_{-3}, +X_{-2}, +X_{-1}, +X > 2/10 >$$

generates only one meaningful 10-variate subset

$$\boldsymbol{X} = \{+X, +X_{-1}, +X_{-2}, +X_{-3}, X_{-4}, X_{-5}, X_{-6}, X_{-7}, X_{-8}, X_{-9}\}$$

which is converted into a set $\boldsymbol{Z} = \{Z^0, Z^1, Z^2\}$ of 3 converted variables whose data matrix is calculated as $\boldsymbol{Z} = \boldsymbol{XW}$,
where $\boldsymbol{W} = (W^{0\prime}, W^{1\prime}, W^{2\prime}, \cdots, W^{9\prime})'$ and $W^k = (1, k, (k)^2)$ for $k = 0, 1, 2, \cdots, 9$. $\boldsymbol{Z}$ is used as regressors. The estimated regression coefficients of X, X_{-1} and X_{-2} must be positive for scientifical reasonableness in estimation.

Example 6: Case of $L = 1$, $K = 9$, $M = 8 < K + 1 = 10$, and $P = 3$ in which the degree of a polynomial is fixed but an optimal far-end point is eventually determined among K =7,8,9.

$$< 8/3 < +X, X_{-1}, X_{-2}, X_{-3}, X_{-4}, X_{-5}, X_{-6}, X_{-7}, X_{-8}, X_{-9} >>$$

which generates 3 meaningful subsets

$$\boldsymbol{X}_1 = \{+X, X_{-1}, X_{-2}, X_{-3}, X_{-4}, X_{-5}, X_{-6}, X_{-7}\};$$
$$\boldsymbol{X}_2 = \{+X, X_{-1}, X_{-2}, X_{-3}, X_{-4}, X_{-5}, X_{-6}, X_{-7}, X_{-8}\}; \text{ and}$$
$$\boldsymbol{X}_3 = \{+X, X_{-1}, X_{-2}, X_{-3}, X_{-4}, X_{-5}, X_{-6}, X_{-7}, X_{-8}, X_{-9}\}$$

each of which is converted into 3 possible sets of $\boldsymbol{Z}_i = \{Z_i^0, Z_i^1, Z_i^2\}$ through $\boldsymbol{Z}_i = \boldsymbol{X}_i\boldsymbol{W}_i$ for all $i = 1, 2, 3$ with the data matrices $\boldsymbol{W}_1 = (W_1^{0\prime}, W_1^{1\prime}, W_1^{2\prime}, \cdots, W_1^{7\prime})'$; $\boldsymbol{W}_2 = (W_1^{0\prime}, W_2^{1\prime}, W_2^{2\prime}, \cdots, W_2^{8\prime})'$; and $\boldsymbol{W}_3 = (W_3^{0\prime}, W_3^{1\prime}, W_3^{2\prime}, \cdots, W_3^{9\prime})'$ and $W_i^{k_i} = (1, k_i, (k_i)^2, \cdots, (k_i)^3)$ for $k_i = 0, 1, 2, \cdots, 6 + i$ and $i = 1, 2, 3$. $\boldsymbol{Z}_i$ is used as regressors. The estimated regression coefficient of X must be positive for scientific reasonableness in estimation, whereas it does not matter whether those of X_{-1}, X_{-2}, $\cdots$, X_{-9} assume positive or negative values.

(II) To-Be-Determined Degree of a Polynomial

Suppose that a degree of a polynomial is consecutively changed from P to Q if $P \leq Q$ or from Q to P if $P > Q$ and an optimal degree $\hat{p}$ is eventually determined among $p = P^*, P^* + 1, \cdots, Q^*$, where $1 \leq P, Q < K$, $P^* = \min\{P, Q\}$ and $Q^* = \max\{P, Q\}$. Let $R = Q^* - P^* + 1$. We postulate that (i) Almon variables X, X_{-1}, X_{-2}, $\cdots$, X_{-K} are enclosed within $< M/P - Q <$ and $>>$ from the $< M/P - Q <$ side or within $<<$ and $> P - Q/M >$ from the $> P - Q/M >$ side like

$$< M/P - Q < \Diamond X, \Diamond X_{-1}, \Diamond X_{-2}, \cdots, \Diamond X_{-(M-1)}, \Diamond X_{-M}, \cdots, \Diamond X_{-K} >> \quad (4.44)$$

or

$$<< \Diamond X_{-K}, \cdots, \Diamond X_{-M}, \Diamond X_{-(M-1)}, \cdots, \Diamond X_{-2}, \Diamond X_{-1}, \Diamond X > P - Q/M >, \quad (4.45)$$

(ii) the $(K - M + 2)$ meaningful subsets are generated as in (4.42) and then (iii) each, denoted by i, of the $(K - M + 2)$ meaningful subsets in (4.44) or (4.45) is converted into the following R possible converted-variable sets

$$\boldsymbol{Z}_{i1} = \{Z_{i1}^0, Z_{i1}^1, Z_{i1}^2, \cdots, Z_{i1}^{P^*}\}; \;\; \boldsymbol{Z}_{i2} = \{Z_{i2}^0, Z_{i2}^1, Z_{i2}^2, \cdots, Z_{i2}^{P^*+1}\};$$

$$\cdots\cdots; \quad \text{and} \quad \boldsymbol{Z}_{iR} = \{Z^0_{iR}, Z^1_{iR}, Z^2_{iR}, \cdots, Z^{Q^*}_{iR}\} \text{ for } \boldsymbol{X}_i \tag{4.46}$$

through $\boldsymbol{Z}_{ir} = \boldsymbol{X}_i \boldsymbol{W}_{ir}$ for $r = 1, 2, \cdots, R$ and $i = 1, 2, \cdots, K - M + 2$, where

$$\boldsymbol{W}_{i1} = (W^{0\prime}_{i1}, W^{1\prime}_{i1}, W^{2\prime}_{i1}, \cdots, W^{(M-1)\prime}_{i1})'; \quad \boldsymbol{W}_{i2} = (W^{0\prime}_{i2}, W^{1\prime}_{i2}, W^{2\prime}_{i2}, \cdots, W^{M\prime}_{i2})';$$

$$\cdots\cdots; \text{ and } \boldsymbol{W}_{iR} = (W^{0\prime}_{iR}, W^{1\prime}_{iR}, W^{2\prime}_{iR}, \cdots, W^{K\prime}_{iR})' \text{ for } \boldsymbol{X}_i \text{ for } i = 1, 2, \cdots, K - M + 2 \tag{4.47}$$

where $W^{k_i}_{ir} = (1, k_i, (k_i)^2, \cdots, (k_i)^{P^*+r-1})$ for $k_i = 0, 1, 2, \cdots, M + i - 2$ for $i = 1, 2, \cdots, K - M + 2$. Needless to say, $R \times (K - M + 2)$ possible converted-variable sets are generated in total.

It must be kept in mind that $P - Q$ of $< M/P - Q <$ and $> P - Q/M >$ does not imply a subtraction but implies a degree range of a polynomial specified by a user. If the far-end point K of Almon variables X, X_{-1}, X_{-2}, $\cdots$, X_{-K} is known, M must be equal to $K + 1$ in (4.44) or (4.45). In this case, the optimal degree $\hat{p}$ of a polynomial must be determined, if $P \neq Q$. If $0 < M < K + 1$ and $P \neq Q$ in (4.44) or (4.45), then not only the optimal degree $\hat{p}$ of a polynomial but also the optimal far-end point $\hat{K}$ must be determined at the same time.

Example 7: Case of $L = 1$, $K = 11$, $M = 12 = K + 1$, $P = 2$ and $Q = 3$ in which the far-end point is known but an optimal degree of polynomial is eventually determined between $p = 2$ and $p = 3$.

$$< 12/2-3 < -X, X_{-1}, X_{-2}, +X_{-3}, X_{-4}, X_{-5}, +X_{-6}, X_{-7}, X_{-8}, X_{-9}, X_{-10}, X_{-11} >>$$

which is equivalent to

$$<< X_{-11}, X_{-10}, X_{-9}, X_{-8}, X_{-7}, +X_{-6}, X_{-5}, X_{-4}, +X_{-3}, X_{-2}, X_{-1}, -X > 3-2/12 >$$

generates only one meaningful 12-variate subset

$$\boldsymbol{X} = \{-X, X_{-1}, X_{-2}, +X_{-3}, X_{-4}, X_{-5}, +X_{-6}, X_{-7}, X_{-8}, X_{-9}, X_{-10}, X_{-11}\}$$

which is converted into two possible converted-variable sets

$$\boldsymbol{Z}_1 = \{Z^0_1, Z^1_1, Z^2_1\} \text{ for } p = 2 \text{ and } \boldsymbol{Z}_2 = \{Z^0_2, Z^1_2, Z^2_2, Z^3_2\} \text{ for } p = 3$$

whose data matrices are calculated as $\boldsymbol{Z}_r = \boldsymbol{X}\boldsymbol{W}_r$ for $r = 1, 2$ where

$$\boldsymbol{W}_1 = (W^{0\prime}_1, W^{1\prime}_1, W^{2\prime}_1, \cdots, W^{11\prime}_1)' \text{ and } \boldsymbol{W}_2 = (W^{0\prime}_2, W^{1\prime}_2, W^{2\prime}_2, \cdots, W^{11\prime}_2)'$$

and $W^k_r = (1, k, (k)^2, \cdots, (k)^{r+1})$ for $k = 0, 1, 2, \cdots, 11$ and $r = 1, 2$. $i = 1$ is omitted. $\boldsymbol{Z}_r$ is used as regressors. The estimated regression coefficient of X must be negative and those of X_{-3} and X_{-6} must be positive for scientific reasonableness in estimation.

Example 8: Case of $L = 1$, $K = 10$, $M = 9 < K + 1 = 11$, $P = 2$, and $Q = 4$

so that $R = Q - P + 1 = 3$ and $K - M + 2 = 3$ in which an optimal far-end point is determined among $K = 8, 9, 10$ and an optimal degree is eventually determined among $p = 2, 3, 4$.

$$< 9/2 - 4 < +X, +X_{-1}, X_{-2}, X_{-3}, X_{-4}, X_{-5}, X_{-6}, X_{-7}, X_{-8}, X_{-9}, X_{-10} >>$$

generates 3 meaningful subsets

$$\boldsymbol{X}_1 = \{+X, +X_{-1}, X_{-2}, \cdots, X_{-7}, X_{-8}\}; \quad \boldsymbol{X}_2 = \{+X, +X_{-1}, X_{-2}, \cdots, X_{-8}, X_{-9}\};$$
$$\text{and } \boldsymbol{X}_3 = \{+X, +X_{-1}, X_{-2}, \cdots, X_{-9}, X_{-10}\} \quad \text{for } Y$$

each of which is converted into 3 possible converted-variable sets

$$\boldsymbol{Z}_{11} = \{Z_{11}^0, Z_{11}^1, Z_{11}^2\},\ \boldsymbol{Z}_{12} = \{Z_{12}^0, Z_{12}^1, Z_{12}^2, Z_{12}^3\} \text{ and } \boldsymbol{Z}_{13} = \{Z_{13}^0, Z_{13}^1, Z_{13}^2, Z_{13}^3, Z_{13}^4\} \text{ for } \boldsymbol{X}_1;$$
$$\boldsymbol{Z}_{21} = \{Z_{21}^0, Z_{21}^1, Z_{21}^2\},\ \boldsymbol{Z}_{22} = \{Z_{22}^0, Z_{22}^1, Z_{22}^2, Z_{22}^3\} \text{ and } \boldsymbol{Z}_{23} = \{Z_{23}^0, Z_{23}^1, Z_{23}^2, Z_{23}^3, Z_{23}^4\} \text{ for } \boldsymbol{X}_2; \text{ and}$$
$$\boldsymbol{Z}_{31} = \{Z_{31}^0, Z_{31}^1, Z_{31}^2\},\ \boldsymbol{Z}_{32} = \{Z_{32}^0, Z_{32}^1, Z_{32}^2, Z_{32}^3\} \text{ and } \boldsymbol{Z}_{33} = \{Z_{33}^0, Z_{33}^1, Z_{33}^2, Z_{33}^3, Z_{33}^4\} \text{ for } \boldsymbol{X}_3$$

whose data matrices are calculated as $\boldsymbol{Z}_{ir} = \boldsymbol{X}_i \boldsymbol{W}_{ir}$ for $r = 1, 2, 3$ and $i = 1, 2, 3$ where

$$\boldsymbol{W}_{11} = (W_{11}^{0\prime}, W_{11}^{1\prime}, W_{11}^{2\prime}, \cdots, W_{11}^{8\prime})', \quad \boldsymbol{W}_{12} = (W_{12}^{0\prime}, W_{12}^{1\prime}, W_{12}^{2\prime}, \cdots, W_{12}^{9\prime})' \text{ and}$$
$$\boldsymbol{W}_{13} = (W_{13}^{0\prime}, W_{13}^{1\prime}, W_{13}^{2\prime}, \cdots, W_{13}^{10\prime})' \text{ for } \boldsymbol{X}_1; \quad \boldsymbol{W}_{21} = (W_{21}^{0\prime}, W_{21}^{1\prime}, W_{21}^{2\prime}, \cdots, W_{21}^{8\prime})',$$
$$\boldsymbol{W}_{22} = (W_{22}^{0\prime}, W_{22}^{1\prime}, W_{22}^{2\prime}, \cdots, W_{22}^{9\prime})' \text{ and } \boldsymbol{W}_{23} = (W_{23}^{0\prime}, W_{23}^{1\prime}, W_{23}^{2\prime}, \cdots, W_{23}^{10\prime})' \text{ for } \boldsymbol{X}_2;$$
$$\text{and } \boldsymbol{W}_{31} = (W_{31}^{0\prime}, W_{31}^{1\prime}, W_{31}^{2\prime}, \cdots, W_{31}^{8\prime})', \quad \boldsymbol{W}_{32} = (W_{32}^{0\prime}, W_{32}^{1\prime}, W_{32}^{2\prime}, \cdots, W_{32}^{9\prime})' \text{ and}$$
$$\boldsymbol{W}_{33} = (W_{33}^{0\prime}, W_{33}^{1\prime}, W_{33}^{2\prime}, \cdots, W_{33}^{10\prime})' \text{ for } \boldsymbol{X}_3$$

and $W_{ir}^{k_i} = (1, k_i, (k_i)^2, \cdots, (k_i)^{r+1})$ for $r = 1, 2, 3$, $k_i = 0, 1, 2, \cdots, 7{+}i$ and $i = 1, 2, 3$. The estimated regression coefficients of X and X_{-1} must be positive for scientific reasonableness in estimation.

4.4 Nested Case

Nested combinatorial or sequential variables may be considered, though they are seldom used in real research. We postulate that they are enclosed within (and) like a set of grouped variables.

Example 9:

$$< 1 < (< 1 < +X_1, X_2 > 2 >), X_3, (-X_4, +X_5), (< 1 < +X_6, +X_7 >>) > 1 >$$

regards $\{< 1 < +X_1, X_2 > 2 >\}$; $\{X_3\}$; $\{-X_4, +X_5\}$; and $\{< 1 < +X_6, +X_7 >>\}$ as equivalent to each other and eventually generates

$$\{+X_1\}; \ \{X_2\}; \ \{+X_1, X_2\}; \ \{X_3\}; \ \{-X_4, +X_5\}; \ \{+X_6\}; \text{ and } \{+X_6, +X_7\}.$$

Needless to say, exactly the same meaningful subsets can be also generated by $< 1 < +X_1, X_2, (+X_1,\ X_2), X_3, (-X_4, +X_5), +X_6, (+X_6, +X_7) > 1 >$.

4.5 Illustrative Examples for Variable Selection

Now we are in a position to illustrate a variable selection method for ADLR in the System **OEPP** by full-scale examples. Following the notation defined in Section 3, we postulate that a dependent variable Y and all possible classified explanatory variables $\boldsymbol{X}$ are entered in a functional form and a constant term, denoted by X_0, is entered as itself in a functional form, if needed, and included in all meaningful subsets.

Let $\downarrow$ imply "actions by a computer" such as "generation, estimation, evaluation, ranking, analysis, prediction or policy simulation through the System **OEPP**" and $\Downarrow$ imply "actions by a user". The integer j reflecting the degree of certainty and confidence in his professional, statistical and data-analytic knowledge and model building experience is specified and a scientific, statistical and data-analytic criterion set is loaded together with a functional form with the data by a user:

Example 10:

$$Y = F(X_0, < 1 < +X_1^0, (+X_2^0, +X_3^0) > 2 >, << X_{-10}^1, X_{-9}^1, \cdots, X_{-1}^1, -X^1 > 2/11 >)$$

or

$$Y = F(X_0, < 11/2 < -X^1, X_{-1}^1, \cdots, X_{-9}^1, X_{-10}^1 >>, < 1 < +X_1^0, (+X_2^0, +X_3^0) > 2 >)$$

$$\downarrow$$

$$\begin{aligned} \boldsymbol{X}_1 &= \{X_0, +X_1^0, -X^1, X_{-1}^1, \cdots, X_{-9}^1, X_{-10}^1\} \text{ with } p = 2; \\ \boldsymbol{X}_2 &= \{X_0, +X_2^0, +X_3^0, -X^1, X_{-1}^1, \cdots, X_{-9}^1, X_{-10}^1\} \text{ with } p = 2; \text{ and} \\ \boldsymbol{X}_3 &= \{X_0, +X_1^0, +X_2^0, +X_3^0, -X^1, X_{-1}^1, \cdots, X_{-9}^1, X_{-10}^1\} \text{ with } p = 2 \text{ for } Y \end{aligned}$$

$$\downarrow$$

$$\{X_0, +X_1^0, Z_{11}^{10}, Z_{11}^{11}, Z_{11}^{12}\} \text{ and } \boldsymbol{X}_1; \ \{X_0, +X_2^0, +X_3^0, Z_{21}^{10}, Z_{21}^{11}, Z_{21}^{12}\} \text{ for } \boldsymbol{X}_2;$$
$$\text{and } \{X_0, +X_1^0, +X_2^0, +X_3^0, Z_{31}^{10}, Z_{31}^{11}, Z_{31}^{12}\} \text{ for } \boldsymbol{X}_3$$

$$\downarrow$$

$$\begin{aligned} \hat{Y}_{11} &= \hat{a}_{11}^{00} + \hat{a}_{11}^{01} X_1^0 + \hat{b}_{11}^{10} Z_{11}^{10} + \hat{b}_{11}^{11} Z_{11}^{11} + \hat{b}_{11}^{12} Z_{11}^{12} \\ \hat{Y}_{21} &= \hat{a}_{21}^{00} + \hat{a}_{21}^{01} X_2^0 + \hat{a}_{21}^{02} X_3^0 + \hat{b}_{21}^{10} Z_{21}^{10} + \hat{b}_{21}^{11} Z_{21}^{11} + \hat{b}_{21}^{12} Z_{21}^{12} \\ \hat{Y}_{31} &= \hat{a}_{31}^{00} + \hat{a}_{31}^{01} X_1^0 + \hat{a}_{31}^{02} X_2^0 + \hat{a}_{31}^{03} X_3^0 + \hat{b}_{31}^{10} Z_{31}^{10} + \hat{b}_{31}^{11} Z_{31}^{11} + \hat{b}_{31}^{12} Z_{31}^{12} \end{aligned}$$

$$\downarrow$$

$$\begin{aligned} \hat{Y}_{11} &= \hat{a}_{11}^{00} + \hat{a}_{11}^{01} X_1^0 + \hat{a}_{11}^{10} X^1 + \hat{a}_{11}^{11} X_{-1}^1 + \hat{a}_{11}^{12} X_{-2}^1 + \cdots + \hat{a}_{11}^{1,10} X_{-10}^1 \\ \hat{Y}_{21} &= \hat{a}_{21}^{00} + \hat{a}_{21}^{01} X_2^0 + \hat{a}_{21}^{02} X_3^0 + \hat{a}_{21}^{10} X^1 + \hat{a}_{21}^{11} X_{-1}^1 + \hat{a}_{21}^{12} X_{-2}^1 + \cdots + \hat{a}_{21}^{1,10} X_{-10}^1 \\ \hat{Y}_{13} &= \hat{a}_{31}^{00} + \hat{a}_{31}^{01} X_1^0 + \hat{a}_{31}^{02} X_2^0 + \hat{a}_{31}^{03} X_3^0 + \hat{a}_{31}^{10} X^1 + \hat{a}_{31}^{11} X_{-1}^1 + \hat{a}_{31}^{12} X_{-2}^1 + \cdots + \hat{a}_{31}^{1,10} X_{-10}^1 \end{aligned}$$

$$\downarrow$$

Evaluation of all regression subequations by means of scientific conditions, statistical hypothesis tests and data-analytic tests

$\downarrow$

Ranking of the regression subequations which have passed all scientific conditions, statistical hypothesis tests and data-analytic tests by means of a fitting measure

$\downarrow$

The selection of the first j practically best regression subequations if they exist

$\downarrow$

The analysis, prediction and/or policy simulation by each of the first j practically best regression subequations

$\downarrow$ if $j = 1$ or $\Downarrow$ if $j > 1$

The (first) practically best regression subequation if $j = 1$ or the user's selection of the ultimately practically best regression subequation among the first j practically best regression subequations if $j > 1$

$\Downarrow$

Presentation of the practically best regression equation and the related analysis, prediction and/or policy simulation in a refereed journal, a book, a conference, a seminar, an executive business meeting, etc. Then, final evaluation by referees, editors, audiences, professors, presidents, critics, readers, etc.

Example 11: Case of no variable selection.

$$Y = F(X_0, << 9/2 < +X^1, X^1_{-1}, X^1_{-2}, X^1_{-3}, \cdots, X^1_{-7}, X^1_{-8} >>, << X^2_{-10}, X^2_{-9}, X^2_{-8}, \cdots, X^2_{-1}, -X^2 > 3/11 >)$$

$$\downarrow$$

$$\{X_0, +X^1, X^1_{-1}, X^1_{-2}, X^1_{-3}, \cdots, X^1_{-8}, X^2_{-10}, X^2_{-9}, X^2_{-8}, \cdots, X^2_{-1}, -X^2\} \text{ for } Y$$

$$\downarrow$$

$$\{X_0, Z^{10}, Z^{11}, Z^{12}, Z^{23}, Z^{22}, Z^{21}, Z^{20}\} \text{ with } p_1 = 2 \text{ and } p_2 = 3$$

$$\downarrow$$

$$\hat{Y} = \hat{a}^{00} + \hat{b}^{10} Z_1^{10} + \hat{b}^{11} Z^{11} + \hat{b}^{12} Z^{12} + \hat{b}^{23} Z^{23} + \hat{b}^{22} Z^{22} + \hat{b}^{21} Z^{21} + \hat{b}^{20} Z^{20}$$

$$\downarrow$$

$$\begin{aligned}\hat{Y} &= \hat{a}^{00} + \hat{a}^{10}X^1 + \hat{a}^{11}X^1_{-1} + \hat{a}^{12}X^1_{-2} + \cdots + \hat{a}^{17}X^1_{-7} + \hat{a}^{18}X^1_{-8} \\ &+\hat{a}^{2,10}X^2_{-10} + \hat{a}^{29}X^2_{-9} + \cdots + \hat{a}^{22}X^2_{-2} + \hat{a}^{21}X^2_{-1} + \hat{a}^{20}X^2\end{aligned}$$

$$\downarrow$$

$\hat{a}^{10} > 0$ and $\hat{a}^{20} < 0$ must be met for scientific reasonableness in estimation.

Example 12: We omit the procedure at and after the stage of evaluation.

$$Y = F(X_0, < 1 < +X^0_1, -X^0_2 >>, << X^1_{-9}, X^1_{-8}, \cdots, X^1_{-1}, +X^1 > 2 - 4/10 >)$$

or

$$Y = F(X_0, < 10/2 - 4 < +X^1, X^1_{-1}, \cdots, X^1_{-8}, X^1_{-9} >>, << -X^0_2, +X^0_1 > 1 >)$$

$$\downarrow$$

$$\boldsymbol{X}_1 = \{X_0, +X^0_1, +X^1, X^1_{-1}, X^1_{-2}, \cdots, X^1_{-9}\} \text{ with } p = 2, 3, 4 \text{ and}$$
$$\boldsymbol{X}_2 = \{X_0, +X^0_1, -X^0_2, +X^1, X^1_{-1}, X^1_{-2}, \cdots, X^1_{-9}\} \text{ with } p = 2, 3, 4 \text{ for } Y$$

$$\downarrow$$

$$\{X_0, +X^0_1, Z^{10}_{11}, Z^{11}_{11}, Z^{12}_{11}\},\ \{X_0, +X^0_1, Z^{10}_{12}, Z^{11}_{12}, Z^{12}_{12}, Z^{13}_{12}\} \text{ and}$$
$$\{X_0, +X^0_1, Z^{10}_{13}, Z^{11}_{13}, Z^{12}_{13}, Z^{13}_{13}, Z^{14}_{13}\} \text{ for } \boldsymbol{X}_1;$$
$$\text{and } \{X_0, +X^0_1, -X^0_2, Z^{10}_{21}, Z^{11}_{21}, Z^{12}_{21}\},\ \{X_0, +X^0_1, -X^0_2, Z^{10}_{22}, Z^{11}_{22}, Z^{12}_{22}, Z^{13}_{22}\}$$
$$\text{and } \{X_0, +X^0_1, -X^0_2, Z^{10}_{23}, Z^{11}_{23}, Z^{12}_{23}, Z^{13}_{23}, Z^{14}_{23}\} \text{ for } \boldsymbol{X}_2$$

$$\downarrow$$

$$\hat{Y}_{11} = \hat{a}^{00}_{11} + \hat{a}^{01}_{11}X^0_1 + \hat{b}^{10}_{11}Z^{10}_{11} + \hat{b}^{11}_{11}Z^{11}_{11} + \hat{b}^{12}_{11}Z^{12}_{11}$$
$$\hat{Y}_{12} = \hat{a}^{00}_{12} + \hat{a}^{01}_{12}X^0_1 + \hat{b}^{10}_{12}Z^{10}_{12} + \hat{b}^{11}_{12}Z^{11}_{12} + \hat{b}^{12}_{12}Z^{12}_{12} + \hat{b}^{13}_{12}Z^{13}_{12}$$
$$\hat{Y}_{13} = \hat{a}^{00}_{13} + \hat{a}^{01}_{13}X^0_1 + \hat{b}^{10}_{13}Z^{10}_{13} + \hat{b}^{11}_{13}Z^{11}_{13} + \hat{b}^{12}_{13}Z^{12}_{13} + \hat{b}^{13}_{13}Z^{13}_{13} + \hat{b}^{14}_{13}Z^{14}_{13}$$
$$\hat{Y}_{21} = \hat{a}^{00}_{21} + \hat{a}^{01}_{21}X^0_1 + \hat{a}^{02}_{21}X^0_2 + \hat{b}^{10}_{21}Z^{10}_{21} + \hat{b}^{11}_{21}Z^{11}_{21} + \hat{b}^{12}_{21}Z^{12}_{21}$$
$$\hat{Y}_{22} = \hat{a}^{00}_{22} + \hat{a}^{01}_{22}X^0_1 + \hat{a}^{02}_{22}X^0_2 + \hat{b}^{10}_{22}Z^{10}_{22} + \hat{b}^{11}_{22}Z^{11}_{22} + \hat{b}^{12}_{22}Z^{12}_{22} + \hat{b}^{13}_{22}Z^{13}_{22}$$
$$\hat{Y}_{23} = \hat{a}^{00}_{23} + \hat{a}^{01}_{23}X^0_1 + \hat{a}^{02}_{23}X^0_2 + \hat{b}^{10}_{23}Z^{10}_{23} + \hat{b}^{11}_{23}Z^{11}_{23} + \hat{b}^{12}_{23}Z^{12}_{23} + \hat{b}^{13}_{23}Z^{13}_{23} + \hat{b}^{14}_{23}Z^{14}_{23}$$

$$\downarrow$$

$$\hat{Y}_{11} = \hat{a}^{00}_{11} + \hat{a}^{01}_{11}X^0_1 + \hat{a}^{10}_{11}X^1 + \hat{a}^{11}_{11}X^1_{-1} + \hat{a}^{12}_{11}X^1_{-2} + \cdots + \hat{a}^{19}_{11}X^1_{-9}$$
$$\hat{Y}_{12} = \hat{a}^{00}_{12} + \hat{a}^{01}_{12}X^0_1 + \hat{a}^{10}_{12}X^1 + \hat{a}^{11}_{12}X^1_{-1} + \hat{a}^{12}_{12}X^1_{-2} + \cdots + \hat{a}^{19}_{12}X^1_{-9}$$
$$\hat{Y}_{13} = \hat{a}^{00}_{13} + \hat{a}^{01}_{13}X^0_1 + \hat{a}^{10}_{13}X^1 + \hat{a}^{11}_{13}X^1_{-1} + \hat{a}^{12}_{13}X^1_{-2} + \cdots + \hat{a}^{19}_{13}X^1_{-9}$$

$$\hat{Y}_{21} = \hat{a}_{21}^{00} + \hat{a}_{21}^{01}X_1^0 + \hat{a}_{21}^{02}X_2^0 + \hat{a}_{21}^{10}X^1 + \hat{a}_{21}^{11}X_{-1}^1 + \hat{a}_{21}^{12}X_{-2}^1 + \cdots + \hat{a}_{21}^{19}X_{-9}^1$$
$$\hat{Y}_{22} = \hat{a}_{22}^{00} + \hat{a}_{22}^{01}X_1^0 + \hat{a}_{22}^{02}X_2^0 + \hat{a}_{22}^{10}X^1 + \hat{a}_{22}^{11}X_{-1}^1 + \hat{a}_{22}^{12}X_{-2}^1 + \cdots + \hat{a}_{22}^{19}X_{-9}^1$$
$$\hat{Y}_{23} = \hat{a}_{23}^{00} + \hat{a}_{23}^{01}X_1^0 + \hat{a}_{23}^{02}X_2^0 + \hat{a}_{23}^{10}X^1 + \hat{a}_{23}^{11}X_{-1}^1 + \hat{a}_{23}^{12}X_{-2}^1 + \cdots + \hat{a}_{23}^{19}X_{-9}^1$$

Example 13: We omit the procedure at and after the stage of evaluation.

$$Y = F(X_0, < 8/2 - 3 < +X^1, -X_{-1}^1, +X_{-2}^1, -X_{-3}^1, \cdots, -X_{-7}, +X_{-8}^1, -X_{-9}^1 >>)$$

$$\downarrow$$

$$\boldsymbol{X}_1 = \{X_0, +X^1, -X_{-1}^1, +X_{-2}^1, -X_{-3}^1, \cdots, +X_{-6}^1, -X_{-7}^1\} \text{ with } p = 2, 3;$$
$$\boldsymbol{X}_2 = \{X_0, +X^1, -X_{-1}^1, +X_{-2}^1, -X_{-3}^1, \cdots, -X_{-7}^1, +X_{-8}^1\} \text{ with } p = 2, 3; \text{ and}$$
$$\boldsymbol{X}_3 = \{X_0, +X^1, -X_{-1}^1, +X_{-2}^1, -X_{-3}^1, \cdots, +X_{-8}^1, -X_{-9}^1\} \text{ with } p = 2, 3 \text{ for } Y$$

$$\downarrow$$

$$\{X_0, Z_{11}^{10}, Z_{11}^{11}, Z_{11}^{12}\} \text{ and } \{X_0, Z_{12}^{10}, Z_{12}^{11}, Z_{12}^{12}, Z_{12}^{13}\} \text{ for } \boldsymbol{X}_1;$$
$$\{X_0, Z_{21}^{10}, Z_{21}^{11}, Z_{21}^{12}\}, \text{ and } \{X_0, Z_{22}^{10}, Z_{22}^{11}, Z_{22}^{12}, Z_{22}^{13}\} \text{ for } \boldsymbol{X}_2; \text{ and}$$
$$\{X_0, Z_{31}^{10}, Z_{31}^{11}, Z_{31}^{12}\} \text{ and } \{X_0, Z_{32}^{10}, Z_{32}^{11}, Z_{32}^{12}, Z_{32}^{13}\} \text{ for } \boldsymbol{X}_3$$

$$\downarrow$$

$$\hat{Y}_{11} = \hat{a}_{11}^{00} + \hat{b}_{11}^{10}Z_{11}^{10} + \hat{b}_{11}^{11}Z_{11}^{11} + \hat{b}_{11}^{12}Z_{11}^{12}$$
$$\hat{Y}_{12} = \hat{a}_{12}^{00} + \hat{b}_{12}^{10}Z_{12}^{10} + \hat{b}_{12}^{11}Z_{12}^{11} + \hat{b}_{12}^{12}Z_{12}^{12} + \hat{b}_{12}^{13}Z_{12}^{13}$$
$$\hat{Y}_{21} = \hat{a}_{21}^{00} + \hat{b}_{21}^{10}Z_{21}^{10} + \hat{b}_{21}^{11}Z_{21}^{11} + \hat{b}_{21}^{12}Z_{21}^{12}$$
$$\hat{Y}_{22} = \hat{a}_{22}^{00} + \hat{b}_{22}^{10}Z_{22}^{10} + \hat{b}_{22}^{11}Z_{22}^{11} + \hat{b}_{22}^{12}Z_{22}^{12} + \hat{b}_{22}^{13}Z_{22}^{13}$$
$$\hat{Y}_{31} = \hat{a}_{31}^{00} + \hat{b}_{31}^{10}Z_{31}^{10} + \hat{b}_{31}^{11}Z_{31}^{11} + \hat{b}_{31}^{12}Z_{31}^{12}$$
$$\hat{Y}_{32} = \hat{a}_{32}^{00} + \hat{b}_{32}^{10}Z_{32}^{10} + \hat{b}_{32}^{11}Z_{32}^{11} + \hat{b}_{32}^{12}Z_{32}^{12} + \hat{b}_{32}^{13}Z_{32}^{13}$$

$$\downarrow$$

$$\hat{Y}_{11} = \hat{a}_{11}^{00} + \hat{a}_{11}^{10}X^1 + \hat{a}_{11}^{11}X_{-1}^1 + \hat{a}_{11}^{12}X_{-2}^1 + \cdots + \hat{a}_{11}^{16}X_{-6}^1 + \hat{a}_{11}^{17}X_{-7}^1$$
$$\hat{Y}_{12} = \hat{a}_{12}^{00} + \hat{a}_{12}^{10}X^1 + \hat{a}_{12}^{11}X_{-1}^1 + \hat{a}_{12}^{12}X_{-2}^1 + \cdots + \hat{a}_{12}^{16}X_{-6}^1 + \hat{a}_{12}^{17}X_{-7}^1$$
$$\hat{Y}_{21} = \hat{a}_{21}^{00} + \hat{a}_{21}^{10}X^1 + \hat{a}_{21}^{11}X_{-1}^1 + \hat{a}_{21}^{12}X_{-2}^1 + \cdots + \hat{a}_{21}^{17}X_{-7}^1 + \hat{a}_{21}^{18}X_{-8}^1$$
$$\hat{Y}_{22} = \hat{a}_{22}^{00} + \hat{a}_{22}^{10}X^1 + \hat{a}_{22}^{11}X_{-1}^1 + \hat{a}_{22}^{12}X_{-2}^1 + \cdots + \hat{a}_{22}^{17}X_{-7}^1 + \hat{a}_{22}^{18}X_{-8}^1$$
$$\hat{Y}_{31} = \hat{a}_{31}^{00} + \hat{a}_{31}^{10}X^1 + \hat{a}_{31}^{11}X_{-1}^1 + \hat{a}_{31}^{12}X_{-2}^1 + \cdots + \hat{a}_{31}^{18}X_{-8}^1 + \hat{a}_{31}^{19}X_{-9}^1$$
$$\hat{Y}_{32} = \hat{a}_{32}^{00} + \hat{a}_{32}^{10}X^1 + \hat{a}_{32}^{11}X_{-1}^1 + \hat{a}_{32}^{12}X_{-2}^1 + \cdots + \hat{a}_{32}^{18}X_{-8}^1 + \hat{a}_{32}^{19}X_{-9}^1$$

$$\downarrow$$

$(-1)^k \times \hat{a}_{ir}^{1k} > 0$ for all $k = 0, 1, 2, \cdots, 6 + i$, $i = 1, 2, 3$ and $r = 1, 2$ must be met for scientific reasonableness in estimation.

5 Sufficiency of Treble Variable Classifications

The scientific variable classification for ADLR, which consists of (i) single vs grouped variable classification, (ii) combinatorial vs sequential variable classification and (iii) non-Almon sequential vs Almon sequential variable classification, is needed to solve the first j ADLR-best subset problems in a run of a computer. Thus, if all possible explanatory variables are classified by the professional knowledge related to research in question, the first at most j practically best regression subequations can be selected after evaluation and ranking when they exist. Now we must show that treble variable classifications are sufficient to generate all meaningful subsets in ADLR for any kind of research in a run of a computer. Suppose that a user wants to estimate each of the following $I \times \Pi_{\ell=1}^{L}(K_\ell - M_\ell + 2)$ regression equation candidates in the primitive procedure, corresponding to all meaningful subsets, by removing and/or adding possible explanatory variables on the basis of the professional knowledge related to the research at hand and by altering the degrees of polynomials:

$$Y = a_i^{00} + \boldsymbol{X}_i^{0*} A_i^0 + \sum_{\ell=1}^{L} \boldsymbol{X}_{k_\ell}^{\ell} A_{k_\ell}^{\ell} + U \quad \text{for } k_\ell = M_\ell - 1, M_\ell, \cdots, K_\ell \text{ and } i = 1, 2, \cdots, I \tag{5.48}$$

where $\boldsymbol{X}_i^0 = \{X_0, \boldsymbol{X}_i^{0*}\}$; he knows all variables in $\boldsymbol{X}_i^{0*}$; $\boldsymbol{X}_{k_\ell}^{\ell} = \{X^\ell, X_{-1}^\ell, X_{-2}^\ell, \cdots, X_{k_\ell}^\ell\}$ for $k_\ell = M_\ell - 1, M_\ell, \cdots, K_\ell$ with $0 < M_\ell - 1 \leq K_\ell$ for $\ell = 1, 2, \cdots, L$; the degrees $p_\ell = P_\ell, P_\ell + 1, \cdots, Q_\ell$ of a polynomial are applied to $\boldsymbol{X}_{k_\ell}^{\ell}$ with $0 < P_\ell \leq Q_\ell < M_\ell - 1$ for all ℓ; $M_\ell - 1$ =possible nearest far-end point; and K_ℓ =possible farthest-end point. Suppose that $\boldsymbol{X}^\ell = \{X^\ell, X_{-1}^\ell, X_{-2}^\ell, \cdots, X_{-(M_\ell-1)}^\ell, X_{-M_\ell}^\ell, \cdots, X_{-K_\ell}^\ell\}$ for all $\ell = 1, 2, \cdots, L$. In a run of a computer, he can make a computer generate all the above $I \times \Pi_{\ell=1}^{L}(K_\ell - M_\ell + 2)$ regression equation candidates by loading the following functional form:

$$\begin{aligned} Y &= F(X_0, < 1 < (\boldsymbol{X}_1^{0*}), (\boldsymbol{X}_2^{0*}), \cdots, (\boldsymbol{X}_I^{0*}) > 1 >, << \boldsymbol{X}^1 > P_1 - Q_1/M_1 >, \\ &\quad << \boldsymbol{X}^2 > P_2 - Q_2/M_2 >, \cdots, << \boldsymbol{X}^L > P_L - Q_L/M_L >) \end{aligned} \tag{5.49}$$

and estimate $I \times \Pi_{\ell=1}^{L}(K_\ell - M_\ell + 2)(Q_\ell^* - P_\ell^* + 1)$ regression subequations, where all $\boldsymbol{X}_i^{0*}$'s are treated as sets of grouped and combinatorial variables; $M_\ell - 1$ is the possible nearest far-end point; and $P_\ell - Q_\ell$ shows the possible polynomial-degree range for the ℓ-th Almon variable set.

Finally, it can be concluded that treble variable classifications in ADLR are not only necessary but also sufficient for the generation of only all meaningful subsets, no matter what research is conducted. In other words, a researcher can clear the first condition [I] in the j-th best subset problem for ADLR, using treble variable classifications. It is easy to master them.

6 Concluding Remarks

It is quite difficult to a priori know a best set of explanatory variables for a dependent variable before estimation and evaluation, because some of data problems (unavailable data, proxy data, inconsistent observation dates of data in a model, etc.) and/or estimation problems (the considerably good but slightly ambiguous principle of least squares, linearization of explanatory variables by the Taylor expansion, unknown functional form, multicollinearity, etc.) often occur. There are many cases in which the best regression equation cannot be determined solely by econometrics. Scientifically reasonable signs and/or magnitudes of (the sums of) regression coefficients and well tracking the turning points of the data of a dependent variable by the estimates are sometimes decisively important. Therefore, informatic (or knowledge-based) and computational techniques to solve the first j-th best subset problems for various estimation methods are useful.

When the number of all possible explanatory variables many of which are consecutively lagged explanatory variables exceeds the sample size of time series data or the consecutively lagged explanatory variables cause multicollinearity, ADLR may be useful. A variable selection problem for ADLR has not been concretely formulated in the literature. The author proposed an informatic and computational method to solve the first j ADLR-best subset problems as a variable selection problem for ADLR and suggested to regard as the practically best regression equation (i) a solution to the (first) ADLR-best subset problem ($j = 1$) or (ii) the one selected among at most first j solutions to the first j ADLR-best subset problems by a user's own new criterion or by comparing them with each other ($j > 1$). A suitable positive integer j specified by a user depends on his professional knowledge, appropriate statistical significance levels, data-analytic criteria and model building experiences. The informatic and computational techniques installed in the Intellectual Statistical System **OEPP** are an easy, quick, resources-saving and high quality method for econometric analysis and/or forecasting.

References

Akaike, H., Information theory and an extension of the maximum likelihood principle, Proceedings of the 2nd international symposium information theory, B.N. Petrov and F. Csaki (eds), pp. 267-281, Akademiai Kiado, Budapest, 1973.

Almon, S., The distributed lag between capital approximations and expenditures, Econometrica, vol. 30, pp. 178-196, 1965.

Durbin, J., Testing for serial correlation in least squares regression when some of the regressors are lagged dependent variables, Econometrica, vol. 38, pp. 410-421, 1970.

Durbin, J. and Watson, G.S., Testing for serial correlation in least squares regression, I, Biometrika, vol. 37, pp. 409-428, 1950.

Durbin, J. and Watson, G.S., Testing for serial correlation in least squares regression, II, Biometrika, vol. 38, pp. 159-178, 1951.

Farebrother, R.W., The Durbin-Watson test for serial correlation when there is no intercept in the regression, Econometrica, vol. 48, pp. 1553-1563, 1980.

Godfrey, L.G., Testing against general autoregressive and moving average error models when the regressors include lagged dependent variables, Econometrica, vol. 46, pp. 1293-1302, 1978.

Jarque, C.M. and A.K. Bera, Efficient tests for normality, homoscedasticity and serial independence of regression residuals, *EconomicsLetters*, Vol. 6, pp. 255-259, 1980.

Onishi, H., A variable selection procedure for econometric models, J. Computational Statistics & Data Analysis, vol. 1, no.2, pp. 85-95, 1983.

Onishi, H., An efficient method for building a prefectural government official reduction model by the Researcher System OEPP, *J. Japanese Society of Computational Statistics*, vol. 7, no.1, pp. 119-139, 1994.

Onishi, H., Researcher System OEPP for Econometrics, I and II, Version 2, OEPP Institute, 3-208-2, Takezono, Tsukuba, Ibaraki-Prefecture, Japan, 1996.

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, pp. 1989-1996, 19??.

Wallis, K.F., Testing for fourth order for autocorrelation in quarterly regression equations, Econometrica, vol. 40, pp. 617-636, 1972.