A study of MT systems applied to multivariate regression

Shota Nakayama*, Suguru Sekine, Yasushi Nagata

Waseda University, 3-4-1 Okubo, Shinjuku-ku Tokyo, 169-8555, Japan

*shota_nakayama@fuji.waseda.jp

Abstract:

Taguchi’s T-method belong to the MT System, which is a representative method in quality engineering. Another relevant technique is multiple single regression (MSR), which incorporates the concept of the resistance line. These methods are used to make predictions when there is one output. In addition, multivariate multiple regression analysis is used as a prediction method when there are multivariate outputs.

This study proposes methods that apply the T-method and MSR to multiple outputs. The proposed methods are: "independent multivariate T-method" (iMvT), "multivariate T-method" (MvT), "simple multivariate T-method" (MvT+), "independent multivariate multiple single regression" (iMMSR), "multivariate multiple single regression" (MMSR), and "simple multivariate multiple single regression" (MMSR+). iMvT and iMMSR combine predictions simply, MvT and MMSR apply generalized least squares, and MvT+ and MMSR+ apply Moore–Penrose’s generalized inverse matrix. We compared the accuracy of the proposed methods with that of multivariate multiple regression analysis. Artificial data analysis was performed by changing the number of training data and correlation. We also performed analyses on the actual data. The results indicate that the proposed methods are more accurate than the existing methods.

Keywords

Mahalanobis-Taguchi System, T-method, Multiple Single Regression, Multivariate Regression

1. Introduction

Multivariate linear models (MLM) are data analysis techniques, wherein multiple outputs for multiple items exist. This method uses multiple regression analysis for each output and has the feature that the relationship between outputs is not considered. We focused on two methods. The first is the T-method proposed by Taguchi (2005). The T-method is used for prediction or estimation in fields, such as economics and weather forecasting. Subsequently, improved methods of the T-method were proposed by Inoh et al. (2012). Another relevant technique is multiple single regression (MSR), which incorporates the concept of the resistance line (Watanabe et al. 1985) into multiple regression analysis. Performance comparisons of several methods, including MSR, were conducted by Maeda (2017).

The purpose of this study is to propose methods that apply both the T-method and MSR to multiple outputs. In addition, we compared our method’s prediction accuracy with that of MLM.

This study is structured as follows: Section 2 describes the conventional methods; Section 3 explains the proposed methods based on the T-method; Section 4 describes the proposed methods based on MSR; Section 5 compares the existing and proposed methods; Section 6 presents the conclusions and proposals for future work.

2. Existing method

2.1 Multivariate Linear Models (MLM)

We explain a multiple regression model with a single output based on Nagata and Munechika (2001). The sample size is \( n \), the number of items is \( p \), and the number of outputs is \( 1 \). The \( n \times (p + 1) \) matrix with \( 1 \) vector in the first column of the items is \( X \), the \( n \times 1 \) matrix of the output is \( y \), the \( \beta \) is a \( (p + 1) \times 1 \) matrix of the
Multivariate regression from MT system, Nakayama et al.

coefficient, and \( \varepsilon \) is the error term of an \( n \times 1 \) matrix. In the multiple regression analysis, the assumed multiple regression model is defined as follows:

\[
y = X\beta + \varepsilon, \tag{2.1}
\]

where \( T \) is the transposition of the matrix. The least-squares estimator of the regression coefficient \( \beta \) and the maximum likelihood estimator of variance \( \sigma^2 \) become:

\[
\hat{\beta} = (X^T X)^{-1} X^T y, \tag{2.2}
\]

\[
\hat{\sigma}^2 = \frac{1}{n} (y - X\hat{\beta})^T (y - X\hat{\beta}). \tag{2.3}
\]

Next, we explain multivariate linear models (MLM) with multiple outputs based on Cassandra (2013), instead of multiple outputs in multiple regression analysis. Let \( n \) be the sample size, \( p \) be the number of items, and \( q \) be the number of outputs. The output \( Y \) is the \( n \times q \) matrix, the regression coefficient \( B \) is the \((p + 1) \times q \) matrix, and the error term \( \varepsilon \) is the \( n \times q \) matrix. The model assumed in MLM is defined as follows:

\[
Y = XB + \varepsilon, \tag{2.4}
\]

The least-squares estimator of \( B \) and the maximum likelihood estimator of the variance \( \Sigma = var(\varepsilon) \) become

\[
\hat{B} = (X^T X)^{-1} X^T Y, \tag{2.5}
\]

\[
\hat{\Sigma} = \frac{1}{n} (Y - XB)^T (Y - XB). \tag{2.6}
\]

The features of the MLM calculation results were described by Hastie et al. (2009). Equations (2.2) and (2.5) show that the least-squares estimator of MLM is the same as in the multiple regression analysis. Therefore, the coefficient for the \( s \)-th output is the same as the least-squares estimator in the regression from \( X \) to \( y_x \). That is, multiple outputs do not affect each other’s least-squares estimator. Evidently, multiple output relations are not considered in the estimation of regression coefficient \( B \) of the MLM.

2.2 Taguchi’s T-method

In the T-method, the unit space is determined based on the data, for which the output value is close to the average of the entire dataset. In accordance with Taguchi (2005) and Tatebayashi et al. (2008), we demonstrate a procedure for analyzing the T-method.

**Step 1: Determination of the unit space**

Let the sample size be \( n \), the number of items be \( p \), and the number of outputs be \( 1 \). The \( n \times p \) item matrix is represented as \( X \) and the \( n \times 1 \) output matrix is represented as \( y \). From these data, we select \( a \) members with values close to the average output value as members of the unit space. We calculate the average value \( \bar{y} \) from the outputs of \( a \) members and the averages \( \bar{x}_1, \bar{x}_2, \ldots, \bar{x}_p \) from each item of the \( a \) members.

The remaining \( l = N - a \) members, not selected for the unit space, are defined as the signal space. The members of this signal space are normalized as follows:

\[
X_{ij} = x_{ij} - \bar{x}_j (i = 1, 2, \ldots, l; j = 1, 2, \ldots, p), \tag{2.7}
\]

\[
Y_{ij} = y_{ij} - \bar{y}_j (i = 1, 2, \ldots, l), \tag{2.8}
\]

**Step 2: Calculation of \( \beta \) and \( \eta \) for each item**

We calculate the proportionality constant \( \beta_j \) and the signal-to-noise (SN) ratio \( \eta_j \) for each item as follows:

Effective divisor 

\[
r = Y^T Y, \tag{2.9}
\]

Total variation 

\[
S_{Tj} = X_j^T X_j, \tag{2.10}
\]

Variation of the proportion term 

\[
S_{\beta j} = (Y_j^T X_j)^2 / r, \tag{2.11}
\]

Error variation 

\[
S_{e j} = S_{Tj} - S_{\beta j}, \tag{2.12}
\]

Error variance 

\[
V_{e j} = S_{e j} / (l - 1), \tag{2.13}
\]

Proportion term 

\[
\beta_j = Y_j^T X_j / r, \tag{2.14}
\]

[DOI : 10.17929/tp.j.7.10] Copyright © 2021 Journal of the Japanese Society for Quality Control. All rights reserved.
SN ratio \( \eta_j = \begin{cases} \frac{1}{r} \frac{(S_{\beta j} - V_{ej})}{V_{ej}} & \text{if } S_{\beta j} > V_{ej}, \\ 0 & \text{otherwise.} \end{cases} \) (2.15)

**Step 3: Calculation of an integrated estimated output value for each signal member**

Each integrated estimated output value \( \hat{y}_i \) becomes

\[
\hat{y}_i = \frac{\eta_{1j} X_{i1} + \eta_{2j} X_{i2} + \cdots + \eta_{pj} X_{ip}}{\eta_1 + \eta_2 + \cdots + \eta_p}.
\] (2.16)

**Step 4: Calculation of the previously normalized overall estimated output value**

After normalizing the output value for the signal data, the previously normalized integrated estimated output value \( \hat{y}_i \) is calculated as follows:

\[
\hat{y}_i = \hat{y}_i + \bar{y}.
\] (2.17)

Using a large amount of data for the unit space leads to a reduction in the sample size of the signal space. Therefore, the sample size in the unit space may be one.

Inoh et al. (2012) proposed two improved methods, namely the Ta-method and Tb-method. In these methods, the concept of a unit space is not required. The normalization in the Ta-method is performed using the average of each item and the output from the whole dataset. When the average value of each item and output from all the data are represented as \( \bar{x}_j (j = 1, 2, \cdots, p) \), \( \bar{y} \), the normalization of the signal data becomes

\[
X_{ij} = x_{ij} - \bar{x}_j \quad (i = 1, 2, \cdots, n; j = 1, 2, \cdots, p),
\] (2.18)

\[
Y_i = y_i - \bar{y} \quad (i = 1, 2, \cdots, n).
\] (2.19)

### 2.3 Multiple Single Regression (MSR)

Maeda (2017) assessed the accuracy of the MSR method. MSR uses the concept of fitting a resistance straight line (Watanabe et al. 1985). In multiple regression analysis, if multicollinearity is strong, the solution cannot be obtained, or the estimation is unstable and the accuracy is poor. The T-method has a problem, in that its prediction accuracy is lower than that of the multiple regression analysis when the collinearities between items are weak; MSR overcomes this shortcoming.

MSR predicts \( \hat{y} = X\hat{\beta} \), as in the multiple regression analysis without error terms. However, unlike multiple regression analysis, because there is no intercept, \( X \) is an \( n \times p \) matrix, and \( \beta \) is a \( p \times 1 \) matrix. A simple regression is then performed on the residuals, as in fitting a resistance line. The proportionality constant \( r_j \) is updated by weighting it with the SN ratio and adding it to the regression coefficient \( \beta \), as in the T-method. Thus, it is a method aimed at performance improvement by combining the resistance straight line, multiple regression analysis, and the T-method. The MSR calculation procedure is described below, in accordance with Maeda (2017).

First, data centering is performed on \( x_{ij} \) and \( y_i \). Data centering is done by subtracting the average, which is similar to equations (2.18) and (2.19). The initial value of the regression coefficient is defined as follows:

\[
\hat{\beta} = 0.
\] (2.20)

Currently, the predicted amounts of output and the residual are given by

\[
\hat{y} = X\hat{\beta},
\] (2.21)

\[
z = y - \hat{y},
\] (2.22)

where \( z \in \mathbb{R}^n \) is the residual vector. As described later, stopping rules are set in the MSR. When the stopping rule is satisfied, the repetition calculation of the MSR is stopped. If the stopping rule is not satisfied, the calculation continues.

In MSR, a single regression is performed on the items and residuals. We calculate the proportionality constant \( r_j \) of the residual \( z \), the centralized item \( X_{ij} \), and the SN ratio \( \eta_j \) as the reliability of the regression. Before estimating the proportionality constant \( r_j \) and SN ratio \( \eta_j \), we solve the following equations:

\[
S_T = z'z,
\] (2.23)

\[
S_{\beta j} = \frac{(z'X_{ij})^2}{X_{ij}'X_{ij}}
\] (2.24)

We calculate the proportionality constant \( r_j \) using equations (2.18) and (2.22) and the SN ratio \( \eta_j \) using equations (2.23) and (2.24). The calculations are as follows:
Multivariate regression from MT system, Nakayama et al.

\[ y_j = (X_j^T X_j)^{-1} X_j^T z, \]  
\[ \eta_j = \frac{S_{\beta_j}}{S_T - S_{\beta_j}}. \]

(2.25)

(2.26)

Finally, we calculate \( \hat{\beta} \) using \( \alpha \). The element of \( \alpha \) is a value obtained by weighting the proportionality constant \( \gamma_j \) of the residual by the SN ratio, which represents the reliability of the item. It follows that

\[ \alpha_j = \frac{\eta_j}{\sum_{m=1}^{p} \eta_m} \gamma_j, \]

\[ \hat{\beta} = \beta + \alpha. \]

(2.27)

(2.28)

**Stopping rule 1: When the residual error gradually decreases**

The stopping rule is executed separately from the above calculation procedure using the jackknife (JK) method to prevent overlearning. Specifically, one piece of data is removed, and the regression coefficient \( \hat{\beta} \) is estimated from the remaining \( n - 1 \) data. Then, the predicted value is calculated for the removed data. Let \( \hat{y}_{i}^{*} \) be the predicted value when the \( i \)-th datum is removed and let \( z_i^{*} \) be the residual. These are given by

\[ \hat{y}_{i}^{*} = (\hat{\beta}_{i}^{*}, \hat{\beta}_{2}^{*}, \ldots, \hat{\beta}_{n}^{*})^T, \]

\[ z_i^{*} = (z_i^{*}, z_{i2}^{*}, \ldots, z_{in}^{*})^T. \]

(2.29)

(2.30)

The residual when the number of repetitions is \( N \) is \( z_i^{(N)} \). The stopping rule is defined as follows:

\[ z_i^{(N)} - z_i^{(N+1)} \leq \delta z_i^{(N)T} z_i^{(N)}, \]

(2.31)

where \( \delta \) is a small value. The repetition stopped when equation (2.31) is satisfied.

**Stopping rule 2: Other conditions**

When executing under stopping rule 1, when \( S_T \) in equation (2.23) becomes 0, and when \( S_{\beta j} = 0 \) in all items in equation (2.24), \( \alpha \) is set to 0 in this iteration and subsequent iterations. If these situations occur, not at the stopping rule, but at the time of repetition using all data, the repetition ends.

**3. Proposed methods based on T-method**

**3.1 Independent Multivariate the T-method (iMvT)**

The independent multivariate T-method (iMvT) is an extension of the T-method with multiple outputs. Essentially, iMvT is equivalent to the combination of prediction results obtained by the conventional T-method. As a result, we obtain each integrated estimated output value \( \hat{y}_{is} \), similar to equation (2.16), and this becomes

\[ \hat{y}_{is} = \frac{X_{i1s} \beta_{1s} + X_{i2s} \beta_{2s} + \ldots + X_{ip} \beta_{ps}}{H_{1s} + H_{2s} + \ldots + H_{ps}}, \]

(3.1)

where \( \mathbf{X} \in \mathbb{R}^{n \times p} \) is the item matrix, \( \mathbf{B} \in \mathbb{R}^{p \times q} \) is the proportionality constant, \( \mathbf{H} \in \mathbb{R}^{p \times q} \) is the SN ratio, and \( \mathbf{Y} \in \mathbb{R}^{n \times q} \) is the output matrix.

**3.2 Multivariate T-method (MvT)**

The multivariate T-method (MvT) makes predictions while considering the relationship between output and the relationship between items. The calculation procedure for MvT is as follows:

MvT assumes a model similar to the T-method. In MLM, the model assumes a regression from \( \mathbf{X} \) to \( \mathbf{Y} \). In an MvT system, the model assumes a regression from \( \mathbf{Y} \) to \( \mathbf{X} \). The MvT model is as follows:

\[ \mathbf{X} = \mathbf{Y} \beta + \mathbf{E}, \]

(3.2)

where \( \mathbf{X} \in \mathbb{R}^{n \times p} \) is the item matrix, \( \mathbf{B} \in \mathbb{R}^{p \times q} \) is the proportionality constant, \( \mathbf{Y} \in \mathbb{R}^{n \times q} \) is the output matrix, \( \mathbf{E} \in \mathbb{R}^{n \times q} \) is the error matrix, \( \mathbf{e} \in \mathbb{R}^{q} \) is the error vector, \( \mathbf{0} \in \mathbb{R}^{q} \) is the mean vector, and \( \mathbf{\Sigma} \in \mathbb{R}^{q \times q} \) is the covariance matrix. Based on this model, we multiply \( \mathbf{\Sigma}^{-1/2} \) on both sides of equation (3.2):

\[ \mathbf{X} \mathbf{\Sigma}^{-1/2} = \mathbf{Y} \mathbf{\Sigma}^{-1/2} + \mathbf{E}' \mathbf{e}' \sim N(\mathbf{0}, \mathbf{I}), \]

(3.3)

and calculate it in the same way as in the least-squares method. It follows that
\[
X \Sigma^{-1/2} (B \Sigma^{-1/2})^T = Y \Sigma^{-1/2} (B \Sigma^{-1/2})^T, \quad (3.4)
\]
\[
\tilde{Y} B \Sigma^{-1} \tilde{B} = X \Sigma^{-1} \tilde{B}, \quad (3.5)
\]
\[
\tilde{Y} = X \Sigma^{-1} \tilde{B} (B \Sigma^{-1} \tilde{B})^{-1}. \quad (3.6)
\]
Furthermore, part of the right side of equation (3.6) is defined as follows:
\[
A^- = \Sigma^{-1} B^T (B \Sigma^{-1} B)^{-1}. \quad (3.7)
\]
\(\mathbf{A}^-\) corresponds to the regression coefficient, and the estimator is given by
\[
\hat{A}^- = \Sigma^{-1} B^T (B \Sigma^{-1} B)^{-1}. \quad (3.8)
\]
Currently, the estimator of the SN ratio \(\hat{H}\) is given by
\[
\hat{H} = \left(\hat{A}^- \Sigma \hat{A}^-\right)^{-1}. \quad (3.9)
\]
Based on equations (3.6) and (3.9), weighting is performed with the SN ratio, and the total prediction amount is calculated as follows:
\[
\hat{Y}' = \frac{\hat{Y} \hat{H}}{\text{tr}(\hat{H})}. \quad (3.10)
\]

3.3 Simple Multivariate T-method (MvT+)

The simple multivariate T-method (MvT+) differs from MvT in that it does not consider the relationship of items and rather the relationship of outputs. The calculation procedure is as follows:

Unlike MvT, MvT+ estimates without multiplying \(\Sigma^{-1/2}\) on both sides of equation (3.2). We have
\[
\hat{Y} = X B^T (B B^T)^{-1}, \quad (3.11)
\]
part of the right side of the above equation is defined as follows:
\[
B^- = B^T (B B^T)^{-1}, \quad (3.12)
\]
where \(B^- \in \mathbb{R}^{p \times q}\) corresponds to the regression coefficient and the estimator is given by
\[
\hat{B}^- = B^T (B \hat{B}^T)^{-1}. \quad (3.13)
\]
Currently, the estimator of the SN ratio \(\hat{H}\) becomes
\[
\hat{H} = \left(B^- \Sigma B^-\right)^{-1}. \quad (3.14)
\]
Based on equations (3.11) and (3.14), weighting is performed with the SN ratio, and the total prediction amount is calculated as follows:
\[
\hat{Y}' = \frac{\hat{Y} \hat{H}}{\text{tr}(\hat{H})}. \quad (3.15)
\]

3.4 Improved Multivariate T-method System

We propose improved methods for the MvT system, similar to the Ta-method of Inoh et al. (2012). We call these the multivariate Ta-method (MvTa) system. Specifically, we propose the independent multivariate Ta-method (iMvTa), MvTa, and the simple multivariate Ta-method (MvTa'). When the average values of each item and output in the data are \(\bar{x}_i, \bar{y}_s\), the signal data are normalized as follows:
\[
X_{ij} = x_{ij} - \bar{x}_j \quad (i = 1, 2, \ldots, n; \ j = 1, 2, \ldots, p), \quad (3.16)
\]
\[
Y_{is} = y_{is} - \bar{y}_s \quad (i = 1, 2, \ldots, n; \ s = 1, 2, \ldots, q). \quad (3.17)
\]

4. Proposed methods based on MSR

4.1 Independent Multivariate Multiple Single Regression (iMMSR)

Independent multivariate multiple single regression (iMMSR) applies MSR to multivariate data. This new method is similar to iMvT in the MvT system. The calculation procedure is as follows: First, when \(X \in \mathbb{R}^{n \times p}\) is the item matrix and \(Y \in \mathbb{R}^{n \times q}\) is the output matrix, data centering is performed on \(X_{ij}\) and \(Y_{is}\). \(X_{ij}\) and \(Y_{is}\) are
defined as follows:

\[
X_{ij} = x_{ij} - \bar{x}_j \quad (i = 1, 2, \cdots, n; j = 1, 2, \cdots, p),
\]

\[
Y_{is} = y_{is} - \bar{y}_s \quad (i = 1, 2, \cdots, n; s = 1, 2, \cdots, q).
\]

The initial value of the estimator \( \mathbf{B} \in \mathbb{R}^{p \times q} \) of the regression coefficient is \( \mathbf{0} \), as shown in

\[
\hat{\mathbf{B}} = \mathbf{0}.
\]

The predicted amounts of the outputs and residuals are obtained from the regression coefficient and \( \mathbf{X} \). It follows that

\[
\hat{\mathbf{Y}} = \mathbf{X} \hat{\mathbf{B}},
\]

\[
\mathbf{Z} = \mathbf{Y} - \hat{\mathbf{Y}}.
\]

The stopping rules are set in the iMMSR. The details are provided in Section 3.4. When the stopping rule is satisfied, we stop the iterative calculation of the iMMSR. If the stopping rule is not satisfied, we proceed to the next step.

In iMMSR, a single regression is performed for items and residuals. The regression coefficient \( \Gamma_{js} \) of the residual \( \mathbf{Z} \) and the centered item \( \mathbf{X}_j \), and the SN ratio \( H_{js} \) as the reliability of the regression are calculated. Before estimating the proportionality constant \( \Gamma_{js} \) and the SN ratio \( H_{js} \), we calculate

\[
\begin{align*}
S_{\mathbf{r}_s} &= \mathbf{Z}_s' \mathbf{Z}_s, \\
S_{\mathbf{B}_{js}} &= \frac{(\mathbf{Z}_s' \mathbf{X}_j)^2}{\mathbf{Z}_s' \mathbf{Z}_s}.
\end{align*}
\]

We estimate the proportionality constant \( \Gamma_{js} \) using equations (4.1) and (4.5) and the SN ratio \( H_{js} \) using equations (4.6) and (4.7). \( \Gamma_{js} \) and \( H_{js} \) are given by:

\[
\Gamma_{js} = (\mathbf{X}_j' \mathbf{X}_j)^{-1} \mathbf{X}_j' \mathbf{Z}_s,
\]

\[
H_{js} = \frac{S_{\mathbf{B}_{js}}}{S_{\mathbf{r}_s} - S_{\mathbf{B}_{js}}}. 
\]

Finally, we calculate \( \hat{\mathbf{B}} \) using \( \mathbf{A} \). The \((j, s)\) element \( A_{js} \) of \( \mathbf{A} \) is a value obtained by weighting the proportionality constant \( \Gamma_{js} \) of the residual by the SN ratio, which represents the reliability of the item. These are estimated as follows:

\[
A_{js} = \frac{H_{js}}{\sum_{m=1}^{p} H_{ms}} \Gamma_{js},
\]

\[
\hat{\mathbf{B}} = \mathbf{B} + \mathbf{A}.
\]

After calculating equation (4.11), we return to equation (4.4) and execute the calculation. In this way, iMMSR repeatedly calculates until the stopping rule is satisfied.

**4.2 Multivariate Multiple Single Regression (MMSR)**

Multivariate multiple single regression (MMSR) is a regression method that considers the relationship between residuals and the relationships between items. In addition, because there is a problem, wherein the inverse matrix cannot be calculated, regularization is performed when calculating the inverse matrix. The calculation procedure is as follows: Equations (4.1) to (4.7) in the iMMSR are omitted because they are the same as in the iMMSR.

Similar to iMMSR, MMSR performs regression on items and residuals, but the calculation procedure is similar to that of MvT. We calculate the regression coefficient of the residual \( \mathbf{Z} \), the centered item \( \mathbf{X} \), and the SN ratio \( \mathbf{H} \) as the reliability of the regression. The MMSR model is as follows:

\[
\mathbf{X} = \mathbf{Z}' + \mathbf{E},
\]

where \( \mathbf{X} \in \mathbb{R}^{n \times p} \) is the item matrix, \( \mathbf{Z} \in \mathbb{R}^{n \times q} \) is the residual matrix, \( \mathbf{I} \in \mathbb{R}^{q \times p} \) is the proportionality constant, \( \mathbf{E} \in \mathbb{R}^{n \times p} \) is the error matrix, \( \mathbf{e} \in \mathbb{R}^{p} \) is the error vector, \( \mathbf{0} \in \mathbb{R}^{p} \) is the mean vector, and \( \mathbf{\Sigma} \in \mathbb{R}^{p \times p} \) is the covariance matrix. Based on this model, MMSR calculates the regression coefficients. The regression coefficient is calculated by applying the generalized least squares (GLS) method. More details on GLS can be found in Amemiya (1985). Based on this model, we multiply \( \mathbf{\Sigma}^{-1/2} \) on both sides of equation (4.12):

\[
\mathbf{X} \mathbf{\Sigma}^{-1/2} = \mathbf{Z} \mathbf{\Sigma}^{-1/2} + \mathbf{E}',
\]

\[\text{DOI: 10.17929 tsp.7.10} \]
By multiplying both sides of the equation by \( \Sigma^{-1/2} \), the error \( \varepsilon_i \) in equation (4.12) becomes \( \tilde{\varepsilon}_i \), and the variance-covariance matrix becomes the identity matrix. Thereafter, we put it in the same way as in the least-squares method. It follows that

\[
X^\top \Sigma^{-\frac{1}{2}} \tilde{\varepsilon} = 2 \tilde{\Omega} \Sigma^{-\frac{1}{2}} \tilde{\varepsilon},
\]

However, according to equation (4.15), \( \tilde{\varepsilon}^\top \tilde{\Omega} \tilde{\varepsilon} \) is not a regular matrix, so the inverse matrix cannot be calculated. Thus, as in ridge regression, we add a regularization term to \( \tilde{\varepsilon}^\top \tilde{\Omega} \tilde{\varepsilon} \) as follows:

\[
\tilde{Z} = X \tilde{\varepsilon}^{-1} \tilde{\varepsilon} \left( \tilde{\varepsilon}^{-1} \tilde{\varepsilon}^\top + \lambda I \right)^{-1},
\]

where \( \lambda \) is a regularization parameter, and it is necessary to search for the optimal value. Then, the part corresponding to the regression coefficient on the right-hand side is defined as follows:

\[
\tilde{\Phi} = \tilde{\Sigma}^{-1} \tilde{\varepsilon} \left( \tilde{\varepsilon}^{-1} \tilde{\varepsilon}^\top + \lambda I \right)^{-1}.
\]

The SN ratio is then calculated using the obtained regression coefficients. As the SN ratio is an index indicating the robustness of the output, the SN ratio \( H \) is given by:

\[
H = V(Z)^{-1}.
\]

However, because calculating the inverse matrix of \( V(Z) = \Phi^\top \Sigma \Phi \) is not possible, regularization must be performed. Regularization is performed by adding an identity matrix. It follows that

\[
H = (\Phi^\top \Sigma \Phi + I)^{-1},
\]

and the estimator of the SN ratio \( \tilde{H} \) is given by

\[
\tilde{H} = (\Phi^\top \Sigma \Phi + I)^{-1}.
\]

Finally, as in the iMMSR, we calculate \( \tilde{B} \) using \( A \). These are estimated as follows:

\[
A = \frac{\Phi \tilde{H}}{\text{tr}(H)},
\]

\[
B = \tilde{B} + A.
\]

After calculating equation (4.22), we execute the return calculation in equation (4.4). The MMSR also performs repeated calculations until the stopping rule is satisfied.

### 4.3 Simple Multivariate Multiple Single Regression (MMSR+)

Simple multivariate multiple single regression (MMSR+) considers only the relationship between residuals. Unlike MMSR, it does not convert the variance-covariance matrix of errors to the identity matrix. The procedure is shown below, but equations (4.1) to (4.8) in the iMMSR are omitted because they are the same as in the iMMSR.

MMSR+ assumes the model of equation (4.12), as well as MMSR. MMSR+ estimates the regression coefficients based on this model. In MMSR, \( \Sigma^{-1/2} \) is multiplied on both sides, and the variance-covariance matrix of the error is used as the unit matrix, but MMSR+ does not perform the operation. Thereafter, equation (4.23) is calculated using the least-squares method. It follows that

\[
X \tilde{\Phi}^\top = 2 \tilde{\Omega} \tilde{\Phi}^\top,
\]

However, \( \tilde{\Phi}^\top \) cannot be calculated because the matrix is not regular. Therefore, the regularization term is added, and the calculation is performed as follows:

\[
\tilde{Z} = X \tilde{\varepsilon} \tilde{\Phi}^\top \left( \tilde{\Phi}^\top \tilde{\Phi}^\top + \lambda I \right)^{-1},
\]

where \( \lambda \) is a parameter, and it is necessary to search for the optimum value, like MMSR. The part corresponding to the regression coefficient on the right-hand side is given by:

\[
\tilde{\Gamma}^{-1} = \tilde{\Phi}^\top \left( \tilde{\Phi}^\top \tilde{\Phi}^\top + \lambda I \right)^{-1},
\]

where \( \tilde{\Gamma}^{-1} \) is a regularization of the Moore-Penrose type pseudoinverse of \( \tilde{\Phi} \). More details on the Moore-Penrose type pseudoinverse can be found in Kanaya (2005). Next, the SN ratio \( \tilde{H} \) is calculated using the regression coefficients obtained. As the SN ratio is an index that indicates the robustness of the output, \( H \) is defined as follows:

\[
H = V(Z)^{-1}.
\]

However, because it is not possible to calculate the inverse matrix of \( V(Z) = \Gamma^{-\top} \Sigma \Gamma^{-1} \), regularization must be performed. Regularization is performed by adding an identity matrix. It follows that
Currently, the estimator of the SN ratio $\hat{H}$ is given by

$$H = \left( (\Gamma^{-}\Sigma^{-} + I)^{-} \right).$$

(4.27)

4.4 Stopping rules

Stopping rules are required for iMMSR, MMSR, and MMSR*, as shown in Sections 4.1, 4.2, and 4.3. Here, three stopping rules are considered: when the residuals converge, when the total variation and the regression coefficient variation are zero, and when the number of repetitions exceeds a certain number.

Stopping rule 1: When the update of the residual converges

The stopping rule performs calculations that are not directly related to the prediction to prevent overlearning. Specifically, one data point is removed and the estimated regression coefficient $\hat{B}$ is calculated from the remaining $n - 1$ data. The predicted amount is calculated for the removed data. Let $\hat{Y}_{i}^*$ be the predicted value when the $i$-th datum is removed, and let $Z_{i}^*$ be the residual, with $i = 1, \cdots, n, s = 1, \cdots, q$, respectively. It follows that

$$\hat{Y}^* = \left( \begin{array}{c} \hat{y}_{11}^* \hat{y}_{21}^* \cdots \hat{y}_{n1}^* \\ \vdots \vdots \vdots \\ \hat{y}_{1q}^* \hat{y}_{2q}^* \cdots \hat{y}_{nq}^* \end{array} \right)^T, \quad Z^* = \left( \begin{array}{c} Z_{11}^* Z_{21}^* \cdots Z_{n1}^* \\ Z_{12}^* Z_{22}^* \cdots Z_{n2}^* \\ \vdots \vdots \vdots \\ Z_{1q}^* Z_{2q}^* \cdots Z_{nq}^* \end{array} \right)^T.$$

(4.31)

Then, let $Z_i^{(N)}$ be the residual when the number of repetitions is $N$. The small value is set to $\delta$, and the repetition is stopped when:

$$|\text{tr}(Z^T(N)Z(N)) - \text{tr}(Z^{(N+1)}T(Z^{(N+1)})| \leq \delta \cdot \text{tr}(Z^{(N)}TZ^{(N)}),$$

(4.32)

is satisfied. This is considered when the repetition of the residuals converges.

Stopping rule 2: When all fluctuations or regression coefficient fluctuations are 0

When executing in JK, when the $S_{ij}$ in equation (4.6) become zero, and when $S_{ij} = 0$ in all items in equation (4.7), the complement set is $A = 0$ and 0 in subsequent iterations. If these situations occur during repetition using all the data, the repetition is terminated. In this study, the program is approximated by $10^{-5}$.

Stopping rule 3: When the number of repetitions exceeds a certain number

The MSR analytically proves that the residual monotonically decreases (Maeda, 2017). However, there is no guarantee that the proposed methods monotonically decrease. Therefore, the calculations may not converge. As such, the calculation is stopped when the number of repetitions was increased.

5. Performance Comparison of Each Method

5.1 Evaluation Index

Comprehensive prediction errors (CPE) were used as a new evaluation index for the simulation in this study. The index that applies prediction errors (PE) to data with multiple output is the CPE. First, PE was explained along with Inoh et al. (2012), and then a new index CPE was defined. The PE for the $s$-th output is defined as follows:

$$\text{PE}_s = E[(\hat{y}'_s - y'_s)^2].$$

(5.1)

This evaluates the goodness of fit of the test data. $y'$ is the value of the test data and $\hat{y}'$ is the predicted amount of the test data. The expected value of the square of the prediction error is calculated by taking the square of the difference between $y'$ and $\hat{y}'$ and taking the average of the number of simulations. The smaller the value of PE, the smaller the error, thus, the smaller the value, the better the prediction accuracy.

In contrast, in this study, evaluation with CPE was necessary to measure the goodness of fit for multivariate...
data. This aims to make the output dimensionless. Therefore, the PEₘ of the s-th output are divided by the variance of the s-th output, and the CPE is calculated by adding them. Thus, the CPE is defined as follows:

\[
CPE = E \left[ \frac{\sum_{s=1}^{q} PEₘ_s}{\sigma_s^2} \right].
\]

(5.2)

5.2 Performance Comparison 1 (Linear Multiple Regression Model)

In Performance Comparison 1, a simulation using a linear multiple regression model was performed, and the evaluation was performed using data in which items were generated independently and data in which correlation was generated between items. There are two outputs and five items. Five types of training data (5, 10, 20, 30, and 100) were prepared and examined. The error variance was set to a large value; this was done to prevent the difference from disappearing when the number of training data increases. The number of test data was 100, and the number of simulations was 1,000. The comparison methods are: MLM, iMvT, MvT, MvT+, iMvTa, MvTa, MvTa+, iMMSR, MMSR, and MMSR+. The stopping rule \( \delta \) of iMMSR, MMSR, and MMSR+ and the regularization parameter \( \lambda \) of MMSR and MMSR+ were optimized by grid search (Raschka and Mirjalili, 2018).

The models in the linear multiple regression were given by

\[
Y = XB + E,
\]

(5.3)

\[
X = (x_1, x_2, \cdots, x_n)^T, \quad x_i \sim N(\mu_X, \Sigma_X), \quad \mu_X = (1,1,1,1,1)^T, \quad \Sigma_X = \begin{pmatrix} 1 & \rho_X^2 & \rho_X^2 & \rho_X^2 & \rho_X^2 \\ \rho_X^2 & 1 & \rho_X^2 & \rho_X^2 & \rho_X^2 \\ \rho_X^2 & \rho_X^2 & 1 & \rho_X & \rho_X \\ \rho_X^2 & \rho_X^2 & \rho_X & 1 & \rho_X \\ \rho_X^2 & \rho_X^2 & \rho_X & \rho_X & 1 \end{pmatrix}, \quad \rho_X \in \{0, 0.7\},
\]

(5.4)

\[
B = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \end{pmatrix}^T,
\]

(5.5)

\[
E = (e_1, e_2, \cdots, e_n)^T, \quad e_i \sim N(0, \Sigma), \quad \Sigma \in \{R, 5R\}, \quad R = \begin{pmatrix} 1 & \rho_E^2 \\ \rho_E & 1 \end{pmatrix}, \quad \rho_E \in \{0, 0.7\}.
\]

(5.6)

This model is called “Model 1.” Thus, when \( \rho_X \) and \( \rho_E \) are both 0, there is no correlation, and when both are 0.7, there are correlated data. In addition, \( \Sigma \) determines the error size. Table 5.1 shows the simulation results for uncorrelated data, and Table 5.2 shows those for correlated data. When the number of training data is small, calculation cannot be performed and calculation of the predicted amount may not be performed. In the following tables, a CPE with a “*” indicates that the calculation could not be performed.

According to Table 5.1, the linear multiple regression data without correlation show that the accuracy of iMMSR is good when the number of training data is small, but the accuracy of MLM is best when the number of training data is large. When the number of training data is small, the features of iMMSR that repeat a single regression are applied well. The prediction accuracy of MLM is good when the amount of learning data is large, and the estimation of MLM improves as the number of learning data increases as the data are generated by the linear multiple regression model.

| Method   | Number of training data (\( \Sigma = R \)) | Number of training data (\( \Sigma = 5R \)) |
|----------|------------------------------------------|------------------------------------------|
|          | 5  | 10 | 20 | 30 | 100 | 5  | 10 | 20 | 30 | 100 |
| MLM      | -  | 1.193 | 0.537 | 0.451 | 0.368 | -  | 3.399 | 1.593 | 1.344 | 1.099 |
| iMvT     | 7.814 | 3.606 | 2.750 | 2.330 | 1.650 | 8.289 | 5.623 | 5.498 | 5.025 | 4.102 |
| MvT      | -  | 1.673 | 0.884 | 0.723 | 0.370 | -  | 3.338 | 1.968 | 1.698 | 1.451 |
| MvT+     | 5.039 | 2.122 | 1.286 | 1.017 | 0.656 | 5.116 | 2.817 | 2.071 | 1.830 | 1.513 |
| iMvTa    | 6.605 | 2.560 | 1.498 | 1.115 | 0.569 | 8.118 | 4.771 | 3.859 | 3.310 | 2.366 |
| MvTa     | -  | 1.494 | 0.697 | 0.536 | 0.394 | -  | 3.054 | 1.775 | 1.518 | 1.323 |
| MvTa+    | 4.826 | 1.939 | 1.072 | 0.775 | 0.448 | 4.765 | 2.497 | 1.821 | 1.571 | 1.344 |
| iMMSR    | 4.724 | 1.114 | 0.538 | 0.451 | 0.368 | 4.567 | 2.381 | 1.573 | 1.341 | 1.100 |
| MMSR     | -  | 1.174 | 0.537 | 0.452 | 0.368 | -  | 2.544 | 1.569 | 1.334 | 1.100 |
| MMSR+    | 5.085 | 1.179 | 0.540 | 0.453 | 0.371 | 4.585 | 2.384 | 1.567 | 1.344 | 1.101 |
5.3 Performance Comparison 2 (Model Behind the T-method)

In Performance Comparison 2, a simulation was performed using a model behind the T-method, and evaluation was performed using data with independently generated outputs and data with correlations generated between outputs. Conditions, such as simulation settings and the comparison methods, were similar to those in Performance Comparison 1.

Equations (5.7) to (5.10) show a model behind the T-method. It follows that

\[ \mathbf{X} = \mathbf{YB} + \mathbf{E}, \]  

\[ \mathbf{Y} = (\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_n)^T, \mathbf{y}_i \sim N(\mu_y, \Sigma_y), \mu_y = (1,1)^T, \Sigma_y = \begin{pmatrix} 1 & \rho_y \\ \rho_y & 1 \end{pmatrix}, \rho_y \in \{0, 0.7\}, \]  

\[ \mathbf{B} = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \end{pmatrix}, \]  

\[ \mathbf{E} = (\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n)^T, \mathbf{e}_i \sim N(\mathbf{0}, \Sigma), \Sigma \in \{\mathbf{R}, 5\mathbf{R}\}, \mathbf{R} = \begin{pmatrix} 1 & \rho_E & \rho_E^2 & \rho_E^3 \\ \rho_E & 1 & \rho_E & \rho_E^2 \\ \rho_E^2 & \rho_E & 1 & \rho_E^2 \\ \rho_E^3 & \rho_E^2 & \rho_E & 1 \end{pmatrix}, \rho_E \in \{0, 0.7\}. \]  

This model is called “Model 2.” This means that when \( \rho_y \) and \( \rho_E \) are both 0, there is no correlation, and when both are 0.7, correlated data exist. Table 5.3 shows the simulation results for uncorrelated data and Table 5.4 shows the simulation results for correlated data.

Tables 5.3 and 5.4 show that the accuracies of MMSR* are the best in many cases. MMSR* not only considers the relationship between the residuals, but also performs calculations, such as the T-method, in the calculation. However, when both the number of training data and the size of the error are small, MtTa is the best method.
5.4 Performance Comparison 3 (Actual Data)

In Performance Comparisons 1 and 2, simulations were performed on artificially generated data. In Performance Comparison 3, we prepared two sets of real data, analyzed them using the existing methods and the proposed methods, and compared the performance.

Protein consumption data were obtained from DoSS@d(a) and results from protein consumption data in Europe with a sample size of 25 consisting of red meat, white meat, eggs, milk, fish, cereals, starchy foods, beans, fruits, and vegetables. We predicted the consumption of red and white meat from the consumption of non-meat proteins with a sample size of 25 consisting of red meat, white meat, eggs, milk, fish, cereals, starchy foods, beans, fruits, and vegetables. We predicted the consumption of red and white meat from the consumption of non-meat proteins and found a weak correlation. Next, we used DoSS@d(b) for the sports test data, which consisted of a sports test with a sample size of 104, including back strength, grip strength, 50m running, long jump, handball throw, suspension, repetitive side jump, and vertical jump. The back muscle strength and grip strength were predicted from the measurement results for each event. Strong correlations were observed. The number of test data was five for protein consumption data and 50 for sports test data. The comparison methods and the number of simulations were the same as in Section 5.2 Performance Comparison 1. The simulation results of the protein consumption data are shown in Table 5.5 and the simulation results of the sports test data are shown in Table 5.6.

| Method  | Number of training data (\(\Sigma = R\)) | Number of training data (\(\Sigma = 5R\)) |
|---------|----------------------------------------|----------------------------------------|
| MLM     | 12.221                                 | 6.805                                  |
| iMvT    | 5.343                                  | 7.899                                  |
| MvT     | 6.500                                  | 6.625                                  |
| MvT^a   | 3.289                                  | 3.974                                  |
| iMvTa   | 4.546                                  | 5.963                                  |
| MvTa    | 5.975                                  | 6.757                                  |
| MvTa^a  | 2.927                                  | 3.968                                  |
| iMMsR   | 2.530                                  | 2.536                                  |
| MmsR    | 2.973                                  | 2.847                                  |
| MmsR^a  | 2.494                                  | 2.651                                  |

Table 5.5 shows that the accuracy of iMMsR and MmsR^a is high for real data with low correlation. Next, Table 5.6 shows that iMMsR is effective for real data with high correlation. In actual data, the results are not the same as those in Comparisons 1 and 2. This is because of the unique relationship of the data.

6. Conclusions and Future Works

6.1 Conclusions

In Section 5, we compared the proposed method with existing methods. The effective methods obtained by classifying the simulation results are summarized in Table 6.1. As shown in Table 6.1, MvT and MmsR are not effective in many cases. The reason why MvT is ineffective is that the features of MvT conflict with the features of the T-method system. MvT requires a large amount of training data to perform calculations because of the estimates of the variance-covariance matrix. However, it is generally said that the T-method system is effective when there is little training data. MmsR is ineffective because there was little relationship between residuals and items. MmsR considers the relationship between the residuals and the items in the calculation, but we believe that
Table 6.1 Effective methods in our simulations

| Ratio of training data to variables | Low     | High     |
|------------------------------------|---------|----------|
|                                    | small error | big error | small error | big error |
| 1~4                                | MvT+ / iMMSR | iMMSR / MMSR⁺ | iMMSR / MMSR⁺ | iMMSR / MMSR⁺ |
| 4~10                               | MLM / MvT⁺ | MMSR⁺ | MMSR⁺ | MMSR⁺ |
| 10~20                              | MLM / MMSR⁺ | MLM / MMSR⁺ | MLM / iMMSR | MMSR⁺ |

this procedure is not meaningful.

6.2 Future Works

We suggest two future studies. The first should focus on the Ta-method. At present, the average value of each item is used when centering, but it may be improved by centering based on the SN ratio, as in the Tb-method proposed by Inoh et al. (2012). Second, we focus on the stopping rules. The update of the residual may not converge and may oscillate in stopping rule 1 of the proposed methods. Currently, we have a limit on the number of times, such as in stopping rule 2, but there may be a better way.

References

Amemiya, T. (1985): Advanced Econometrics, Harvard University Press
Cassandra, Q. (2013): Multivariate Multiple Regression with Applications to Powerlifting Data, University of Minnesota – Duluth, Department of Mathematics and Statistics
Doss@d, Data(a): Sports test data, http://mo161.soci.ous.ac.jp/@d/DoDStat/sports/sports.csv
Doss@d, Data(b): Protein, http://mo161.soci.ous.ac.jp/@d/DoDStat/protein/protein.csv
Hastie, T., Tibshirani, R. and Friedman, J. (2009): The Elements of Statistical Learning, Springer
Inoh, J., Nagata, Y., Horita, K. and Mori, A. (2012): Prediction accuracies of improved Taguchi’s T methods compared to those of multiple regression analysis, Journal of the Japanese Society for Quality Control, Vol. 42, No. 2, pp. 103-115. (in Japanese)
Kanaya, K. (2005): Optimization Mathematics You Can Understand, Kyoritsu Press. (in Japanese)
Maeda, M. (2017): Proposal of a new regression method using the concept of T-method (1), Journal of the Japanese Society for Quality Control, Vol. 47, No. 2, pp. 71-80. (in Japanese)
Nagata, Y. and Munechika, M. (2001): Introduction to multivariate analysis, Science Company. (in Japanese)
Raschka, S. and Mirjalili, V. (2018): Python Machine Learning - 2nd Edition, Packt Publishing
Taguchi, G. (2005): Objective function and generic function (6) – Prediction by Taguchi methods –, Journal of the Quality Engineering Society, Vol. 13, No. 3, pp. 5-10. (in Japanese)
Tatebayashi, K., Tejima, S. and Hasegawa, R. (2008): Introduction to MT System, Nikkagiren Press. (in Japanese)
Watanabe, H., Suzuki, N., Yamada, F. and Ohtsuka, Y. (1985): Introduction to Exploratory Data Analysis, Asakura Press. (in Japanese)

Acknowledgements:

We would like to thank the anonymous referees for their valuable comments. This work was partly supported by JSPS Grants-in-Aid for Scientific Research Grant Number 18K11202.

Authors’ biographical notes:

Shota Nakayama is a graduate student in the Department of Industrial and Management Systems Engineering, the Graduate School of Creative Science and Engineering at Waseda University.

Suguru Sekine is a graduate student in the Department of Industrial and Management Systems Engineering, the Graduate School of Creative Science and Engineering at Waseda University.
Yasushi Nagata is a professor in the Department of Industrial and Management Systems Engineering, Creative Science and Engineering at Waseda University.

[DOI: 10.17929/tqs.7.10]

Received, April 30, 2020
Revised: April 10, 2021
Accepted: June 7, 2021