No. 962

Statistical Philosophy, Computational Techniques and Practically-Best Regression Equation

by

Haruo Onishi

December 2001

Statistical Philosophy, Computational Techniques and Practically-Best Regression Equation

Haruo ONISHI[*]

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

SUMMARY

Regression analysis is one of the most useful statistical methods on which users rely. A variable selection problem for regression analysis must be concretely formulated and scientifically, statistically and data-analytically solved. A regression equation must be consistent with the scientific knowledge, including natural logic and correct common sense, related to research at hand and then statistically and data-analytically best. Scientific knowledge is used for meanings or roles of explanatory variables in a regression equation and the signs and/or magnitudes of their regression coefficients, including the Schur stability condition. Typical statistical tests consist of hypothesis testings about normality, regression coefficients, serial correlation, residual outliers, equal regression coefficients in different sample subgroups and homoscedasticity. Data-analytic investigations are standardized residuals, turning points and fitting. These depend on types of data, lagged dependent variables and structural changes during an estimation period. The j-th OLS-best subset problem is formulated as a standard variable selection problem for regression analysis and a knowledge-based variable selection method is proposed to solve it in a run of a computer, where integer j depends on user's scientific, statistical and data-analytic knowledge and experience in model-building. The software **OEPP** can handle knowledge-based variable selection methods for various estimation methods.

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

[*] 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. 11:00, November 25 (Sunday), 2001. File olspaper.tex.

1 Introduction

Even if a user is not familiar with linear programming, he or she can always obtain the optimal solution by solving a linear programming problem by the symplex or Karmarkar method and use it to accomplish the research. On the other hand, users of regression analysis or econometrics are not so fortunate. Theoretical statistics has been advanced and computational environments have been drastically improved. Despite them, defining a standard variable selection problem for regression analysis and how it can be solved in a run of a computer have not been thoroughly discussed in the literature. Users of the variable selection methods such as stepwise regression, forward selection, backward elimination, t- or F-directed search and mini-max regret principle obtain statistical models as the solutions which are different from each other in most cases and often cannot be interpreted scientifically. As a result, they may become confused or disappointed or even endanger their jobs. At the same time, lots of resources, especially brain labor, are wasted everyday, everywhere in the world. All possible regressions are not efficient and effective for variable selection. The principle of regression analysis is easy to understand. However, it is more difficult than expected to search for a scientifically reasonable and statistically and data-analytically best regression equation. Therefore, every user has to load the data and variables of regression equation candidates into a computer one at a time, cites percentiles of various statistical tests from a statistical book, compares the test statistics with them and scrutinizes the results or predictions. He has to repeat this process until he can find the best regression equation. It is of great importance to stop this absurd primitive procedure and offer a method to search for the best regression equation, although a statistical estimation problem is much more complicated than a mathematical programming problem.

The purpose of this paper is to concretely formulate a standard variable selection problem for regression analysis, develop a standard knowledge-based variable selection method to solve it by taking advantage of present high speed, large memory and cheap calculation cost of a computer, install the algorithm in the Intellectual Statistical System **OEPP**[1] and demonstrate an example. It must be kept in mind that the phrase "a run of a computer" implies that once a user inputs all necessary information in a computer, he can be released from his inessential, laborious and time-consuming work subject to the computer and use his resources for more creative work.

In Section 2, statistical philosophy is discussed. In Section 3, a standard variable selection problem is formulated. In Section 4, remarks on a solution, diagnosis and prediction are addressed to. Processing of professional knowledge is described in

[1] The **OEPP** is software which can handle the Onishi knowledge-based variable selection methods for (C)OLS, (C)GLS, (C)ADLR, (C)BCT, (C)2SLS, (C)2SPC, LIML, LIPC, etc. and consists of about 90,000 lines of algorithm in total written in FORTRAN 77. The "C" of (C) implies constrained. It is available for FACOM or IBM-compatible main frame machines, UNIX workstations and personal computers.

Section 5 and a demonstration is made in Section 6. Section 7 draws up concluding remarks.

2 Statistical Philosophy

2.1 Some Definitions

Let us call the knowledge established in the science(s) related to research in question, including the information about the situation covered by the research at hand (for example, commercial customs in the economy at hand), natural logic and correct common sense, **professional knowledge**. The method to be proposed searches for a **practically best regression equation** which is defined as **not only scientifically reasonable but also statistically and data-analytically best** on the condition of the appropriate scientific, statistical and data-analytic criteria specified by a user. It must be kept in mind that it is impossible to impeccably formulate a variable selection problem for any kind of estimation or, in other words, to perfectly define the best regression equation, as long as we follow the Neyman-Pearson approach to statistics. For instance, the optimal or correct significance levels for statistical tests are usually unavailable, although a solution to a variable selection problem depends on the significance levels. Furthermore, we define the following terms: a **meaningful subset** as a subset which includes all necessary explanatory variables for a dependent variable but excludes any unnecessary, redundant and/or contradiction-causing explanatory variables from the viewpoints of professional knowledge or, in other words, a subset which consists of only explanatory variables representing a behavioral, institutional, technical or natural-law-based relation to a dependent variable and a **regression subequation** as the regression equation of a meaningful subset in the j-th OLS-best subset problem. Thus, a subset which does not include at least one necessary explanatory variable or includes at least one unnecessary, redundant or contradiction-causing explanatory variable is meaningless.

2.2 Mission of Statistics

Sciences are divided into pure and applied sciences. Furthermore, applied sciences may be classified into substantial and methodological sciences. Substantial sciences are directly related to achieving or improving human desires or goals. Economics, business science, sociology, environmental science and engineering are examples of substantial sciences. On the other hand, methodological sciences are not directly related to achieving or improving human desires or goals but support substantial sciences and help to yield useful achievements for them. Statistics, mathematical programming, queueing theories and fuzzy method belong to methodological sciences. Substantial sciences can advance by themselves but in a much faster, broader and deeper way with methodological sciences. Strictly speaking, methodological sciences become useful and then valuable, only when they are used for substantial sciences, that is, they fulfill the missions for substantial sciences. New methodologies may be able to shed light on veiled or unexplored areas of substantial sciences which the substantial sciences cannot unveil by themselves. Accordingly, both substantial and methodological sciences are needed. One cannot choose between substantial and methodological sciences just like between pure and applied sciences. Econometrics is a branch of statistics in which regression analysis is highly advanced to cope with difficult situations in which some of the assumptions made in regression analysis are not met. Statistics and econometrics are not pure but applied mathematics to analyze the data of substantial sciences. Data analysis is made with a finite number of data. Since expectation and asymptotic theory are based on mathematics of infinite numbers of sampling and data, their concepts help data analysis not directly but maybe for conscience' sake, because no data-analytic method has been so far discovered to deal with infinite numbers of sampling and data. The stepwise regression, forward selection, backward elimination, t- or F-directed search and minimax regret principle methods are not methodologies but mathematical exercises. Informatic econometrics which covers theories, methodologies and algorithms of both econometrics and professional knowledge must fulfill the mission to solve socio-economic problems.

An outcome is determined by a research purpose, whereas factors or causes believed to affect the outcome are observed through the mechanism or structure covered by the research in question. A dependent variable represents an outcome and must have its own data set. Also, an explanatory variable represents a factor and must have its own data set. For simplicity, we assume that the unique data set of an outcome is available. Usually, all factors have their own respective unique data sets. However, various data problems have been encoutered in actual research. There is a case in which the data of some factor are not available at all. In this case, some appropriate proxy factors with the data sets may be introduced through trial and error, because a user cannot collect or measure the needed data from the viewpoints of time, cost, privacy, etc. There is another case in which a factor is well known and the observations of its components are available but it is difficult to uniquely determine a suitable data set for the factor. For instance, let a factor be agricultural labor represented by a data set L, L_1=working hours of farmers, L_2=working hours of their wives, L_3=working hours of their parents and L_4=working hours of their children. It is easy to calculate $L_* = L_1 + L_2 + L_3 + L_4$. Should we define L_* as L? In this case, the working efficiencies of farmers' wives, parents and children are evaluated to be as efficient as the farmers but the farmers with strong responsibilities for their families and rich experiences in farming work more efficiently and effectively than the others. Let E_2, E_3 and E_4 be relative efficiencies in comparison with the farmers, where $1 > E_2 \geq E_3 \geq E_4 > 0$. We can define $L_{**} = L_1 + E_2L_2 + E_3L_3 + E_4L_4$. Unfortunately, it is not easy to measure E_2, E_3 and E_4 which are easily improved by mechanization, electrification, computerization and/or automation of equipment capital (rice planters, tractors, harvesters, driers, storage equipment, etc.) and the nationwide data of E_2, E_3 and E_4 are unavailable. E_2, E_3 and E_4 may be guessed from experimental samples or local surveys. In this case, this factor is treated by letting an explanatory variable assume each of the alternative data set candidates L_* and L_{**} or introducing four explanatory variables with componentwise data sets L_1, L_2, L_3 and L_4 each through trial and error. Furthermore, the observation dates of many kinds of data used in a regression equation are often not consistent or the same. For instance, in a cross-sectional regression equation of a certain year, the populations in prefectures (or states) of the census data may be unavailable for that year, because the census is conducted every 5 years. Data of different years may be used.

A mathematical relation between an outcome and factors is expressed as an equation of a dependent variable with possible explanatory variables (candidates). A hypothesis (or a set of hypotheses) is set up on the basis of the equation by a user. When the hypothesis is substantiated, the possible explanatory variables are determined as explanatory variables and a structure, law, rule or theory is established through the estimated equation. Suppose that 25 possible explanatory variables are taken into consideration for a dependent variable. Then, 33,554,431 possible nonempty subsets exist. It is easy to mechanically estimate all of the 33,554,431 regression subequations by the all-possible-regressions method. However, it is quite laborious for a user to distinguish the regression subequations of meaningful subsets from those of meaningless subsets. Eventually, a best one must be selected from among the regression subequations of all meaningful subsets.

The starting point of a user's research is to obtain a regression equation which approximates the true function between a dependent variable and explanatory variables as closely as possible. Then, he can promote the research by using it. It is not easy to obtain such a regression equation but possible by realizing computational techniques to be referred to in this paper. For instance, it was observed in Japan during the high economic growth period that many farmers gave up agriculture, left villages for cities and became factory workers so that the number of farmers has been drastically reduced. But the agricultural production changed but did not decrease, due to more use of equipment capital, fertilizers, insecticides and herbicides in addition to more R&D for agriculture. As a result, a simple correlation coefficient between agricultural production and labor was revealed to be negative. If this simple correlation coefficient is superficially interpreted, it may be concluded that the harder farmers work, the less the agricultural production will be. Unless a regression equation for agricultural production is well estimated and carefully scrutinized, the regression coefficient of agricultural labor assumes a negative estimate. In reality, the labor of farmers increases the agricultural production ceteris paribus. Thus, a contradiction occurs between regression analysis and reality. The estimated regression coefficient of labor must be positive. This is an important point. Statistics and econometrics does not require a positive regression coefficient as a scientifically reasonable condition but professional knowledge about agriculture does.

Suppose that a dependent variable Y with its data y_t's of $|y_t| \gg 0$ for all or most

t is well estimated but we purposely add a new explanatory variable X_* with its artificial data $y_t + (-1)^t$'s for all t to the estimation, where the correlation between Y and X_* is extremely strong. The true regression coefficient a_* of X_* must be zero, i.e., $a_* = 0$, becasue there is no such a factor in the system or structure in question that X_* represents. If a perfect estimation method is available and applied, we must have the estimate $\widehat{a}_*$ of $\widehat{a}_* = 0$. However, if one of the proposed estimation methods, say, ordinary least squares, is applied, $\widehat{a}_* \neq 0$ ($\widehat{a}_* > 0$ in this case) is estimated and becomes significant. This is a contradiction. Even an explanatory variable X which actually represents a factor and whose regression coefficient was significant in the previous estimation may be rejected in the new estimation, if there is some degree of the correlation between X_* and X and the correlation between Y and X is weaker than that between Y and X_*. It is impossible to reject this regression equation by statistics. Only the way to reject it is to evaluate it with the professional knowledge that such a factor does not exist in the system or form the structure. For instance, it is clear that the economies of all countries in the world are strongly or weakly connected with each other through international trades. Even if the Japanese consumption is explained better with the income of a third country than with the Japanese income, such a consumption function is not trusted, because the Japanese cannot use the income of that third country, although the economic relation between Japan and the third country exists. It must be rejected by the professional knowledge or information about the international economy. Thus, in general, professional knowledge is needed to search for a best regression equation in addition to statistical tests and data-analytic investigation. The words "data-analytic investigation" implies here checking how well the partial-test (and final-test) estimates fit the observations of a dependent variable and track the turning points of the observations of a dependent variable.

Let $Y = \boldsymbol{X}A + U$ be a true regression equation where Y =dependent variable or a $(T \times 1)$-vector of its data; $\boldsymbol{X}$ =set of a constant term and all K explanatory variables or a $\{T \times (K+1)\}$-matrix of their data; $A = \{(K+1) \times 1\}$-vector of the regression coefficients of $\boldsymbol{X}$; $U = (T \times 1)$-vector of disturbances; and T =sample size. Many principles of minimizing objective functions representing total effects or indexes of all residuals with respect to regression coefficients can be considered. For instance, minimizing $\sum_{t=1}^{T} \alpha_t |Y_t - \widehat{Y}_t|^\ell$ (or even $\Pi_{t=1}^{T} |Y_t - \widehat{Y}_t|^\ell$) with respect to $\widehat{A}$ for any integer ℓ of $\ell = 1, 2, 3, \cdots$ is reasonable, where $\widehat{Y}_t = X_t'\widehat{A}$, $X_t' = t$-th row of $\boldsymbol{X}$ and $\alpha_t = t$-th known positive weight. This means that unless multicollinearity occurs, many regression equation candidates for $Y = \boldsymbol{X}A + U$ are estimable, even if $\boldsymbol{X}$ is meaningful and uniquely correct. The cases of $\ell = 3, 4, 5, \cdots$ lead to complicated calculation of $\widehat{A}$ so that the statistical properties become ambiguous and statistical hypothesis testings become almost impossible in the present state of the arts. The case of $\ell = 1$ is rarely used in actual applications and does not yield better performance. The principle of ordinary least squares, abbreviated as OLS hereafter, is the case of $\ell = 2$ and $\alpha_t = 1$ for all t and has advantages of simple calculation of $\widehat{A}$ and easy applications of statistical tests. The Gauss-Markov theorem, i.e., BLUE (Best Linear Unbiased Estimator) of OLS has a weak point. Minimum variance (best) and unbiasedness are statistically desirable characteristics and have no problems. How about linearity? Even when a regression equation is linear with respect to regression coefficients, an estimator for A does not need to be linear with respect to Y for a best regression equation. However, both minimum variance and unbiasedness deeply depend on the linearity about Y. If the theorem holds regardless of any functional forms of Y, it may be used confidently and lead to a best regression equation.

Suppose that (1) $K+1=T$ holds, (2) all regression coefficients of A are significant and (3) there are many other minor factors which affect Y slightly. A perfectly-fitted regression equation is always estimated by (1) so that $\widehat{Y}=Y$, though disturbances exist by (3). Needless to say, A cannot be estimated, if $|\boldsymbol{X}'\boldsymbol{X}|=0$. This may state that OLS has inherently an odd characteristic. Many statistics like an adjusted or a doubly adjusted coefficient of determination, AIC, SIC and BIC have been proposed for the principle of parsimony but it has not been known which one is perfect or best. If the number of explanatory variables is one or two, OLS works well, because it is possible to plot the data of a dependent and one or two explanatory variables in a two- or three-dimensional diagram, see the degrees of correlation between them and guess even a functional form. Such a case is often seen in an engineering field where, for instance, temperature, pressure, mositure and inputs of the other materials can be fixed at certain levels, the inputs of one or two materials are changed and then the output is measured. Although OLS works well in a simple case, it should not be overestimated as a perfect estimation method. In social sciences like economics for free market economies in which most factors cannot be controlled except by laws or regulations and the observations result from a general equilibrium or situation, many factors are usually taken into consideration in regression or econometric analysis. In a case of many possible explanatory variables, it is not easy to uniquely specify a single regression equation which definitely becomes best after estimation and evaluation. Nonetheless, the authors of a considerable number of positive economics papers of regression analysis and econometrics issued in refereed journals for social sciences have introduced single regression equations without examining other possible regression equation candidates and have drawn up their conclusions. They insist that their conclusions are correct and criticize others'. However, their different and conflicting conclusions imply that the truth is still in the dark.

A statistically desirable characteristic of unbiasedness does not say anything about what the true values of A are and even the ranges in which the true values of A must exist, if $\boldsymbol{X}$ is not proved to be a best subset. In actual research, a user usually estimates $\widehat{A}$ only once. ,Unbiasedness $\mathcal{E}(\widehat{A})=A$ does not mean $A=\widehat{A}$. Many people, even well-educated people, unconsciously start to believe that A is $\widehat{A}$. It is quite mistaken. Let $i=1,2$; $\boldsymbol{X}_i=i$-th alternative meaningful subset for agricultural production Y; a_{k_i} =true regression coefficient of the common explanatory variable of "labor" in $\boldsymbol{X}_i$; $\widehat{a}_{k_i}$ =estimate of a_{k_i} with $\widehat{a}_{k_1}>0$ and $\widehat{a}_{k_2}<0$ which are both significant in the hypothesis testing of H_0: $a_{k_i}=0$ against H_1: $a_{k_i}\neq 0$;

$\widehat{s}_i$ =estimate of the standard deviation σ_i of a disturbance term U corresponding to $\boldsymbol{X}_i$; and t_i =two-tailed percentile of a t-distribution of the same significance level corresponding to $\boldsymbol{X}_i$. The confidence intervals for a_{k_1} and a_{k_2} are as follows:

$$0 < \widehat{a}_{k_1} - t_1\widehat{s}_1\sqrt{(\boldsymbol{X}_1'\boldsymbol{X}_1)^{-1}_{k_1k_1}} \leq a_{k_1} \leq \widehat{a}_{k_1} + t_1\widehat{s}_1\sqrt{(\boldsymbol{X}_1'\boldsymbol{X}_1)^{-1}_{k_1k_1}} \tag{2.1}$$

and

$$\widehat{a}_{k_2} - t_2\widehat{s}_2\sqrt{(\boldsymbol{X}_2'\boldsymbol{X}_2)^{-1}_{k_2k_2}} \leq a_{k_2} \leq \widehat{a}_{k_2} + t_2\widehat{s}_2\sqrt{(\boldsymbol{X}_2'\boldsymbol{X}_2)^{-1}_{k_2k_2}} < 0 \tag{2.2}$$

where $(\boldsymbol{X}_i'\boldsymbol{X}_i)^{-1}_{k_ik_i} = (k_i, k_i)$-th diagonal element of $(\boldsymbol{X}_i'\boldsymbol{X}_i)^{-1}$ for $i = 1, 2$. In which confidence interval (2.1) or (2.2) is the true value of the regression coefficient of the explanatory variable "labor"? Can a confidence interval provide the range in which the true value of the regression coefficient of the explanatory variable "labor" exist without knowing which subset of $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$ becomes best? Unfortunately, all terms of $\widehat{a}_{k_i}$, t_i, $\widehat{s}_i$ and $(\boldsymbol{X}_i'\boldsymbol{X}_i)^{-1}_{k_ik_i}$ in both the upper and lower bounds of each confidence interval depend on $\boldsymbol{X}_i$. Only when $\boldsymbol{X}_i$ is determined as a best subset, can the upper and power bounds of a confidence interval make sense. Unless a priori known knowledge or information about a_{k_i} is available, a confidence interval does not provide a condition to determine a best subset. Users are quite interested in and sensitive about the signs and magnitudes of estimated regression coefficients, because an opposite sign of a regression coefficient completely changes the interpretation of a regression equation. If a priori known knowledge about $a_{k_i} > 0$ for $i = 1, 2$ is available, then $\boldsymbol{X}_1$ should be kept with the confidence interval (2.1). $\boldsymbol{X}_2$ is regarded as misspecified and ignored. $a_{k_i} > 0$ becomes an important condition to determine a best subset. Consistently speaking, if $a_{k_i} > 0$ for all i is a priori known, a one-tailed t-test should be applied to the hypothesis testing of H$_0$: $a_{k_i} = 0$ against H$_1$: $a_{k_i} > 0$, where an F test cannot be used. The confidence interval for a_{k_i} is

$$\widehat{a}_{k_i} - t_i^*\widehat{s}_i\sqrt{(\boldsymbol{X}_i'\boldsymbol{X}_i)^{-1}_{k_ik_i}} \leq a_{k_i} \tag{2.3}$$

where t_i^* =one-tailed percentile of a t-distribution. If $\widehat{a}_{k_i} > t_i^*\widehat{s}_i\sqrt{(\boldsymbol{X}_i'\boldsymbol{X}_i)^{-1}_{k_ik_i}}$, then H$_1$ should be adopted and $\boldsymbol{X}_i$ survives. If $\widehat{a}_{k_i} \leq t_i^*\widehat{s}_i\sqrt{(\boldsymbol{X}_i'\boldsymbol{X}_i)^{-1}_{k_ik_i}}$, then H$_0$ should be maintained so that $\boldsymbol{X}_i$ should be regarded as misspecified and ignored. Suppose that Y_{-1} =previous time's agricultural production in $\boldsymbol{X}_i$ for $i = 1, 2$ and a_{ℓ_i} =regression coefficient of Y_{-1}. Y_{-1} affects Y, because the crops harvested in the previous time absorbed lots of mold nutrients and minerals from plots and reduce the current agricultural production ceteris paribus. If $a_{\ell_i} > 0$ is true, no fertilizers are needed. $a_{\ell_i} < 0$ is clear. It is natural to perform the following hypothesis testing: H$_0$: $a_{\ell_i} = 0$ against H$_1$: $a_{\ell_i} < 0$. The.confidence interval of a_{ℓ_i} is

$$a_{\ell_i} \leq \widehat{a}_{\ell_i} + t_i^*\widehat{s}_i\sqrt{(\boldsymbol{X}_i'\boldsymbol{X}_i)^{-1}_{\ell_i\ell_i}}. \tag{2.4}$$

If $\widehat{a}_{\ell_i} < -t_i^*\widehat{s}_i\sqrt{(\boldsymbol{X}_i'\boldsymbol{X}_i)^{-1}_{\ell_i\ell_i}}$, then H$_1$ should be adopted and $\boldsymbol{X}_i$ survives. If $\widehat{a}_{\ell_i} \geq -t_i^*\widehat{s}_i\sqrt{(\boldsymbol{X}_i'\boldsymbol{X}_i)^{-1}_{\ell_i\ell_i}}$, then H$_0$ should be maintained so that $\boldsymbol{X}_i$ should be regarded as misspecified and ignored. In general, if a priori known knowledge for A is available, it must be used to determine the confidence intervals for A.

All test statistics are calculated by the partial-test estimates $\widehat{Y}$ and used for hypothesis testing. If at least one lagged dependent variable is an explanatory variable, a user is interested mainly in the performance of the final-test estimates $\widehat{\widehat{Y}}$. Test statistics are not calculated by $\widehat{\widehat{Y}}$. When all turning points of the observations of a dependent variable Y are not well tracked by the partial-test estimates $\widehat{y}_t$ of Y for some t, a slight departure of a final-test estimate $\widehat{\widehat{y}}_t$ from the first turning point observation y_t of Y tends to promote the departure of $\widehat{\widehat{y}}_t$ from y_t after t through a lagged dependent variable in a regression equation. This is serious from the viewpoint of prediction. Thus, this discrepancy between uses of partial-test and final-test estimates may yield unsatisfactory results. Estimation problems may occur from the reasonable but not-necessarily-perfect characteristics of OLS, multicollinearity or strong correlation of the data of explanatory variables and incompetence of OLS to determine the true values of regression coefficients without knowing the best subset. Thus, it is of great importance to recognize that a "statistically best" regression equation selected solely by statistics or econometrics is not always "scientifically reasonable". Occurrence of some of the data and estimation problems and/or difficulty of uniquely specifying a single regression equation which definitely becomes best usually incur a variable selection problem, i.e., a practically best regression equation problem.

Whenever the Neymann-Pearson approach is taken to hypothesis testings of various statistical tests (normal test, t-test, F test, χ^2 test, Durbin-Watson serial correlation test, Jarque-Bera normality test, Chow equal coefficients test, Goldfeld-Quandt heteroscedasticity test, etc.), a regression equation always becomes subjective, because optimal or best significance levels for them are not known and appropriate significance levels are subjectively set by a user. We cannot say that a regression equation is absolutely best. A best regression equation can be subjectively selected by using professional knowledge, statistical hypothesis testings and data-analytic investigation and can make contributions to substantial sciences. Such a best regression equation is not guaranteed to be absolutely best or globally optimal so that it is called the "practically best regression equation" under the scientific conditions, significance levels for statistical tests and data-analytic conditions inputed by a user. In the process of variable selection, meaningfulness of explanatory variables must have priority over the principle of parsimony and, furthermore, scientific reasonableness of regression coefficients must be investigated before the applications of statistical tests and data-analytic conditions. Subsequently, the meanings or roles of explanatory variables for a dependent variable in a practically best regression equation are guaranteed.

3 A Standard Variable Selection Problem for Regression Analysis

3.1 Notation

We introduce the following notation for time series, cross-sectional and longitudinal (or panel or time-series-and-cross-sectional-pooled) data but explain the notation of criteria in Subsection 3.2 and some in the j-th OLS-best subset problem in Subsection 3.4:

$\hat{}$, $\tilde{}$ ≡indicators for partial and final tests, respectively;

$S \equiv$ number of all cross-sectional units where $s = 1, 2, \cdots, S$ for $S \geq 1$;

$T \equiv$ number of all estimation times where $t = 1, 2, \cdots, T$ for $T \geq 1$;

$P \equiv ST =$ sample size or number of all data of each variable[2] where $P \gg 1$;

$p \equiv S(t-1) + s$ for all s for each of all t where $p = 1, 2, \cdots, P$;

$K \equiv$ number of all possible nonconstant explanatory variables where $k = 1, 2, \cdots, K$;

$y_{st} \equiv \{S(t-1) + s\}$-th datum[3] of a dependent variable for all s and t;

$Y \equiv$ dependent variable or $Y = (Y_1', Y_2', \cdots, Y_T')' = (P \times 1)$-vector of its data y_{st}'s for all s and t;

$Y_{-\ell} \equiv \ell$-th time lagged dependent variable or $Y_{-\ell} = (y_{1,1-\ell}, y_{2,1-\ell}, \cdots, y_{S,1-\ell}, y_{1,2-\ell}, y_{2,2-\ell}, \cdots, y_{S,2-\ell}, \cdots\cdots, y_{1,T-\ell}, y_{2,T-\ell}, \cdots, y_{S,T-\ell})' = (P \times 1)$-vector of its data $y_{s,t-\ell}$'s for all s and t where $\ell = 1, 2, 3, \cdots$;

$X_0 \equiv$ constant term or $X_0 = (1, 1, \cdots, 1)' = (P \times 1)$-vector of 1's;

y ≡arithmetic average of Y, i.e., $y = \sum_{s=1}^{S} \sum_{t=1}^{T} y_{st}/ST = Y'X_0/P$;

$\tau_k \equiv$ known time lag number of a lagged variable of the k-th possible explanatory variable (candidate) if $\tau_k = 1, 2, 3, \cdots$ or $\tau_k = 0$ for the k-th current explanatory variable;

$x^k_{s,t-\tau_k} \equiv \{S(t-1)+s\}$-th datum of the k-th possible explanatory variable concerned with cross-sectional unit s and time $t - \tau_k$ where $x^k_{s,t-0} \equiv x^k_{st}$;

[2] Time series data are characterized by $S = 1$ and $T > 1$, cross-sectional data by $S > 1$ and $T = 1$ and longitudinal data by $S > 1$ and $T > 1$. In the System **OEPP**, longitudinal data are read in the order of $S(t-1) + s$ for all s for each of all t by the parameter FTCD=@ of the ESTI command, while those in the order of $T(s-1) + t$ for all t for each of all s by the parameter FTCD=$.

[3] Let Z be a variable with its observations z_{st}'s and Y be a dependent variable whose data y_{st}'s are transformed by $y_{st} = \ln(z_{st})$ for all s and t. Thus, $Y = \ln(Z)$. Here, Y is called a **transformed dependent variable** concerned with Z, whereas Z is called the **original dependent variable** of Y. The partial-test and/or final-test estimates of Z are eventually desired in most research. The **inverse transformation** means the calculation of $\widehat{z}_{st} = \exp(\widehat{y}_{st})$ and/or $\widetilde{z}_{st} = \exp(\widetilde{y}_{st})$ for all s and t to seek the partial-test estimates $\widehat{z}_{st}$'s and/or final-test estimates $\widetilde{z}_{st}$'s of the original dependent variable Z from the partial-test estimates $\widehat{y}_{st}$'s and/or final-test estimates $\widetilde{y}_{st}$'s of the transformed dependent variable Y, respectively.

$X^k_{t-\tau_k} \equiv (x^k_{1,t-\tau_k}, x^k_{2,t-\tau_k}, \cdots, x^k_{S,t-\tau_k})' = (S \times 1)$-th vector of its data $x^k_{s,t-\tau_k}$'s, where $X^k_{t-0} \equiv X^k_t$;

$X_k \equiv k$-th possible explanatory variable or $X_k = (X^{k\prime}_{1-\tau_k}, X^{k\prime}_{2-\tau_k}, \cdots, X^{k\prime}_{T-\tau_k})' = (P \times 1)$-vector of its data $x^k_{s,t-\tau_k}$'s;

$\boldsymbol{X} \equiv \{X_0, X_1, X_2, \cdots, X_K\}$ = set of all possible explanatory variables or $\boldsymbol{X} = (X_0, X_1, X_2, \cdots, X_K) = \{P \times (K+1)\}$-matrix of their data where if $Y_{-\ell} \in \boldsymbol{X}$, then one of X_k's for all k must be $Y_{-\ell}$;

$i \equiv$ number assigned to a subset or submatrix of $\boldsymbol{X}$ where $i = 1, 2, 3, \cdots, 2^K - 1$;

$\boldsymbol{X}_i \equiv i$-th subset of $\boldsymbol{X}$ or i-th $\{P \times (K_i + 1)\}$-submatrix of $\boldsymbol{X}$, composed of their data $x^{k_i i}_{s,t-\tau_{k_i}}$'s for $k_i = 0, 1, 2, \cdots, K_i$ where $X_0 \in \boldsymbol{X}_i$ for all i if $X_0 \in \boldsymbol{X}$;

$K_i \equiv$ number of nonconstant explanatory variables in $\boldsymbol{X}_i$;

$a_k \equiv$ true regression coefficient corresponding to X_k;

$A \equiv (a_0, a_1, a_2, \cdots, a_K)' = \{(K+1) \times 1\}$-vector of true regression coefficients corresponding to $\boldsymbol{X}$;

$A_i \equiv (a_{0i}, a_{1i}, a_{2i}, \cdots, a_{K_i i})' = \{(K_i + 1) \times 1\}$-vector of true regression coefficients of $\boldsymbol{X}_i$;

$\widehat{A}_i \equiv (\widehat{a}_{0i}, \widehat{a}_{1i}, \widehat{a}_{2i}, \cdots, \widehat{a}_{K_i i})'$ = OLS-estimate of A_i which is estimated, if $\boldsymbol{X}_i$ is meaningful;

$\widehat{\mathcal{C}}(\widehat{A}_i) \equiv$ estimated covariance matrix of $\widehat{A}_i$;

$\widehat{\mathcal{V}}(\widehat{a}_{k_i i}) \equiv$ estimated variance of $\widehat{a}_{k_i i}$, i.e., the $(k_i + 1, k_i + 1)$-element of $\widehat{\mathcal{C}}(\widehat{A}_i)$;

$\widehat{s}_{k_i i} \equiv$ estimated standard deviation (or standard error) of $\widehat{a}_{k_i i}$;

$\widehat{t}_{k_i i} \equiv t$-ratio of $\widehat{a}_{k_i i}$;

$\widehat{y}^i_{st} \equiv$ partial-test-OLS estimate of y_{st} based on $\boldsymbol{X}_i$;

$\widehat{Y}^i_t \equiv (\widehat{y}^i_{1t}, \widehat{y}^i_{2t}, \cdots, \widehat{y}^i_{St})'$ for all t;

$\widehat{Y}_i \equiv (\widehat{Y}^{i\prime}_1, \widehat{Y}^{i\prime}_2, \cdots, \widehat{Y}^{i\prime}_T)'$ = partial-test-OLS estimate of Y based on $\boldsymbol{X}_i$;

$\widehat{e}^i_{st} \equiv y_{st} - \widehat{y}^i_{st}$ = partial-test residual of y_{st} based on $\boldsymbol{X}_i$;

$\widehat{E}^i_t \equiv (\widehat{e}^i_{1t}, \widehat{e}^i_{2t}, \cdots, \widehat{e}^i_{St})'$ for all t;

$\widehat{E}_i \equiv (\widehat{E}^{i\prime}_1, \widehat{E}^{i\prime}_2, \cdots, \widehat{E}^{i\prime}_T)' = (P \times 1)$-vector of partial-test residuals based on $\boldsymbol{X}_i$;

$\widetilde{y}^i_{st}$ = final-test estimate of y_{st} based on $\boldsymbol{X}_i$, if at least one $Y_{-\ell} \in \boldsymbol{X}$ for some ℓ;

$\widetilde{Y}^i_t \equiv (\widetilde{y}^i_{1t}, \widetilde{y}^i_{2t}, \cdots, \widetilde{y}^i_{St})'$ for all t;

$\widetilde{Y}_i \equiv (\widetilde{Y}^{i\prime}_1, \widetilde{Y}^{i\prime}_2, \cdots, \widetilde{Y}^{i\prime}_T)'$ = final-test estimate of Y based on $\boldsymbol{X}_i$;

$\widetilde{e}^i_{st} \equiv y_{st} - \widetilde{y}^i_{st}$ = final-test residual from y_{st} based on $\boldsymbol{X}_i$;

$\widetilde{E}^i_t \equiv (\widetilde{e}^i_{1t}, \widetilde{e}^i_{2t}, \cdots, \widetilde{e}^i_{St})'$ for all t;

$\widetilde{E}_i \equiv (\widetilde{E}^{i\prime}_1, \widetilde{E}^{i\prime}_2, \cdots, \widetilde{E}^{i\prime}_T)' = (P \times 1)$-vector of final-test residuals based on $\boldsymbol{X}_i$;

$u_{st} \equiv$ disturbance of cross-sectional unit s and time t;

$U_t \equiv (u_{1t}, u_{2t}, \cdots, u_{St})' = (S \times 1)$-vector of disturbances u_{st}'s at time t;

$U \equiv$ disturbance term or $U = (U'_1, U'_2, \cdots, U'_T)' = (P \times 1)$-vector of disturbances u_{st}'s for all s and t;

$\sigma^2, \sigma \equiv$ unknown true variance and standard deviation of u_{st} for all s and t, respectively, where $\mathcal{V}(u_{st}) = \sigma^2$ for all s and t;

$\widehat{\sigma}_i^2, \widehat{\sigma}_i \equiv$ OLS-estimates of σ^2 and σ based on $\boldsymbol{X}_i$, respectively;

$\widehat{R}_i^2 \equiv$ (unadjusted) coefficient of determination of $\widehat{Y}_i$;

$\widehat{\mathcal{R}}_i^2 \equiv$ (Theil) adjusted coefficient of determination of $\widehat{Y}_i$;

$\widetilde{IC}_i \equiv$ (Theil) inequality coefficient of $\widetilde{Y}_i$;

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

$\mathrm{I}_n \equiv (n \times n)$-identity matrix for any positive integer n;

$T^\star \equiv$possible nearest time lag number (known) of all possible lagged dependent variables $Y_{-\ell}$'s if they are in $\boldsymbol{X}$ where $T^\star$ is a positive integer, for example, $T^\star = 1$ for annual times series or longitudinal data and $T^\star = 4$ for quarterly times series or longitudinal data;

$T^{\star\star} \equiv$possible farthest time lag number (known) of all possible lagged dependent variables $Y_{-\ell}$'s if they are in $\boldsymbol{X}$ where $T^{\star\star}$ is a positive integer and $T^{\star\star} \geq T^\star$;

$T_i^\star \equiv$possible nearest time lag number of all lagged dependent variables $Y_{-\ell}$'s in $\boldsymbol{X}_i$ if $Y_{-\ell} \in \boldsymbol{X}_i$ so that $Y_{-T_i^\star} \in \boldsymbol{X}_i$ where $T_i^\star \geq T^\star > 0$;

$T_i^{\star\star} \equiv$possible farthest time lag number of all lagged dependent variables $Y_{-\ell}$'s in $\boldsymbol{X}_i$ if $Y_{-\ell} \in \boldsymbol{X}_i$ so that $Y_{-T_i^{\star\star}} \in \boldsymbol{X}_i$ where $T^{\star\star} \geq T_i^{\star\star} \geq T_i^\star \geq T^\star$;

$N \equiv$number of aggregate linear constraints imposed on A;

$\boldsymbol{C} \equiv \{N \times (K+1)\}$-matrix of known coefficients of N aggregate linear constraints related to A;

$\boldsymbol{c} \equiv (N \times 1)$-vector of known values of N aggregate linear constraints;

$N_i \equiv \mathrm{rank}(\boldsymbol{C}_i) =$ number of (individual) linear constraints imposed on A_i which are automatically derived from N aggregate linear constraints[4] where $N_i = 0$ unless constraints are imposed on A_i;

$\boldsymbol{C}_i \equiv \{N_i \times (K_i + 1)\}$-submatrix of $\boldsymbol{C}$, composed of known coefficients of N_i constraints;

$\boldsymbol{c}_i \equiv (N_i \times 1)$-vector of known values of N_i linear constraints;

$P_i \equiv P - K_i - 1 + N_i =$ number of degrees of freedom where $P_i \gg 1$.

[4] Let us introduce some examples of linear constraints on regression coefficients $\widehat{a}_{k_i i}$'s to be estimated. (1) $\sum_{k_i=1}^{K_i} \widehat{a}_{k_i i} = 0$ for no money illusion in a Cobb-Douglas-type demand function estimated with nominal prices and income and (2) $\sum_{k_i=1}^{K_i} \widehat{a}_{k_i i} = 1$ for constant returns to scale in a Cobb-Douglas-type production function.

3.2 Criteria for Scientific Conditions, Statistical and Data-Analytic Tests

A user has to specify the following criteria, if needed, for scientific conditions, statistical and data-analytic tests to solve the j-th OLS-best subset problem[5] :

$\alpha_h^1 =$ a priori known lower bound of the h-th magnitude condition about regression coefficients due to the professional knowledge related to the research at hand;

$\alpha_h^2 =$ a priori known upper bound of the h-th magnitude condition about regression coefficients due to the professional knowledge related to the research at hand;

$\beta =$ significance level (100β %) of a one- or two-tailed t-test for testing hypotheses about regression coefficients ($0 < \beta \ll 1$);

$\gamma =$ significance level (100γ %) of the Durbin-Watson serial correlation test or the Durbin h-test ($0 < \gamma \ll 1$);

$\varepsilon =$ tolerance level for standardized residual test;

$\eta =$ significance level (100η %) of a χ^2-distribution for the Jarque-Bera normality test ($0 < \eta \ll 1$);

$\nu =$ significance level (100ν %) of a two-tailed t-test for a residual outlier ($0 < \nu \ll 1$);

$\psi =$ significance level (100ψ %) of an F distribution for the Chow equal coefficients test ($0 < \psi \ll 1$);

$\omega =$ significance level (100ω %) of an F distribution for the Goldfeld-Quandt homoscedasticity test ($0 < \omega \ll 1$);

$\zeta_p^1 =$ value ($100\zeta_p^1$ %) to define a turning point at $y_{st} \neq 0$ in a partial test ($0 < \zeta_p^1 \ll 1$);

$\zeta_p^2 =$ value to define a turning point at $y_{st} = 0$ in a partial test;

$\zeta_f^1 =$ value ($100\zeta_f^1$ %) to define a turning point at $y_{st} \neq 0$ in a final test ($0 < \zeta_f^1 \ll 1$);

$\zeta_f^2 =$ value to define a turning point at $y_{st} = 0$ in a final test;

$\theta =$ minimum tolerance level of an adjusted coefficient of determination ($0 \ll \theta < 1$);

$\delta =$ maximum tolerance level of the inequality coefficient ($0 < \delta \ll 1$);

$\mathcal{Q} = \{\alpha_h^1$'s, α_h^2's, β, γ, δ, ε, ζ_p^1, ζ_p^2, ζ_f^1, ζ_f^2, η, θ, ν, ψ, $\omega\} =$ criterion set where if some of the scientific conditions and statistical and data-analytic tests are not employed, the corresponding criteria must be eliminated.

[5] The percentiles of normal, χ^2, t-, F tests and the lower and upper limits of the Durbin-Watson serial correlation tests of appropriate degrees of freedom at specified significance levels are automatically obtained and compared with the corresponding test statistics in the System **OEPP**.

Unfortunately, there are no optimal or best criteria to be specified for statistical and data-analytic tests. A user has to use subjectively-most-appropriate or conventionally-used criteria for them. Lower bounds α_h^1's and/or upper bounds α_h^2's of magnitude conditions depend on the professional knowledge related to the research in question. For example, empirically often used criterion values, which may be of course not absolutely appropriate, are β = significance level 0.01 $\sim$ 0.1 (1 $\sim$ 10 %) for a t-test; γ = significance level 0.01 or 0.05 (1 or 5 %) for the Durbin-Watson serial correlation test with subjective acceptance of an inconclusive case; ε =1.65, 2.65 and 2 for standardized residual tolerance levels for times series, cross-sectional and longitudinal data, respectively; η = significance level 0.01 $\sim$ 0.05 (1 $\sim$ 5 %) for the Jarque-Bera normality test; ν = significance level 0.05 $\sim$ 0.1 (5 $\sim$ 10 %) for an outlier t-test; ψ = significance level 0.01 $\sim$ 0.05 (1 $\sim$ 5 %) for the Chow equal coefficients test; ω = significance level 0.01 $\sim$ 0.05 (1 $\sim$ 5 %) for the Goldfeld-Quandt homoscedasticity test; ζ_p^1, ζ_f^1 = slope 0.005 (0.5 %) to define a turning point in partial and final tests, respectively; ζ_p^2, ζ_f^2 = values depending on the data of Y; θ =0.8 $\sim$ 0.9 tolerance level of an adjusted coefficient of determination for ordinary time series or longitudinal data, respectively, where "ordinary" implies that the data of a dependent variable are neither ratios nor rates; θ =0.7 $\sim$ 0.8 tolerance level of an adjusted coefficient of determination for ordinary cross-sectional data, or θ =0.5 $\sim$ 0.6 tolerance level of an adjusted coefficient of determination when the data of a dependent variable are measured by ratios or rates, regardless of the type of data; and δ = 0.3 tolerance level of an inequality coefficient for ordinary time series or longitudinal data or δ = 0.5 tolerance level of an inequality coefficient when the data of a dependent variable are measured by ratios or rates, whether time series or longitudinal data.

3.3 Assumptions

(1) A user must have not necessarily perfect but sufficient professional knowledge about the science related to his research at hand. He must be able to introduce the set $\boldsymbol{X}$ of all possible explanatory variables for the dependent variable Y through the professional knowledge and then classify $\boldsymbol{X}$ to make a computer to generate only the meaningful subsets from $\boldsymbol{X}$ for his research (see Section 5). Furthermore, he must have the professional knowledge about the signs and/or magnitudes of regression coefficients and the nature (for example, stability and structural change or difference) of the system or mechanism focused on, if any.

(2) The functional forms of the regression subequations of all meaningful subsets are linear with respect to the regression coefficients.

(3) The sample size must exceed the number of the constant term and the possible explanatory variables in the smallest meaningful subset of $\boldsymbol{X}$. It is desirable

that the sample size exceeds sufficiently the number of the constant term and the possible explanatory variables in the largest meaningful subset of $\boldsymbol{X}$.

(4) The disturbance term U is normally distributed as $U \sim \mathcal{N}(\mathbf{0}_P, \sigma^2\mathbf{I}_P)$, regardless of i.

(5) $\mathrm{abs}(|\boldsymbol{X}_i'\boldsymbol{X}_i|) > \epsilon$ or $\mathrm{r}(\boldsymbol{X}) = K_i + 1$ for at least one meaningful subset $\boldsymbol{X}_i$ where ϵ=preset or user-specified inverse-matrix-existence criterion value[6] .

(6) $\boldsymbol{X}$ is nonstochastic or independent of U if $\boldsymbol{X}$ is stochastic.

(7) The principle of minimizing the sum, $(Y-\boldsymbol{X}_i\widehat{A}_i)'(Y-\boldsymbol{X}_i\widehat{A}_i)$, of squared errors with respect to $\widehat{A}_i$ is suitable for the research, regardless of i.

3.4 The J-th OLS-Best Subset Problem as a Standard Variable Selection Problem

So far, the best regression equation estimated by OLS has not been defined in the literature. A variable selection problem for OLS has not been concretely proposed yet. Onishi (1983) proposed a procedure to search for the practically best regression equation for OLS in a run of a computer. However, the effectiveness of the procedure was limited. We reformulate the j-th OLS-best subset problem for regression analysis, where a solution is called the j-th practically best regression subequation as follows.

The J-th OLS-Best Subset Problem

Obtain the j-th practically best regression subequation $\widehat{Y}_i = \boldsymbol{X}_i\widehat{A}_i$ as a solution in a run of a computer by searching for subset $\boldsymbol{X}_i$ from set $\boldsymbol{X}$ of all possible explanatory variables specified for the dependent variable Y and estimating the true regression coefficient vector A_i of $\boldsymbol{X}_i$ and the true variance σ^2 and standard deviation σ of the disturbance term U and calculating the standard deviations and t-ratios of $\widehat{A}_i$ and other important statistics under the criterion set $\mathcal{Q}$, i.e., $\{\alpha_h^1$'s, α_h^2's, β, γ, δ, ε, ζ_p^1, ζ_p^2, ζ_f^1, ζ_f^2, η, θ, ν, ψ, $\omega\}$ specified by a user such that

[I] $\boldsymbol{X}_i$ is meaningful for Y from the viewpoints of the professional knowledge related to the research at hand;

[II] $\widehat{A}_i$, $\widehat{\mathcal{C}}(\widehat{A}_i)$, $\widehat{\mathcal{V}}(\widehat{a}_{k_i i})$, $\widehat{Y}_i$, $\widehat{E}_i$, $\widehat{\sigma}_i^2$, $\widehat{\sigma}_i$, $\widehat{s}_{k_i i}$ and $\widehat{t}_{k_i i}$ are calculated as follows:

$$\widehat{Y}_i = \boldsymbol{X}_i\widehat{A}_i, \quad \widehat{E}_i = Y - \widehat{Y}_i, \quad \widehat{\sigma}_i^2 = \frac{\widehat{E}_i'\widehat{E}_i}{P_i},$$

[6] If $\mathrm{abs}(|\boldsymbol{X}_i'\boldsymbol{X}_i|) \le \epsilon$ for a preset or user-specified small ϵ, the computer regards $\boldsymbol{X}_i$ as having multicollinearity, ignores $\boldsymbol{X}_i$ and goes to the next meaningful subset. The total number of occurrences of multicollinearity is printed by the System **OEPP**, if multicollinearity happens.

$$\widehat{\sigma}_i = \sqrt{\widehat{\sigma}_i^2}, \quad \widehat{s}_{k_i i} = \sqrt{\widehat{\mathcal{V}}(\widehat{a}_{k_i i})} \quad \text{and} \quad \widehat{t}_{k_i i} = \frac{\widehat{a}_{k_i i}}{\widehat{s}_{k_i i}},$$

(i) in a case where no constraints are imposed on A_i

$$\widehat{A}_i = (\boldsymbol{X}_i'\boldsymbol{X}_i)^{-1}\boldsymbol{X}_i'Y, \quad \widehat{\mathcal{C}}(\widehat{A}_i) = \widehat{\sigma}_i^2(\boldsymbol{X}_i'\boldsymbol{X}_i)^{-1},$$

or (ii) in a case where a linear constraint set $\boldsymbol{C}_i A_i = \boldsymbol{c}_i$ is imposed on A_i

$$\widehat{A}_i = (\boldsymbol{X}_i'\boldsymbol{X}_i)^{-1}\left[\boldsymbol{X}_i'Y + \boldsymbol{C}_i'\{\boldsymbol{C}_i(\boldsymbol{X}_i'\boldsymbol{X}_i)^{-1}\boldsymbol{C}_i'\}^{-1}\{\boldsymbol{c}_i - \boldsymbol{C}_i(\boldsymbol{X}_i'\boldsymbol{X}_i)^{-1}\boldsymbol{X}_i'Y\}\right],$$

$$\widehat{\mathcal{C}}(\widehat{A}_i) = \widehat{\sigma}_i^2\left[(\boldsymbol{X}_i'\boldsymbol{X}_i)^{-1} - (\boldsymbol{X}_i'\boldsymbol{X}_i)^{-1}\boldsymbol{C}_i'\{\boldsymbol{C}_i(\boldsymbol{X}_i'\boldsymbol{X}_i)^{-1}\boldsymbol{C}_i'\}^{-1}\boldsymbol{C}_i(\boldsymbol{X}_i'\boldsymbol{X}_i)^{-1}\right];$$

[**III**] $\widehat{A}_i$ must satisfy all or some of the following scientific conditions (*i*), (*ii*) and/or (*iii*) concerned with regression coefficients, if required from the viewpoints of the professional knowledge related to the research[7] :

(*i*) in a case where the following a priori known lower and/or upper bounds of arithmetic expressions of regression coefficients must be met:

$$\alpha_{h_i}^1 \leq f_{h_i}(\widehat{A}_i), \quad f_{h_i}(\widehat{A}_i) \leq \alpha_{h_i}^2 \quad \text{or} \quad \alpha_{h_i}^1 \leq f_{h_i}(\widehat{A}_i) \leq \alpha_{h_i}^2 \quad \text{for } h_i = 1, 2, \cdots, H_i$$

where $f_{h_i}(\widehat{A}_i)$ which denotes a function of $\widehat{A}_i$ (which is linear with respect to $\widehat{A}_i$ in most cases) and $\alpha_{h_i}^1 \leq f_{h_i}(\widehat{A}_i)$, $f_{h_i}(\widehat{A}_i) \leq \alpha_{h_i}^2$ and $\alpha_{h_i}^1 \leq f_{h_i}(\widehat{A}_i) \leq \alpha_{h_i}^2$ are generated from aggregate magnitude conditions $\alpha_h^1 \leq f_h(\widehat{A})$, $f_h(\widehat{A}) \leq \alpha_h^2$ and $\alpha_h^1 \leq f_h(\widehat{A}) \leq \alpha_h^2$ for $h = 1, 2, \cdots, H$ loaded into the computer, respectively;

[7] The Schur stability condition (*ii*) and the concavity or convexity (*iii*) can be eventually expressed by (*i*). However, it is clear and convenient to introduce them in an easily usable way. In the System **OEPP**, the a priori known sign conditions are inputed in a functional form in which a dependent variable and all possible classified explanatory variables with/without the positive or negative signs, i.e., + or −, of their regression coefficients are entered. For instance, $Y = F(X_0 < 2 < +X_1, -X_2 > 2 >< 0 < X_3, +X_4, X_5 > 3 ><< X_6, X_7, X_8, X_9, +X_{10} > 1 >)$ where the a priori known signs of the regression coefficients of X_1, X_4 and X_{10} are positive, that of X_2 is negative and those of X_0, X_3, X_5, X_6, X_7, X_8 and X_9 are unavailable, implying that they can assume either positive or negative real values. To be logically interpreted, a regression subequation with some or all of X_1, X_2, X_4 and X_{10} must have the corresponding a priori signed regression coefficients. It is regarded as unsatisfactory, otherwise.

To load the regression coefficients related to magnitude conditions into a computer, their variable notation are used. Only aggregate magnitude conditions corresponding to A are loaded into a computer through $f_h(\boldsymbol{X})$ and the individual magnitude conditions corresponding to A_i needed to evaluate regression subequations are automatically derived by the computer. Accordingly, $f_{h_i}(\widehat{A}_i)$ is recognized through $f_{h_i}(\boldsymbol{X}_i)$ which is derived from $f_h(\boldsymbol{X})$ which implies $f_h(\widehat{A})$ for $h = 1, 2, \cdots, H$. For instance, in consumption function $Y = a_0 + a_1X_1 + a_2X_2 + \cdots + U$, marginal propensity to consume $\partial Y/\partial X_1 = a_1$ is positive but less than 1, if consumers behave rationally, where Y =consumption in value and X_1 =net income. Subsequently, $0 < X_1 < 1$ or $1 > X_1 > 0$ is inputed. Hence, $0 < \widehat{a}_{1i} < 1$ must be met in the j-th practically best regression subequation.

(*ii*) in a case where the following Schur stability condition[8] consisting of $T_i^{\star\star}$ strict inequalities must be met, if (i) $Y_{-\ell} \in \boldsymbol{X}_i$ for some integer $\ell \geq 1$, (ii) $\boldsymbol{X}_i$ is expressed as (3.14) in **Case 2** of [**XI**] defined below and (iii) it is known that Y does not diverge by itself[9] as $t \to +\infty$, i.e., the absolute values of all $T_i^{\star\star}$ eigenvalues of the characteristic equation of the $T_i^{\star\star}$-th order of ℓ

$$\ell^{T_i^{\star\star}} + \widehat{b}_{T_i^{\star\star}-T_i^{\star},i}\ell^{T_i^{\star\star}-T_i^{\star}} + \widehat{b}_{T_i^{\star\star}-T_i^{\star}-1,i}\ell^{T_i^{\star\star}-T_i^{\star}-1} + \cdots + \widehat{b}_{1i}\ell + \widehat{b}_{0i} = 0 \qquad (3.5)$$

for (3.14) in **Case 2** of [**XI**] must be less than 1[10] :

$$|\widehat{\boldsymbol{A}}_{k_i i}| \equiv \begin{vmatrix} \widehat{\boldsymbol{A}}^{*}_{k_i i} & \widehat{\boldsymbol{A}}^{**\prime}_{k_i i} \\ \widehat{\boldsymbol{A}}^{**}_{k_i i} & \widehat{\boldsymbol{A}}^{*\prime}_{k_i i} \end{vmatrix} > 0 \quad \text{for } k_i = 1, 2, \cdots, T_i^{\star\star}$$

for

$$\widehat{\boldsymbol{A}}^{*}_{k_i i} \equiv \begin{pmatrix} 1 & \widehat{b}_{T_i^{\star\star}-1,i} & \widehat{b}_{T_i^{\star\star}-2,i} & \cdots & \widehat{b}_{k_i i} \\ 0 & 1 & \widehat{b}_{T_i^{\star\star}-1,i} & \cdots & \widehat{b}_{k_i+1,i} \\ 0 & 0 & 1 & \cdots & \widehat{b}_{k_i+2,i} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \end{pmatrix}$$

and

$$\widehat{\boldsymbol{A}}^{**}_{k_i i} \equiv \begin{pmatrix} \widehat{b}_{0i} & \widehat{b}_{1i} & \widehat{b}_{2i} & \cdots & \widehat{b}_{T_i^{\star\star}-k_i,i} \\ 0 & \widehat{b}_{0i} & \widehat{b}_{1i} & \cdots & \widehat{b}_{T_i^{\star\star}-k_i-1,i} \\ 0 & 0 & \widehat{b}_{0i} & \cdots & \widehat{b}_{T_i^{\star\star}-k_i-2,i} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & \widehat{b}_{0i} \end{pmatrix}$$

where $\widehat{b}_{T_i^{\star\star}-k_i,i} \equiv 0$ for $k_i = 1, 2, \cdots, T_i^{\star} - 1$ if $T_i^{\star} \geq 2$; $\widehat{b}_{k_i i} \equiv -\widehat{a}_{K_i-k_i,i}$ for $k_i = 0, 1, 2, \cdots, T_i^{\star\star} - T_i^{\star}$; and $\widehat{\boldsymbol{A}}_{k_i i} = \{2(T_i^{\star\star} - k_i + 1) \times 2(T_i^{\star\star} - k_i + 1)\}$-matrix for $k_i = 0, 1, 2, \cdots, T_i^{\star\star} - T_i^{\star}$;
and/or
(*iii*) if the regression subequation $Y = \boldsymbol{X}_i A_i + U$ is expressed as $Y = a_{0i} +$

[8] The Schur stability condition is related to a difference equation, while the Routh-Hurwitz stability condition is related to a differential equation. The author was not able to get the Schur's original paper but learned his stability condition in Hibino (1962). Needless to say, $\ell^{T_i^{\star\star}}$, $\ell^{T_i^{\star\star}-T_i^{\star}}$, $\ell^{T_i^{\star\star}-T_i^{\star}-1}$, $\cdots$, ℓ and the constant term $\widehat{b}_{0i}$ in (3.5) correspond to $\widehat{y}^{i}_{st}$, $y_{s,t-T_i^{\star}}$, $y_{s,t-T_i^{\star}-1}$, $\cdots$, $y_{s,t-T_i^{\star\star}+1}$ and $y_{s,t-T_i^{\star\star}}$ in (3.14), respectively.

[9] Suppose, for example, that the current real consumption Y is affected by the current real income X and the past consumptions Y_{-1} and Y_{-2} as an inertia effect of eating habit and real incomes are kept constant over time where $\widehat{Y} = \widehat{a}_0 + \widehat{a}_1 X + \widehat{a}_2 Y_{-1} + \widehat{a}_3 Y_{-2}$. It is unrealistic and impossible that the consumption becomes unlimited as time passes.

[10] A solution part to the difference equation (3.14) converges to zero.

$\boldsymbol{Z}_i \varPsi_i \boldsymbol{Z}_i' + \boldsymbol{Z}_i \varPhi_i + U$[11] and must be concave or convex with respect to before-transformation explanatory variables $\boldsymbol{Z}_i$ from the viewpoints of research at hand, then the estimated $\widehat{\varPsi}_i$ must be positive definite or negative definite, respectively, depending on the scientific knowledge of the research,
where $\widehat{\varPsi}_i$ is positive definite, if all eigenvalues or the determinants of all principal submatrices of $\widehat{\varPsi}_i$ are positive and $\widehat{\varPsi}_i$ is negative definite, if all eigenvalues of $\widehat{\varPsi}_i$ are negative;

[IV] the Jarque-Bera normality test statistic $\widehat{JB}_i$ must satisfy the following inequality in order to maintain the null hypothesis $\mathrm{H}_0 : U$ is normally distributed with the expectation $\mathcal{E}(U) = \boldsymbol{0}_P$ at a 100η % significance level of a χ^2 test[12] [13] :

$$\widehat{JB}_i = P\left\{\frac{\widehat{\mathcal{S}}_i^2}{6} + \frac{(\widehat{\mathcal{K}}_i - 3)^2}{24}\right\} \le \chi_2^2(\eta)$$

for

$$\widehat{\mathcal{S}}_i^2 = \frac{\{\sum_{s=1}^S \sum_{t=1}^T (\widehat{e}_{st}^i)^3/P\}^2}{(\widehat{E}_i'\widehat{E}_i/P)^3} \quad \text{and} \quad \widehat{\mathcal{K}}_i = \frac{\sum_{s=1}^S \sum_{t=1}^T (\widehat{e}_{st}^i)^4/P}{(\widehat{E}_i'\widehat{E}_i/P)^2}$$

where $\chi_2^2(\eta) \doteq \eta$ percentile of a χ^2 distribution 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 at a 100β % significance level of a t-test, depending on the purpose of the research, if necessary[14]

[11] For instance, an average cost function is concave. Suppose that $\widehat{Y}_i = \widehat{a}_{0i} + \widehat{a}_{1i}X_1 + \widehat{a}_{2i}X_3 + \widehat{a}_{3i}X_5 + \widehat{a}_{4i}X_7 + \widehat{a}_{5i}X_9 + \widehat{a}_{6i}X_{11} + \widehat{a}_{7i}X_{12} + \widehat{a}_{8i}X_{13}$ for $\boldsymbol{X}_i = \{X_0, X_1, X_3, X_5, X_7, X_9, X_{11}, X_{12}, X_{13}\}$ means $\widehat{Y}_i = \widehat{a}_{0i} + \widehat{a}_{1i}Z_1^2 + \widehat{a}_{2i}Z_3^2 + \widehat{a}_{3i}Z_5^2 + \widehat{a}_{4i}Z_1Z_3 + \widehat{a}_{5i}Z_3Z_5 + \widehat{a}_{6i}Z_1 + \widehat{a}_{7i}Z_3 + \widehat{a}_{8i}Z_5$ and must be concave, then

$$\widehat{\varPsi}_i = \begin{pmatrix} \widehat{a}_{1i} & \widehat{a}_{4i}/2 & 0 \\ \widehat{a}_{4i}/2 & \widehat{a}_{2i} & \widehat{a}_{5i}/2 \\ 0 & \widehat{a}_{5i}/2 & \widehat{a}_{3i} \end{pmatrix}$$

must be positive definite with respect to $\boldsymbol{Z}_i = \{Z_1, Z_3, Z_5\}$, that is, $\widehat{\varPsi}_i$ must satisfy the following conditions: (i) $\widehat{a}_{1i} > 0$, (ii) $\widehat{a}_{2i} > 0$, (iii) $\widehat{a}_{3i} > 0$, (iv) $\widehat{a}_{1i}\widehat{a}_{2i} - \widehat{a}_{4i}^2/4 > 0$, (v) $\widehat{a}_{2i}\widehat{a}_{3i} - \widehat{a}_{5i}^2/4 > 0$ and (vi) $\widehat{a}_{1i}\widehat{a}_{2i}\widehat{a}_{3i} - \widehat{a}_{3i}\widehat{a}_{4i}^2/4 - \widehat{a}_{1i}\widehat{a}_{5i}^2/4 > 0$, where $\varPhi_i = (\widehat{a}_{6i}, \widehat{a}_{7i}, \widehat{a}_{8i})'$. $\boldsymbol{X}_i$ is actually used to estimate $Y = \boldsymbol{X}_i A_i + U$ but a user must know the following relations: $X_1 = Z_1^2$, $X_3 = Z_3^2$, $X_5 = Z_5^2$, $X_7 = Z_1Z_3$, $X_9 = Z_3Z_5$, $X_{11} = Z_1$, $X_{12} = Z_3$ and $X_{13} = Z_5$.

[12] $\widehat{\mathcal{S}}_i$ and $\widehat{\mathcal{K}}_i - 3$ correspond to skewness and kurtosis, respectively. For instance, $\chi_2^2(0.1) \doteq 4.60517$; $\chi_2^2(0.05) \doteq 5.99147$; $\chi_2^2(0.025) \doteq 7.377776$; and $\chi_2^2(0.01) \doteq 4.605170$.

[13] The Shapiro-Wilk normality test is alternatively available (Shapiro and Wilk, 1965).

[14] If the null hypothesis H_0 is expected to be maintained, $\mathrm{G}_d'\boldsymbol{X} = \mathrm{g}_d$ for $\mathrm{g}_d = (\mathrm{g}_0^d, \mathrm{g}_1^d, \mathrm{g}_2^d, \cdots, \mathrm{g}_K^d)' \neq \boldsymbol{0}_{K+1}$ with known elements g_k^d's for all $k = 0, 1, 2, \cdots, K$ is loaded into a computer from which $\mathrm{G}_{id_i}'\boldsymbol{X}_i = \mathrm{g}_{d_i}$ is derived, implying $\mathrm{G}_{id_i}'A_i = \mathrm{g}_{d_i}$, where $\mathrm{g}_k^d X_k$'s of only nonzero g_k^d's are actually inputed and then G_d is formed. The alternative hypothesis $\mathrm{H}_1 : \mathrm{G}_{id_i}'A_i \neq \mathrm{g}_{d_i}$ is recognized through $\mathrm{G}_{id_i}'\boldsymbol{X}_i \# \mathrm{g}_d$ which is derived from $\mathrm{G}_d'\boldsymbol{X} \# \mathrm{g}_d$ loaded into a computer. It is ruled in the System **OEPP** that if $\mathrm{G}_d'\boldsymbol{X} > \mathrm{g}_d$, $\mathrm{G}_d'\boldsymbol{X} < \mathrm{g}_d$ or $\mathrm{G}_d'\boldsymbol{X} \# \mathrm{g}_d$ is loaded into a computer, the corresponding d_i-th alternative hypothesis H_1 for A_i is expected to be adopted for the j-th practically best regression subequation, whereas if $\mathrm{G}_d'\boldsymbol{X} = \mathrm{g}_d$ is loaded, the

[15] [16] : for $d_i = 1, 2, \cdots, D_i$,

(i) to adopt H_1 for $\mathrm{H}_0 : \mathrm{G}'_{id_i} A_i = \mathrm{g}_{d_i}$ against $\mathrm{H}_1 : \mathrm{G}'_{id_i} A_i \neq \mathrm{g}_{d_i}$,

$$\frac{|\mathrm{G}'_{id_i}\widehat{A}_i - \mathrm{g}_{d_i}|}{\widehat{S}_{id_i}} > t_{P_i}(\beta/2),$$

(ii) to adopt H_1 for $\mathrm{H}_0 : \mathrm{G}'_{id_i} A_i = \mathrm{g}_{d_i}$ against $\mathrm{H}_1 : \mathrm{G}'_{id_i} A_i > \mathrm{g}_{d_i}$,

$$\frac{\mathrm{G}'_{id_i}\widehat{A}_i - \mathrm{g}_{d_i}}{\widehat{S}_{id_i}} > t_{P_i}(\beta),$$

(iii) to adopt H_1 for $\mathrm{H}_0 : \mathrm{G}'_{id_i} A_i = \mathrm{g}_{d_i}$ against $\mathrm{H}_1 : \mathrm{G}'_{id_i} A_i < \mathrm{g}_{d_i}$,

$$\frac{\mathrm{g}_{d_i} - \mathrm{G}'_{id_i}\widehat{A}_i}{\widehat{S}_{id_i}} > t_{P_i}(\beta),$$

or

(iv) to maintain H_0 for $\mathrm{H}_0 : \mathrm{G}'_{id_i} A_i = \mathrm{g}_{d_i}$ against $\mathrm{H}_1 : \mathrm{G}'_{id_i} A_i \neq \mathrm{g}_{d_i}$ where at least 2 elements are usually not zero in G_{id_i},

$$\frac{|\mathrm{G}'_{id_i}\widehat{A}_i - \mathrm{g}_{d_i}|}{\widehat{S}_{id_i}} \leq t_{P_i}(\beta/2),$$

corresponding d_i-th null hypothesis H_0 for A_i is expected to be maintained for the j-th practically best regression subequation, depending on the purpose of the research.

[15] For instance, the hypothesis testing of H_0: $a_{k_i i} = 0$ against H_1: $a_{k_i i} \neq 0$ concerned with regression coefficient $a_{k_i i}$ can be made by $\mathrm{G}_{id_i} = (0'_{k_i}, 1, 0'_{K_i - k_i})'$ and $\mathrm{g}_{d_i} = 0$. Since $\widehat{S}_{id_i} = \sqrt{\mathrm{G}'_{id_i}\widehat{\mathcal{C}}(\widehat{A}_i)\mathrm{G}_{id_i}} = \sqrt{\widehat{\mathcal{V}}(\widehat{a}_{k_i i})} = \widehat{s}_{k_i i}$, we have $|\mathrm{G}'_{id_i}\widehat{A}_i - \mathrm{g}_{d_i}|/\widehat{S}_{id_i} = |\widehat{a}_{k_i i}|/\widehat{s}_{k_i i} = |\widehat{t}_{k_i i}| = t$-ratio in the condition [**II**].

As explained in Section 5, the hypothesis testing for each regression coefficient is closely tied up with the sign attached to the corresponding (possible) explanatory variable in a functional form for estimation in the System **OEPP**. Suppose that $Y = F(X_0, < 3 < X_1, +X_2, -X_3 > 3 >, \cdots)$ for variable selection or $Y = F(X_0, X_1, +X_2, -X_3, \cdots)$ for estimation of a single regression equation is inputed. Each of all meaningful subsets or a meaningful set always has the non-signed variable X_1 and the signed variables $+X_2$ and $-X_3$ and the others like $\boldsymbol{X}_i$ or $\boldsymbol{X} = \{X_0, X_1, +X_2, -X_3, \cdots\}$. In variable selection, it is expected to adopt H_1 for (i) $\mathrm{H}_0 : a_{1i} = 0$ against $\mathrm{H}_1 : a_{1i} \neq 0$ set up by $X_1 \in \boldsymbol{X}_i$, (ii) $\mathrm{H}_0 : a_{2i} = 0$ against $\mathrm{H}_1 : a_{2i} > 0$ set up by $+X_2 \in \boldsymbol{X}_i$ and (iii) $\mathrm{H}_0 : a_{3i} = 0$ against $\mathrm{H}_1 : a_{3i} < 0$ set up by $-X_3 \in \boldsymbol{X}_i$. In estimation of a single regression equation, it is expected to adopt H_1 for (i) $\mathrm{H}_0 : a_1 = 0$ against $\mathrm{H}_1 : a_1 \neq 0$ set up by $X_1 \in \boldsymbol{X}$, (ii) $\mathrm{H}_0 : a_2 = 0$ against $\mathrm{H}_1 : a_2 > 0$ set up by $+X_2 \in \boldsymbol{X}$ and (iii) $\mathrm{H}_0 : a_3 = 0$ against $\mathrm{H}_1 : a_3 < 0$ set up by $-X_3 \in \boldsymbol{X}$. The command RCHT with the parameter of a significance level (or a percentile) conducts all necessary hypothesis testings for the regression coefficients of all (possible) explanatory variables (and a constant term).

[16] It is possible to use a different significance level $100\beta_d$ % for each hypothesis testing. In this case, $t_{P_i}(\beta/2)$ and $t_{P_i}(\beta)$ must be replaced with $t_{P_i}(\beta_{d_i}/2)$ and $t_{P_i}(\beta_{d_i})$, respectively. Thus, $\mathcal{Q} = \{\alpha_h^1\text{'s}, \alpha_h^2\text{'s}, \beta_d\text{'s}, \gamma, \delta, \varepsilon, \zeta_p^1, \zeta_p^2, \zeta_f^1, \zeta_f^2, \eta, \theta, \nu, \psi, \omega\}$.

for

$$(\widehat{S}_{id_i})^2 = \mathrm{G}'_{id_i}\widehat{C}(\widehat{A}_i)\mathrm{G}_{id_i} \quad \text{and} \quad \widehat{S}_{id_i} = \sqrt{(\widehat{S}_{id_i})^2}$$

where $\mathrm{G}_{id_i} = \{(K_i+1)\times 1\}$-vector of known coefficients of the d_i-th hypothesis concerned with A_i; g_{d_i} = a priori known value of the d_i-th hypothesis; $t_{P_i}(\beta/2)$ = percentile of a two-tailed t-test of a 100β % significance level with P_i degrees of freedom; $t_{P_i}(\beta)$ = percentile of a one-tailed t-test of a 100β % significance level with P_i degrees of freedom; and $\mathrm{G}'_{id_i}A_i \neq \mathrm{g}_{d_i}$, $\mathrm{G}'_{id_i}A_i > \mathrm{g}_{d_i}$ and $\mathrm{G}'_{id_i}A_i < \mathrm{g}_{d_i}$ are generated from aggregate hypothetical relations $\mathrm{G}'_dA \neq \mathrm{g}_d$, $\mathrm{G}'_dA > \mathrm{g}_d$ and $\mathrm{G}'_dA < \mathrm{g}_d$ loaded into the computer for $d = 1, 2, \cdots, D$, respectively;

[VI] the Durbin-Watson serial correlation test statistic $\widehat{DW}_i$ defined below must satisfy the following inequality at a 100γ % significance level, if time series data are used, i.e., $S = 1$ (Durbin and Watson, 1950, 1951, 1971):

(i) if no lagged dependent variables are included in $\boldsymbol{X}_i$ for $T > 6$, then

$$\widehat{DW}_i > d_i(\gamma) \quad \text{if} \quad \widehat{DW}_i \leq 2 \quad \text{or} \quad 4 - \widehat{DW}_i \geq d_i(\gamma) \quad \text{if} \quad \widehat{DW}_i > 2$$

for annual data ($m = 1$) or for quarterly data ($m = 4$) (K. F. Wallis, 1972)

$$\widehat{DW}_i = \frac{\sum_{t=1+m}^{T}(\widehat{e}^{\,i}_{1t} - \widehat{e}^{\,i}_{1,t-m})^2}{\sum_{t=1}^{T}(\widehat{e}^{\,i}_{1t})^2}$$

or

(ii) if $Y_{-\ell} \in \boldsymbol{X}_i$ and $T \cdot (\widehat{\sigma}^{\,i}_{y_\ell})^2 < 1$ (Durbin, 1970 and Godfrey, 1978), then

$$|h_i| = \frac{|\sum_{t=1+m}^{T}\widehat{e}^{\,i}_{1t}\widehat{e}^{\,i}_{1,t-m}|}{\sum_{t=1+m}^{T}(\widehat{e}^{\,i}_{1t})^2}\sqrt{\frac{T}{1 - T(\widehat{\sigma}^{\,i}_{y_\ell})^2}} \leq h(\gamma)$$

where $d_i(\gamma) = d^{u}_{T,K_i+1-N_i}(\gamma)$ if an inconclusive case is regarded as subjectively unacceptable or $d_i(\gamma) = d^{\ell}_{T,K_i+1-N_i}(\gamma)$ if an inconclusive case is regarded as subjectively acceptable; $d^{u}_{T,K_i+1-N_i}(\gamma)$ = upper limit of the Durbin-Watson serial correlation test of a 100γ % significance level with (T, K_i+1-N_i) degrees of freedom; $d^{\ell}_{T,K_i+1-N_i}(\gamma)$ = its lower limit; $(\widehat{\sigma}^{\,i}_{y_\ell})^2$ = estimated variance of the regression coefficient of the ℓ-th time lagged dependent variable $Y_{-\ell}$ (often $\ell = 1$ for annual data and $\ell = 4$ for quarterly data); and $h(\gamma)$ = percentile of a normal test, for example $h(0.05) = 1.645$ and $h(0.01) = 2.325$;

[VII] the Chow equal coefficients test statistic $\widehat{C}_i$ must satisfy the following inequality at a 100ψ % significance level of an F test, if (i) it is considered that structural changes (e.g., oil crises) may have happened M_c times or structural difference (e.g., cold versus hot areas) may have existed M_c cross-sectional subgroups, (ii) the Chow equal coefficients test needs to be applied and (iii) all regression coefficients ($\widehat{A}_i^{*m}$ for $m = 1, 2, \cdots, M_c + 1$, expressed in column vectors) estimated with M_c sample submatrices of $(Y, \boldsymbol{X}_i)$ satisfy the a priori known sign and/or magnitude conditions[17] like [**III**] when they are available[18] :

$$\widehat{C}_i = \frac{(\widehat{E}_i'\widehat{E}_i - \widehat{E}_i^{+\prime}\widehat{E}_i^{+})/\{M_c(K_i + 1 - N_i)\}}{\widehat{E}_i^{+\prime}\widehat{E}_i^{+}/\{P - (M_c + 1)(K_i + 1 - N_i)\}} \leq F_{P-(M_c+1)(K_i+1-N_i)}^{M_c(K_i+1-N_i)}(\psi)$$

where (i) if it is considered that structural changes which affect A_i may have happened M_c times, then the estimation period of T times must be partitioned into $(M_c + 1)$ T_c^m-subperiods for $ST_c^m > K_i + 1 - N_i$ and $T = \sum_{m=1}^{M_c+1} T_c^m$ for $m = 1, 2, \cdots, M_c + 1$ or (ii) if it is considered that structural changes which affect A_i may have happened in M_c cross-sectional-unit subgroups, then all data y_{st}'s and $x_{s,t-\tau_{k_i}}^{k_i i}$'s for all s, t and k_i must be rearranged in Y and $\boldsymbol{X}_i$ by the ordering rule $T(s-1)+t$ for all s and t and, furthermore, S cross-sectional units must be partitioned into $(M_c + 1)$ S_c^m-cross-sectional-unit subgroups for $TS_c^m > K_i + 1 - N_i$ and $S = \sum_{m=1}^{M_c+1} S_c^m$ for $m = 1, 2, \cdots, M_c + 1$; $P^m = ST_c^m$ or $S_c^m T$ for all m; $(Y, \boldsymbol{X}_i)$ is partitioned into $(M_c + 1)$ $\{P^m \times (K_i + 2)\}$-submatrices $(Y^m, \boldsymbol{X}_i^m)$ for all m so that $Y = (Y^{1\prime}, Y^{2\prime}, \cdots, Y^{(M_c+1)\prime})'$ and $\boldsymbol{X}_i = (\boldsymbol{X}_i^{1\prime}, \boldsymbol{X}_i^{2\prime}, \cdots, \boldsymbol{X}_i^{(M_c+1)\prime})'$; $\widehat{E}_i^{+} = (\widehat{E}_i^{1\prime}, \widehat{E}_i^{2\prime}, \cdots, \widehat{E}_i^{(M_c+1)\prime})'$; $\widehat{E}_i^m = (P^m \times 1)$-vector of residuals resulted from regressing Y^m on $\boldsymbol{X}_i^m$ with/without constraint set $\boldsymbol{C}_i A_i^{*m} = \boldsymbol{c}_i$ for each m; and $F_{P-(M_c+1)(K_i+1-N_i)}^{M_c(K_i+1-N_i)}(\psi) = \psi$ percentile of an F distribution with $\{M_c(K_i + 1 - N_i), P - (M_c + 1)(K_i + 1 - N_i)\}$ degrees of freedom;

[**VIII**] the Goldfeld-Quandt homoscedasticity test statistic $\widehat{GQ}_i$ concerned with variance σ^2 of the disturbance term U must satisfy the following inequality at a 100ω % significance level of an F test, if (i) it is considered that structural changes or differences may have affected σ^2, (ii) the Goldfeld-Quandt

[17] These sign and/or magnitude conditions are not necessarily the same as those in the above [**III**]. It is sometimes observed that the regression coefficient of the variable "real income" for the consumption of interior goods is positive while the economy is poor. However, it becomes negative, after the people become rich because they have a tendency to buy the superior substitutes. This can be explained as follows. Let the P samples be partitioned into two sample groups ST_c^1 and ST_c^2 such as $P = ST_c^1 + ST_c^2$ and $a_{k_i i}^{*}$, $a_{k_i i}^{**}$ and $a_{k_i i}$ be the above regression coefficients for the ST_c^1-, ST_c^2- and P-sample groups, respectively. The sign conditions are $a_{k_i i}^{*} > 0$ in the first ST_c^1 samples, $a_{k_i i}^{**} < 0$ in the last ST_c^2 samples and $a_{k_i i} \neq 0$, i.e., no sign condition on $a_{k_i i}$ in the whole P samples. As far as hypothesis testing is concerned, $\mathrm{H}_0 : a_{k_i i}^{*} = 0$ against $\mathrm{H}_1 : a_{k_i i}^{*} > 0$ in the first ST_c^1 samples, $\mathrm{H}_0 : a_{k_i i}^{**} = 0$ against $\mathrm{H}_1 : a_{k_i i}^{**} < 0$ in the last ST_c^2 samples and $\mathrm{H}_0 : a_{k_i i} = 0$ against $\mathrm{H}_1 : a_{k_i i} \neq 0$ in the whole P samples may be plausible. For instance, the rice consumption in Japan before and after around 1962 is an example. If one of the a priori known sign and/or magnitude conditions is not met in one of $(M_c + 1)$ regressions, the Chow equal coefficients test should not be applied. The same logic is available for the condition [**VIII**].

[18] A case may happen in which some of the regression coefficients in A_i are affected by structural changes but the remaining are not. Chow also referred to this case. However, it may be difficult to a priori know in actual research which regression coefficients are affected by the structural changes and which regression coefficients are not.

homoscedasticity test needs to be applied and (iii) all regression coefficients ($\widehat{A}_i^*$ and $\widehat{A}_i^{**}$ are expressed in column vectors) estimated with M_v sample submatrices of $(Y, \boldsymbol{X}_i)$ satisfy the a priori known sign and/or magnitude conditions like [**III**] when they are available:

$$\widehat{GQ}_i = \frac{\widehat{E}_i^{*\prime}\widehat{E}_i^*}{\widehat{E}_i^{**\prime}\widehat{E}_i^{**}} \leq F_{Q-K_i-1+N_i}^{Q-K_i-1+N_i}(\omega)$$

where it is considered that structural changes may have happened and affected the variances $\mathcal{V}(u_{st})$ of the disturbance term; $(Y, \boldsymbol{X}_i)$ is rearranged into $(Y_*, \boldsymbol{X}_{*i})$ by the order of the unknown but user-properly-guessed magnitudes of variances $\mathcal{V}(u_{st})$'s and then partitioned into three submatrices $(Y^*, \boldsymbol{X}_i^*)$, $(Y_c^*, \boldsymbol{X}_{ci}^*)$ and $(Y^{**}, \boldsymbol{X}_i^{**})$ in such a way that the dimension Q of $(Y^*, \boldsymbol{X}_i^*)$ is equal to that of $(Y^{**}, \boldsymbol{X}_i^{**})$ for user-specified Q such that $Q > K_i+1-N_i$, $P = 2Q+Q_c$ and $Q_c \geq 0$ so that $Y_* = (Y^{*\prime}, Y_c^{*\prime}, Y^{**\prime})'$ and $\boldsymbol{X}_{*i} = (\boldsymbol{X}_i^{*\prime}, \boldsymbol{X}_{ci}^{*\prime}, \boldsymbol{X}_i^{**\prime})'$; $\widehat{E}_i^* = (Q \times 1)$-vector of residuals resulted from regressing Y^* on the first Q samples $\boldsymbol{X}_i^*$ of $\boldsymbol{X}_{*i}$ with/without constraint set $\boldsymbol{C}_i A_i^* = \boldsymbol{c}_i$; $\widehat{E}_i^{**} = (Q \times 1)$-vector of residuals resulted from regressing Y^{**} on the last Q samples $\boldsymbol{X}_i^{**}$ of $\boldsymbol{X}_{*i}$ with/without constraint set $\boldsymbol{C}_i A_i^{**} = \boldsymbol{c}_i$; $(Y_c^*, \boldsymbol{X}_{ci}^*)$ =central submatrix to be omitted; $Q_c =$ first dimension of $\dim(Y_c^*, \boldsymbol{X}_{ci}^*) = \{Q_c \times (K_i + 2)\}$; and $F_{Q-K_i-1+N_i}^{Q-K_i-1+N_i}(\omega) = \omega$ percentile of an F distribution of $(Q - K_i - 1 + N_i, Q - K_i - 1 + N_i)$ degrees of freedom;

[**IX**] (*i*) $\widehat{E}_i$ must satisfy the following residual outlier t-test[19] [20] (Sawa, 1979) at a 100ν % significance level for a rather small sample size, if necessary:

$$\max_{s,t} \widehat{OT}_{st}^{\,i} \leq t_{P_i-1}(\nu/2P) \quad \text{for } s = 1, 2, \cdots, S \text{ and } t = 1, 2, \cdots, T$$

for

$$\widehat{OT}_{st}^{\,i} = \frac{|\widehat{e}_{st}^{\,i}|/\sqrt{1 - \widehat{q}_{pp}^{\,i}}\ .}{\sqrt{\widehat{E}_i'\widehat{E}_i - (\widehat{e}_{st}^{\,i})^2/(1 - \widehat{q}_{pp}^{\,i})}/\sqrt{P_i - 1}} \quad \text{for } p = S(t-1) + s$$

[19] $\widehat{E}_i = \{\mathrm{I}_P - \boldsymbol{X}_i(\boldsymbol{X}_i'\boldsymbol{X}_i)^{-1}\boldsymbol{X}_i'\}U \sim \mathcal{N}\left(\mathbf{0}_P, \sigma^2\{\mathrm{I}_P - \boldsymbol{X}_i(\boldsymbol{X}_i'\boldsymbol{X}_i)^{-1}\boldsymbol{X}_i'\}\right)$ in an unconstrained case or $\widehat{E}_i = [\mathrm{I}_P - \boldsymbol{X}_i(\boldsymbol{X}_i'\boldsymbol{X}_i)^{-1}\boldsymbol{X}_i' + \boldsymbol{X}_i(\boldsymbol{X}_i'\boldsymbol{X}_i)^{-1}\boldsymbol{C}_i'\{\boldsymbol{C}_i(\boldsymbol{X}_i'\boldsymbol{X}_i)^{-1}\boldsymbol{C}_i'\}^{-1}\boldsymbol{C}_i \times(\boldsymbol{X}_i'\boldsymbol{X}_i)^{-1}\boldsymbol{X}_i']U \sim \mathcal{N}\left(\mathbf{0}_P, \sigma^2\ [\mathrm{I}_P - \boldsymbol{X}_i(\boldsymbol{X}_i'\boldsymbol{X}_i)^{-1}\boldsymbol{X}_i' - \boldsymbol{X}_i(\boldsymbol{X}_i'\boldsymbol{X}_i)^{-1}\boldsymbol{C}_i' \times\{\boldsymbol{C}_i(\boldsymbol{X}_i'\boldsymbol{X}_i)^{-1}\boldsymbol{C}_i'\}^{-1}\boldsymbol{C}_i(\boldsymbol{X}_i'\boldsymbol{X}_i)^{-1}\boldsymbol{X}_i']\right)$ in a constrained case. It turns out that $\overline{e}_{st}^{\,i} \equiv \widehat{e}_{st}^{\,i}/\sigma\sqrt{1 - \widehat{q}_{pp}^{\,i}} \sim \mathcal{N}(0, 1)$ for $p = S(t-1) + s$.

[20] The larger S and/or T become, the smaller will $\nu/2P$ become. Hence, $t_{P_i-1}(\nu/2P)$ becomes larger so that the outlier t-test becomes weaker and a real outlier cannot be detected, even if it exists. One remedy is to use the number of residuals exceeding a user-specified appropriate ε in (*ii*) of [**IX**] as a substitute for the P of $\nu/2P$ in (*i*).

and/or
(ii) all standardized residuals $\widehat{\widehat{e}}^{i}_{st}$'s defined below must not exceed the user-specified criterion value ε[21] , if necessary[22] :

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

for

$$\widehat{\widehat{e}}^{i}_{st} \equiv \frac{\widehat{e}^{i}_{st}}{\widehat{\sigma}_i \sqrt{1 - \widehat{q}^{i}_{pp}}} \quad \text{for } p = S(t-1) + s$$

where $t_{P_i - 1}(\nu/2P) = \nu/2P$ percentile of a two-tailed t-test with $P_i - 1$, i.e., $P - K_i + N_i - 2$ degrees of freedom; $\widehat{q}^{i}_{pp} = (p, p)$-, i.e., $\{S(t-1)+s, S(t-1)+s\}$-diagonal element of $\boldsymbol{X}_i(\boldsymbol{X}'_i \boldsymbol{X}_i)^{-1}\boldsymbol{X}'_i$ in an unconstrained case or $\boldsymbol{X}_i(\boldsymbol{X}'_i\boldsymbol{X}_i)^{-1} \times \boldsymbol{X}'_i - \boldsymbol{X}_i(\boldsymbol{X}'_i\boldsymbol{X}_i)^{-1}\boldsymbol{C}'_i\{\boldsymbol{C}_i(\boldsymbol{X}'_i\boldsymbol{X}_i)^{-1}\boldsymbol{C}'_i\}^{-1}\boldsymbol{C}_i(\boldsymbol{X}'_i\boldsymbol{X}_i)^{-1}\boldsymbol{X}'_i$ in a constrained case;

[X] $\widehat{Y}_i$ must satisfy the following turning point test defined by user-specified ζ^1_p and ζ^2_p, if (i) time series or longitudinal data are used ($T \geq 3$) and (ii) it is necessary[23] :
if

$$(y_{st} - y_{s,t-n})(y_{s,t+n} - y_{st}) < 0 \tag{3.6}$$

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

[22] 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.

[23] It is quite important to track turning points of y_{st}'s by $\widehat{y}_{st}$'s (and $\widetilde{y}_{st}$'s) in applications, especially, in prediction. A turning point at time t can be mathematically defined by the inequality (3.6)., However, from the practical viewpoint, (3.7) and (3.8) should be introduced. Let us consider 2 cases: 3 annual data (dollars) of consumption $(y^a_{s,t-1}, y^a_{st}, y^a_{s,t+1})$ = (999999999, 1000000000, 999999998) and 3 annual data (%) of unemployment rate $(y^b_{s,t-1}, y^b_{st}, y^b_{s,t+1}) = (3, 2, 4)$ for some s where $n = 1$. Both y^a_{st}'s and y^b_{st}'s satisfy (3.6), because $(y^a_{st} - y^a_{s,t-1})(y^a_{s,t+1} - y^a_{st}) = -2 < 0$ and $(y^b_{st} - y^b_{s,t-1})(y^b_{s,t+1} - y^b_{st}) = -2 < 0$. However, the $\wedge$-turning point indicated by y^a_{st}'s is so flat that the impact of such a turning point on the economy is not serious. Empirically speaking, it is almost impossible to well track such a flat turning point by a regression equation. On the other hand, the $\vee$-turning point indicated by y^b_{st}'s is so steep that the impact on the society is large. In order to disregard a flat turning point, (3.7) and (3.8) are introduced. For instance, if $\zeta^1_p = 0.01$ (1 %) is specified, only the $\vee$-turning point of y^b_{st}'s is required to be tracked well by the j-th practically best regression subequation, because $\min\{|1 - y^a_{s,t-1}/y^a_{st}|, |1 - y^a_{s,t+1}/y^a_{st}|\} = 0.000000001 < \zeta^1_p = 0.01$ (disregarding the $\wedge$-turning point by y^a_{st}'s) and $\min\{|1 - y^b_{s,t-1}/y^b_{st}|, |1 - y^b_{s,t+1}/y^b_{st}|\} = 0.5 > \zeta^1_p = 0.01$ (regarded as the $\vee$-turning point by y^b_{st}'s).

It is easy to find the j-th practically best regression subequation, if the data, $y_{s1}, y_{s2}, \cdots, y_{sT}$, of a dependent variable increase or decrease monotonically over time for each s, implying that no turning point exists. However, it is quite difficult to search for the j-th practically best regression subequation, if there are many $\wedge$- and/or $\vee$-turning points. The final-test estimates $\widetilde{y}^i_{s1}, \widetilde{y}^i_{s2}, \cdots, \widetilde{y}^i_{sT}$ go away from $y_{s1}, y_{s2}, \cdots, y_{sT}$ after the time of the first turning point which the partial-test estimates $\widehat{y}^i_{s1}, \widehat{y}^i_{s2}, \cdots, \widehat{y}^i_{sT}$ failed to well track, if $Y_{-1} \in \boldsymbol{X}_i$.

and

$$\min\left\{\left|1-\frac{y_{s,t-n}}{y_{st}}\right|,\left|1-\frac{y_{s,t+n}}{y_{st}}\right|\right\}\geq\zeta_p^1 \quad \text{for } y_{st}\neq 0 \tag{3.7}$$

or

$$\min\{|y_{s,t-n}|,|y_{s,t+n}|\}\geq\zeta_p^2 \quad \text{for } y_{st}=0, \tag{3.8}$$

then

$$(y_{st}-y_{s,t-n})(\widehat{y}_{st}^i-\widehat{y}_{s,t-n}^i)>0 \tag{3.9}$$

and

$$(y_{s,t+n}-y_{st})(\widehat{y}_{s,t+n}^i-\widehat{y}_{st}^i)>0 \tag{3.10}$$

for $s=1,2,\cdots,S$, $t=1+n,2+n,\cdots,T-n$ and $n=T_i^*,T_i^*+1,\cdots,T_i^{**}$ when $Y_{-T_i^*}\in\boldsymbol{X}_i$ and $Y_{-T_i^{**}}\in\boldsymbol{X}_i$ or $n=1$ when no lagged dependent variables are in $\boldsymbol{X}_i$;

and

[**XI**] one of the following three cases must be met with respect to a fitting measure[24] specified by the user:

(i) **Case 1 in which no lagged dependent variables are included in** $\boldsymbol{X}$**:** the specified adjusted coefficient of determination $\widehat{\mathcal{R}}_i^2$ of $\widehat{Y}_i$ defined below is greater than or equal to θ and the j-th largest among the specified adjusted coefficients of determination of the subsets which satisfy (i) all those applied from among the conditions [**I**] to [**X**] and (ii) (3.11) in [**XI**], depending on the type of data (time series, cross-sectional or longitudinal data):

$$\widehat{\mathcal{R}}_i^2=\max\left\{1-\frac{(1-\widehat{R}_i^2)(P-1)}{P_i},0\right\}\geq\theta \tag{3.11}$$

or

$$\widehat{R}_i^2=1-\frac{\widehat{E}_i'\widehat{E}_i}{(Y-yX_0)'(Y-yX_0)}\geq\theta, \tag{3.12}$$

(ii) **Case 2[25] in which at least one lagged dependent variable is an**

[24] $0\leq\widehat{\mathcal{R}}_i^2\leq 1$. $\widehat{\mathcal{R}}_i^2=1$ implies perfect fitting of $\widehat{Y}_i$ (partial-test estimate) to Y, whereas $\widehat{\mathcal{R}}_i^2=0$ implies the worst fitting. As $\widehat{\mathcal{R}}_i^2\to 1$, the degree of fitting monotonically increases.

The Akaike information criterion statistic $\widehat{AIC}_i=P\{\ln 2\pi+1+\ln(\widehat{E}_i'\widehat{E}_i/P)\}+2(K_i-N_i+2)$ can be used instead of $\widehat{\mathcal{R}}_i^2$ and the phrase "the j-th largest" must be replaced with "the j-th smallest" (Akaike, 1973). $P(\ln 2\pi+1)+4$ is not essential, because it is constant where a_{0i} and σ^2 are counted as parameters to be determined in K_i-N_i+2. $-\infty<\widehat{AIC}_i<+\infty$. $\widehat{AIC}_i=-\infty$ implies perfect fitting of $\widehat{Y}_i$ to Y, whereas $\widehat{AIC}_i=+\infty$ implies the worst fitting. As $\widehat{AIC}_i\to-\infty$, the degree of fitting monotonically increases.

[25] Suppose that Y is a transformed dependent variable, Z is the original dependent variable of Y with its observations z_{st}'s and $\widetilde{Z}_i$ is the i-th final-test estimate of the original dependent variable

important variable[26] **in X** so that all meaningful subsets include such a lagged dependent variable where we assume that, without loss of generality, $\boldsymbol{X}_i = \{X_0, X_{1i}, X_{2i}, \cdots, X_{K_i^* i}, Y_{-T_i^*}, Y_{-(T_i^*+1)}, \cdots, Y_{-T_i^{**}}\}$ for $T_i^{**} \geq T_i^* \geq 1$; $X_{k_i i} \equiv$ k_i-th explanatory variable in $\boldsymbol{X}_i$ or $X_{k_i i} \equiv (X_{1-\tau_{k_i}}^{k_i i\prime}, X_{2-\tau_{k_i}}^{k_i i\prime}, \cdots, X_{T-\tau_{k_i}}^{k_i i\prime})' = (P \times 1)$-vector and $X_{t-\tau_{k_i}}^{k_i i} \equiv (x_{1,t-\tau_{k_i}}^{k_i i}, x_{2,t-\tau_{k_i}}^{k_i i}, \cdots, x_{S,t-\tau_{k_i}}^{k_i i})' = (S \times 1)$-vector of its data $x_{s,t-\tau_{k_i}}^{k_i i}$'s for $k_i = 1, 2, \cdots, K_i^*$; $K_i = K_i^* + T_i^{**} - T_i^* + 1$; $\tau_{k_i} \equiv$ known time lag number if $\tau_{k_i} > 0$ or 0 otherwise ($X_{k_i i}$ is a current variable); $k_i^* \equiv K_i^* - T_i^* + 1$; $\widehat{b}_{k_i i} \equiv -\widehat{a}_{K_i - k_i, i}$ for $k_i = 0, 1, 2, \cdots, T_i^{**} - T_i^*$; and the i-th estimated regression subequation is

$$
\begin{aligned}
\widehat{y}_{st}^i &= \widehat{a}_{0i} + \sum_{k_i=1}^{K_i} \widehat{a}_{k_i i} x_{s,t-\tau_{k_i}}^{k_i i} = \widehat{a}_{0i} + \sum_{k_i=1}^{K_i^*} \widehat{a}_{k_i i} x_{s,t-\tau_{k_i}}^{k_i i} + \sum_{t_i=T_i^*}^{T_i^{**}} \widehat{a}_{k_i^*+t_i, i} y_{s,t-t_i} \quad (3.13) \\
&\equiv \widehat{a}_{0i} + \sum_{k_i=1}^{K_i^*} \widehat{a}_{k_i i} x_{s,t-\tau_{k_i}}^{k_i i} - \sum_{t_i=T_i^*}^{T_i^{**}} \widehat{b}_{T_i^{**}-t_i, i} y_{s,t-t_i} : \quad (3.14)
\end{aligned}
$$

the inequality coefficient[27] $\widetilde{IC}_i$ of $\widetilde{Y}_i$ defined below to measure the degree of fitting of the final-test estimates $\widetilde{Y}_i$ to Y must not exceed the criterion value δ and is the j-th lowest among the inequality coefficients of the subsets which satisfy (i) all those applied from among the conditions [I] to [X], (ii) (3.11) in [XI], (iii) (3.15) in [XI] and (iv) a turning point test for $\widetilde{Y}_i$ characterized by (3.16) to (3.20), depending on the type of data (time series or longitudinal data):

$$
\widetilde{IC}_i = \min\left\{ \frac{\sqrt{(Y - \widetilde{Y}_i)'(Y - \widetilde{Y}_i)/P}}{\sqrt{\widetilde{Y}_i'\widetilde{Y}_i/P + Y'Y/P}}, 1 \right\} \leq \delta \qquad (3.15)
$$

Z with its final-test estimates $\widetilde{z}_{st}^i$'s where $Y = f(Z)$ and $y_{st} = f(z_{st})$ so that $Z = f^{-1}(Y)$ and $z_{st} = f^{-1}(y_{st})$. The performance of $\widetilde{Z}_i$ is often more interesting than that of $\widetilde{Y}_i$. In this case, Y, y_{st}, $y_{s,t-n}$, $y_{s,t+n}$, $\widetilde{y}_{st}^i$, $\widetilde{y}_{s,t-n}^i$ and $\widetilde{y}_{s,t+n}^i$ can be replaced with Z, z_{st}, $z_{s,t-n}$, $z_{s,t+n}$, $\widetilde{z}_{st}^i$, $\widetilde{z}_{s,t-n}^i$ and $\widetilde{z}_{s,t+n}^i$ in (3.15) to (3.20), respectively.

[26] As explained in Subsection 5.2, a set of all possible nonconstant explanatory variables specified for a dependent variable can be always classified into some sets of combinatorial (explanatory) variables and/or some sets of sequential (explanatory) variables. When at least one of the variables in a set must be selected in all meaningful subsets, the variables in this set are called "important variables". On the other hand, if it is possible to generate all meaningful subsets even when none of the variables in a set is selected, the variables in this set are called "optional variables". When all possible lagged dependent variables Y_{-T^*}, $Y_{-(T^*+1)}$, $\cdots$, $Y_{-T^{**}}$ in $\boldsymbol{X}$ are sequential and important, they are classified as $< m < Y_{-T^*}, Y_{-(T^*+1)}, \cdots, Y_{-T^{**}} >>$ or $<< Y_{-T^{**}}, Y_{-(T^{**}-1)}, \cdots, Y_{-T^*} > m >$ for some m of $m = 1, 2, \cdots, T^{**} - T^* + 1$. When all possible lagged dependent variables Y_{-T^*}, $Y_{-(T^*+1)}$, $\cdots$, $Y_{-T^{**}}$ are sequential and optional, they are classified as $< 0 < Y_{-T^*}, Y_{-(T^*+1)}, \cdots, Y_{-T^{**}} >>$ or $<< Y_{-T^{**}}, Y_{-(T^{**}-1)}, \cdots, Y_{-T^*} > 0 >$.

[27] $0 \leq \widetilde{IC}_i \leq 1$. $\widetilde{IC}_i = 0$ implies perfect fitting of $\widetilde{Y}_i$ (final-test estimate) to Y, whereas $\widetilde{IC}_i = 1$ implies the worst fitting. As $\widetilde{IC}_i \to 0$, the degree of fitting monotonically increases.

for $\widetilde{Y}_i$ satisfying the following turning point test defined by user-specified ζ_f^1 and ζ_f^2:
if

$$(y_{st} - y_{s,t-n})(y_{s,t+n} - y_{st}) < 0 \tag{3.16}$$

and

$$\min\left\{\left|1 - \frac{y_{s,t-n}}{y_{st}}\right|, \left|1 - \frac{y_{s,t+n}}{y_{st}}\right|\right\} \geq \zeta_f^1 \quad \text{for } y_{st} \neq 0 \tag{3.17}$$

or

$$\min\{|y_{s,t-n}|, |y_{s,t+n}|\} \geq \zeta_f^2 \quad \text{for } y_{st} = 0, \tag{3.18}$$

then

$$(y_{st} - y_{s,t-n})(\widetilde{y}^i_{st} - \widetilde{y}^i_{s,t-n}) > 0 \tag{3.19}$$

and

$$(y_{s,t+n} - y_{st})(\widetilde{y}^i_{s,t+n} - \widetilde{y}^i_{st}) > 0 \tag{3.20}$$

for $s = 1, 2, \cdots, S$, $t = 1 + n, 2 + n, \cdots, T - n$ and $n = T_i^*, T_i^* + 1, \cdots, T_i^{**}$,

where $\widetilde{Y}_i$ is calculated as follows:

$$\begin{aligned} \widetilde{y}^i_{st} &= \widehat{y}^i_{st} = \widehat{a}_{0i} + \sum_{k_i=1}^{K_i^*} \widehat{a}_{k_i i} x^{k_i i}_{s,t-\tau_{k_i}} - \sum_{t_i=T_i^*}^{T_i^{**}} \widehat{b}_{T_i^{**}-t_i,i} y_{s,t-t_i} \\ &\quad \text{for } s = 1, 2, \cdots, S \text{ and } t = 1, 2, \cdots, T_i^*, \end{aligned}$$

$$\begin{aligned} \widetilde{y}^i_{st} &= \widehat{a}_{0i} + \sum_{k_i=1}^{K_i^*} \widehat{a}_{k_i i} x^{k_i i}_{s,t-\tau_{k_i}} - \sum_{t_i=T_i^*}^{t-1} \widehat{b}_{T_i^{**}-t_i,i} \widetilde{y}^i_{s,t-t_i} - \sum_{t_i=t}^{T_i^{**}} \widehat{b}_{T_i^{**}-t_i,i} y_{s,t-t_i} \\ &\quad \text{for } s = 1, 2, \cdots, S \text{ and } t = T_i^* + 1, T_i^* + 2, \cdots, T_i^{**} \text{ if } T_i^* < T_i^{**} \end{aligned}$$

and

$$\begin{aligned} \widetilde{y}^i_{st} &= \widehat{a}_{0i} + \sum_{k_i=1}^{K_i^*} \widehat{a}_{k_i i} x^{k_i i}_{s,t-\tau_{k_i}} - \sum_{t_i=T_i^*}^{T_i^{**}} \widehat{b}_{T_i^{**}-t_i,i} \widetilde{y}^i_{s,t-t_i} \\ &\quad \text{for } s = 1, 2, \cdots, S \text{ and } t = T_i^{**} + 1, T_i^{**} + 2, \cdots, T \end{aligned}$$

or

(iii) Case 3 in which all lagged dependent variables are optional variables in X so that some meaningful subsets include at least one lagged dependent variable but the others do not include any lagged dependent variables: the specified adjusted coefficient of determination $\widehat{\mathcal{R}}_i^2$ of $\widehat{Y}_i$ is greater than or equal to θ and the j-th largest among the specified adjusted coefficients of determination of the subsets which satisfy (i) all those applied from among the conditions [**I**] to [**X**] and (*ii*) (3.11) in [**XI**] and, furthermore, (*iii*) (3.15) in [**XI**] and (*iv*) a turning point test for $\widetilde{Y}_i$ characterized by (3.16) to (3.20) only if the subsets include at least one lagged dependent variable, depending on the type of data (time series or longitudinal data). □

A standard variable selection problem for regression analysis or OLS is defined as the (first) OLS-best subset problem under an appropriate criterion set $\mathcal{Q}$ where $j = 1$. In the usage of time series data, a standard variable selection problem is characterized by the conditions [**I**] to [**XI**]. In the usage of cross-sectional data, a standard variable selection problem is characterized by the conditions [**I**] to [**V**], [**VII**] to [**IX**] and **Case 1** in [**XI**]. In the usage of longitudinal data, a standard variable selection problem is characterized by the conditions [**I**] to [**V**] and [**VII**] to [**XI**].

If a user has a new scientific condition(s), statistical test(s) and/or data-analytic test(s) which are not cited in the j-th OLS-best subset problem, he can add them to a standard variable selection problem before the last condition [**XI**] and then develop software.

4 Remarks on a Solution, Diagnosis and Prediction

One should not offer users unsatisfactory regression equations as the best. A distrust against statistics and econometrics may be instilled in users' minds. Therefore, thorough scrutiny is needed to search for the practically best regression equation. Although many statistical tests have been proposed, there is no guidance on how systematically they should be applied for regression analysis. There is no comprehensive evaluation function of various statistical tests. If a user has not only sufficient professional knowledge about his research but also rich experiences in model building and knows appropriate criterion values, then a solution to the (first) OLS-best subset problem ($j = 1$) should be adopted as the practically best regression equation. However, if he has a new test not cited here or does not know appropriate criterion values, he should solve, for example, the first 3 or 5 OLS-best subset problems ($j = 1, 2, 3$ or $j = 1, 2, 3, 4, 5$) in a run of a computer and determine the practically best regression equation by using the new test or comparing the selected (at most) 3 or 5 practically best regression subequations together with their calculated test statistics with each other. This may be an easy job because the selected regression subequations have already passed all of the conditions in the third or fifth OLS-best subset problem. Of course, if a severe criterion value is set for a statistical or data-analytic test, no practically best regression subequation may exist. In this case, good software automatically gives a user a message about the condition up till which at least one regression subequation has passed all early conditions and the next condition in which all regression subequations failed to pass. The message becomes a good diagnosis for him. It is suggested to solve the first j OLS-best subset problems under rather weak criterion values at first. Then, if he comes to notice more appropriate criterion values, he should resolve the (first) OLS-best subset problem ($j = 1$) with new criterion values. Thus, he must be an expert in the field of his research.

If a user or programmer wants to develop software to solve the j-th OLS-best subset problem, he should enable it to handle more flexible criteria in addition to the criteria in the j-th OLS-best subset problem. For instance, software had better have the function of being able to apply a t-test of a $100\beta_k$ % significance level or a directly loaded percentile (critical point value) t_k^{β} to the k-th regression coefficient for all $k = 1, 2, \cdots, K$ where $0 < \beta_k \ll 1$. Furthermore, the software should be designed to solve the first j OLS-best subset problems for each of M different dependent variables (for instance, consumption, investment, import, production, etc. in an economic system) for any j and M where M kinds of the j-th OLS-best subset problem are solved in a run of a computer.

Needless to say, the estimated regression coefficients, the estimates and predicted values of a dependent variable, residuals, t-ratios and even test statistics are functions of the criterion values set by a user, i.e., $\widehat{a}_{k_i i} = f^{a}_{k_i i}(\mathcal{Q})$, $\widehat{y}^{i}_{st} = f^{yi}_{st}(\mathcal{Q})$, $\widehat{e}^{i}_{st} = f^{ei}_{st}(\mathcal{Q})$, $\widehat{t}_{k_i i} = f^{t}_{k_i i}(\mathcal{Q})$, $\widehat{DW}_i = f_i^{DW}(\mathcal{Q})$, etc. Accordingly, if one of the criterion values is altered, a different j-th practically best regression subequation may be obtained. Conventional criterion values are usually specified. It must be kept in mind that the phrase "practically best" is used in this sense.

The conditions [**I**] and [**III**] stem from scientific requirements, the conditions [**II**], [**IV**] to [**IX**] and [**XI**] are statistical and the condition [**X**] is rather data-analytic. With respect to each of all meaningful subsets, all conditions are sequentially investigated in the order of the conditions [**I**] to [**XI**]. The order of the conditions [**I**] to [**XI**] in the j-th OLS-best subset problem seems to be plausible. First of all, the condition [**I**] must be satisfied for any kind of research. Then as the second condition [**II**], the regression coefficients and their variances, standard deviations and t-ratios, variance and standard deviation of a disturbance term and important test statistics should be calculated only for the subsets which were judged meaningful for the research. The condition [**III**] handles hypotheses about regression coefficients which cannot be put or are difficult to be put under statistical tests and should be cleared before all statistical and data-analytic tests are applied. The order of the conditions [**IV**] to [**X**] does not matter. However, the condition [**IV**] should be applied before the conditions [**V**], [**VII**], [**VIII**] and [**IX**], because they are related to a normal distribution. It is convenient to apply the condition [**VI**] before the conditions [**VII**], [**VIII**] and [**IX**], because a serial correlation test is always used for time series data. The conditions [**VII**] and [**VIII**] are usually used when the external information is available for structural changes or differences. The condition [**X**] is optional and should be used whenever at least one lagged dependent variable is an explanatory one. If the condition [**X**] is not met by the partial test, the turning point(s) will not be tracked well in the final test and the prediction will fail.

The condition [XI] for fitting must come at last. It must be noted that the conditions [I] and [III] to [X] are used to simply judge whether or not calculated values or test statistics satisfy specified criteria (simply, pass or failure) but not to see how good they are. The condition [XI] examines how relatively good a degree of fitting is if it satisfies a certain level indicated by θ and/or δ and is used to rank the regression subequations of the subsets which have passed all conditions applied prior to it, of course depending on the type of data, the presence or absence of a lagged dependent variable, etc. Needless to say, a prediction must be made by the practically best regression equation.

If a subset is meaningless, the computer ignores it. If the regression subequation of a meaningful subset does not satisfy a condition, then the computer checks whether or not it has ever satisfied most conditions, memorizes the aborted condition if so, and indicates until which scientific, statistical or data-analytic condition was satisfied as part of a diagnosis in the printout if the practically best regression subequation does not exist.

Whenever a t-test for hypothesis testing on regression coefficients, the Chow equal coefficients test, the Goldfeld-Quandt homoscedasticity test and/or an outlier t-test are employed, the Jarque-Bera normality test should be applied, because they are based on a normal distribution. Roughly speaking, if a 100η % Jarque-Bera normality test, a 100β % hypothesis testing, a 100γ % Durbin-Watson serial correlation test or Durbin h-test, a 100ψ % Chow equal coefficients test, a 100ω % Goldfeld-Quandt homoscedasticity test and a 100ν % outlier t-test are applied, the practically best regression equation comes to be obtained with the **total significance level** 100ϕ %, where $\phi = 1 - (1-\eta)(1-\beta)(1-\gamma)(1-\psi)(1-\omega)(1-\nu)$. If $\eta = \beta = \gamma = \psi = \omega = \nu = 0.05$, then the total significance level ϕ or 100ϕ % becomes 0.2649 or 26.49 %, respectively. In other words, there is a 26.49 % risk that the practically best regression equation was actually not best.

5 Processing of Professional Knowledge

5.1 A Priori Known Signs of Regression Coefficients

So far statistics has not been able to solve the j-th OLS-best subset problem. We need informatic statistics or econometrics to solve it in which not only statistics but also professional knowledge or information are systematically interwoven. It is decisively important to create a methodology and algorithm to satisfy the conditions [I] and [III] in the j-th OLS-best subset problem, no matter what research is conducted. Fortunately, it is possible. The information about the signs of some regression coefficients is often available from the professional knowledge related to the research at hand. A user may want to test hypotheses about the signs of some regression coefficients. For instance, it is unrealistic that production becomes optimal ceteris paribus, when people do not work at all if the estimated regression coefficient of the variable "labor" is negative in the production function or the demand quantity for personal computers is drastically increased, when the price is set extremely high ceteris paribus if the estimated regression coefficient of the variable "price" is positive in the demand function. There are many situations as above in which the signs of the regression coefficients of some explanatory variables are a priori known. In these situations, 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 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.

5.2 Scientific Variable Classifcation

Double variable classifications needs to be applied to all possible nonconstant explanatory variables. They are (*i*) single or grouped variables and (*ii*) combinatorial or sequential variables and must be employed to solve the j-th OLS-best subset problem for regression analysis.

5.2.1 Single or Grouped Variables

A **single variable** is defined as one which has its meanings or role by itself in interpreting the regression equation in which it is included. Most explanatory variables are usually treated as single variables. On the other hand, a **grouped variable** is defined as one which cannot have its clear meanings or role by itself but can have it only when it is used together with the other appropriate variables. Such grouped variables are called **a set of grouped variables**. The grouped variables in a set usually represent (*i*) complementary relations such as a pair of patients' expenditures on medical doctors' services and prescribed medicines bought in pharmacies or (*ii*) a choice of an aggregate variable or all componentwise variables such as a total score, X, of an entrance examination (single variable) versus scores, X_1, X_2, $\cdots$, X_K, of K individual subjects (grouped variables) like mathematics, economics, biology and literature where $X = \sum_{k=1}^{K} X_k$.

In order to distinguish a single variable with/without the sign from a set of grouped variables with/without the signs of their regression coefficients, we postulate that the latter is enclosed within parentheses (and) like $(\diamondsuit X_1, \diamondsuit X_2, \cdots, \diamondsuit X_K)$ and, furthermore, in variable selection, it is treated just like a single variable and the parentheses are ignored in the subsets which include the set of grouped variables.

Let $\mathcal{X}$ represent a single variable with/without the sign of its regression coefficient, $\diamondsuit X$, or a set of grouped variables with/without the signs of their regression coefficients, $(\diamondsuit X_1, \diamondsuit X_2, \cdots, \diamondsuit X_K)$, for some $K = 2, 3, \cdots$. $\mathcal{X}$ is called a **condensed variable** and introduced to compactly express variable classifications. For instance, if $\mathcal{X} = +X$, then $\mathcal{X}$ means variable X whose regression coefficient must be positive, implying that it is a priori known or to be hypothetically tested that an increase in X increases the dependent variable, ceteris paribus, and vice versa. Furthermore, if a user wants to apply a t-test, then $+X$ requires a one-tailed t-test but not a two-tailed t-test, implying that an F test is inappropriate. If $\mathcal{X} = X$, then $\mathcal{X}$ means variable X whose regression coefficient can assume a positive or negative value and requires a two-tailed t-test or an F test.

Example 1:

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

implies that $\mathcal{X}_1 = -X_1$, $\mathcal{X}_2 = (+X_2, +X_3)$, $\mathcal{X}_3 = +X_4$, $\mathcal{X}_4 = X_5$ and $\mathcal{X}_5 = (-X_6, X_7, +X_8)$. There are 8 explanatory variables X_1 to X_8, consisting of 3 single variables and 2 sets of 2 and 3 grouped variables so that there are 5 condensed variables $\mathcal{X}_1$ to $\mathcal{X}_5$. The a priori known signs of the regression coefficients of X_1 and X_6 are negative and those of X_2, X_3, X_4 and X_8 are positive but those of X_5 and X_7 are unavailable.

For instance, if $\mathcal{X}_1$, i.e., the single variable X_1 is selected in a meaningful subset of variables X_1 to X_8, the estimated regression coefficient of X_1 must be negative for the regression subequation to become scientifically reasonable with respect to sign conditions. If $\mathcal{X}_2$, i.e., a set of the grouped variables X_2 and X_3 is selected in a meaningful subset, then the signs of the estimated regression coefficients of X_2 and X_3 must be positive for the regression subequation to become scientifically reasonable with respect to sign conditions. Since X_2 and X_3 are not single but grouped variables, any subsets which include only X_2 or X_3 become meaningless. If $\mathcal{X}_3$, i.e., the single variable X_4 is selected in a meaningful subset, the estimated regression coefficient of X_4 must be positive for the regression subequation to become scientifically reasonable with respect to sign conditions. If $\mathcal{X}_4$, i.e., the single variable X_5 is selected in a meaningful subset, it does not matter whether the estimated regression coefficient of X_5 is positive or negative for the regression subequation to become scientifically reasonable with respect to sign conditions. Finally, $\mathcal{X}_5$, i.e., a set of grouped variables X_6, X_7 and X_8 must be in a meaningful subset, if selected, and the estimated regression coefficients of X_6 and X_8 must be negative and positive, respectively, for the meaningful subset to become scientifically reasonable with respect to sign conditions.

If X_k for $k = 1, 2, \cdots, 8$, i.e., $\mathcal{X}_\kappa$ for $\kappa = 1, 2, \cdots, 5$ must be selected in a combinatorial way, the number of all nonempty meaningful subsets concerned with X_1, X_2, $\cdots$, and X_8 becomes 31 $(= 2^5 - 1)$, which correspond to all possible nonempty subsets concerned with $\mathcal{X}_\kappa$ for $\kappa = 1, 2, \cdots, 5$. Since the number of all possible nonempty subsets concerned with X_k's is 255 $(= 2^8 - 1)$, that of all meaningless subsets becomes 224.

If X_k for $k = 1, 2, \cdots, 8$ must be selected from the left-hand side in a sequential (and additional) way, the number of all nonempty subsets is 8. However, the number of all nonempty meaningful subsets is 5, corresponding to $\mathcal{X}_\kappa$ for $\kappa = 1, 2, \cdots, 5$ which are sequentially (and additionally) selected from the left-hand side. They are $\{-X_1\}$,

$\{-X_1, +X_2, +X_3\}$, $\{-X_1, +X_2, +X_3, +X_4\}$, $\{-X_1, +X_2, +X_3, +X_4, X_5\}$ and $\{-X_1, +X_2, +X_3, +X_4, X_5, -X_6, X_7, +X_8\}$.

Needless to say, if X_k for $k = 1, 2, \cdots, 8$ must be selected from the right-hand side in a sequential way, the number of all nonempty meaningful subsets is also 5, but the meaningful subsets differ from those selected from the left-hand side. They are $\{-X_6, X_7, +X_8\}$, $\{X_5, -X_6, X_7, +X_8\}$, $\{+X_4, X_5, -X_6, X_7, +X_8\}$, $\{+X_2, +X_3, +X_4, X_5, -X_6, X_7, +X_8\}$ and $\{-X_1, +X_2, +X_3, +X_4, X_5, -X_6, X_7, +X_8\}$.

5.2.2 Combinatorial or Sequential Condensed Variables

Here general variable classifications are referred to. However, basic variable classifications which are the simplest of general variable classifications are mainly used in actual applications.

● Generalized Combinatorial Condensed Variables

K condensed variables $\mathcal{X}_1$, $\mathcal{X}_2$, $\cdots$, $\mathcal{X}_K$ are focused on. Let L be a positive integer not exceeding K so that $1 \leq L \leq K$, M_ℓ and N_ℓ for $\ell = 1, 2, \cdots, L$ be nonnegative integers not exceeding K for all ℓ such that (i) $M_1 + N_1 > 0$ if $L = 1$ and (ii) $[M_j^*, N_j^*] \cap [M_\ell^*, N_\ell^*] = \emptyset$ for $M_\ell^* \equiv \min\{M_\ell, N_\ell\}$, $N_\ell^* \equiv \max\{M_\ell, N_\ell\}$ with allowable $M_1 = N_1 = 0$ if $2 \leq L \leq K$. We postulate that the classification form

$$< M_1 < M_2 < \cdots < M_L < \mathcal{X}_1, \mathcal{X}_2, \cdots, \mathcal{X}_K > N_L > \cdots > N_2 > N_1 > \qquad (5.21)$$

(i) generates $\sum_{\ell=1}^{L} \sum_{m_\ell=M_\ell^*}^{N_\ell^*} \begin{pmatrix} K \\ m_\ell \end{pmatrix}$ meaningful subsets of $\emptyset$ (if some $M_\ell^* = 0$) and $\{\mathcal{X}_{m_1}, \mathcal{X}_{m_2}, \cdots, \mathcal{X}_{m_i}\}$ for $m_1, m_2, \cdots, m_i = 1, 2, \cdots, K$, $m_i \neq m_j$, $i \neq j$, $i, j = M_\ell^*, M_\ell^*+1, \cdots, N_\ell^*$ and $\ell = 1, 2, \cdots, L$ and, furthermore, (ii) investigates whether or not the estimated regression coefficients of the variables in all meaningful nonempty subsets meet the + or − signs, if included in $\mathcal{X}_k$'s for $k = 1, 2, \cdots, K$, when (5.21) is used for estimation, where $\mathcal{X}_1$, $\mathcal{X}_2$, $\cdots$, $\mathcal{X}_K$ are called **a set of combinatorial condensed variables** or simply **a set of combinatorial variables.**

A user introduces condensed variables $\mathcal{X}_1$, $\mathcal{X}_2$, $\cdots$, $\mathcal{X}_K$ and knows the appropriate integers for M_ℓ and N_ℓ for all ℓ from the viewpoints of the professional knowledge. It is impossible for him to implement his research well, otherwise. It must be noted that the positions of $\mathcal{X}_k$'s in (5.21) do not matter. Needless to say, if $M_1 = 0$ or $N_1 = 0$ for $L = 1$ or $M_1 = 0$ and/or $N_1 = 0$ for $L = 2, 3, \cdots$, then an empty subset becomes meaningful with respect to these condensed variables. Case $L = 1$ in (5.21), that is, $< M_1 < \mathcal{X}_1, \mathcal{X}_2, \cdots, \mathcal{X}_K > N_1 >$, which is called **basic combinatorial variable classification**, is used in most research. If $0 < M_\ell, N_\ell \leq K$ for all ℓ, then $\mathcal{X}_1$, $\mathcal{X}_2$, $\cdots$, $\mathcal{X}_K$ are called **combinatorial and important** condensed variables. If $M_\ell = 0$ or $N_\ell = 0$ for some ℓ (usually, $\ell = 1$), then $\mathcal{X}_1$, $\mathcal{X}_2$, $\cdots$, $\mathcal{X}_K$ are called **combinatorial and optional** condensed variables.

Example 2: Case of $K = 3$, $L = 1$, $M_1 = 1$, $N_1 = 2$, $\mathcal{X}_1 = X_1$, $\mathcal{X}_2 = +X_2$ and $\mathcal{X}_3 = -X_3$:

$$< 1 < X_1, +X_2, -X_3 > 2 >$$

(i) generates the following 6 meaningful subsets with respect to X_1, $+X_2$ and $-X_3$: 1) $\{X_1\}$; 2) $\{+X_2\}$; 3) $\{-X_3\}$; 4) $\{X_1, +X_2\}$; 5) $\{X_1, -X_3\}$; and 6) $\{+X_2, -X_3\}$ and (ii) investigates the signs of the regression coefficients of these meaningful subsets, where $\begin{pmatrix} 3 \\ 1 \end{pmatrix} + \begin{pmatrix} 3 \\ 2 \end{pmatrix} = 3 + 3 = 6$. For instance, the regression coefficients of X_2 and X_3 in $\{+X_2, -X_3\}$ must be positive and negative, respectively, as preconditions for application of statistical and data-analytic tests. The integers 1 and 2, the + sign (of the regression coefficient) of variabe X_2 and the − sign of variable X_3 reflect professional knowledge. On the other hand, the a priori known sign is unavailable for X_1 so that the regression coefficient of X_1 can assume a positive or negative value. Therefore, the other subsets $\emptyset$ and $\{X_1, +X_2, -X_3\}$ are regarded as meaningless. Needless to say, the following 11 classification forms are equivalent to the above: (1) $< 1 < X_1, -X_3, +X_2 > 2 >$, (2) $< 1 < +X_2, X_1, -X_3 > 2 >$, (3) $< 1 < +X_2, -X_3, X_1 > 2 >$, (4) $< 1 < -X_3, X_1, +X_2 > 2 >$, (5) $< 1 < -X_3, +X_2, X_1 > 2 >$, (6) $< 2 < X_1, +X_2, -X_3 > 1 >$, (7) $< 2 < X_1, -X_3, +X_2 > 1 >$, (8) $< 2 < +X_2, X_1, -X_3 > 1 >$, (9) $< 2 < +X_2, -X_3, X_1 > 1 >$, (10) $< 2 < -X_3, X_1, +X_2 > 1 >$ and (11) $< 2 < -X_3, +X_2, X_1 > 1 >$.

Example 3: Case of $K = 3$, $L = 2$, $M_1 = 1$, $M_2 = 3$, $N_1 = 1$ and $N_2 = 3$:

$$< 1 < 3 < (-X_1, X_2), +X_3, +X_4 > 3 > 1 >$$

(i) generates the following 4 meaningful subsets: 1) $\{-X_1, X_2\}$; 2) $\{+X_3\}$; 3) $\{+X_4\}$; and 4) $\{-X_1, X_2, +X_3, +X_4\}$, which are obtained as $< 1 < 3 < \mathcal{X}_1, \mathcal{X}_2, \mathcal{X}_3 > 3 > 1 >$ generating $\{\mathcal{X}_1\}$, $\{\mathcal{X}_2\}$, $\{\mathcal{X}_3\}$ and $\{\mathcal{X}_1, \mathcal{X}_2, \mathcal{X}_3\}$ and (ii) investigates the signs of the regression coefficients of these 4 meaningful subsets, where $\mathcal{X}_1 = (-X_1, X_2)$, $\mathcal{X}_2 = +X_3$ and $\mathcal{X}_3 = +X_4$.

● Generalized Sequential Condensed Variables

Let M, L, K, J, I, H, G and F be nonnegative integers not exceeding K and K =number of condensed sequential variables. The two kinds of a sequential variable classification form are introduced.[28] .

[28] For example, R&D for improving the breed of a fruit like apple or orange measured as expenditures on R&D usually starts contributing the production in a society after a certain time period (say, after 10 to 12 years), continues contributing during some period (say, a period of 15 to 20 years) and then terminates its role when the new breeding technology is completely settled into plant capital and widely disseminated or replaced with the latest technology. Lagged R&D expenditures may be treated by a sequential variable classification. Since R&D is consecutively conducted,

(i) Case of Sequential and Important Condensed Variables for 1$\leq M \leq$K

The classification form

$$< M < \underline{L < \underline{J < \underline{I < \underline{H < \underline{G \leq}}}}} \mathcal{X}_1, \mathcal{X}_2, \cdots, \mathcal{X}_M, \cdots, \mathcal{X}_K \underline{\underline{\underline{\underline{\underline{\geq} >} >} >} >} > >>$$

or

$$<< \underline{< \underline{< \underline{< \underline{< \underline{\leq}}}}} \mathcal{X}_K, \cdots, \mathcal{X}_M, \cdots, \mathcal{X}_2, \mathcal{X}_1 \underline{\underline{\underline{\underline{\underline{\geq G} > H} > I} > J} > L} > M > \quad (5.22)$$

$$\text{for } 0 < M < K, 1 \leq L \leq (K-M)/H+1,$$
$$1 \leq H \leq K, 0 \leq I < K, 1 \leq J < K \text{ and } 1 \leq G < K$$

(*i*) generates at most JL meaningful subsets $\{\mathcal{X}_\kappa, \mathcal{X}_{\kappa+1}, \mathcal{X}_{\kappa+2}, \cdots, \mathcal{X}_\lambda\}$ for $\kappa = 1+G(j-1)+I(\ell-1)$, $\lambda = \min\{M+G(j-1)+(H+I)(\ell-1), K\}$, $\ell = 1, 2, \cdots, L$ and $j = 1, 2, \cdots, J$ and, furthermore, (*ii*) investigates whether or not the estimated regression coefficients of the variables in all meaningful subsets meet the + or − signs, if included in $\mathcal{X}_k$'s for $k = 1, 2, \cdots, K$, when (5.22) is used for estimation. Five pairs of the underlined parts are omissible orderly from the G side toward the L side together with the corresponding underlined <'s or >'s in the opposite side, the number of <'s must be equal to that of >'s, and $L = K - M + 1$, $J = 1$, $I = 0$, $H = 1$ and $G = 1$ are set as defaults, if not inputed.

Let us derive the JL meaningful subsets of $< M < L < J < \mathcal{X}_1, \mathcal{X}_2, \cdots, \mathcal{X}_M, \cdots, \mathcal{X}_K >>>>$ or $<<<< \mathcal{X}_K, \cdots, \mathcal{X}_M, \cdots, \mathcal{X}_2, \mathcal{X}_1 > J > L > M >$ in a case of $1 \leq M < K$, $L = K-M-J+2$ and $1 < J < K$. They are (1-1) $\{\mathcal{X}_1, \mathcal{X}_2, \cdots, \mathcal{X}_M\}$, (1-2) $\{\mathcal{X}_1, \mathcal{X}_2, \cdots, \mathcal{X}_{M+1}\}$, $\cdots$, (1-L) $\{\mathcal{X}_1, \mathcal{X}_2, \cdots, \mathcal{X}_{M+L-1}\}$ corresponding to $j = 1$; (2-1) $\{\mathcal{X}_2, \mathcal{X}_3, \cdots, \mathcal{X}_{M+1}\}$, (2-2) $\{\mathcal{X}_2, \mathcal{X}_3, \cdots, \mathcal{X}_{M+2}\}$, $\cdots$, (2-L) $\{\mathcal{X}_2, \mathcal{X}_3, \cdots, \mathcal{X}_{M+L}\}$ corresponding to $j = 2$; $\cdots\cdots$, (J-1) $\{\mathcal{X}_J, \mathcal{X}_{J+1}, \cdots, \mathcal{X}_{M+J-1}\}$, ($J$-2) $\{\mathcal{X}_J, \mathcal{X}_{J+1}, \cdots, \mathcal{X}_{M+J}\}$, $\cdots$, (J-L) $\{\mathcal{X}_J, \mathcal{X}_{J+1}, \cdots, \mathcal{X}_K\}$ corresponding to $j = J$. If $J = 1$, then only (1-1) to (1-L) are derived as the meaningful subsets with $L = K-M+1$. $\mathcal{X}_1$, $\mathcal{X}_2$, $\cdots$, $\mathcal{X}_K$ in (5.22) are called **a set of sequential and important condensed variables**.

Example 4: Case of $K = 4$, $M = 2$, $L = 3$ and $J = 2$,

$$< 2 < 3 < 2 < +X_1, +X_2, -X_3, X_4 >>>>,$$

or

$$<<<< X_4, -X_3, +X_2, +X_1 > 2 > 3 > 2 >$$

the above example can be expressed as $< 15 < 5 < 3 < +X_{-10}, +X_{-11}, \cdots, +X_{-21} >>>>$ or $<<<< +X_{-21}, \cdots, +X_{-11}, +X_{-10} > 3 > 5 > 15 >$ where X_{-t} =R&D of t years ago for $t = 10, 11, \cdots, 21$ and the + sign in front of X_{-t} indicates that the regression coefficient of X_{-t} must be positive, implying that the R&D of t years ago brings a good effect on the current production. R&D for agriculture, forestry and fishery (say, tuna breeding) should be evaluated in a longer term.

(i) generates the following 5 meaningful subsets: 1) $\{+X_1, +X_2\}$; 2) $\{X_1, +X_2, -X_3\}$; 3) $\{+X_1, +X_2, -X_3, X_4\}$; 4) $\{+X_2, -X_3\}$; and 5) $\{+X_2, -X_3, X_4\}$ and (ii) investigates the signs of the regression coefficients of these meaningful subsets. $I = 0$, $H = 1$ and $G = 1$ are set as defaults.

(ii) Case of Sequential and Optional Condensed Variables

The classification form

$$< 0 < \underline{L < \underline{J < \underline{I < \underline{H < \underline{G < \underline{F \leq}}}}}} \mathcal{X}_1, \mathcal{X}_2, \cdots, \mathcal{X}_F, \cdots, \mathcal{X}_K \underline{\underline{\underline{\underline{\underline{\underline{\geq} >} >} >} >} >} >>$$

or

$$<< \underline{< \underline{< \underline{< \underline{< \underline{< \underline{\leq}}}}}} \mathcal{X}_K, \cdots, \mathcal{X}_F, \cdots, \mathcal{X}_2, \mathcal{X}_1 \underline{\underline{\underline{\underline{\underline{\underline{\geq F} > G} > H} > I} > J} > L} > 0 > \tag{5.23}$$

$$\text{for } 1 \leq L \leq (K-F)/H + 1,\ 1 \leq H \leq K,$$
$$0 \leq I < K,\ 1 \leq J < K,\ 1 \leq G < K \text{ and } 1 \leq F < K$$

(i) generates at most $(JL+1)$ meaningful subsets $\emptyset$ and $\{\mathcal{X}_\kappa, \mathcal{X}_{\kappa+1}, \mathcal{X}_{\kappa+2}, \cdots, \mathcal{X}_\lambda\}$ for $\kappa = 1 + G(j-1) + I(\ell-1)$, $\lambda = \min\{F + G(j-1) + (H+I)(\ell-1), K\}$, $\ell = 1, 2, \cdots, L$ and $j = 1, 2, \cdots, J$ and, furthermore, (ii) investigates whether or not the estimated regression coefficients of the variables in all meaningful nonempty subsets meet the + or − signs, if included in $\mathcal{X}_k$'s for $k = 1, 2, \cdots, K$, when (5.23) is used for estimation. Six pairs of the underlined parts are omissible orderly from the F side toward the L side together with the corresponding underlined <'s or >'s in the opposite side, the number of <'s must be equal to that of >'s, and $L = K$, $J = 1$, $I = 0$, $H = 1$, $G = 1$ and $F = 1$ are set as defaults, if not inputed. (5.23) corresponds to (5.22) but accepts $\emptyset$ as meaningful.

Example 5: Case of $K = 4$ (and $L = 4$, $J = 1$, $I = 0$, $H = 1$, $G = 1$ and $F = 1$):

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

(i) generates the following 5 meaningful subsets: 1) $\emptyset$; 2) $\{+X_1\}$; 3) $\{+X_1, -X_2\}$; 4) $\{+X_1, -X_2, X_3\}$; and 5) $\{+X_1, -X_2, X_3, +X_4\}$ and, furthermore, (ii) investigates the signs of the estimated regression coefficients of the nonempty meaningful subsets.

Nonnegative integers M, L, J, I, H, G and F in (5.22) and (5.23) reflect part of the professional knowledge needed in research. The simplest case of $< M < \mathcal{X}_1, \mathcal{X}_2, \cdots, \mathcal{X}_K >>$ or $<< \mathcal{X}_K, \cdots, \mathcal{X}_2, \mathcal{X}_1 > M >$ for $M = 0, 1, 2, \cdots, K$ for (5.22) or (5.23), which is called **basic sequential variable classification,** is often used in actual research. For instance, consecutively lagged variables or power variables in a polynomial are classified as basic sequential variables. The positions of sequential condensed variables $\mathcal{X}_1$, $\mathcal{X}_2$, $\cdots$, $\mathcal{X}_K$ in (5.22) and (5.23) are decisively important.

$\mathcal{X}_1$, $\mathcal{X}_2$, $\cdots$, $\mathcal{X}_K$ in (5.23) are called **a set of sequential and optional condensed variables.**

Finally, let us give an example of a combination of sets of combinatorial and sequential variables.

Example 6:

$$< 1 < +X_1, -X_2, (+X_1, X_3), (-X_2, +X_4) > 1 > < 0 < X_5, X_6 >>$$

First of all, $< 1 < +X_1, -X_2, (+X_1, X_3), (-X_2, +X_4) > 1 >$ generates the following 4 partially meaningful subsets: a-1) $\{+X_1\}$; a-2) $\{-X_2\}$; a-3) $\{+X_1, X_3\}$; and a-4) $\{-X_2, +X_4\}$. Secondly, $< 0 < X_5, X_6 >>$ generates the following 3 partially meaningful subsets: b-1) $\emptyset$; b-2) $\{X_5\}$; and b-3) $\{X_5, X_6\}$. Finally, we have the following 12 (=4×3) meaningful subsets derived from the combination of a-1) to a-4) and b-1) to b-3): 1) $\{+X_1\}$; 2) $\{-X_2\}$; 3) $\{+X_1, X_3\}$; 4) $\{-X_2, +X_4\}$; 5) $\{+X_1, X_5\}$; 6) $\{-X_2, X_5\}$; 7) $\{+X_1, X_3, X_5\}$; 8) $\{-X_2, +X_4, X_5\}$; 9) $\{+X_1, X_5, X_6\}$; 10) $\{-X_2, X_5, X_6\}$; 11) $\{+X_1, X_3, X_5, X_6\}$; and 12) $\{-X_2, +X_4, X_5, X_6\}$. Furthermore, it is checked whether or not the estimated regression coefficients of the above 12 meaningful subsets coincide with the corresponding specified signs. It must be noted that although variables X_1 and X_2 with their signs appear twice, all 12 meaningful subsets do not have two X_1's and/or X_2's so that there are no redundancies in the meaningful subsets.

5.3 Standard Functional Form

All meaningful subsets generated by the forms (5.21) and/or (5.22) or (5.23) are equivalent to each other before evaluation for their regression subequations. Let X_0 be a constant term. We postulate that X_0 is not enclosed within any symbols, a dependent variable is expressed as a function of a constant term, if needed, and a set of all possible classified explanatory variables and X_0 is automatically included in all derived meaningful subsets.

It is convenient for users to introduce a standard functional form $Y = F(X_0[\boldsymbol{X}])$ for regression analysis where Y =dependent variable and $[\boldsymbol{X}]$ =set of all possible explanatory variables classified with the professional knowledge related to the research in question. Let us give an example:

Example 7: With the Schur stability condition concerned with Y, Y_{-1}, Y_{-2} and Y_{-3}:

$$\begin{aligned} Y \;=\; & F(X_0 \;\; < 2 < +X_1, -X_2 > 2 > \; < 0 < X_3 > 1 > \\ & <<<< Y_{-3}, Y_{-2}, +Y_{-1} > 2 > 2 > 1 >) \end{aligned} \tag{5.24}$$

which is equivalent to $Y = F(X_0 < 2 < +X_1, -X_2 > 2 > < 1 < 2 < 2 < +Y_{-1}, Y_{-2}, Y_{-3} >>>> < 1 < X_3 > 0 >)$ and $Y = F(X_0 < 2 < +X_1, -X_2 > 2 > < 0 < X_3 > 1 >< 1 < +Y_{-1}, Y_{-2}, (Y_{-2}, +Y_{-1}), (Y_{-3}, Y_{-2}) > 1 >)$.

$< 2 < +X_1, -X_2 > 2 >$ yields only one partially meaningful subset (i) $\{+X_1, -X_2\}$ with respect to $+X_1$ and $-X_2$. $< 0 < X_3 > 1 >$ yields 2 partially meaningful subsets (ii-1) $\emptyset$ and (ii-2) $\{X_3\}$ with respect to X_3.

It is clear that $M = 1$, $L = 2$ and $J = 2$. The defaults are $I = 0$, $H = 1$, $G = 1$ and $F = 1$. Thus, $j = 1, 2$ and $\ell = 1, 2$, $\kappa = 1 + 1 \cdot (j - 1) + 0 \cdot (\ell - 1) = j$, $\lambda = 1 + 1 \cdot (j - 1) + (1 + 0)(\ell - 1) = j + \ell - 1$. Accordingly, $j = 1$ and $\ell = 1$ lead to $\kappa = 1$ and $\lambda = 1$ and yield (iii-1) $\{+Y_{-1}\}$. $j = 1$ and $\ell = 2$ lead to $\kappa = 1$ and $\lambda = 2$ and yield (iii-2) $\{Y_{-2}, +Y_{-1}\}$. $j = 2$ and $\ell = 1$ lead to $\kappa = 2$ and $\lambda = 2$ and yield (iii-3) $\{Y_{-2}\}$. $j = 2$ and $\ell = 2$ lead to $\kappa = 2$ and $\lambda = 3$ and yield (iii-4) $\{Y_{-3}, Y_{-2}\}$. All 8 meaningful subsets consist of [1] $\{X_0\}$, (i), (ii-1) and (iii-1); [2] $\{X_0\}$, (i), (ii-1) and (iii-2); [3] $\{X_0\}$, (i), (ii-1) and (iii-3); [4] $\{X_0\}$, (i), (ii-1) and (iii-4); [5] $\{X_0\}$, (i), (ii-2) and (iii-1); [6] $\{X_0\}$, (i), (ii-2) and (iii-2); [7] $\{X_0\}$, (i), (ii-2) and (iii-3); and [8] $\{X_0\}$, (i), (ii-2) and (iii-4). Finally, we have the following 8 meaningful subsets: [1] $\{X_0, +X_1, -X_2, +Y_{-1}\}$; [2] $\{X_0, +X_1, -X_2, Y_{-2}, +Y_{-1}\}$; [3] $\{X_0, +X_1, -X_2, Y_{-2}\}$; [4] $\{X_0, +X_1, -X_2, Y_{-3}, Y_{-2}\}$; [5] $\{X_0, +X_1, -X_2, X_3, +Y_{-1}\}$; [6] $\{X_0, +X_1, -X_2, X_3, Y_{-2}, +Y_{-1}\}$; [7] $\{X_0, +X_1, -X_2, X_3, Y_{-2}\}$; and [8] $\{X_0, +X_1, -X_2, X_3, Y_{-3}, Y_{-2}\}$.

The estimated regression subequations $\widehat{Y}_\kappa$ for $\kappa = 1, 2, \cdots, 8$ of the above 8 meaningful subsets can be expressed as follows:

$$\begin{array}{lrcl}
[1] & \widehat{Y}_1 & = & \widehat{a}_{01} + \widehat{a}_{11}X_1 + \widehat{a}_{21}X_2 + \widehat{a}_{31}Y_{-1}, \\
[2] & \widehat{Y}_2 & = & \widehat{a}_{02} + \widehat{a}_{12}X_1 + \widehat{a}_{22}X_2 + \widehat{a}_{32}Y_{-2} + \widehat{a}_{42}Y_{-1}, \\
[3] & \widehat{Y}_3 & = & \widehat{a}_{03} + \widehat{a}_{13}X_1 + \widehat{a}_{23}X_2 + \widehat{a}_{33}Y_{-2}, \\
[4] & \widehat{Y}_4 & = & \widehat{a}_{04} + \widehat{a}_{14}X_1 + \widehat{a}_{24}X_2 + \widehat{a}_{34}Y_{-3} + \widehat{a}_{44}Y_{-2}, \\
[5] & \widehat{Y}_5 & = & \widehat{a}_{05} + \widehat{a}_{15}X_1 + \widehat{a}_{25}X_2 + \widehat{a}_{35}X_3 + \widehat{a}_{45}Y_{-1}, \\
[6] & \widehat{Y}_6 & = & \widehat{a}_{06} + \widehat{a}_{16}X_1 + \widehat{a}_{26}X_2 + \widehat{a}_{36}X_3 + \widehat{a}_{46}Y_{-2} + \widehat{a}_{56}Y_{-1}, \\
[7] & \widehat{Y}_7 & = & \widehat{a}_{07} + \widehat{a}_{17}X_1 + \widehat{a}_{27}X_2 + \widehat{a}_{37}X_3 + \widehat{a}_{47}Y_{-2}, \\
[8] & \widehat{Y}_8 & = & \widehat{a}_{08} + \widehat{a}_{18}X_1 + \widehat{a}_{28}X_2 + \widehat{a}_{38}X_3 + \widehat{a}_{48}Y_{-3} + \widehat{a}_{58}Y_{-2}.
\end{array}$$

As preconditions for the applications of statistical and data-analytic tests, these 8 regression subequations must satisfy the following sign conditions specified in the functional form (5.24): (i) $\widehat{a}_{1\kappa} > 0$ for all $\kappa = 1, 2, \cdots, 8$ by $+X_1$, (ii) $\widehat{a}_{2\kappa} < 0$ for all $\kappa = 1, 2, \cdots, 8$ by $-X_2$ and, furthermore, (iii) $\widehat{a}_{31} > 0$, $\widehat{a}_{42} > 0$, $\widehat{a}_{45} > 0$ and $\widehat{a}_{56} > 0$ by $+Y_{-1}$. Suppose that the Schur stability condition must be met by solutions of the characteristic equations of the regression subequations. For example, the characteristic equation of the second regression subequation in [2] ($\kappa = 2$) becomes

$$\ell^2 - \widehat{a}_{42}\ell - \widehat{a}_{32} = 0,$$

derived from $Y - \widehat{a}_{42}Y_{-1} - \widehat{a}_{32}Y_{-2} = 0$ where $\widehat{b}_{12} = -\widehat{a}_{42}$ and $\widehat{b}_{02} = -\widehat{a}_{32}$. We have $T_2^{*} = 1$, $T_2^{**} = 2$ and $k_2 = 1, 2$. The following Schur stability condition must be satisfied for the second regression subequation before statistical and data-analytical tests are applied:

for $k_2 = 1$

$$|\widehat{\boldsymbol{A}}_{12}| = \begin{vmatrix} \widehat{\boldsymbol{A}}^{*}_{12} & \widehat{\boldsymbol{A}}^{**\prime}_{12} \\ \widehat{\boldsymbol{A}}^{**}_{12} & \widehat{\boldsymbol{A}}^{*\prime}_{12} \end{vmatrix} = \begin{vmatrix} 1 & -\widehat{a}_{42} & -\widehat{a}_{32} & 0 \\ 0 & 1 & -\widehat{a}_{42} & -\widehat{a}_{32} \\ -\widehat{a}_{32} & -\widehat{a}_{42} & 1 & 0 \\ 0 & -\widehat{a}_{32} & -\widehat{a}_{42} & 1 \end{vmatrix} > 0,$$

which leads to $1 - 2\widehat{a}_{32}\widehat{a}_{42}^2 - 2\widehat{a}_{32}^2 - \widehat{a}_{42}^2 + \widehat{a}_{32}^4 - \widehat{a}_{32}^2\widehat{a}_{42}^2 = (1+\widehat{a}_{32})^2(1-\widehat{a}_{32}-\widehat{a}_{42})(1-\widehat{a}_{32}+\widehat{a}_{42}) > 0$,
and for $k_2 = 2$

$$|\widehat{\boldsymbol{A}}_{22}| = \begin{vmatrix} \widehat{\boldsymbol{A}}^{*}_{22} & \widehat{\boldsymbol{A}}^{**\prime}_{22} \\ \widehat{\boldsymbol{A}}^{**}_{22} & \widehat{\boldsymbol{A}}^{*\prime}_{12} \end{vmatrix} = \begin{vmatrix} 1 & -\widehat{a}_{32} \\ -\widehat{a}_{32} & 1 \end{vmatrix} > 0,$$

which leads to $1 - \widehat{a}_{32}^2 > 0$.

Then, the estimated regression coefficients $\widehat{a}_{32}$ and $\widehat{a}_{42}$ must satisfy (i) $-1 < \widehat{a}_{32} < 0$ or $0 < \widehat{a}_{32} < 1$ and (ii) $-(1-\widehat{a}_{32}) < \widehat{a}_{42} < 1-\widehat{a}_{32}$ from the viewpoints of the Schur stability condition. Finally, if the regression subequation of the second meaningful subset $\{X_0, +X_1, -X_2, Y_{-2}, +Y_{-1}\}$ has (i) $\widehat{a}_{12} > 0$, (ii) $\widehat{a}_{22} < 0$, (ii) $-1 < \widehat{a}_{32} < 0$ or $0 < \widehat{a}_{32} < 1$ and (iv) $0 < \widehat{a}_{42} < 1 - \widehat{a}_{32}$ which are required from both the sign and Schur stability conditions, it becomes scientifically reasonable and ready for statistical and data-analytic tests. The cases of all other regression subequations are omitted.

Unless variable selection is needed, a single regression equation is estimated by $Y = F(\Diamond X_0, \Diamond X_1, \Diamond X_2, \cdots, \Diamond X_K)$. If $\Diamond X_k = +X_k$ or $-X_k$ for some $k = 0, 1, 2, \cdots, K$, then the estimated regression coefficient $\widehat{a}_k$ of X_k is examined whether $\widehat{a}_k > 0$ or $\widehat{a}_k < 0$, respectively and a one-tailed t-test for its regression coefficient a_k is conducted, if requested. On the other hand, if $\Diamond X_k = X_k$ for some k, then $\widehat{a}_k$ of X_k can be either positive or negative so that a two-tailed t-test for its regression coefficient a_k is conducted, if requested.

5.4 Sufficiency of Scientific Variable Classification

In Subsection 5.2, double variable classifications were proposed to efficiently generate only all meaningful subsets from a given set of all possible explanatory variables specified for a dependent variable. They are regarded as necessary conditions for an efficient generation of only all meaningful subsets. However, we have to prove that double variable classifications are sufficient to always generate only all meaningful subsets but not any meaningless subsets in a run of a computer, no matter what research is conducted.

Suppose that a user wants to estimate and evaluate the following I possible regression equation candidates one at a time to search for the practically best one:

$$Y = a_{0i} + \boldsymbol{X}_i A_i + U \quad \text{for } i = 1, 2, \cdots, I \tag{5.25}$$

where a_{0i} =true regression coefficient of a constant term in the i-th regression equation candidate; $\boldsymbol{X}_i$ =set or row vector of (nonconstant) explanatory variable candidates, which he knows concretely, in the i-th regression equation candidate; A_i =column vector of true regression coefficients of $\boldsymbol{X}_i$; and U =disturbance term. Then all I possible regression equation candidates can be generated and estimated and, furthermore, the signs of all estimated regression coefficients are checked by the following functional form in a run of a computer:

$$Y = F(\Diamond X_0 < 1 < (\Diamond \boldsymbol{X}_1), (\Diamond \boldsymbol{X}_2), \cdots, (\Diamond \boldsymbol{X}_I) > 1 >) \tag{5.26}$$

which generates

$$\{\Diamond X_0, \Diamond \boldsymbol{X}_1\}, \ \{\Diamond X_0, \Diamond \boldsymbol{X}_2\}, \ \cdots, \ \{\Diamond X_0, \Diamond \boldsymbol{X}_I\} \quad \text{for } Y, \tag{5.27}$$

estimates

$$\widehat{Y}_i = \widehat{a}_{0i} + \boldsymbol{X}_i \widehat{A}_i \quad \text{for } i = 1, 2, \cdots, I, \tag{5.28}$$

and investigates whether or not all $\widehat{A}_i$'s coincide with the corresponding a priori known signs where $\Diamond X_0$ = constant term with/without the a priori known sign of its regression coefficient and $\Diamond \boldsymbol{X}_i = \boldsymbol{X}_i$ with/without the a priori known signs of their regression coefficients. Although the entry becomes long, the variable sets of all I regression equation candidates which correspond to all I meaningful subsets can be generated by (5.26) and estimated as (5.28) in a run of a computer. It is clear that all $\Diamond \boldsymbol{X}_i$'s for $i = 1, 2, \cdots, I$ are treated as combinatorial and grouped variable sets.

We conclude that scientific variable classification characterized by double variable classifications, single or grouped and combinatorial or sequential, is not only necessary but also sufficient to generate only all meaningful subsets from a given set of all possible explanatory variables specified for a dependent variable in a run of a computer, no matter what research is conducted.

5.5 Derivation of Individual Magnitude Conditions from an Aggregate Magnitude Condition

It is unrealistic that consumption with constant income over time becomes infinitely large or negative or oscillates between positive and negative values, as time passes or a rice production becomes treble, if all inputs which can be bought with money are doubled under the same weather condition, implying that increasing returns to scale in the rice production are unusually high. Thus, there are situations in which the a priori known + or − signs of the regression coefficients of explanatory variables are not sufficient and the regression coefficient of an explanatory variable or the sum of the regression coefficients of some explanatory variables must exist in a certain range. Such a range is called a magnitude condition. It is convenient to express a regression coefficient with the notation of the corresponding possible explanatory variable (and a constant term), when an aggregate magnitude condition is loaded into a computer. The individual magnitude condition for each of all meaningful subsets can be easily derived from an aggregate magnitude condition. An aggregate magnitude condition corresponds to a set of all possible explanatory variables for a dependent variable. Let us denote an aggregate linear magnitude condition[29] by

$$CX > \alpha^1, \quad CX < \alpha^2 \quad \text{or} \quad \alpha^1 < CX < \alpha^2 \tag{5.29}$$

where α^1 =known lower bound, α^2 =known upper bound and $C = \{1 \times (K+1)\}$-vector of known coefficients. Since the explanatory variables having zero coefficients can be removed from the above, we can rewrite the above aggregate magnitude conditions as follows:

$$C^*X^* > \alpha^1, \quad C^*X^* < \alpha^2 \quad \text{or} \quad \alpha^1 < C^*X^* < \alpha^2 \tag{5.30}$$

which is actually loaded into a computer where all elements of C^* are non-zero. It should be noticed that an aggregate magnitude condition does not usually make sense as a magnitude condition but each derived individual magnitude condition must make sense. If the i-th meaningful subset X_i is derived, it is possible to select the elements corresponding to X_i from C^* and enter zeros into the explanatory variables which are in X_i but not in X^*. Let us denote such a magnitude condition by

$$C_iX_i > \alpha^1, \quad C_iX_i < \alpha^2 \quad \text{or} \quad \alpha^1 < C_iX_i < \alpha^2. \tag{5.31}$$

Then X_i is replaced with estimated $\widehat{A}_i$ in the process of evaluation. $C_i\widehat{A}_i$ can be easily calculated and is compared with α^1 and/or α^2.

If $C_i\widehat{A}_i > \alpha^1$, $C_i\widehat{A}_i < \alpha^2$ or $\alpha^1 < C_i\widehat{A}_i < \alpha^2$, then the regression equation is regarded as satisfactory with respect to this magnitude condition and furthermore examined with the other magnitude conditions, if any, statistical and data-analytic criteria. On the other hand, if $C_i\widehat{A}_i < \alpha^1$ or $C_i\widehat{A}_i > \alpha^2$, then the regression equation is regarded as unsatisfactory and cannot become a candidate to solve the j-th OLS-best subset problem.

This technique can be easily used for the derivation of all individual hypothetical relations and constraints of regression coefficients from an aggregate hypothetical relation and constraint, respectively.

5.6 Reviewing Function

Sometimes a user thinks before estimation and evaluation that some special or his favorite meaningful subset may be best and feels uneasy whether or not the software really solved the first j OLS-best subset problems, when his prediction was

[29] Division, multiplication, square root, cubic root, absolute value, logarithm, exponential and power can be used for a magnitude condition in the System **OEPP**.

wrong. It is user-friendly to install a subsystem in software by which he can see the estimated regression subequations of his favorite meaningful subsets and all reasons why they were not satisfactory. We would like to introduce the function reviewing the estimated regression subequations of his favorite meaningful subsets $\boldsymbol{X}_\ell$'s for $\ell = 1, 2, \cdots, L$ and $L \geq 1$ installed in the System **OEPP**. The following referential functional forms are inputed:

$$Y = F(\boldsymbol{X}_\ell) \quad \text{for} \quad \ell = 1, 2, \cdots, L,$$

then $\widehat{Y}_\ell = \boldsymbol{X}_\ell \widehat{A}_\ell$ is printed together with important test statistics and all unsatisfactory reasons, if this regression subequation was unsatisfactory. He can be convinced.

6 Demonstration by Intellectual Statistical System OEPP

The numbers of civil servants in the tax divisions in 46 prefectural governments except for the Tokyo metropolitan government are focused on. The administrative functions of the Tokyo metropolitan government differ from those of the other 46 prefectural governments. Thus, the sample size is 46 and cross-sectional data in 1996 are used. We use the following variable notation:

Y =number of civil servants in the tax divisions (unit: persons); X_0 =constant term; X_1 =population registered in basic residence census (1,000 persons); X_2 =number of households registered in basic residence census (1,000 households); X_3 =administrated area (km^2); X_4 =habitable area (=administrated area minus mountainous and lacustrine areas) (km^2); X_5 =population in the third industrial sector (100 persons); X_6 =number of vehicles (1,000 vehicles); X_7 =number of inspections of newly-built structures of houses, buildings, etc. (cases); X_8 =number of canvassers who persuade tax payers to pay their taxes to local governments by credit systems (persons); X_9 =number of restaurants and cafeterias (restaurants or cafeterias); and X_{10} =number of business offices (offices).

To search for the statistically and data-analytically best regression equation through all possible regressions only by statistical and data-analytically tests without using the professional knowledge about the local administration, the following functional form was inputed:

$$Y = F(X_0 < 1 < X_1, X_2, X_3, X_4, X_5, X_6, X_7, X_8, X_9, X_{10} > 10 >) \qquad (6.32)$$

together with the following statistical and data-analytic criteria:

β = significance level 0.1 (10 %) for a t-test for the regression coefficients of all non-constant explanatory variables;
η = significance level 0.05 (5 %) for the Jarque-Bera normality test;
ψ = significance level 0.05 (5 %) for the Chow equal coefficients test with the Kantou

(Eastern) cultural area group (27 prefectures) versus the Kansai (Western) cultural area group (19 prefectures);

$\omega =$ significance level 0.05 (5 %) for the Goldfeld-Quandt homoscedasticity test with northern 15 prefectures (snowy area) versus southern 15 prefectures (typhoon-often-hit area);

$\nu =$ significance level 0.05 (5 %) for a t-test for detecting a residual outlier;

$\varepsilon = 2.65$ standardized residual tolerance level with the subjective allowance of a maximum of 2 violations due to cross-sectional data;

$\theta = 0.7$ minimum tolerance level for an adjusted coefficient of determination,

where the total significancel level due to β, η, ψ, ω and ν is 0.26694 or 26.694 %.

There are 1,023 possible nonempty subsets because there are 10 different nonconstant explanatory variables in (6.32), where $2^{10} - 1 = 1023$. We set $j = 1$ to solve the (first) OLS-best subset problem. Then, the following statistically and data-analytically best regression subequation was searched for by a notebook-type PC (VAIO, Sony) in the CPU time of about 1 second:

The Scientifically Unreasonable but Statistically and Data-Analytically Best Regression Equation

$$\begin{array}{ccccc}
\widehat{Y}_{488} & = & 30.17807 & -0.2288400X_1 & +0.5683528X_2 \\
(S.ERR.) & & (14.53501) & (0.5609610 \times 10^{-1}) & (0.1160882) \\
(T\text{-}RATIO) & & (2.076233) & (-4.079427) & (4.895871) \\
 & & & & \\
 & & +0.1754694X_6 & -0.1529770 \times 10^{-2}X_8 & +0.1897389 \times^{-1} X_9 \\
 & & (0.3426833 \times 10^{-1}) & (0.6831756 \times 10^{-3}) & (0.2028726\times^{-2}) \\
 & & (5.120455) & (-2.239205) & (9.352613)
\end{array} \qquad (6.33)$$

$$\widehat{R}^2_{488} = 0.9793,\ \widehat{\mathcal{R}}^2_{488} = 0.9767,\ \widehat{AIC}_{488} = 490.0118,\ \widehat{SD}_{488} = 46.8393,\ \widehat{V}^2_{488} = 2193.92,$$
$$\widehat{DF}_{488} = 40,\ \widehat{JB}_{488} = 4.703,\ \widehat{OT}_{488} = 3.169,\ \widehat{CS}_{488} = 2.147,\ \widehat{GQ}_{488} = 2.171,$$
$$\widehat{SR}_{12,488} = -2.86162,\ \widehat{SR}_{39,488} = 2.83930 \text{ and } TSL = 0.26694,$$

where $\widehat{Y}_{488}$ =estimate of Y by the 488-th subset; $(S.ERR.)$ = standard errors of the estimated regression coefficients; $(T\text{-}RATIO)$ = t-ratios of the estimated regression coefficients; $\widehat{R}^2_{488}$ =coefficient of determination; $\widehat{\mathcal{R}}^2_{488}$ =adjusted coefficient of determination; $\widehat{AIC}_{488}$ =Akaike information criterion; $\widehat{SD}_{488}$, $\widehat{V}^2_{488}$ =estimated standard deviation and variance of a disturbance term, respectively; $\widehat{DF}_{488}$ =degrees of freedom, $\widehat{JB}_{488}$ =Jarque-Bera normality test statistic; $\widehat{OT}_{488}$ =test statistic of a t-test for a residual outlier: $\widehat{CS}_{488}$ =Chow equal coefficients test statistic; $\widehat{GQ}_{488}$ =Goldfeld-Quandt homoscedasticity test statistic; $\widehat{SR}_{t,488}$ =standardized residual at sample number t which is 2.65 or greater in absolute value and TSL =total significance level defined as $1 - (1 - \beta)(1 - \eta)(1 - \psi)(1 - \omega)(1 - \nu)$.

Let us scientifically evaluate (6.33) which is already known as statistically and data-analytically best under the current criteria. The regression coefficient -0.2288400 of X_1 is negative, implying that the larger the population, the less will the work collecting residence taxes be or the less civil servants will be needed in the tax divisions ceteris paribus. This contradicts the fact. The selection of variables X_1 and X_2 shows a sort of redundancy. Corporate tax is one of the most important sources for local governments' revenues and is a necessary variable. Variable X_{10} which represents the source of corporate tax is not selected. Therefore, (6.33) is unreasonable and unsatisfactory from the viewpoints of local governments' administrations. We regard (6.33) as scientifically unreasonable, although it is statistically and data-analytically best. It must be kept in mind that a two-tailed t-test of 10 % signficance level was applied to each of the regression coefficients of explanatory variables X_1, X_2, X_6, X_8 and X_9 in (6.33).

Let us input the administrative knowledge about various taxes (the work of collecting residence and corporate taxes is the main one, many tax-collecting branches are needed if administrated areas are broad, other taxes dealing incidently with the work of main taxes are optional, and income tax belongs to the work of the central government so that it is not here referrred to and so on) in the functional form (6.32). The administrative knowledge is reflected in all possible explanatory variables in the following functional form:

$$\begin{aligned} Y &= F(X_0 \ < 1 < +X_1, +X_2 > 1 > \ < 1 < +X_3, +X_4 > 1 > \\ &\quad < 0 < +X_5, +X_6, +X_7, -X_8, +X_9 > 5 > \ < 1 < +X_{10} > 1 >). \quad (6.34) \end{aligned}$$

No magnitude conditions on regression coefficients are a priori known. Since the data are cross-sectional, a Schur stability condition and a turning point test are not needed. The number of all meaningful subsets is only 128, where $2 \times 2 \times 2^5 \times 1 = 128$. Hence, the remaining 895 nonempty subsets out of all 1,023 possible subsets are regarded as meaningless. We set $j = 1$ to solve the OLS-best subset problem under the same statistical and data-analytic criteria as before. Then, the following practically best regression subequation was searched for (with the regression table and graphs of $\widehat{Y}_{118}$, Y and $\widehat{E}_{118}$ which are not shown here) in the CPU time of only 470 milliseconds at the cost of 1 Japanese yen (less than 1 US cent!):

The Practically Best (or Scientifically Reasonable and Statistically and Data-Analytically Best) Regression Equation

$$\begin{array}{rclll}
\widehat{Y}_{118} & = & -14.11679 & +0.831384 \times 10^{-1} X_2 & +0.1260718 \times 10^{-2} X_3 \\
(\mathit{S.ERR.}) & & (12.41127) & (0.3534358 \times 10^{-1}) & (0.6762401 \times 10^{-3}) \\
(\mathit{T\text{-}RATIO}) & & (-1.137417) & (2.352291) & (1.864305) \\
 & & -0.1017291 \times 10^{-2} X_8 & +0.2421174 \times 10^{-2} X_{10} & \\
 & & (0.6874731 \times 10^{-3}) & (0.2467794 \times 10^{-3}) & \\
 & & (-1.479754) & (9.811087) &
\end{array} \qquad (6.35)$$

$\widehat{R}^2_{118} = 0.9771$, $\widehat{\overline{R}}^2_{118} = 0.9749$, $\widehat{AIC}_{118} = 492.5980$, $\widehat{SD}_{118} = 48.6293$, $\widehat{V}^2_{118} = 2364.81$, $\widehat{DF}_{118} = 41$, $\widehat{JB}_{118} = 0.6168$, $\widehat{OT}_{488} = 3.256$, $\widehat{CS}_{118} = 1.1410$, $\widehat{GQ}_{118} = 2.787$, $\widehat{SR}_{11,118} = -2.93068$, $\widehat{SR}_{22,118} = 2.70445$ and $TSL = 0.26694$.

A one-tailed t-test of 10 % significance level was applied to each of the regression coefficients of explanatory variables X_2, X_3, X_8 and X_{10} in (6.35). We decided to regard the practically best regression subequation (6.35) as the practically best regression equation for the number of civil servants in the tax divisions of local governments with a 26.694 % risk of committing a type I error or with a 26.694 % risk that (6.35) was actually not best and actually used (6.35) together with the practically best regression equations for the other divisions of local governments, which are not cited here, for the local government civil servants reduction reform policy. In total, about 300,000 civil servants of the local governments in Japan have been reduced during a decade.

7 Concluding Remarks

In applications of regression analysis, the professional knowledge or information based on the science(s) related to the research in question is usually needed in addition to statistical knowledge. Many users of regression analysis have suffered by ineffective variable selection methods like the forward selection, backward elimination, stepwise regression, t- or F-directed search, mini-max regret methods and so on. These methods mislead the users into accepting wrong and ridiculous conclusions. Users have been waiting for a method to concretely and rigorously solve a variable selection problem from both statisticians' and users' viewpoints.

To shed some light on the problem, Onishi concretely formulated the j-th OLS-best subset problem for regression analysis as standard, proposed a knowledge-based variable selection method to solve it, completed the algorithm in the Intellectual Statistical System **OEPP** and demonstrated how to solve it. It is easy to extend the proposed method to the Aitken generalized least squares, Almon distributed lag regression, Box-Cox transformation, probit model, two stage least squares, limited information maximum likelihood and other estimation methods. The System **OEPP** is quite effective for teaching beginners, including students, up to advanced applied researchers, including public policy makers and business strategists, regression analysis. They can learn how important it is to employ the professional knowledge about their research and specify appropriate criteria for statistical and data-analytic tests. Then, statistics and econometrics become more useful sciences than at present, attract smarter students into these fields and convince the public that statistics and econometrics eventually help to raise their welfare through substantial sciences.

Statistical software must be earth- and user-friendly, i.e., should not waste resources but release users from inessential, laborious and time-consuming inputting work and give them more opportunities to concentrate on creative work. It is recommended that statistical software should be designed not only to obtain the practically best regression equation but also to answer user's questions or clear his doubts. For instance, if a favorite meaningful subset which he guesses or believes will become the practically best before estimation is not actually selected as the practically best, all reasons why it is not practically best should be printed in the printout, when he requests them. Thus, his doubt disappears. Even now, nobody can compete with computers for calculation and data retrieval. It is better to develop creative methodologies and algorithms for knowledge-based variable selection methods for various kinds of estimations and make a PC solve user's statistical problems. It must be kept in mind that if a user wants to conduct a good research project, he must be an expert in his research field.

References

[1] 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), 267-281, Akademiai Kiado, Budapest (1973).

[2] Beale, E.M.L., Kendall, M.G. and D.W. Mann, The discarding of variables in multivariate analysis, *Biometrika*, **54** 357-365 (1967).

[3] Chow, G.C., Tests of equality between sets of coefficients in two linear regressions, *Econometrica*, **28**, 591-605 (1960).

[4] Chow, G.C., *Econometrics*, McGraw-Hill, 1983.

[5] Daniel, C. and F.S. Wood, *Fitting Equations to Data*, John Wiley, 1971.

[6] Draper, N.R. and Smith, H., *Applied Regression Analysis*, 2nd ed., John Wiley, 1981.

[7] Durbin, J., Testing for serial correlation in least squares regression when some of the regressors are lagged dependent variables, *Econometrica*, **38**, 410-421 (1970).

[8] Durbin, J. and Watson, G.S., Testing for serial correlation in least squares regression, I, *Biometrika*, **37**, 409-428 (1950).

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

[10] Durbin, J. and Watson, G.S., Test for serial correlation in least squares regression, III, *Biometrika*, **58**, 1-42 (1971).

[11] Farebrother, R.W., The Durbin-Watson test for serial correlation when there is no intercept in the regression, *Econometrica*, **48**, 1553-1563 (1980).

[12] Fisher, F.M., Tests on equality between sets of coefficients in two linear regressions: an expository note, *Econometrica*, **28**, 361-366 (1970).

[13] Furnival, G.M., All possible regressions with less computation, *Technometrics*, **13**, 403-408 (1971).

[14] Garside, M.J., The best subset in multiple regression analysis, *Appl. Statist.*, **14**, 196-200 (1965).

[15] Godfrey, L.G., Testing against general autoregressive and moving average error models when the regressors include lagged dependent variables, *Econometrica*, **46**, 1293-1302 (1978).

[16] Goldfeld, S.M. and Quandt, R.E., Some tests for homoscedasticity, *J. Amer. Statist. Assoc.*, **60**, 539-547 (1965).

[17] Gorman, J.W. and R.J. Toman, Selection of variables for fitting equation, *Technometrics*, **8**, 27-51 (1966).

[18] Haux, R., *Expert Systems in Statistics*, ed. by R. Haux, Gustav Fischer, 1986.

[19] Hibino, I., *Doutai Kiezai no Bunseki* in Japanese (*Dynamic Economic Analysis* in English), 6-th ed., Doubunkan, Japan, 1962.

[20] Hocking, R.R., The analysis and selection of variables in linear regression, *Biometrics*, **32**, 1-49 (1976).

[21] Hocking, R.R. and R.N. Leslie, Selection of the best subset in regression analysis, *Technometrics*, **6**, 531-540 (1967).

[22] Jarque, C.M. and A.K. Bera, Efficient tests for normality, homoscedasticity and serial independence of regression residuals, *Economics Letters*, **6**, 255-259 (1980).

[23] Johnston, J., *Econometric Methods*, 3rd ed., McGraw-Hill, 1984.

[24] Judge, G.G., Griffiths, W., Hill, R.C., and Lee, T.-C., *The Theory and Practice of Econometrics*, John Wiley, 1980.

[25] Kelejian, H.H. and Oates, W.E., *Introduction to Econometrics – Principles and Applications*, Harper & Row, 1974.

[26] Kitagawa, T., *Toukei Jyohou Ron, I* and *II* in Japanese, (*Statistical Information Theory, I* and *II* in English), Kyouritsu Shuppan, Tokyo, 1987.

[27] La Motte, L.R. and R.R. Hocking, Computational efficiency in the selection of regression variables, *Technometrics*, **12**, 83-93 (1970).

[28] Linhart, H. and Zucchini, W., *Model Selection*, John Wiley, 1986.

[29] Maddala, G.S., *Econometrics*, McGraw-Hill, 1977.

[30] Mallows, C.L., Some comments on Cp, *Technometrics*, **15**, 661-675 (1973).

[31] Mantel, N., Why stepdown procedures in variable selection, *Technometrics*, **12**, 621-625 (1970).

[32] Miller, A.J., *Subset Selection in Regression*, Chapman and Hall, London, 1990.

[33] Onishi, H., A variable selection procedure for econometric models, *J. Computa. Statist. Data Analy.*, **1** 2, 85-95 (1983).

[34] Onishi, H., Variable selection in the Researcher System OEPP and governmental use, *Statistica Applicata*, vol. 6, no. 3, pp. 309-338, 1994, invited and presented in Conference on Public Statistical Information Systems, Siena, Italy, 1993.

[35] Onishi, H., A user knowledge-based variable selection method for limited information maximum likelihood using principal components, *Computational Statistics & Data Analysis*, vol. 19, no. 4, pp. 379-399, 1995.

[36] Onishi, H., *Intellectual Statistical System OEPP for Regression Analysis and Econometrics*, OEPP Institute, 3-208-2, Takezono, Tsukuba, Ibaraki, Japan, 2001.

[37] Pau, L.F., *Artificial Intelligence in Economics and Management*, North-Holland, 1986.

[38] Pindyck, R.S. and Rubinfeld, D.L., *Econometric Models and Economic Forecasts*, McGraw-Hill, 1976.

[39] Savin, N.E. and White, K.J., The Durbin-Watson test for serial correlation with extreme sample sizes or many regressors, *Econometrica*, **45** 8, 1989-1996 (1977).

[40] Sawa, T., Information criterion for discriminating among alternative regression models, *Econometrica*, **46**, 1273-1292 (1978).

[41] Sawa, T., *Kaiki Bunseki* in Japanese (*Regression Analysis* in English), Asakura Shoten, Japan, 1979.

[42] Schatzoff, M., Tsao, R. and Fienberg, S., Efficient calculation of all possible regressions, *Technometrics*, **10**, 769-779 (1968).

[43] Shapiro, S.S. and Wilk, M.B., An analysis of variance test for normality (complete samples), *Biometrika*, **52**, 591-611 (1965).

[44] Shapiro, S.S., Wilk, M.B. and Chen, H.J., A comparative study of various tests of normality, *Amer. Statist. Assoc.*, **63**, 1343-1372 (1968).

[45] Theil, H., *Economic Forecast and Policy*, North-Holland, 1961.

[46] Theil, H., *Principles of Econometrics*, North-Holland, 1979.

[47] Wallis, K.F., Testing for fourth order for autocorrelation in quarterly regression equations, *Econometrica*, **40**, 617-636 (1972).

[48] Weisberg, A., *Applied Linear Regression*, John Wiley, 1980.