Constrained Coefficient Component Regression Analysis and Its Application

Koki Kyo and Hideo Noda
Department of Economics, Asahikawa University, Hokkaido, JAPAN

ABSTRACT
In this paper we propose a new approach to regression analysis. In the proposed approach the basic model is transformed into several independent partial models through orthogonal transformations of dependent variable and regression coefficients. In the transformed model, we treat the coefficient components, which are defined in a set of linear combinations of the regression coefficients, as the parameters instead of the regression coefficients. Because every coefficient component is included in an independent partial model, each can be manipulated independently. Hence, model building no longer must be considered a combination of coefficient components. This not only reduces the number of conceivable candidate models, making model selection easier, but it also leads to an improvement in estimates of parameters and the possibility of finding better models. We also analyze the relationship between the productivity of rice farming and the temperature in Hokkaido as an illustrative application of the proposed method. The estimation results suggest the advantages of the proposed approach.

Key words: Multivariate regression model, Model selection, Constrained coefficient component approach, Minimum AIC procedure

1. INTRODUCTION

One of the salient issues of model building in regression analysis has been the selection of explanatory variables. This problem is treated as an issue of determining whether each available explanatory variable can be included or not in a model to be built. It can also be treated as an issue of parameter manipulation, which is a question of whether or not to forcibly set some regression coefficients as zero, in a basic model that takes in all the available explanatory variables. For that reason, we call this the constrained coefficient approach in this paper.

When we set a model via the constrained coefficient approach, however, the number of models considered for selection increases rapidly as the number of explanatory variables increases. Because of that fact, if the number of available explanatory variables is large, model selection, together with model building, becomes a difficult challenge of regression analysis. At times, this problem has become quite serious. For instance, Sala-i-Martin was greatly exercised by a variable selection problem in cross-country economic growth regression with more than 60 explanatory variables [1]. In general, model selection procedures such as backward elimination, forward selection, and stepwise regression are used for convenience [2], [3]. In order to mitigate the difficulties in the constrained coefficient approach, Sala-i-Martin, Doppelhofer and Miller (2004) introduced and employed the Bayesian averaging of classical estimates approach to determine the significance of variables in cross-country growth regressions [4]; Suzuki, Goto, and Tawara (2000) proposed a Bayesian averaging approach to regression analysis for optimal prediction [5]. As mentioned in Suzuki, Goto, and Tawara (2000), however, the development of an algorithm that is both simple and versatile remains a challenge [5]. In addition, their method is not suited for the construction of confidence intervals for parameters, and it is difficult to apply it to structural analysis.

Another important point is that the conventional methods of model selection have completely omitted information on the explanatory variables to be excluded from the model. Thus, the previous studies that have looked at the minimization of such information loss may be insufficient. Consequently, while model selection must involve consideration of an enormous group of candidate models, the possibility of better models becoming available remains unclear. Furthermore, the conventional constrained coefficient approach treats explanatory variables as a choice between two alternatives - whether or not each is to be employed in the model. If an explanatory variable that should be taken into consideration is removed from the model, the degrees of freedom will be excessive,
and therefore the variation in the dependent variable cannot be fully explained, and biases in the estimates of parameters and the predictive value of the dependent variable become larger. Furthermore, from the viewpoint of structural and effective analysis, rather than judging whether each explanatory variable relates to the dependent variable, the analysis of the direction and strength of the influence of each explanatory variable on the dependent variable is considered more important. Therefore, simply excluding the available explanatory variables from the model is not necessarily the best way to improve the model.

In this paper a new approach to regression model building called the constrained coefficient component approach is proposed as an alternative method to the conventional constrained coefficient approach. The proposed approach is executed based on a basic model that includes all meaningful and available explanatory variables. The key idea is that it transforms the basic model into several independent partial models through orthogonal transformation of the dependent variable and regression coefficients. In the transformed model, we treat the coefficient components, which are defined in a set of linear combinations of the regression coefficients, as the parameters instead of the regression coefficients. Because every coefficient component is included in an independent partial model, each can be manipulated independently. Hence, model building no longer must be considered a combination of coefficient components. This not only reduces the number of conceivable candidate models, making model selection easier, but it also leads to an improvement in estimates of parameters and the possibility of finding better models.

A method related to the proposed approach is principal components regression. The principal components regression method seems to have been first proposed by Tintner (1945) [6]. A systematic introduction to the principal components regression method and related references can be found in Basilevsky (1994) [7]. It should be emphasized that the constrained coefficient component approach proposed in this paper is related to principal components regression, but there are remarkable distinctions between these two kinds of approaches. First, the former is a methodology for regression analysis, including model selection and model estimation, but the latter is a technique for mitigating the difficulties in a regression model with multicollinearity. Second, the former is applied as a method for model selection based on the minimum AIC procedure, but the latter is executed based on the singular values of the design matrix of the model under consideration.

The remainder of the paper is organized as follows. In Section 2, the basic model is defined. Further, as preliminaries to the following sections the process of transforming the basic model into independent model components is shown, and formulation of the conventional approach is also given. In Section 3, we introduce theoretical considerations, formulation and procedures for the constrained coefficient component approach proposed. Further, the relation between the proposed and the conventional approaches is mentioned, and performance evaluation for the proposed approach is achieved by several lemmas, theorems and an illustrative example. In Section 4, the relationship between the productivity of rice farming in Hokkaido and the temperature is analyzed as an illustrative application of the proposed method. In Section 5 we summarize the main findings.

2. SETTING AND FORMULATION

2.1 Basic Model

Suppose that there are m variables, \( x_1, x_2, \ldots, x_m \), which presumably explain a random variable, say \( y \). A statistical model that expresses the variation of \( y \) by a linear combination of these variables is called a linear regression model. In a linear regression model, while \( y \) is the subject of statistical analysis, it depends on the variables \( x_1, x_2, \ldots, x_m \). So, \( y \) is called the dependent variable, and \( x_1, x_2, \ldots, x_m \) are called explanatory variables. We consider here a linear regression model for a sample, \( y_1, y_2, \ldots, y_n \), of size \( n \) for \( y \) as

\[
y_i = \sum_{j=1}^{m} x_{ij} \beta_j + \varepsilon_i, \quad (i = 1, 2, \ldots, n) \tag{1}
\]

where \( x_{i1}, x_{i2}, \ldots, x_{im} \) are the values of the explanatory variables corresponding to the \( i \)-th sample value \( y_i \), and \( \beta_1, \beta_2, \ldots, \beta_m \) are unknown parameters which are called regression coefficients. Further, \( \sum_{j=1}^{m} x_{ij} \beta_j \) is called the regression equation for \( y_i \), and \( \varepsilon_i \), which is called the error term, as it is an error for expressing \( y_i \) in terms of the regression equation.

As generally practiced, the following assumptions are made for the model in Eq. (1):

(a) \( x_{ij} \) are deterministic for all \( i, j \).
(b) Expectation and variance of $\varepsilon_i$ are $E\{\varepsilon_i\} = 0$ and $\text{Var}\{\varepsilon_i\} = \sigma^2$ respectively for all $i$.

(c) $\varepsilon_i$ and $\varepsilon_i'$ are independent of each other for all $i \neq i'$.

(d) $\varepsilon_i$ are normally distributed for all $i$.

Assumption (a) is commonly used to simplify the definition of the model. Assumptions (b) and (c) assure that the least-squares estimates of regression coefficients become the best linear unbiased estimates (BLUE), which is commonly referred to as the Gauss-Markov theorem [3]. Furthermore, it is well known that under assumption (d), the least-squares estimates of the parameters are equivalent with the maximum likelihood estimates.

We call a model the basic model when all candidates for the explanatory variables are taken into account. In this paper, we consider the model in Eq. (1) the basic model. In the model expressed in Eq. (1), the degree of freedom is $n-m$, and the regression coefficients $\beta_1, \beta_2, \ldots, \beta_m$ and the error variance $\sigma^2$ can be regarded as unknown parameters. Thus, there are $m+1$ unknown parameters. It is known that the degree of freedom, $n-m$, must be greater than 1, that is, $n > m + 1$, in order to get meaningful parameter estimation.

The model in Eq. (1) can be expressed in a vector-matrix form as follows:

$$ y = X \beta + \varepsilon, \quad (2) $$

where $y = (y_1, y_2, \ldots, y_n)^T$ denotes the dependent variable vector, $\beta = (\beta_1, \beta_2, \ldots, \beta_m)^T$ denotes the regression coefficient vector, $\varepsilon = (\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n)^T$ is the error vector, and $X = (x_{ij})$ is an $n \times m$ matrix, which is called the design matrix.

As implied by assumptions (b), (c), and (d) for the model in Eq. (1), the error vector $\varepsilon$ has an $n$ variate normal distribution with mean $0_n$ and covariance matrix $\sigma^2 I_n$. That is,

$$ \varepsilon \sim N(0_n, \sigma^2 I_n), \quad (3) $$

where $0_n$ denotes an $n$ dimensional vector of zeros, and $I_n$ represents the identity matrix of order $n$. Furthermore, because this article does not take up multicollinearity as a problem, we assume that the design matrix $X$ is a full-rank matrix with rank $m$.

### 2.2 Defining Coefficient Component

We will employ Householder transformation for the design matrix in order to reduce the dimension of the model relevant to the regression coefficient vector $\beta$ and use orthogonal transformations to express the model in a form of independent components.

For the reduction of model dimension, a Householder transformation is applied as

$$ H y = H X \beta + H \varepsilon, \quad (4) $$

where $H$ is an orthogonal matrix with dimension $n$. By using the Householder transformation the design matrix $X$ is transformed as

$$ H X = \begin{bmatrix} T & O \end{bmatrix}, \quad (5) $$

and $H y$ is calculated simultaneously. In Eq. (5), $T$ is an upper triangular matrix with dimension $m$, and $O$ is a null matrix with all elements being zeros. Note that the design matrix $X$ is a full-rank matrix, and hence the matrix $T$ in Eq. (5) is also a full-rank matrix. The algorithms of the Householder transformation for regression analysis were introduced in detail in [8]~[10].

Now, we express the orthogonal matrix $H$ in the form $H^* = \begin{bmatrix} H_1^* & H_2^* \end{bmatrix}$. Here, $H_1$ and $H_2$ are the $m \times n$ matrix and $(n-m) \times n$ matrix, respectively. Then, the model in Eq. (4) is rewritten as follows:

$$ w = T \beta + \eta, \quad (6) $$

$$ e = H_2 \varepsilon, \quad (7) $$

where

$$ w = (w_1, w_2, \ldots, w_m)^T = H_1 y, \quad (8) $$

$$ \eta = (\eta_1, \eta_2, \ldots, \eta_m)^T = H_1 \varepsilon, \quad (9) $$

$$ e = (e_1, e_2, \ldots, e_{n-m})^T = H_2 y. \quad (10) $$

From Eq. (3) and the property of an orthogonal matrix, it can be verified that $\eta$ and $e$ are independent of each other and that their distributions are as follows:

$$ \eta \sim N(0_m, \sigma^2 I_m), \quad (11) $$

$$ e \sim N(0_{n-m}, \sigma^2 I_{n-m}). \quad (12) $$

Thus, the dimension of the model relevant to $\beta$ is reduced to the least.

Further, in order to transform the model in Eq. (6) into independent components, we consider the following singular value decomposition of matrix $T$:

$$ T = U^* D V, \quad (13) $$

where $U$ and $V$ are both orthogonal matrices with dimension $m$, and $D = \text{diag}(d_1, d_2, \ldots, d_m)$ is a diagonal matrix with the diagonal elements
\(d_1, d_2, \ldots, d_m\) being the singular values of matrix \(T\). Obviously, because matrix \(T\) is nonsingular, matrix \(D\) is also nonsingular. Therefore, \(d_i \neq 0 \ (i = 1, 2, \ldots, m)\) can be assured. Incidentally, in a general algorithm, the condition \(d_i > 0\) is imposed on each singular value. This can be assured by skillfully setting the matrices \(U\) and \(V\), but we will not adhere to this convention because it is not the purpose of this paper. While a general algorithm for singular value decomposition has been introduced by Golub and Loan (1989) [11], we will carry out singular value decomposition by making it an eigenvalue problem that is solved through the use of the fact that matrix \(T\) is a nonsingular square matrix.

From Eq. (13), we find the following relations:

\[
T^*T = V^*D^2V, \quad (14)
\]

\[
TT^* = U^*D^2U. \quad (15)
\]

In Eqs. (14) and (15), the matrices \(T^*T\) and \(TT^*\) are both positive definite symmetric matrices with dimension \(m\). Therefore, each has \(m\) positive real eigenvalues. It can also be seen from Eqs. (14) and (15) that matrices \(T^*T\) and \(TT^*\) have the same eigenvalues, say \(d_1^*, d_2^*, \ldots, d_m^*\). When these eigenvalues differ from one another, the orthogonal matrices \(V\) and \(U\) can be derived respectively by resolving Eqs. (14) and (15) as two eigenvalue problems. Then, Eq. (13) can be rewritten as \(D = UTV^*\), and hence each singular value can be calculated by

\[
d_i = u_i^*Tv_i \quad (i = 1, 2, \ldots, m), \quad (16)
\]

where \(u_i\) and \(v_i\) represent the vector of the \(i\)-th column of matrices \(U^*\) and \(V^*\), respectively.

Moreover, by using the orthogonal matrices \(V\) and \(U\) we define an \(m\)-dimensional vector \(\alpha\) by

\[
\alpha = (\alpha_1, \alpha_2, \ldots, \alpha_m)^T = V\beta, \quad (17)
\]

which is as an orthogonal transformation of \(\beta\), and define an \(m\)-dimensional vector \(z\) by

\[
z = (z_1, z_2, \ldots, z_m)^T = U\omega = UH_1y, \quad (18)
\]

which is \(m\) orthogonal component of the dependent variable vector \(y\). Then, the model in Eq. (6) can be rewritten as

\[
z = D\alpha + \psi, \quad (19)
\]

where

\[
\psi = (\psi_1, \psi_2, \ldots, \psi_m)^T = U\eta. \quad (20)
\]

Obviously, each \(\alpha_i\) in Eq. (17) is a linear combination of regression coefficients in the basic model, and therefore we can call each a coefficient component. Correspondingly, vector \(\alpha\), defined by Eq. (17), is called a coefficient component vector. It is obvious that \(\psi\) is independent of \(\epsilon\) and that its distribution is

\[
\psi \sim N(0_m, \sigma^2I_m). \quad (21)
\]

The model in Eq. (2) is equivalent to the model in Eqs. (6) and (7), and furthermore it is also equivalent to the model defined in Eq. (19) together with Eq. (7). In particular, \(z\) in the model in Eq. (19) plays an important role as the pivot for defining the model relevant to \(\beta\) or equivalently \(\alpha\). Thus, \(z\) is referred to as a pivot vector, and each of its elements is called a pivot component. Incidentally, the model in Eq. (19) can be expressed in components as

\[
z_i = d_i\alpha_i + \psi_i \quad (i = 1, 2, \ldots, m). \quad (22)
\]

The following facts are evident from the model in Eq. (22): The model involving each pivot component has one coefficient component, and each component of the model is independent of the others. Thus, each coefficient component can be manipulated independently.

### 2.3 Conventional Approach

In the conventional approach of model building, when a certain explanatory variable, say \(x_i\), is to be included in a candidate model, the corresponding regression coefficient, \(\beta_i\), is treated as a parameter, and when it is to be excluded from the candidate model, the constraint \(\beta_i = 0\) is imposed. Here, a constraint on the regression coefficient \(\beta_i\) is expressed as follows:

\[
\text{Re}(\beta_i) : \beta_i = 0 \quad (i = 1, 2, \ldots, m), \quad (23)
\]

and for all of the regression coefficients in the basic model, a set called the coefficient constraint set, is defined as

\[
\mathcal{B} \equiv \{\text{Re}(\beta_1), \text{Re}(\beta_2), \ldots, \text{Re}(\beta_m)\}. \quad (24)
\]

Because each subset of the coefficient constraint set \(\mathcal{B}\) corresponds to one candidate model, the number of candidate models is equivalent to the number of all different subsets of the coefficient constraint set \(\mathcal{B}\), which is \(2^m\).
3. PROPOSED APPROACH

3.1 Consideration and Formulation

We now propose a new method called the constrained coefficient component approach for building regression models. In this approach, a model is constructed by manipulating the coefficient components in the model in Eq. (19) instead of regression coefficients.

Similarly to the conventional approach, manipulation for the coefficient components can be done in two ways. One is to impose a zero constraint on the coefficient components, and the other is to treat them as unknown parameters. Specifically, for \( i = 1, 2, \ldots, m \), the former is to constrain the coefficient component \( \alpha_i \) by

\[
\alpha_i = 0, \quad (25)
\]

and the latter is to regard it as an unknown parameter and estimate it by

\[
\hat{\alpha}_i = \frac{z_i}{d_i}, \quad (26)
\]

which is the same as the maximum likelihood estimate.

In the conventional constrained coefficient approach the degrees of freedom of a model are controlled by constraining some specific regression coefficients to be zeros. This is natural from the viewpoint of explanatory variable selection. Considering coefficient components in light of adjustment of the degrees of freedom, however, the possibility exists that a constraint with a constant other than zero might be placed. Hence, the importance of constraining with zero as in Eq. (25) is not obvious. This can be interpreted from the viewpoint of the stability of prediction for the dependent variable. In other words, constraining the relevant coefficient components rather than constraining them with another constant can minimize the variance of the prediction for the dependent variable, or equivalently, that for the pivot components. In this paper, such rules for setting up a model will be called the principle of minimum variance prediction (see Appendix A for the explanation in detail).

By analogy with the expression of Eqs. (23) and (24), the constraint on the coefficient components in Eq. (25) is expressed in the following form:

\[
\text{Re}(\alpha_i) : \alpha_i = 0 \quad (i = 1, 2, \ldots, m). \quad (27)
\]

Then, the set of constraints on all coefficient components that are applicable to the model in Eq. (19) is defined as

\[
\mathcal{A} = \{\text{Re}(\alpha_1), \text{Re}(\alpha_2), \ldots, \text{Re}(\alpha_m)\}. \quad (28)
\]

Furthermore, if the coefficient component constraints in a specific subset in the coefficient component constraint set \( \mathcal{A} \) is applied to the model in Eq. (19), then a candidate model with reasonable degrees of freedom can be controlled. Although models that can be constructed as just described are as many as \( 2^m \) in total, if the priority of constraints in the coefficient component constraint set \( \mathcal{A} \) is definable based on a criterion such as AIC, then those that are not sufficiently competitive are eliminated from the candidate models a priori. In this way, since the number of the conceivable candidate models is reduced, it will facilitate the simplification of the model selection process. Then, the next important issue becomes the possibility and method of defining the priority of each coefficient component constraint in the coefficient component constraint set \( \mathcal{A} \).

3.2 Fundamental Procedure

Suppose that in the set, \( \{|z_1|, |z_2|, \ldots, |z_m|\} \), no identical members exist of the absolute values of each pivot component in the model in Eq. (19). The series of \( z_1, z_2, \ldots, z_m \) is rearranged in descending order of \( |z_1|, |z_2|, \ldots, |z_m| \) and expressed as

\[
z_1^*, z_2^*, \ldots, z_m^*, \quad (29)
\]

while the series of the coefficient components, \( \alpha_1, \alpha_2, \ldots, \alpha_m \), is rearranged to produce the following:

\[
\alpha_1^*, \alpha_2^*, \ldots, \alpha_m^*. \quad (30)
\]

In addition, for an integer \( k \) that satisfies \( 0 \leq k \leq m \), when \( k > 0 \), \( \mathcal{A}_k \) denotes a subset of \( \mathcal{A} \) defined by

\[
\mathcal{A}_k = \{\text{Re}(\alpha_{m-k+1}^*), \ldots, \text{Re}(\alpha_m^*)\}. \quad (31)
\]

When \( k = 0 \), it is defined as \( \mathcal{A}_0 = \Phi \) with \( \Phi \) representing an empty set. Based on the above notation, the next theorem is developed (see Appendix B for proof).

**Theorem 1** For a given integer \( k \), among the models obtained by applying any \( k \) coefficient component constraints in the coefficient component constraint set \( \mathcal{A} \) to the basic model in Eqs. (19) and (7), the model in which the coefficient component constraints in \( \mathcal{A}_k \) is employed leads to
the minimum AIC value. The corresponding AIC value is given by

$$AIC(A_k) = n \ln(\hat{\sigma}^2(A_k)) + \lambda.$$  \hspace{1cm} (32)

In Eq. (32), $\lambda$ is a constant defined as

$$\lambda = n(\ln{2\pi}) + 2(m-k+1),$$  \hspace{1cm} (33)

and $\hat{\sigma}^2(A_k)$ is the maximum likelihood estimate of the error variance that is given by

$$\hat{\sigma}^2(A_k) = \frac{1}{n} \left[ \sum_{i=m-k+1}^{m} (z_i^*)^2 + e^t \hat{e} \right]$$

$$= \frac{1}{n} \left[ \sum_{i=m-k+1}^{m} (z_i^*)^2 + \sum_{i=1}^{n-m} e_i^2 \right].$$  \hspace{1cm} (34)

where $e = (e_1, e_2, \ldots, e_{n-m})^t$ has been defined by Eq. (10).

As pointed out by Konishi and Kitagawa (2004), and Matsushima (1996) [12], [13], a number of information criteria, including AIC and BIC, etc., can be defined by a general form as follows:

$$c_1 \ln(\hat{\sigma}^2) + c_2,$$  \hspace{1cm} (35)

where $c_1$ and $c_2$ are two constants that are free from sample data, and $\hat{\sigma}^2$ denotes the maximum likelihood estimate of the error variance of a regression model under consideration. From Appendix B, it can be seen that the result expressed by Theorem 1 is always true when we substitute any information criterion defined by the form in Eq. (35) for AIC. Thus, Theorem 1 presents a very general procedure for the proposed approach.

Moreover, Theorem 1 suggests that for a given $A_k$, the conceivable candidate models are solely identified. Thereby, let $M(A_k)$ denote the model to which the coefficient component constraints in the set $A_k$ are applied. In the proposed approach, model selection should be pursued only in the model set $\{M(A_0), M(A_1), \ldots, M(A_m)\}$.

When the minimum AIC procedure is applied in this way, “the best model!” can be selected easily from the model set $\{M(A_0), M(A_1), \ldots, M(A_m)\}$. Thus, whether each coefficient component in the basic model is a parameter or not can be determined based on Theorem 1. Here, we exclude the two extreme cases mentioned previously and consider the value of $k \in \{1, 2, \ldots, m-1\}$. If $M(A_k)$ is the best model, then by the expression defined for Theorem 1 the set of coefficient components to be estimated by the maximum likelihood method is $\{\alpha_1^*, \alpha_2^*, \ldots, \alpha_m^*\}$, and the complementary set is $\{\alpha_{m-k+1}^*, \alpha_{m-k+2}^*, \ldots, \alpha_m^*\}$.

The estimate of the coefficient component vector, $\alpha^* = (\alpha_1^*, \alpha_2^*, \ldots, \alpha_m^*)^t$, for the best model is given by $\hat{\alpha}^* = (\hat{\alpha}_1^*, \hat{\alpha}_2^*, \ldots, \hat{\alpha}_{m-k}^*, 0, \ldots, 0)^t$. Furthermore, the corresponding estimate of the regression coefficient vector $\beta$ by the proposed approach can be obtained from Eq. (10).

### 3.3 Extensional Procedure

Usually, a regression model may be defined with an unknown constant term. In such a case, in order to maintain the results obtained by using the above procedures, we regard $\beta_1$ as the constant term by putting $x_{11} = 1$, and replacing the values of $x_{ij}$ in the basic model in Eq. (1) by the corresponding centered values $\bar{x}_{ij} = x_{ij} - \bar{x}_j$ for $i = 1, 2, \ldots, n$; $j = 2, 3, \ldots, m$. Further, if we define $\alpha_1$ by

$$\alpha_1 = \beta_1 + \sum_{j=2}^{m} \bar{x}_j \beta_j,$$

and define a vector $(\alpha_2, \alpha_3, \ldots, \alpha_m)^t$ as an appropriate orthogonal transformation of $(\beta_2, \beta_3, \ldots, \beta_m)^t$, then we can obtain a model that is defined on $(\alpha_1, \alpha_2, \ldots, \alpha_m)^t$ as the coefficient components. Thus, the procedure for the proposed approach can be extended to the present situation. The procedure for this extensional procedure was first introduced in detail by Kyo (2005) [14]. By the extensional procedure we can not only utilize the proposed approach but also maintain the structure of the original model.

### 3.4 Relation to Conventional Method

For the subsequent discussion, we summarize here the relation between the models derived in order to evaluate the proposed and the conventional approaches. It is obvious from Section 2.2 that the model expressed by Eqs. (6) and (7) and the model expressed by Eqs. (19) and (7) are the fundamentals of the approaches and they are equivalent to each other because both of them are obtained from the same basic model in Eq. (2) by using orthogonal transformations. It should be pointed out that the expressions in Eqs. (3), (11) and (21) are in consistency with each other before modeling strategy is practiced. Moreover, the regression coefficient vector $\beta$, which is the key to practicing the conventional approach, is only related to the model in Eq. (6). On the other hand, the coefficient component vector $\alpha$, which is the key to practicing the proposed approach, is only related
to the model in Eq. (19). Thus, we only focus on Eqs. (6) and (19) to consider the relationship between the two kinds of approaches.

Here, we express each vector in Eqs. (6) and (19) as follows: $w = (w_1, w_2)\trans, \beta = (\beta_1, \beta_2)\trans, \eta = (\eta_1, \eta_2)\trans, z = (z_1, z_2)\trans, \alpha = (\alpha_1, \alpha_2)\trans$ and $\psi = (\psi_1, \psi_2)\trans$. Correspondingly, the upper triangular matrix $T$, diagonal matrix $D$, and orthogonal matrix $V$ are expressed respectively by

$$T = \begin{bmatrix} T_{11} & T_{12} \\ O & T_{22} \end{bmatrix},$$

$$D = \begin{bmatrix} D_1 & O \\ O & D_2 \end{bmatrix},$$

$$V = \begin{bmatrix} V_{11} & V_{12} \\ V_{21} & V_{22} \end{bmatrix}.$$  

If we assume that the vectors $w_2, \beta_2, \eta_2, z_2, \alpha_2$, and $\psi_2$ have the same dimension, say $k$, and the square matrices $T_{22}, D_2$, and $V_{22}$ also have the same dimension, then the dimensions of other partial vectors and partial matrices are obvious. In this paper, we focus on the case that $0 < k < m$. Using the above expressions, the partial model in Eq. (6) can be rewritten as

$$\begin{bmatrix} w_1 \\ w_2 \end{bmatrix} = \begin{bmatrix} T_{11} & T_{12} \\ O & T_{22} \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} + \begin{bmatrix} \eta_1 \\ \eta_2 \end{bmatrix},$$  

(36)

and the partial model in Eq. (19) is

$$\begin{bmatrix} z_1 \\ z_2 \end{bmatrix} = \begin{bmatrix} D_1 & O \\ O & D_2 \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \alpha_2 \end{bmatrix} + \begin{bmatrix} \psi_1 \\ \psi_2 \end{bmatrix}.$$  

(37)

Further, Eq. (17) is rewritten as

$$\begin{bmatrix} \alpha_1 \\ \alpha_2 \end{bmatrix} = \begin{bmatrix} V_{11} & V_{12} \\ V_{21} & V_{22} \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix}.$$  

(38)

Obviously, the models in Eqs. (36) and (37) contain the following relations:

$$w_2 = T_{22}\beta + \eta_2,$$  

(39)

$$z_2 = D_2\alpha + \psi_2.$$  

(40)

As mentioned above, before modeling strategy is practiced the models in Eqs. (39) and (40) are contained in each basic model respectively, so the distributions in Eqs. (11) and (21) can be applied. Hence, the following relations are obtained immediately.

$$E\{w_1^2w_2\} = \beta^\trans T_{22}^\trans T_{22}\beta + k\sigma^2,$$  

(41)

$$E\{z_1^2z_2\} = \alpha^\trans D_2^\trans \alpha + k\sigma^2,$$  

(42)

where $E$ stands for the expectation with respect to the distribution expressed by Eqs. (11) and (21) respectively. From Eqs. (41) and (42), we have

$$E\{w_1^2w_2\} - \beta^\trans T_{22}^\trans T_{22}\beta = E\{z_1^2z_2\} - \alpha^\trans D_2^\trans \alpha.$$  

(43)

Note that the above equation is a very important result for the subsequent discussion.

### 3.5 Performance Evaluation

As mentioned in Section 3.2, we can obtain a very efficient procedure for regression model selection by proposing the constrained coefficient component approach. However, though this is a great advantage of the proposed approach, there is another important issue that must be inspected theoretically in order to show the performance of the proposed approach.

Firstly, we give the following lemma as the basis of the succeeding results (the proof is omitted).

**Lemma 1** Under the basic model, the maximum likelihood estimates $\hat{\alpha}$ and $\hat{\alpha}_2$ for $\alpha_1$ and $\alpha_2$ are given respectively by $\hat{\alpha}_1 = D_1^{-1}z_1$ and $\hat{\alpha}_2 = D_2^{-1}z_2$. Furthermore, for a given value of the error variance $\sigma^2$, $\hat{\alpha}_1 \sim N(\alpha_1, \sigma^2 D_1^{-2})$ and $\hat{\alpha}_2 \sim N(\alpha_2, \sigma^2 D_2^{-2})$ are independent of each other.

Without loss of generality, in the conventional approach we consider the model which is constructed by applying the following constraints:

$$\alpha_2 = 0_k.$$  

(44)

Thus, the model in Eq. (36) is expressed in the form

$$\begin{bmatrix} w_1 \\ w_2 \end{bmatrix} = \begin{bmatrix} T_{11} \\ O \end{bmatrix} \beta_1 + \begin{bmatrix} \eta_1 \\ \eta_2 \end{bmatrix}.$$  

(45)

On the other hand, in the proposed approach, if we put the constraints on $\alpha_2$ in the form

$$\alpha_2 = 0_k,$$  

(46)

then the model in Eq. (37) becomes as follows:

$$\begin{bmatrix} z_1 \\ z_2 \end{bmatrix} = \begin{bmatrix} D_1 \\ O \end{bmatrix} \alpha_1 + \begin{bmatrix} \psi_1 \\ \psi_2 \end{bmatrix}.$$  

(47)

Thus, the model expressed by Eqs. (45) and (7) is constructed in the constrained coefficient approach, while the model expressed by Eqs. (47) and (7) is built in the constrained coefficient-component approach. The next lemma can be derived for these two types of modeling (see Appendix C for proof).
Lemma 2 Under the basic model and the principle of minimum variance prediction, Eq. (46) holds as long as Eq. (34) holds, i.e., Eq. (46) is a sufficient condition of Eq. (44).

Lemma 2 shows that Eq. (46) must hold when Eq. (44) holds, i.e., the model constructed by using the proposed approach contains that constructed by using the conventional approach. This result leads to the next theorem (see Appendix D for proof).

Theorem 2 Let $\hat{\sigma}_1^2(k)$ be the maximum likelihood estimate of error variance for the model in Eqs. (45) and (7) that is constructed in the constrained coefficient approach, and let $\hat{\sigma}_2^2(k)$ be that for the model in Eqs. (47) and (7) that is constructed in the constrained coefficient component approach. Then the following inequality holds if Eq. (46) holds but Eq. (44) does not hold,

$$E[\hat{\sigma}_1^2(k)] > E[\hat{\sigma}_2^2(k)],$$

and the following equality holds if and only if both of Eqs. (46) and (44) hold:

$$E[\hat{\sigma}_1^2(k)] = E[\hat{\sigma}_2^2(k)],$$

where $E$ stands for the expectation with respect to the distribution defined by the basic model.

Theorem 2 together with Lemma 2 implies that on average performance of the proposed approach is better than that of the conventional approach; the worst is the case that these two approaches have the same performance but it is less likely to occur. In addition, by the definition of AIC, we can see that the following relations hold based on Eqs. (48) and (49).

$$E\left[\exp\{AIC_1(k)\}\right] \geq E\left[\exp\{AIC_2(k)\}\right],$$

where $E$ has the same meaning as that in Eqs. (48) and (49), $AIC_1(k)$ and $AIC_2(k)$ denote the AIC values that correspond respectively to $\hat{\sigma}_1^2(k)$ and $\hat{\sigma}_2^2(k)$. The relation in Eq. (50) is an important result which shows that it is possible to obtain better model if the proposed approach is applied.

### Table 1 Results for the conventional approach

<table>
<thead>
<tr>
<th>Explanatory variables applied in models</th>
<th>$x_1$</th>
<th>$x_1, x_2$</th>
<th>$x_1, x_3$</th>
<th>$x_1, x_4$</th>
<th>$x_1, x_2, x_3$</th>
<th>$x_1, x_2, x_4$</th>
<th>$x_1, x_2, x_3, x_4$</th>
</tr>
</thead>
<tbody>
<tr>
<td>The number of constraints</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>The number of parameters</td>
<td>2</td>
<td>3</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>Estimate of $\sigma^2$</td>
<td>1488.51</td>
<td>168.88</td>
<td>284.34</td>
<td>1126.62</td>
<td>92.08</td>
<td>145.68</td>
<td>106.67</td>
</tr>
<tr>
<td>AIC Value</td>
<td>186.87</td>
<td>165.54</td>
<td>175.76</td>
<td>203.30</td>
<td>155.43</td>
<td>164.39</td>
<td>158.15</td>
</tr>
</tbody>
</table>

### Table 2 Results for the proposed approach

<table>
<thead>
<tr>
<th>The number of constraints</th>
<th>3</th>
<th>2</th>
<th>1</th>
<th>0</th>
</tr>
</thead>
<tbody>
<tr>
<td>The number of parameters</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>Estimate of $\sigma^2$</td>
<td>1488.51</td>
<td>96.54</td>
<td>66.89</td>
<td>83.44</td>
</tr>
<tr>
<td>AIC Value</td>
<td>206.87</td>
<td>154.16</td>
<td>148.82</td>
<td>149.76</td>
</tr>
</tbody>
</table>

3.6 Illustrative Example

We give an example in the practice of management as an illustration to show the performance of the proposed approach. The objects in this example are 20 big retail dealers including 11 supermarkets and 9 department stores. Here, we take the proceeds as the dependent variable. The basic model is constructed in a constant term $x_1 = 1$ and three explanatory variables, the number of employees $x_2$, the floor space $x_3$, and a dummy variable $x_4$ which is defined by $x_4 = 0$ for supermarkets and $x_4 = 1$ for department stores. That is, the basic model can be expressed as the model in Eq. (1) with $m = 4$ and $n = 20$. Data used here are cited from Takeuchi and Tsukuda (1990, pp. 92-94) [15].

We use the extensional procedure mentioned in Section 3.3 and only consider here variation of models on $x_2$, $x_3$ and $x_4$ by regarding $b_1$ as an unknown parameter for simplicity. Thus, there are $2^3 = 8$ candidate models for the conventional approach and 4 candidate models for the proposed approach. The estimates of error variance and AIC values for the conventional and the proposed approaches are shown, respectively, in Table 1 and Table 2. It is evident from these Tables that the performance of the proposed approach is better than that of the conventional approach. The models constructed in the two approaches are equivalent to each other when they are defined with the maximum and the minimum degrees of freedom, so in these cases the models have the same performance.
4. APPLICATION

4.1 Model
In this section, as an illustration for the proposed approach, we analyze the relationship between the productivity of rice farming and the temperature in Hokkaido. The model used here is written as

\[ q_i = a s_i^{h_i} \quad (i = 1, 2, \ldots, n), \quad (51) \]

where \( q_i \) represents the crop yields of rice, and \( s_i \) is the planted area in year \( i \). In addition, \( a \) is an unknown constant and \( h_i \) is the rate of increase in the crop yields in terms of the rate of increase per unit of planted area, i.e., the elasticity coefficient of the crop yields of the planted area. The subscript, \( i \), of \( h_i \) indicates that the elasticity coefficient varies from year to year. The model in Eq. (51) will be made as follows if both sides are transformed logarithmically and a disturbance term is added.

\[ \ln\{q_i\} = \ln\{a\} + \ln\{s_i\}h_i + \varepsilon_i, \quad (52) \]

where \( \varepsilon_i \) denotes the error term when we express \( \ln\{q_i\} \) in terms of \( \ln\{a\} + \ln\{s_i\}h_i \). As shown in Eq. (52), \( h_i \) denotes the productivity of the logarithm of rice production with respect to the logarithm of the rice-planted area.

While technical factors in cultivation such as the use of fertilizers and breed improvement are sure to affect the productivity of rice farming directly, the influence of meteorological variables as natural conditions cannot be ignored. As a macro model used for the productivity analysis of rice farming, only those influences by meteorological variables are incorporated into the model in this article. While there are many meteorological variables, including temperature and rainfall, etc., that may be taken into consideration, in Hokkaido the temperature is the most important factor that influences rice farming productivity. Here, we assume that the elasticity coefficient \( h_i \) is a linear function of the monthly average temperature, that is,

\[ h_i = b_0 + \sum_{j=1}^{12} t_j^{(i)} b_j, \quad (53) \]

where \( t_j^{(i)} \) represents the average temperature in Sapporo in month \( j \) of year \( i \), and \( b_0, b_1, \ldots, b_{12} \) are unknown coefficients. The coefficients from \( b_1 \) through \( b_{12} \) indicate the influence of monthly average temperatures on rice farming productivity, so they are called monthly temperature effects. By using Eq. (53), Eq. (52) can be rewritten as follows:

\[ \ln\{q_i\} = \ln\{a\} + \ln\{s_i\}b_0 + \sum_{j=1}^{12} \ln\{s_i\} t_j^{(i)} b_j + \varepsilon_i. \quad (54) \]

Further, by putting \( y_i = \ln\{q_i\}, x_1 = \ln\{a\}, \beta_1 = \ln\{a\}, x_{12} = \ln\{s_i\}, x_{j+2} = \ln\{s_i\} t_j^{(i)} \quad (j = 1, 2, \ldots, 12) \), the model in Eq. (54) can be expressed as the linear regression model in Eq. (1). In this case, the number of explanatory variables is \( m = 14 \) with \( x_{11} = 1 \) corresponding to the constant term.

4.2 Model Selection
Data used for model estimation and selection are from a period of 113 years between 1890 and 2002. Consequently the sample size is \( n = 113 \). Because the number of explanatory variables in the basic model is 14 (including the constant term), the number of corresponding coefficient components will also be 14. Based on the procedure of the constrained coefficient-component approach

\footnote{Reference material source: Report on Survey of Agriculture in Cold Districts, Annual Statistics of Agriculture in Hokkaido and Crops Statistics by the Ministry of Agriculture, Forestry and Fisheries.}
proposed, the values of the pivot components, \( z_1, z_2, \ldots, z_{14} \), for the model in Eq. (19), and the values of \( d_1, d_2, \ldots, d_{14} \) are obtained. As indicated by Theorem 1, the priority of constraints on each coefficient component can be determined by the absolute values of each pivot component \( z_i \). The sum of squared errors of the basic model, \( \sum_{i=1}^{n-m} e_{i}^2 \) is needed to calculate the maximum likelihood estimate of error variance in Eq. (34). In this case, \( \sum_{i=1}^{n-m} e_{i}^2 = 14.3449 \). Table 3 shows the values of \( z_i \) and \( d_i \) as well as the priority of the coefficient component constraints.

Then, based on the priority of constraints on each coefficient component we can build conceivable candidate models by imposing the zero constraint successively on each coefficient component. By applying Eqs. (32), (33), and (34), the values of \( \hat{\sigma}_k^2 \), the estimate of error variance for the model to which \( k \) coefficient components are applied, and \( AIC_k \) can be calculated, and the results are shown in Table 4. It can be seen that the model with the minimum AIC value is that to which the number of coefficient components to be applied is 8. The order of priority of the coefficient components to which constraints will be applied is 8, 10, 13, 5, 12, 6, 4, and 11. Meanwhile, the basic model with no constraints to be added has the estimate of error variance of \( \hat{\sigma}_0^2 = 0.12695 \) and the AIC value of \( AIC_0 = 117.45 \). Moreover, as indicated in Table 4, no clear differences exist among the AIC values of those models to which the number of coefficient component constraints to be applied is between 6 and 10. In an actual application, model selection should be made by examining calculation results from various areas.

### 4.3 Estimation Results

Now, based strictly on the minimum AIC procedure, we select a model to which 8 coefficient component constraints are applied as the best model and estimate the parameters of this model. The parameters of the model defined in Eqs. (51) and (53) correspond one-to-one with the parameters in the estimated regression model. By virtue of that fact, estimated values of such parameters are obtainable from the estimation results of the regression model. The estimation results obtained in this way are as follows:

\[
q_i = 1.7565e_i^{h_i},
\]

\[
h_i = 0.4243 + 0.0007t_i^{(i)} + 0.0021t_i^{(i)} + 0.0020t_5^{(i)} + 0.0031t_4^{(i)} + 0.0067t_3^{(i)} + 0.0088t_3^{(i)} + 0.0092t_2^{(i)} + 0.0068t_1^{(i)} + 0.0024t_1^{(i)} + 0.0009t_{10}^{(i)} - 0.0023t_{11}^{(i)} + 0.0044t_{12}^{(i)}.
\]

These results are very meaningful for analyzing rice farming productivity and yield forecasts. As seen in the model in Eq. (56), the temperature effects are the highest in July, and the values of the months between May and August are notably high. This result agrees with the common knowledge of crop farming. Furthermore, while the temperature effects during the seasons are apparently irrelevant to the rice production and yield relatively small values, the results indicate that almost all estimated temperature effects are positive. This fact implies that there is autocorrelation between the temperatures in different months.

In the foregoing analytical model, the number of explanatory variables is as large as 14. For this reason, selecting a model using the conventional constrained coefficient approach is presumed to be fairly difficult. Further, because the conventional approach cannot allow all of the temperature effects in each month to appear in the model, not only is the comparison of the strength of each effect impossible, but also the estimated and predicted

<table>
<thead>
<tr>
<th>Table 4 Estimates of error variance and AIC values of each candidate model</th>
</tr>
</thead>
<tbody>
<tr>
<td>The number of applied coefficient constraints (k)</td>
</tr>
<tr>
<td>Additional coefficient constraint number</td>
</tr>
<tr>
<td>Estimate of error variance, ( \hat{\sigma}_k^2 )</td>
</tr>
<tr>
<td>The number of free parameters</td>
</tr>
<tr>
<td>AIC_k value</td>
</tr>
<tr>
<td>---------------------------------------------------------------</td>
</tr>
<tr>
<td>The number of applied coefficient constraints (k)</td>
</tr>
<tr>
<td>Additional coefficient constraint number</td>
</tr>
<tr>
<td>Estimate of error variance, ( \hat{\sigma}_k^2 )</td>
</tr>
<tr>
<td>The number of free parameters</td>
</tr>
<tr>
<td>AIC_k value</td>
</tr>
</tbody>
</table>
values accompanying the exclusion of special temperature effects are likely to involve a considerable bias. Such facts should also be considered among the advantages of the newly proposed constrained coefficient approach.

5. CONCLUSION

In this paper, we have proposed the constrained coefficient component approach as a new method in place of the conventional constrained coefficient approach, which constructs a model by selecting explanatory variables for a regression analysis. In the proposed approach, a linear regression model that employs all available explanatory variables is regarded as the basic model, and a set of coefficient components is structured through orthogonal transformations of the regression coefficients. Subsequently, candidate models are constructed by adding constraints to the structured coefficient components, upon which the models are selected and estimated. The following has been explained theoretically for the proposed approach. Using the proposed approach, the number of candidate models can be reduced, and thereby the operation of model selection becomes easier than that in the conventional approach. It is also an advantage of the proposed approach that the procedure can lead to an improvement in estimates of parameters and the possibility of finding better models. Finally, we have applied the proposed approach to analyzing the effects of temperatures on rice farming productivity in Hokkaido, thereby verifying the advantages of the constrained coefficient component approach in that case.

ACKNOWLEDGMENTS

We gratefully acknowledge the valuable comments of the anonymous referees. This study was carried out under ISM Cooperative Research Program (2005-ISM-CRP-2026).

REFERENCES


Appendix A. The Principle of Minimum Variance Prediction

If we impose the constraint $\alpha_i = c$ on the coefficient component $\alpha_i$ by using an appropriate constant $c$, then we have $z_i = c \delta_i + \psi_i$ from the model in Eq. (22). Thus, from the assumptions about
the distribution of $\psi_i$, the expectation of $z_i^2$ is as follows:

$$E(z_i^2) = c^2 d_i^2 + \sigma^2.$$ 

Furthermore, by using the relationship between the variance $\text{Var}(z_i)$ of $z_i$ and its expectation $E(z_i)$, we have the following equation:

$$\text{Var}(z_i) = E(z_i^2) - (E(z_i))^2 = c^2 d_i^2 + \sigma^2 - (E(z_i))^2. \quad (57)$$

Obviously, all of the quantities $(E(z_i))^2$, $d_i^2 > 0$ and $\sigma^2$ in Eq. (57) are free from the constant $c$, and hence we can see that the variance $\text{Var}(z_i)$ of $z_i$ is in proportion to $c^2$. Thus, $\text{Var}(z_i)$ is minimized by putting $c = 0$. Therefore, for each coefficient component $\alpha_i$, the constraint expressed in Eq. (25) gives minimum variance of prediction for $z_i$.

**Appendix B. Proof of Theorem 1**

Here, we redefine a series of pivots by

$$\tilde{z}_1, \tilde{z}_2, \ldots, \tilde{z}_m$$

in a permutation of the pivot series $z_1, z_2, \ldots, z_m$ which is used in Eq. (19), and correspondingly a permutation of $d_1, d_2, \ldots, d_m$ and of $\alpha_1, \alpha_2, \ldots, \alpha_m$ are expressed respectively as $\tilde{d}_1, \tilde{d}_2, \ldots, \tilde{d}_m$ and $\tilde{\alpha}_1, \tilde{\alpha}_2, \ldots, \tilde{\alpha}_m$. Further, let $\tilde{\mathbf{A}}_k$ be a subset of the coefficient component constraint set $\mathbf{A}$ defined by Eq. (28) for $0 < k < m$. In particular, for $k = 0$ it is defined by $\tilde{\mathbf{A}}_0 = \emptyset$, an empty set, and for $k > 0$ it is defined by

$$\tilde{\mathbf{A}}_k \equiv \{ \text{Re}(\tilde{\alpha}_{m-k+1}), \ldots, \text{Re}(\tilde{\alpha}_m) \},$$

where Re($\tilde{\alpha}_i$) is defined by Eq. (27).

As mentioned in Section 3.4, the likelihood for the basic model defined by Eq. (2), or equivalently defined by Eqs. (6) and (7), and that of the model defined by Eqs. (19) and (7) are equivalent to each other. Thus, from Eqs. (7) and (12), the likelihood for the partial model defined by Eq. (7) is

$$L_1(\sigma^2) = (2\pi \sigma^2)^{-\frac{m}{2}} \exp \left\{ -\frac{1}{2\sigma^2} e^T e \right\},$$

and the likelihood for the partial model defined by Eq. (19) as

$$L_2(\alpha, \sigma^2) = (2\pi \sigma^2)^{-\frac{m}{2}} \times \exp \left\{ -\frac{1}{2\sigma^2} \sum_{i=1}^{m} (\tilde{z}_i - \tilde{d}_i \tilde{\alpha}_i)^2 \right\}.$$

Further, since these two partial model are independent of each other, the likelihood for the basic model can be obtained by

$$L(\alpha_0, \alpha, \sigma^2) = L_1(\alpha_0, \sigma^2) \times L_2(\alpha, \sigma^2),$$

and hence the corresponding log-likelihood is

$$\ell(\alpha_0, \alpha, \sigma^2) = \ln\{L(\alpha_0, \alpha, \sigma^2)\} = -\frac{n}{2} \ln\{2\pi \sigma^2\} - \frac{1}{2\sigma^2} \sum_{i=1}^{m} (\tilde{z}_i - \tilde{d}_i \tilde{\alpha}_i)^2 + e^T e.$$ 

For the model under the constraints in the set $\tilde{\mathbf{A}}_k$, the estimates of the free coefficient components are obtained from the following equation:

$$\tilde{z}_i - \tilde{d}_i \tilde{\alpha}_i = 0 \quad (i = 1, 2, \ldots, m - k).$$

Then, the maximum likelihood estimate of $\sigma^2$ is given by

$$\hat{\sigma}^2(\tilde{\mathbf{A}}_k) = \frac{1}{n} \sum_{i=m-k+1}^{m} \tilde{z}_i^2 + e^T e. \quad (59)$$

Let $\tilde{\alpha}(\tilde{\mathbf{A}}_k)$ be the maximum likelihood estimate for the vector of the free coefficient components $\alpha(\tilde{\mathbf{A}}_k)$. Under the constraints in the set $\tilde{\mathbf{A}}_k$, the maximum log-likelihood is given by

$$\ell(\tilde{\alpha}(\tilde{\mathbf{A}}_k), \hat{\sigma}^2(\tilde{\mathbf{A}}_k)) = -\frac{n}{2} \left[ \ln\{2\hat{\sigma}^2(\tilde{\mathbf{A}}_k)\} + 1 \right].$$

According to Sakamoto, Ishiguro and Kitagawa (1986) [10], the general definition of Akaike information criterion, AIC, is given by

$$\text{AIC} = -2(\text{maximum log-likelihood}) + 2(\text{number of parameters}). \quad (60)$$

Obviously, the number of all parameters in the basic model is $m + 1$, and if the constraints in the set $\tilde{\mathbf{A}}_k$ are applied, then $k$ coefficient components are constrained so that the number of free parameters becomes $m - k + 1$. Therefore, applying the definition of AIC in Eq. (60), the value of AIC for the model is calculated by

$$\text{AIC}(\tilde{\mathbf{A}}_k) = -2\ell(\tilde{\alpha}(\tilde{\mathbf{A}}_k), \hat{\sigma}^2(\tilde{\mathbf{A}}_k)) + 2(m - k + 1)
= n \ln\{\hat{\sigma}^2(\tilde{\mathbf{A}}_k)\} + \lambda, \quad (61)$$

where $\lambda$ is defined by Eq. (33).

It is obvious that the value of AIC defined by Eq. (61) is a monotone increasing function of $\hat{\sigma}^2(\tilde{\mathbf{A}}_k)$.
Thus, when the number of the constrained coefficient components \( k \) is fixed, the smaller the value of \( \delta^2(\mathbf{A}_k) \), then the smaller the value of AIC. If we apply the minimum AIC procedure, then it is more desirable to use the constraints that lead to a smaller \( \delta^2(\mathbf{A}_k) \) so that AIC can also get a smaller value. It can be seen from Eq. (59) that when we use the smallest \( k \) elements in the set \( \{ z_{l1}^1, z_{l2}^1, \ldots, z_{lm}^1 \} \) as \( z_{l1}^2, z_{l2}^2, \ldots, z_{km}^2 \), then the value of \( \delta^2(\mathbf{A}_k) \) will be smaller than all other selections for \( z_{l1}^2, z_{l2}^2, \ldots, z_{km}^2 \). Thus, the minimum AIC value can be obtained by replacing Eq. (58) with Eq. (29) and replacing \( \mathbf{A}_k \) with \( \mathbf{A}_h \), which is defined by Eq. (31). Thus, \( \delta^2(\mathbf{A}_h) \) and AIC(\( \mathbf{A}_h \)) defined by Eqs. (59) and (61) will be equivalent to the quantities defined by Eqs. (34) and (32), respectively. Thus, Theorem 1 is proved.

**Appendix C. Proof of Lemma 2**

From the fact that \( \mathbf{V} \) is an orthogonal matrix, Eq. (38) can be rewritten as follows:

\[
\begin{bmatrix}
\beta_1 \\
\beta_2
\end{bmatrix} =
\begin{bmatrix}
\mathbf{V}_{11}^t & \mathbf{V}_{12}^t \\
\mathbf{V}_{21}^t & \mathbf{V}_{22}^t
\end{bmatrix}
\begin{bmatrix}
\alpha_1 \\
\alpha_2
\end{bmatrix}. 
\]

From Eq. (62), we have

\[
\beta_2 = \mathbf{V}_{12}^t \alpha_1 + \mathbf{V}_{22}^t \alpha_2. 
\]

Further, the equation \( \alpha_2 = -\mathbf{V}_{22}^t \mathbf{V}_{12}^t \alpha_1 \) is obtained by applying Eq. (44) to Eq. (63). This equation always holds, and therefore for the maximum likelihood estimates of \( \alpha_1 \) and \( \alpha_2 \) we have \( \alpha_2 = -\mathbf{V}_{22}^t \mathbf{V}_{12}^t \alpha_1 \). If we express the variance-covariance matrix of \( \alpha_2 \), and the covariance matrix of \( \alpha_1 \) and \( \alpha_2 \) respectively by

\[
\text{Var}\{\alpha_2\} = \mathbb{E}\{(\alpha_2 - \alpha_2)(\alpha_2 - \alpha_2)^t\}, \\
\text{Cov}\{\alpha_1, \alpha_2\} = \mathbb{E}\{(\alpha_1 - \alpha_1)(\alpha_2 - \alpha_2)^t\},
\]

then we have

\[
\text{Var}\{\alpha_2\} = -\mathbf{V}_{22}^t \mathbf{V}_{12}^t \text{Cov}\{\alpha_1, \alpha_2\}. 
\]

It can be seen from Lemma 1 that \( \text{Cov}\{\alpha_1, \alpha_2\} = \mathbf{O} \) is an inevitable result. Therefore, from Eq. (64) the following relation is obtained:

\[
\text{Var}\{\alpha_2\} = \mathbf{O}. 
\]

Eq. (65) implies that \( \alpha_2 = \alpha_2 \) is always a constant for any sample data, and so it is necessary to constrain \( \alpha_2 \) by a constant. As mentioned before, the minimum variance prediction can be obtained by using Eq. (46). Hence, Lemma 2 is proved.

**Appendix D. Proof of Theorem 2**

From Appendix B and Section 3.5, we can see that the likelihood for the model defined by Eqs. (45) and (7) is

\[
L(\beta_1, \sigma^2) = (2\pi \sigma^2)^{-\frac{3}{2}} \exp\left\{-\frac{1}{2\sigma^2}(w_i - T_{11}\beta_1)^2 \times (w_i - T_{11}\beta_1) + w_i^2 w_i + e_i^2 e_i\right\},
\]

and the corresponding log-likelihood is

\[
\ell(\beta_1, \sigma^2) = -\frac{n}{2} \ln\{2\pi \sigma^2\} - \frac{1}{2\sigma^2}(w_i - T_{11}\beta_1)^2 \times (w_i - T_{11}\beta_1) + w_i^2 w_i + e_i^2 e_i.
\]

For a given value of \( \sigma^2 \), the maximum likelihood estimate of \( \beta_1 \) is given by \( w_i - T_{11}\beta_1 = 0_{m-1} \). Further, the maximum likelihood estimate of \( \sigma^2 \) is obtained as follows:

\[
\tilde{\sigma}^2(k) = \frac{1}{n}(w_i^2 w_i + e_i^2 e_i).
\]

On the other hand, the likelihood for the model is defined by Eqs. (47) and (7) as

\[
L(\alpha_1, \sigma^2) = (2\pi \sigma^2)^{-\frac{3}{2}} \exp\left\{-\frac{1}{2\sigma^2}(z_i - D_1 \alpha_1)^2 \times (z_i - D_1 \alpha_1) + z_i^2 z_i + e_i^2 e_i\right\},
\]

and the log-likelihood is given by

\[
\ell(\alpha_1, \sigma^2) = -\frac{n}{2} \ln\{2\pi \sigma^2\} - \frac{1}{2\sigma^2}(z_i - D_1 \alpha_1)^2 \times (z_i - D_1 \alpha_1) + z_i^2 z_i + e_i^2 e_i.
\]

Similarly, the maximum likelihood estimate of \( \alpha_1 \) is given by \( z_i - D_1 \alpha_1 = 0_{m-1} \), and that of \( \sigma^2 \) is given by

\[
\tilde{\sigma}^2(k) = \frac{1}{n}(z_i^2 z_i + e_i^2 e_i).
\]

From Eqs. (66) and (67), we have

\[
\tilde{\sigma}^2(k) - \tilde{\sigma}^2(k) = \frac{1}{n}(w_i^2 w_i - z_i^2 z_i).
\]

Moreover, from Eqs. (68) and (43), we have

\[
\mathbb{E}\{\tilde{\sigma}^2(k) - E(\tilde{\sigma}^2(k)\}
\]

\[
= \beta^2 T_{22} T_{22} \beta - \alpha^2 D_2 \alpha.
\]

From Lemma 2, when Eq. (44) holds Eq. (46) must hold, but when Eq. (46) holds Eq. (44) does not necessarily hold. Thus, when Eq. (46) holds but Eq. (44) does not hold, from Eq. (69) we have \( \mathbb{E}\{\tilde{\sigma}^2(k)\} - E(\tilde{\sigma}^2(k)\) = \( \beta^2 T_{22} T_{22} \beta - \alpha^2 D_2 \alpha > 0 \), which leads to the inequality in Eq. (48). On the other hand, when both of Eqs. (46) and (44) hold, we have \( \mathbb{E}\{\tilde{\sigma}^2(k)\} - E(\tilde{\sigma}^2(k)\) = 0 from Eq. (69), which leads to the equality in Eq. (49). Thus, Theorem 2 is proved.