More Accurate Knowledge Search in Technological Development for Robust Parameter Design

DOI: 10.12776/QIP.V26I1.1639

Kosuke Oyama, Masato Ohkubo, Yasushi Nagata

Received: 2022-01-05 Accepted: 2021-03-12 Published: 2022-03-31

ABSTRACT

Purpose: The causality search Taguchi (CS-T) method was proposed to support system selection in a robust parameter design. However, the target of the analysis is likely to be quasi-experimental data. This can be difficult to analyse with the CS-T method. Therefore, this study proposes a new analysis approach that can perform a more accurate knowledge search by applying the instrumental variable.

Methodology/Approach: Using the CS-T method, appropriate knowledge search is difficult with quasi-experimental data, including endogeneity. We examined an analytical process that addresses the endogeneity between mechanism and output by utilizing the control and noise factors that constitute the mechanism as instrumental variables.

Findings: The results show that 1) the proposed method has sufficient practical accuracy, even for quasi-experimental data including endogeneity; and 2) the extracted mechanism is less likely to fluctuate depending on the number of experimental conditions used. Moreover, we can clarify the position of the CS-T and proposed methods in system selection.

Research Limitation/Implication: We perform estimation under the assumption that the threshold is known. However, the extracted mechanism may change depending on the threshold; this requires discussing how to determine them.

Originality/Value of paper: Technological development requires a high degree of engineer sophistication. However, this study's analytical process allows conducting more accurate knowledge search in a realistic and systematic way without requiring a high level of engineer input.

Category: Research paper

Keywords: robust parameter design; Taguchi method; technological development; instrumental variable method; statistical modelling

1 INTRODUCTION

Generally, the developmental activities of quality engineering (Taguchi method) are divided into two areas: technological development and product design. In the Taguchi method, system selection is performed during technological development, whereas in product design, robust parameter design (RPD) and tolerance design are performed. The goal is to achieve both quality improvement and cost reduction. The RPD is a system design method based on experimental design, and it decreases the effects of random and systematic errors that can change the function of the system.

However, it is difficult to achieve quality improvement in RPD in product design if the system selection (technological development) in the front-end process is inadequate. Kawada (2013) reports that the same issue applies to its use in research and development. In many technological developmental processes, the best existing system is selected. Thereafter, the system is developed by improving the output/objective characteristics, such as variance and signal-to-noise ratio (SNR), and obtaining ideas for new control factors and generic functions from the mechanisms that achieve the target values. In other words, knowledge search is very important to understand the relationship between mechanisms and output/objective characteristics.

Hosokawa (2020) cites one-factor experiments and the RPD of the objective function as the knowledge search approaches used in this process.

One-factor experiments can focus on a single mechanism and facilitate its comprehensive analysis. Conversely, besides being inefficient, these experiments reduce the ability to detect mechanisms for improvement because the range of change in the output/objective characteristics and mechanisms is small.

The RPD of the objective function can be performed efficiently to a certain extent, notwithstanding that the range of change in the output/objective characteristics expands by moving many factors. Conversely, despite that the mechanism of improvement is unknown, the reliability of the graph of factorial effects may be compromised by the effects of interactions. According to Hosokawa and Miyagi (2019), only a limited amount of studies has been effectively used, specifically Mashhadi et al. (2016), Gamage, Jayamaha and Grigg (2017), and Göhler, Ebro and Howard (2018).

Thus, there is a growing need for technological development of an efficient and accurate method to understand the mechanism of improvement. Therefore, Hosokawa et al. (2015) proposed the causality search - Taguchi (CS-T) method to solve this problem. This method increases the amplitude of the mechanism and output/objective characteristics by moving the factors assigned to the orthogonal array and estimates the relationship between them using the Ta-method (Inoh et al., 2012) contribution ratio and the overall estimated SNR. The Ta-method is a modification of the Taguchi (T) method.

Because the analysis is performed while repeatedly adding experimental conditions, the experiment can be censored, which holds promise as a methodology for efficient knowledge search. For specific examples, see Hosokawa and Miyagi (2019).

However, the data handled using the CS-T method can be interpreted as quasi-experimental data in which observational and experimental data are mixed. Therefore, problems, such as sensor failure (measurement error) or a mixture of unexpected factors (omitted variables), may occur, and bias due to endogeneity may arise when a regression problem is assumed. Meanwhile, the CS-T method, which employs the Ta-method, a regression technique, is at risk of not extracting the mechanisms that improve the output/objective characteristics because the contribution ratio and the overall estimated SNR are affected.

In this study, we propose an analysis process that can remove the bias caused by endogeneity, which hinders the extraction of mechanisms that improve output/objective characteristics by incorporating the concept of instrumental variables (IVs). Further, the proposed method enables the censoring of experiments as well as the CS-T method, which is expected to facilitate efficient knowledge search.

The remainder of this paper proceeds as follows: Section 2 explains the CS-T method for a static system; Section 3 proposes a new knowledge search method based on the description of the assumed data structure and concepts; Section 4 compares the methods through simulations to clarify their respective positions in technological development; and Section 5 summarizes the study and describes future developments.

2 CAUSALITY SEARCH TAGUCHI METHOD

As previously mentioned, system selection must be performed in the RPD front-end process in product design. Taguchi (1993) states that “the system selection is which system (generic function as a technical means) is chosen as the generic function.” Here, the generic function refers to the technical means of achieving the objective function. As an example of a generic function, Taguchi (2004) cites the chemical reaction of an engine. In conjunction with the determination of generic functions, it is also necessary to develop control factors that constitute the system. This is to determine the limits of improvement of the selected system and evaluate the system properly by moving the control factors.

Subsequently, we outline the CS-T method based on Hosokawa (2020) as an approach to support system selection, using a static characteristic system as an example. For the target data, it is assumed that the relationship between the control factor and the mechanism is known to some extent and that the control factor affects the output/objective characteristics only through the mechanism. In the CS-T method, the factors that represent mechanisms are referred to as effective explanation factors (EEFs), including sensor data, physical property data, and intermediate data in computer-aided engineering, and they cannot be leveled at the time of the experiment neither do they have a target value. In this study, we interpret the EEFs as an intermediate characteristic and proceed with the discussion.

We use the orthogonal array L_8 with four control factors $\mathbf{x}$ as the inner array and one error factor $\mathbf{z}$ as the outer array, assigned at two levels for discussion. When the index $i = 1,2,\dots,I$ $(I = 8)$ denotes the combination of control factors, $j = 1,2$ denotes the combination of noise factors, and $k = 1,\dots,r$ denotes the types of EEFs, an example of the data used in the CS-T method with the observed EEFs s_{ijk} and output y_{ij} are shown in Table 1.

Table 1 – Example of the Data Format Used in the CS-T Method

No.	x_A	x_B	x_C	x_D	$z = 1$					$z = 2$				
					s_1	s_2	$\cdots$	s_r	y	s_1	s_2	$\cdots$	s_r	y
1	1	1	1	1	s_{111}	s_{112}	$\cdots$	s_{11r}	y_{11}	s_{121}	s_{122}	$\cdots$	s_{12r}	y_{12}
2	1	1	2	2	s_{211}	s_{212}	$\cdots$	s_{21r}	y_{21}	s_{221}	s_{222}	$\cdots$	s_{22r}	y_{22}
3	1	2	1	2										
4	1	2	2	1										
5	2	1	1	2	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
6	2	1	2	1										
7	2	2	1	1										
$I = 8$	2	2	2	2	s_{I11}	s_{I12}	$\cdots$	s_{I1r}	y_{I1}	s_{I21}	s_{I22}	$\cdots$	s_{I2r}	y_{I2}

The method consists of four steps.

In the first step, the data are created. Control factors and noise factors are selected and assigned to the orthogonal array to increase the range of changes in the EEFs and output/objective characteristics. Thereafter, an arbitrary number of experimental conditions is randomly selected from the orthogonal array, and experiments are conducted to measure the output and EEFs. Here, the EEFs and outputs are converted into objective characteristics as necessary.

In the second step, the explanatory rate R^2 of the entire EEF is calculated by conducting the Ta-method with the objective characteristic as the objective variable and the EEF as the explanatory variable. This serve as an indicator of the termination of the experiment. Note, however, that the Ta-method is essentially a method that assumes that the explanatory variables arise from the objective variables.

In the third step, the EEF is assigned to the orthogonal array of the two-level system as the first level if it is included in the estimation equation created in the second step, and the second level if it is not. The overall estimated SNR $\hat{\eta}$ is calculated after assigning the factors to the orthogonal array. Thereafter, $\hat{\eta}$ is used to perform an analysis of variance to derive the contribution ratio $\hat{\rho}_k$ of each EEF.

In the fourth step, the decision to terminate the experiment is made. If R^2 is greater than 0.6, and the $\hat{\rho}_k$ of any EEF shows a similar trend more than five times, the experiment is terminated. If the conditions are not met, one unselected experimental condition is randomly selected, the output and the EEFs are measured anew, and the two steps are returned.

3 PROPOSAL OF A NEW METHOD

3.1 Data Structures

Hosokawa et al. (2015) and Hosokawa (2020) do not perform simulations, but only validate the results using real data analysis. Therefore, in this section, we discuss and define the data structure while considering the assumptions of the RPD.

The CS-T method assumes that the control factors affect the output/objective characteristics only through the EEFs. Moreover, the Ta-method is used in the CS-T method. With the aforementioned in mind, we refer to the model of static systems in Myers, Khuri and Vining (1992). Thus, the output y_{ij} is:

$$y_{ij} = \delta_0 + \mathbf{s}_{ij}^T \boldsymbol{\delta} + \varepsilon_{ij},$$
$$\mathbf{s}_{ij} = \begin{pmatrix} s_{ij1} \\ \vdots \\ s_{ijr} \end{pmatrix} \quad (1)$$

where δ_0 is the intercept parameter, $\mathbf{s}_{ij}$ denotes the EEFs, ε_{ij} denotes the error terms, and the r-dimensional vector $\boldsymbol{\delta}$ denotes the vector of coefficients for the control factors. The kth EEF is:

$$s_{ijk} = \beta_{0k} + \mathbf{x}_i^T \boldsymbol{\beta}_k + \mathbf{z}_j^T \boldsymbol{\gamma}_k + \mathbf{x}_i^T \boldsymbol{\Omega}_k \mathbf{z}_j + \varepsilon_{ijk},$$
$$\mathbf{x}_i = \begin{pmatrix} x_{i1} \\ \vdots \\ x_{ip} \end{pmatrix}, \mathbf{z}_j = \begin{pmatrix} z_{j1} \\ \vdots \\ z_{jq} \end{pmatrix}, \mathbf{z}_j \sim N\left(0, \sigma_{\mathbf{z}}^2 \mathbf{I}_q\right) \quad (2)$$

where β_{0k} is the intercept parameter, $\mathbf{x}_i$ denotes the control factors, $\mathbf{z}_j$ denotes the noise factors, the p-dimensional vector $\boldsymbol{\beta}_k$ denotes the vector of coefficients for the control factors, the q-dimensional vector $\boldsymbol{\gamma}_k$ denotes that for the noise factors, the $p \times q$ matrix $\boldsymbol{\Omega}_k$ denotes the matrix of control-by-noise interaction coefficients and ε_{ijk} denotes the error terms. Each level of the noise factor follows a normal distribution $N(0, \sigma_{\mathbf{z}}^2)$. The index $i = 1,2, \dots, I$ denotes the combination of control factors, and $j = 1,2, \dots, J$ denotes the combination of noise factors.

In this model, the EEFs consist of control and noise factors, and the output consists of EEFs. In other words, the measured output is an aggregate of multiple EEFs, and changes in the output are always the result of changes in some EEFs. Moreover, from the perspective that RPD uses the interaction between control and noise factors, it is more natural for the control and noise factors to have a direct interaction relationship.

Upon applying to the analysis of Hosokawa et al. (2015), it is implied that variations in control factors and paper features affect the distance the paper is folded from the reference point through the paper transport speed. Hence, by choosing a combination of control factors that attenuate the effects of variations in paper features, the paper transport speed and the paper folded distance from the reference point are stabilized.

3.2 Concept of the Proposed Method

Causal inference is a useful method for estimating relationships between variables. Causal inference methods include Gaussian graphical modelling (Lauritzen, 1996), linear non-gaussian acyclic models (Shimizu et al., 2006), and propensity scores (Rosenbaum and Rubin, 1983). However, they are not applicable to this problem.

Therefore, in this study, we apply the IV concept, which is one of the causal inferences. Kuroki, Miyakawa and Kawata (2003) interpret this in terms of causal diagrams as being able to deal with unobserved covariates (omitted variables). It is also tolerant to measurement errors. See Bowden and Turkington (1985) for further details.

Although the CS-T method uses orthogonal arrays, they should be treated as quasi-experimental data. Therefore, it is possible that the missing variables or measurement errors are included in the EEFs. If this happens, endogeneity will occur, which will cause a bias in the regression coefficients when assuming a regression model. This is because as the EEFs change, the error terms and omitted variables also change, thereby making it impossible to determine the effect of the EEFs on the output.

The IV method can be used to solve this problem. Furthermore, because the data covered in this study use orthogonal arrays and assume that the control factors affect the output/objective characteristics only through the EEFs, making the control factors the IVs facilitates the IV selection, which is normally difficult.

Additionally, there are many EEFs that are not involved in the output of the data targeted in this study, and it is necessary to select variables for them. Therefore, in the proposed method, the EEFs are reduced by using stability selection (Meinshausen and Bühlmann, 2010), which is a variable selection method.

3.3 Analysis Process of the Proposed Method

In the following sections, the proposed method is explained in detail by assuming a static system.

The first step entails data creation, similar to the CS-T case. However, the level of each control factor should be extracted such that they are not all the same. The use of a mixed-level orthogonal array is desirable because the combination of rows to be extracted affects the amplitude of each EEF.

The second step involves the selection of EEFs to be used in the estimation of the relationships in the third step. Here, the output/objective characteristics are the objective variables, and all the EEFs are the explanatory variables. Thereafter, we use lasso regression with the regularization parameter λ. However, selection is performed using stability selection.

For stability selection, we first prepare B bootstrap samples. We define $\Psi\left\{k \in \hat{S}^{\lambda}(\Psi_b)\right\}$ as the indicator function in which the coefficient of the EEF of variable number k is 1 if it is non-zero and 0 if it is zero in the b-th bootstrap sample Ψ_b. $\hat{S}^{\lambda}$ denotes the selection result of the lasso regression using λ for Ψ_b. Thereafter the choice probability $\widehat{\Pi}_k^{\lambda}(k = 1,\cdots,p)$ is defined as the sum up to $b = 1,\cdots,B$ divided by B. We also treat the set of EEFs larger than an arbitrary threshold $\pi_{thr}(0 < \pi_{thr} < 1)$ as candidate $\hat{S}^{stable}$, as follows:

$$\hat{S}^{stable} = \left\{k\colon \max_{\lambda \in \mathbf{\Lambda}} \widehat{\Pi}_k^{\lambda} \geq \pi_{thr}\right\} \tag{3}$$

where $\mathbf{\Lambda}$ denotes an arbitrary set of λ.

The third step is the estimation of the relationship between the EEFs and the output/objective characteristics using the conditional IV method (Brito and Pearl, 2002). The use of an IV is meant to exclude bias by examining the effect of the factors assigned to the orthogonal array on the output/objective characteristics of changes in the EEFs.

Here, we consider $\hat{S}^{stable}$ as the treatment variable. Then, the control and noise factors as well as $\hat{S}^{stable}$ which are not the treatment variables, are considered as $n \times l$ matrix Z, $\hat{S}^{stable}$ as $n \times m$ matrix X, and the output/objective characteristics as $n \times 1$ matrix Y. Y corresponds to the outcome variable.

Thereafter, the effect of the treatment variable on the outcome variable (partial regression coefficient) in the conditional IV method can be written as:

$$\hat{\beta}_{IV} = (X^T P_Z X)^{-1} X^T P_Z Y \tag{4}$$

where P_Z is the projection matrix.

From the above-stated, we calculate $\hat{\beta}_{IV}$ for all elements of $\hat{S}^{stable}$.

The conditional IV method is used to deal with the exclusion restriction that the IV affects the outcome variable only through the treatment variable.

In the fourth step, the decision to terminate the experiment is made. When the third step has been performed three or more times, the results up to the time when censoring is considered and the results up to the previous time are considered as two groups. Next, we extract the results of the EEFs that have the strongest relationship with the output/objective characteristics at the times when we consider censoring. The significance level is then set and an F-test is performed. The null hypothesis is that "the variances of the two groups are equal" and the alternative hypothesis is that "the variances of the two groups are not equal". If the null hypothesis can be rejected, one unselected experimental condition is randomly selected, the output/objective characteristics and EEFs are measured anew, and the second step is returned. If the null hypothesis cannot be rejected, the experiment is terminated.

4 VERIFICATION BY SIMULATIONS

4.1 Simulations Outline

In this section, simulations are conducted to compare the methods. Further, the position of the CS-T and the proposed methods in technological development is clarified. Subsequently, the control factor is **x**, the noise factor is **z**, the EEF is **s**, and the output is y.

4.2 Simulations Settings

Verification is performed by analysing the data of the two patterns generated according to the data structure defined in Section 3.

The first pattern is quasi-experimental data, where the EEFs associated with y contain unobserved covariates u_1 and u_2. The model equation is defined as:

$$s_k = \begin{cases} \beta_{Ak}x_A + \beta_{BZk}x_B z + \beta_{Zk}z + \varepsilon_k \ \ (k = 1,2,\cdots,9) \\ 5.0(x_A + x_B z + z + 2.0u_1) + \varepsilon_k \ \ (k = 10) \\ \beta_{Ck}x_C + \beta_{DZk}x_D z + \beta_{Zk}z + \varepsilon_k \ \ (k = 11,12,\cdots,19) \\ 5.0(x_C + x_D z + z + 2.0u_2) + \varepsilon_k \ \ (k = 20) \\ \beta_{Ek}x_E + \beta_{FZk}x_F z + \beta_{Zk}z + \varepsilon_k \ \ (k = 21,22,\cdots,30), \end{cases} \tag{5}$$

$$y = 3.0s_{10} + 2.0s_{20} + 7.0(-u_1 + u_2) + \varepsilon_{31}, \tag{6}$$

$$\varepsilon_1, \varepsilon_2, \cdots, \varepsilon_{31}, u_1, u_2 \sim N(0,1), \tag{7}$$

$$\beta_{Ak}, \beta_{BZk}, \beta_{Ck}, \beta_{DZk}, \beta_{Ek}, \beta_{FZk}, \beta_{Zk} \sim unif(1,10) \tag{8}$$

The second pattern is the experimental data, where s_{10} has the most direct effect with y, while s_{20} has the most total effect. The model equation is defined as:

$$s_k = \begin{cases} \beta_{Ak}x_A + \beta_{BZk}x_B z + \beta_{Zk}z + \varepsilon_k & (k = 1,2,\cdots,9) \\ 5.0(x_A + x_B z + z) + \varepsilon_k & (k = 10) \\ \beta_{Ck}x_C + \beta_{DZk}x_D z + \beta_{Zk}z + \varepsilon_k & (k = 11,12,\cdots,19) \\ 50.0(x_C + x_D z + z) + \varepsilon_k & (k = 20) \\ \beta_{Ek}x_E + \beta_{FZk}x_F z + \beta_{Zk}z + \varepsilon_k & (k = 21,22,\cdots,30), \end{cases} \tag{9}$$

$$y = 3.0s_{10} + 2.0s_{20} + \varepsilon_{31}, \tag{10}$$

$$\varepsilon_1, \varepsilon_2, \cdots, \varepsilon_{31} \sim N(0,1), \tag{11}$$

$$\beta_{Ak}, \beta_{BZk}, \beta_{Ck}, \beta_{DZk}, \beta_{Ek}, \beta_{FZk}, \beta_{Zk} \sim unif(1,10) \tag{12}$$

4.3 Method of Analysis

The data used for the simulation were generated using the orthogonal array L_{18}. The factors x_A, x_B, x_C, x_D, x_E, and x_F were assigned to the 2nd, 3rd, 4th, 5th, 6th and 7th columns, respectively. The noise factor z was arranged in the outer array. The first level of the noise factor was 1, and the second level was −1. The first level of the control factor was 1, the second level was 0, and the third level was −1. The number of datasets was 5,000. The initial number of experimental conditions was set to nine.

Columns 1 to 30 of the L_{32} orthogonal array were used for the extraction of EEFs in the CS-T method. The hyperparameters of the adaptive lasso were determined by the leave-one-out using $Mean\ Squared\ Error$ (MSE) as an evaluation criterion. Stability selection was set to $B = 100$, $\mathbf{\Lambda} = (1.3^1, 1.3^2, \cdots, 1.3^{30})$. π_{thr} was varied for each pattern, with 0.95 for patterns 1 and 0.80 for pattern 2, respectively. The significance level of the censoring criterion was set to 5%.

We set accuracy evaluation criteria which is the absolute values of the relationship between y and **s** calculated by each method, sorted in descending order, and the probability that the top two are non-zero and correctly aligned with s_{10} and s_{20}. The above-mentioned evaluation criteria are shown in the bar graph. The leftmost bar shows the results when nine experimental conditions are used, while the rightmost bar shows the results when all experimental conditions are used.

4.4 Simulations Results

First, we focus on the first pattern. As can be seen in Figure 1, the accuracy of the CS-T, adaptive lasso, and SS-LM methods is always less than 50%, which is rather impractical. However, the accuracy of KS-IV is always above 90%, which is a practical result.

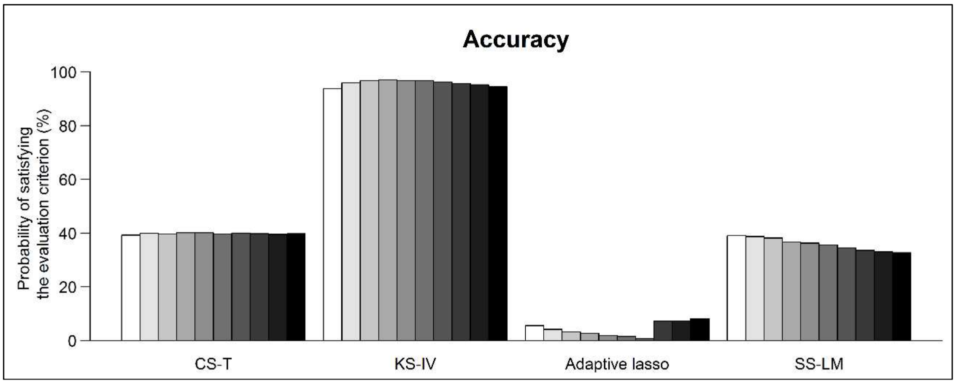

Figure 1 – Pattern 1 Results (Accuracy)

These results show that the CS-T, adaptive lasso, and SS-LM methods are unreliable when omitted variables are mixed. Conversely, the proposed method performs well.

According to the rules of the proposed method and the CS-T method, the probability of the shortest termination is 46.6% at 13 experimental conditions used for the CS-T method and 97.5% at 11 experimental conditions used for KS-IV, indicating that the proposed method can terminate the experiment stably.

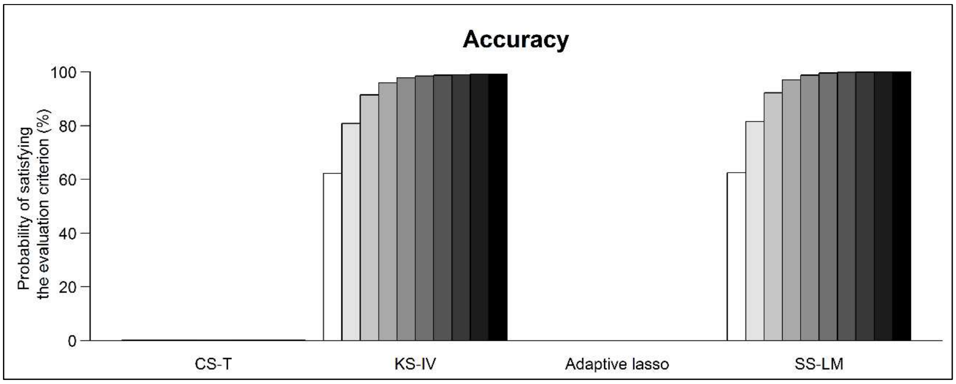

Figure 2 – Pattern 2 Results (Accuracy)

Thereafter, we focus on the second pattern. The rules of the proposed method and the CS-T method, the probability of the shortest termination is 82.8% at 13 experimental conditions used for the CS-T method and 85.5% at 11 experimental conditions used for KS-IV. Moreover, the accuracy of the proposed method and SS-LM increases with the number of experimental conditions used, as can be seen in Figure 2. However, the CS-T and adaptive lasso methods do not estimate the correct relationship.

These results show that the proposed method performs well even with experimental data where the total and direct effects are different.

Based on the above-stated results, we clarify position of the CS-T and proposed methods in system selection. In models where the total and direct effects differ, the accuracy of the CS-T method is approximately zero. This is because the Ta-method is included, thereby adding the effect from **s** to y to the effect from **x** and **z** to **s**, so that the total effect is estimated.

As a confirmation, the probability that s_{20} has the greatest influence on y is shown in Figure 3. For comparison, the other methods calculate the probability by evaluating the standard partial regression coefficient. Figure 3 shows that the CS-T method has a high probability of determining that s_{20} has the greatest effect on y.

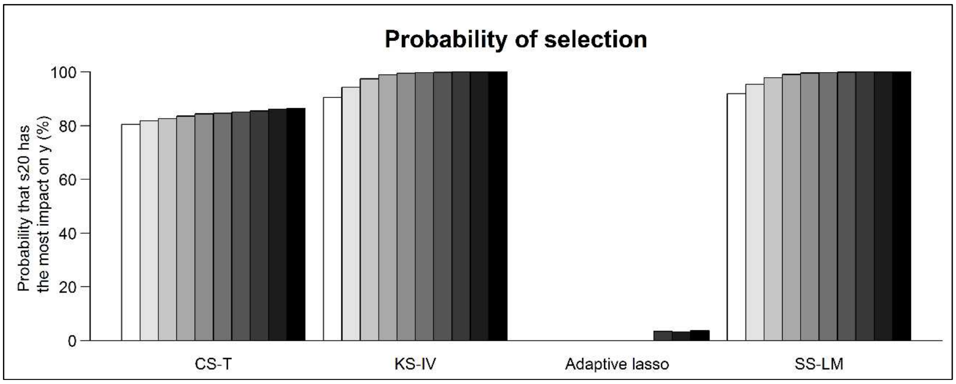

Figure 3 – Pattern 2 Results (Probability that s_{20} has the Most Influence on y)

First, the aim of the proposed and CS-T methods is to convert the extracted **s** into a new control factor or a generic function. For the former, it is always necessary to choose s that has a strong direct effect on y. Ideally, the latter should also consider the effect on **s** when **x** is optimized. In other words, the proposed method is suitable for developing control factors, and the CS-T or the proposed method with standard partial regression coefficients is suitable for developing generic functions.

Nonetheless if there is an omitted variable, it is necessary to choose **s** that has a strong direct effect on y. This is because the effect of **x** and **z** on s is combined with the effect of the omitted variable on **s**.

5 CONCLUSION

The CS-T method proposed by Hosokawa et al. (2015) is a technique for constructing an efficient knowledge search flow for system selection. However, there is room for debate on how to verify the usefulness and deal with the endogenous nature of the data. Therefore, in this study, we attempt to extract mechanisms more accurately and easily by proposing an analysis flow that introduces variable selection by stability selection and the IV concept.

Simulations show that knowledge search using the proposed method has sufficient practical accuracy, even for quasi-experimental data with omitted variables. In addition, the probability of censoring in the shortest time is high. In other words, it can be interpreted that the extracted mechanism is less likely to fluctuate depending on the number of experimental conditions used, and a stable knowledge search can be performed.

Moreover, based on the results of the simulations, we are able to clarify the position of the CS-T and proposed methods in technological development.

Originally, developing control factors and generic functions in technological development requires a high degree of engineer sophistication. However, the analysis process in this study does not require a high level of sense from the engineer, and can be carried out in a realistic and systematic manner. Therefore, it can be expected to establish quality engineering that is both efficient and creative.

Future development entails the determination of the optimal threshold value π_{thr}. Depending on the value of π_{thr}, the EEFs selected in the stability selection will change, which will affect the estimation of the relationship between the EEFs and the output/objective characteristics. In this study, the analysis is conducted with known values, but better results can be expected by changing π_{thr} as the number of experimental conditions used changes.

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.

REFERENCES

Bowden, R.J. and Turkington, D.A., 1985. *Instrumental Variables*. Cambridge: Cambridge University Press.

Brito, C. and Pearl, J., 2002. Generalized instrumental variables. *Uncertainty in Artificial Intelligence*, 18, pp.85-93.

Gamage, P., Jayamaha, N.P. and Grigg, N.P., 2017. Acceptance of Taguchi's Quality Philosophy and Practice by Lean practitioners in apparel manufacturing. *Total Quality Management & Business Excellence*, 28(11-12), pp.1322-1338. DOI: 10.1080/14783363.2015.1135729.

Göhler, S.M., Ebro, M. and Howard, T.J., 2018. Mechanism and coherence of robust design methodology: A robust design process proposal. *Total Quality Management & Business Excellence*, 29(3-4), pp.239-259. DOI: 10.1080/14783363.2016.1180952.

Hosokawa, T. and Miyagi, Z., 2019. Quality engineering-based management: a proposal for achieving total optimisation of large systems. *Total Quality Management & Business Excellence*, 30(1), pp.182-194. DOI: 10.1080/14783363.2019.1665843.

Hosokawa, T., 2020. *Taguchi Method for Technology Development CS-T Method for Exploring Generic Functions*. Tokyo, Japan: Nikkagiren.

Hosokawa, T., Okamuro, A., Sasaki, Y. and Tada, K., 2015. A Proposal of Development Methodology Integrating Parameter design and T-Method. *Journal of the Japanese Society for Quality Control*, 45(2), pp.64-72. DOI: 10.20684/quality.45.2_194.

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*, 42(2), pp.265-277. DOI: 10.20684/quality.42.2_265.

Kawada, N., 2013. Application of Quality Engineering in Research and Development. *J-TREC TECHNICAL REVIEW*, 1, pp.38-45.

Kuroki, M., Miyakawa, M. and Kawata, R., 2003. Instrumental Variable(IV)Selection for Estimating Total Effects in Conditional IV Method. *Japanese Journal of Applied Statistics*, 32(2), pp.89-100. DOI: 10.5023/jappstat.32.89.

Lauritzen, S.L., 1996. *Graphical models*. Oxford: Oxford University Press.

Mashhadi, A.F., Alänge, S., Gustafsson, G. and Roos, L.U., 2016. The Volvo Robust Engineering system: How to make robust design work in an industrial context. *Total Quality Management & Business Excellence*, 27(5-6), pp.647-665. DOI: 10.1080/14783363.2015.1039938.

Meinshausen, N. and Bühlmann, P., 2010. Stability selection. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 72(4), pp.417-473. DOI: 10.1111/j.1467-9868.2010.00740.x.

Myers, R.H., Khuri, A.I. and Vining, G., 1992. Response Surface Alternative to Taguchi Robust Parameter Design Approach. *The American Statistician*, 46(2), pp.131-139. DOI: 10.2307/2684183.

Rosenbaum, P.R. and Rubin, D.B., 1983. The central role of the propensity score in observational studies for causal effects. *Biometrika*, 70(1), pp.41-55. DOI: 10.2307/2335942.

Shimizu, S., Hoyer, P.O., Hyvärinen, A. and Kerminen, A., 2006. A linear non-gaussian acyclic model for causal discovery. *Journal of Machine Learning Research: J.M.L.R*, 7, pp.2003-2030. DOI: 10.5555/1248547.1248619#sec-terms.

Taguchi, G., 1993. System Selection. *Journal of Quality Engineering Society*, 1(5), pp.2-8. DOI: 10.18890/qes.1.5_2.

Taguchi, G., 2004. Objective Function and Generic Function (2):―Product Development and Technology Development―. *Journal of Quality Engineering Society*, 12(4), pp.6-13. DOI: 10.18890/qes.12.5_6.

ABOUT AUTHORS

Kosuke Oyama[0000-0002-1195-8172] (K.O.) – Graduate student, Department of Industrial and Management Systems Engineering, Waseda University, Tokyo, Japan, e-mail: whbcbr1000rr@akane.waseda.jp.

Masato Ohkubo[0000-0002-1885-6945] (M.O.) – Lecturer, Department of Business Administration, Toyo University, Tokyo, Japan, e-mail: ohkubo@toyo.jp.

Yasushi Nagata[0000-0002-4947-1204] (Y.N.) – Prof., Department of Industrial and Management Systems Engineering, Waseda University, Tokyo, Japan, e-mail: ynagata@waseda.jp.

AUTHOR CONTRIBUTIONS

Conceptualization, K.O. and M.O.; Methodology, K.O. and M.O.; Software, K.O.; Validation, K.O.; Formal analysis, K.O.; Investigation, K.O.; Resources, K.O.; Data curation, K.O.; Original draft preparation, K.O.; Review and editing, M.O. and Y.N.; Visualization, K.O.; Supervision, M.O. and Y.N.; Project administration, Y.N.; Funding acquisition, Y.N.

CONFLICTS OF INTEREST

The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.