



Preferences for Supercomputer Resources Using the Logit Model

Hyungwook Shim  and Jaegyoon Hahm*

National Supercomputer Center, Korea Institution of Science and Technology Information, Daejeon 34141, Republic of Korea

Abstract

Public research, which requires large computational resources, utilizes the supercomputers of the National Supercomputing Center in the Republic of Korea. The average utilization rate of resources over the past three years reached 80%. Therefore, to ensure the operational stability of this national infrastructure, specialized centers have been established to distribute the computational demand concentrated in the national centers. It is necessary to predict the computational demand accurately to build an appropriate resource scale. Therefore, it is important to estimate the inflow and outflow of computational demand between the national and specialized centers to size the resources required to construct specialized centers. We conducted a logit model analysis using the probabilistic utility theory to derive the preferences of individual users for future supercomputer resources. This analysis shows that the computational demand share of specialized centers is 59.5%, which exceeds the resource utilization plan of existing specialized centers.

Index Terms: Supercomputer, Computing resource, Logit model, Utility theory, Discrete choice model

I. INTRODUCTION

Korea's national supercomputing resources are divided into the National Supercomputing Center (national center), specialized centers, and unit centers. The national center should forecast future demand, in addition to securing and managing world-class supercomputing resources and promoting national projects related to research for core technology development and the training of professional human resources. Specialized centers should establish, operate, and provide services for supercomputer resources in key areas of supercomputer utilization designated by the government, and perform roles such as the dissemination of research results and management of large-capacity data. Although the role of the Unit Center is not specified by law, there is a plan to introduce a number of private companies as sub-organizations of the specialized center based on the government's

“National Supercomputer Innovation Strategy.”

Recently, the usage rate of supercomputer resources provided by the national center has exceeded 80% on average per year, and the usage rate per hour has often exceeded 90%. Therefore, at the national level, it is impossible to handle the increasing computational demand using limited resources.

To address this, specialized centers were established for each respective field in 2022. A specialized center should provide specialized supercomputer services in each field by building an infrastructure suitable for the computational demand in that field. It is important to predict the appropriate scale of infrastructure for smooth operation, and it is imperative to estimate the computational demand in line with the start of service in 2024. However, the demand for specialized centers is being investigated individually in the field, targeting users with experience using existing supercomputers. For example, a survey on the demand for a spe-


Received 16 July 2023, Revised 16 October 2023, Accepted 19 October 2023

*Corresponding Author Jaegyoon Hahm (E-mail: jaehahm@kisti.re.kr)

National Supercomputer Center, Korea Institution of Science and Technology Information, Daejeon 34141, Republic of Korea

Open Access <https://doi.org/10.56977/jicce.2023.21.4.261>

print ISSN: 2234-8255 online ISSN: 2234-8883

 This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0/>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

Copyright © The Korea Institute of Information and Communication Engineering

cialized center in the field of nuclear fusion targets only users who use supercomputers in that field. However, once a specialized center is established, there is bound to be an influx of users using the resources of the existing national center. This is because for certain people, the utility of providing separate specialized resources and services operated by a specialized center may be greater than that of a national center. Furthermore, specialized centers must allocate a portion of their resources for joint utilization. This policy measures the temporary expansion of resources to the maximum scale when large-scale computational resources beyond the national center are needed. Therefore, it is important to estimate the demand inflow between national and specialized centers to set a joint utilization ratio. To date, the only academic study related to this topic is one that proposed a national center demand management plan [1]. In particular, in the case of Korea, there is a lack of related references because studies related to the inflow and outflow of demand have never occurred because of a single national center resource.

In this situation, utility functions were constructed for the national and specialized centers through the logit model, an individual behavior model. The share ratio was estimated to calculate the new computational demand shared with specialized centers. In addition, the elasticity of each center's usage cost was analyzed to derive the characteristics of individual behaviors regarding the use of supercomputers. The analysis results can be used as basic data for estimating the size of the specialized center infrastructure. Simultaneously, it is possible to prepare countermeasures so that computational demand can flow into specialized centers rather than saturated national centers by considering individual behavioral characteristics.

This paper consists of seven sections. Sections 1 and 2 describes the background and need for the study, and the scholarly distinction between differentiation and progress through an analysis of prior research. Section 3 explains the current state of domestic supercomputer resources. Section 4 describes methodologies such as probability utility theory and the theoretical background of the logit model, and Section 5 discusses the survey subjects and methods. Section 6 derives a selection model, examines the variable selection and its validity of variable selection, and presents the analysis results. Finally, in Section 7, the results are summarized, the viewpoint is clarified, and the final point and pursuit plan are presented.

II. LITERATURE REVIEW

Mitra [2] proposed a disaggregate mode choice model for agricultural freight. To develop the model, they used disaggregated revealed preference (RP) data for grain movement

in elevators. The utility function contains the attributes of the mode. It introduces a binary extreme value model such as the probit and mixed logit models. Based on the estimated McFadden likelihood ratio, the probit model exhibited the best fit. The demand elasticity was calculated to assess the sensitivity of the mode choice probability to changes in shipping cost, elevator capacity, and shipment volume. Waraich [3] proposed a simple parking model and described its implementation in a conventional agent-based traffic simulation. The parking model provides feedback to the traffic simulation such that the entire simulation can respond to spatial differences in parking demand and supply. Scenario simulation results for the city of Zurich, Switzerland show that the model can capture key elements of parking, including capacity and price, and help in designing parking-oriented transportation policies. Ding [4] estimated the travel behavior of individual travelers, divided into several groups based on their personal characteristics. Travelers were grouped by cluster analysis using Statistical Analysis System (SAS) software. A trip to the central business district (CBD) of Nanjing City, China, was selected as a case study. Two travel modes were investigated: public transportation (buses and subways) and vehicles. Personal and travel information were collected through an RP survey and a stated preference (SP) survey. Personal information included sex, occupation, income, and vehicle ownership, whereas travel information included mode choice, walking time, waiting time, ride time, fare, and comfort. The RP/SP received 524 valid responses. Car-sharing-promoting neighborhoods are a new concept in urban development that combines car sharing, sustainable transportation planning, and attractive housing to reduce private car use and improve neighborhood quality. To investigate residents' preferences for such neighborhoods, a statement choice experiment was designed to systematically vary the attributes of neighborhoods promoting car sharing to derive their usefulness for people with a specific socio-demographic profile. The survey was conducted among residents living in a densely populated urban area in the Netherlands. A total of 610 valid responses were obtained. A mixed logit model was estimated to derive the utility of car sharing by facilitating neighborhoods for specific profiles [5]. Carlevschi [6] used data from 142 developing countries to classify monetary policy into three mutually exclusive categories: fixed exchange rates (or hard pegs), inflation targets, and reference groups (soft pegs). As the dependent variable is unordered and categorical, it is estimated as a polynomial probability model in which each country chooses the one that provides the highest utility among the three monetary policies. The multinomial probit model is a choice probability model that evaluates three choices. The probability of selecting an inflation target versus a soft peg is based on independent variables representing country characteristics such as trade openness, vulnerability to external shocks, fis-

cal dominance, and central bank characteristics. It reports two results: the logarithm of the ratio and the logarithm of the ratio of the probability of choosing a fixed exchange rate versus choosing a criterion. Xie [7] proposed the construction of a multilevel dynamic game model to analyze the complex dynamic relationships between consumers, taxis, and government policy-making behavior. Through the inverse induction method, it is concluded that the government should determine the subsidy factor per unit of product as well as the optimal reduction rate for carbon emissions determined by the carbon tax and the operator's judgment. Based on this, with respect to the behavioral choices of consumers, taxis, and governments, this study provides five conclusions on the question of whether governments should levy and subsidize carbon taxes and the magnitude of their financial impositions and subsidies. It also makes value judgments regarding the forms/methods for future taxi operations.

In previous studies, various preference analysis studies have been conducted using a selection model that follows the probability utility theory. Mitra [2] described a choice model for agricultural freight shippers. Waraich [3] suggested a parking model for traffic. Ding [4] estimated travel behaviors, and Wang [5] investigated residents' preferences for carsharing in such neighborhoods. Therefore, it is possible to apply the probability utility theory to the field of supercomputers to represent the computational resources of a specific institution as a utility function and derive a selection model for resource use. However, few studies have been conducted on the selection models for supercomputing resource use. This study develops a utility function for national and specialized centers with domestic supercomputer resources as a selection model and estimates the selection probability for each selection alternative.

III. STATE OF SUPERCOMPUTER RESOURCES

Currently, supercomputer resources operated according to legal grounds can be divided into national and specialized centers. The national center is designated as KISTI (Korea Institute of Science Technology Information) and currently provides supercomputer resources of 25.3PF, and the specialized center plans to build and operate about 480PF of resources by 2031 for 10 fields. As shown in Table 1, the specialized centers will operate with a total of ten institu-

tions for Material/Nano, Life/Health, ICT, Meteorological/Climate/Environment, Autonomous driving, Space, Nuclear fusion/accelerator, Manufacturing base technology, Disaster, and National defense security.

Each institution plans to build infrastructure considering the computational demand by field, and the current scale of infrastructure reflects the results of a demand survey targeting users at the national center. However, the size of the infrastructure of the specialized center to be built by 2031 must consider the new computational demands. This is because new computational demand is rapidly increasing owing to the recent influx of artificial intelligence (AI) computational demand in various fields, and the potential computational demand in each field where supercomputers cannot be used because of the limited resources of the national center must also be considered. The inflow of new computational demands can select resources from the national and specialized centers differently from the past. Therefore, considering the characteristics of each center, users are highly likely to select and use resources with high efficiency. In this case, the demand share ratio between the national and the specialized centers can be estimated through a discrete selection model using the probability utility. The share ratio between the national and specialized Centers for New Computational Demands can contribute to determining a more practical scale for future specialized center infrastructures by field.

IV. Theoretical Background

The demand share ratio for supercomputing resources was estimated using the logit model, one of McFadden's discrete choice models. Until now, there has been no case in which the logit model has been applied to estimate the sharing ratio of supercomputer resources. However, the reason for applying this model for the first time in this study is that the inflow and outflow of demand for the use of national and specialized centers are determined by the utility of individual users. Users must consider usage fees and the available time to use supercomputer resources. In particular, usage time, which is the biggest limitation in using existing national centers, was applied as a highly useful variable that determines the success of the research. Usage fees also have a significant effect on the utility of users who do not receive support from companies or governments. Therefore, the logit model applying the utility theory is the most appropriate for this study and is valuable as an academic attempt.

The logit model is an individual behavior model that can be applied when multiple alternatives exist and follows the principle of maximizing utility, in that an alternative with the highest level of utility is selected from all alternatives available for an individual to select. The degree of utility for each

Table 1. List of Specialized Centers

Material/ Nano	Bio/Health	ICT	Meteorology/ Climate/ Environment	Autonomous driving
Space	Nuclear fusion/ Accelerator	Manufacturing technology	Disaster	Defense/ Security

alternative is expressed by the utility function U_i , which is divided into observable utility V_i and unobservable utility ε_i as shown in (1). i represents an alternative.

$$U_i = V_i + \varepsilon_i \tag{1}$$

The probability of selecting alternative was derived as follows: The probability that individual i chooses alternative i is expressed as (2).

$$P_i(n) = \text{Prob}(U_{ni} > U_{nj}, \forall j, j = 1, \dots, j) \\ = \text{Prob}(\varepsilon_{nj} > \varepsilon_{ni} + V_{ni} - V_{nj}) \tag{2}$$

In (2), because ε cannot be expressed as a definite numerical value, it is assumed to be a random variable with a constant distribution.

In the case of the binomial logit model, ε_i is assumed to be a normal distribution, it becomes a probit model. Assuming that ε_i has a Weibull distribution, it becomes a logit model. The probability of selecting a specific alternative in the logit model is mathematically derived as follows: Assuming that ε_i in (1) has a Weibull distribution, the probability that ε_i is smaller than any constant ε is as in Equation (3).

$$P(\varepsilon_i \leq \varepsilon) = \exp[-\exp(-\varepsilon)] = e^{-e^{-\varepsilon}} \tag{3}$$

Deriving the probability density function of the Weibull distribution from Equation (3) is equivalent to Equation (4).

$$\psi(\varepsilon) = e^{-\varepsilon} \exp(-e^{-\varepsilon}) \tag{4}$$

Here, the probability density function for representing the probability of $\varepsilon_i(=b)$ is as shown in (5).

$$e^{-b} \exp(-e^{-b}) = \exp[-\exp(-b) - b] \tag{5}$$

For example, the probability that alternative a is selected can be calculated from (3) and (5), as shown in (6):

$$P_a = \int_{-\infty}^{\infty} \exp(-b) \exp[-\exp(-b) \sum_{n=1}^J \exp(V_j - V_a)] db \tag{6}$$

When substituted with a variable for the calculation, as in (7), (6) can be rewritten as (8).

$$\exp(-b) = Z, \sum_{n=1}^J \exp(V_j - V_a) = a \tag{7}$$

$$P_a = \int_0^{\infty} \exp(-Za) dz = \frac{1}{a} \tag{8}$$

Based on (6) and (8), the final alternative selection probability is given by (9):

$$P_a = \frac{\exp(V_a)}{\sum_{n=1}^J \exp(V_j)} \tag{9}$$

The observable utility V_j of (9) can be expressed as (10) when it is assumed to be in the form of a linear function, considering alternative characteristics. α is the alternative characteristic constant, β is the coefficient, x is the alternative general variable, and x_j is the alternative characteristic variable.

$$V_j = \alpha + \beta x + \beta_j x_j \tag{10}$$

The logit model test involved checking the sign of the parameters, testing the significance of the parameters, and testing the significance of the model. First, in reviewing the sign of the parameter, we determine whether the sign of the parameter of the independent variable is reasonable in the derived utility function. In general, it is appropriate for travel time and travel cost to have a negative sign for all means, and it is necessary to establish and review individual standards for alternative characteristic variables according to the characteristics of the alternatives. For example, if the alternative special constant for a car is the presence or absence of children, it is reasonable to have a positive sign, because the need for a car is higher than in the case without children. The t-test was used to determine whether the parameter estimates were statistically valid. The significance test of the model uses Mcfadden's LRI (Likelihood Ratio Index), and a value of 0.2 to 04 is judged appropriate [8-9].

V. SURVEY

The survey was conducted using the SP method, and the subjects were students and office workers in the field of science and technology with potential use of supercomputers. Only 180 survey results were used, the reliability of which was recognized in the survey results of 200 participants. Preference surveys on alternatives were presented to respondents in the form of a choice set through conjoint analysis, and a 3-level orthogonal table was used for the choice set for accurate decision-making. The level of each variable was reflected at the $\pm 30\%$ level based on the reference value. The alternative characteristic constant appears as $n-1$ when the number of alternatives is n , and the general variables are set to use time and cost that can reflect the individual characteristics of alternatives. The survey respondents' characteristics are listed in Table 2. In terms of affiliation, enterprises accounted for 58%,

Table 2. Survey characteristics

Response Results	
Affiliation	Enterprise (58%), University (10%), Research Institute (2%), Others (30%)
Field	Material/Nano (6%), Disaster (12%), Meteorology/Climate/Environment (18%), Space (6%), Autonomous driving (2%), Bio/Health (14%), Nuclear fusion/Accelerator (6%), Others (36%)
Purpose	Government R&D (32%), Enterprise R&D (40%), Individual Research (28%), Others (0%)
Fund Scale	~\$799 (34%), \$800~\$3799 (28%), \$3800~\$7499 (16%), \$7500~ (22%)
Possession of Supercomputer	Possession (22%), Non-possession (78%)

other 30%, universities 10%, and research institutes 2%. Excluding the other fields, 18% of the fields were meteorology/climate/environment, 14% were bio/health, and 12% were disasters. For the purpose of using supercomputers, enterprise R&D accounted for 40%, government R&D for 32%, and research funds of \$799 or less were the highest at 34%. 22% of respondents owned supercomputers.

VI. ESTIMATION AND RESULTS

Logit model analysis was performed using SAS Studio’s MDC procedure (PROC MDC function). The utility function of the alternative is given by (11). The utility function of the national center is composed of the alternative characteristic constant α (national center), the coefficient β , and alternative general variables time and cost.

$$\alpha + \beta_{time}time_{NC} + \beta_{cost}cost_{NC} \tag{11}$$

The variables refer to the content of the “National Supercomputer Centre’s R&D Innovation Support Program”. The alternative characteristic constant has values of ‘0’ and ‘1,’ and both time and cost are set as alternative general variables. The alternative general variable time is the unit CPU usage period (days) per \$750, and the cost means the unit CPU usage cost (\$1) per day. Table 3 summarizes these variables.

Table 3. Variable list

Variable	Code	Choice set
Alternative specific constant	α (Dummy)	NC (1), SC (0)
Alternative generic variables	time	3-Level - NC: 187, 267, 346 - SC: 96, 138, 179
	cost	3-Level - NC: 2.5, 3.6, 4.7 - SC: 5.0, 7.2, 9.4

A correlation analysis was performed to analyze the validity of the variable selection of the model. Correlation analysis showed the direction and strength of the mutual relationships between independent variables, and an appropriate Spearman’s correlation coefficient was used when variables had equal interval ratios, scales, and sequence characteristics. As a result of the analysis (Table 4), all correlation coefficients are significant at the 0.01 level (both sides) and calculated at the -0.4 level, so there are some negative correlations between variables, but they do not appear to be at a level that determines inappropriateness for variable selection.

Table 4. Correlation analysis results

		Time	Cost
Time	correlation coefficient	1.000	-.403**
	significance probability	-	.000
	N	180	180
Cost	correlation coefficient	-.403**	1.000
	significance probability	.000	-
	N	180	180

**p<0.01.

Table 5 presents the estimation results of these coefficients. Looking at the estimated coefficient, β_{cost} has a negative (-) sign, meaning that the utility of the alternative decreases as the cost value increases. The β_{time} coefficient has a positive (+) sign, and as the value increases, the utility of the alternative increases. Both β_{cost} and β_{time} were statistically significant at the significance level of 5% as a result of the t-test, and the goodness of fit of the model was found to be within 0.2 to 0.5 of McFadden’s LRI value according to the likelihood ratio test result, so it can be analyzed at an appropriate level.

Table 5. Estimate results

	Estimate	Standard Error	t Value	Approx Pr > t
α	2.710	0.919	2.95	0.003
β_{time}	0.013	0.006	2.06	0.039
β_{cost}	-0.877	0.224	-3.92	<.000

McFadden’s LRI: 0.495

Using the coefficient estimation results, the utility function of (11) can be expressed as (12).

$$2.71 + 0.013 \cdot time_{NC} - 0.877 \cdot cost_{NC} \tag{12}$$

Table 6 shows the computational demand shares for the national and specialized center using the utility function in (12). The share of the new computational demand was 40.5% for national centers and 59.5% for specialized centers, which was 19% higher for specialized centers than for national centers.

Table 6. Share rate

	Share Rate (%)
National Center	40.5
Specialized Center	59.5

The change in the probability of selecting a national center at the cost of using a national center means direct elasticity, and the change in the probability of selecting a specialized center according to a change in the cost of using a national

center represents cross-elasticity. As shown in Table 7, in the case of the group using the national center, the elasticity of the cost of using the National Center and the specialized center was relatively similar, approximately 10 times higher. This means that when using a specialized center, sensitivity to the cost of using a specialized center is much higher than sensitivity to the cost of using a national center.

Table 7. Elasticity calculation result for usage cost

	Usage cost for National Center	Usage cost for National Center
National Center	-2.145*	-2.454
Specialized Center	-0.144	-1.460*

*Direct Elasticity

VII. DISCUSSION AND CONCLUSION

The share ratio of the national and specialized centers of the new supercomputer computational demand was derived, and elasticity was analyzed. The analysis showed that the share rate of specialized centers was high, but 40% of the demand could still depend on national centers. When specialized centers for each field were established, the results were very different from those of the previous plan, in which most of the demand would move to specialized centers owing to professional services, speed of resource use, and convenience. From the standpoint of the government managing and operating supercomputer resources in the future, it will be necessary to consider ways to induce new computational demand for specialized centers and the additional expansion of national center resources.

This study introduces a logit model for the first time as a methodology for estimating the appropriate size of supercomputer infrastructure to be built in the future. This is meaningful because it suggests the factors that affect the share rate by deriving the share rate and elasticity of the National Center and specialized center for the new computational demand. However, it can be said that the limitation is that the number of domestic supercomputer users is relatively small, the survey was conducted in a state where awareness was low at the start of the specialized center, and sufficient survey response results were not secured. In addition,

much work remains to be done to estimate the infrastructure size of each specialized center by estimating the share rate at the National Center and Specialized Center levels, rather than the share rate for each of the 10 fields. In the future, we plan to make efforts to build an economical infrastructure by referring to the analysis results and applying the methodology to additionally analyze the individual share ratios of specialized centers at the stage of establishing an operation plan for the 10 specialized centers.

ACKNOWLEDGEMENTS

This study was funded by KISTI (K-23-L02-C03).

REFERENCES

- [1] H. W. Shim and J. G. Hahm, "A study on demand management plans for National Supercomputer resources," *Technology in Society*, vol. 75, p. 102376, Sep. 2023. DOI: 10.1016/j.techsoc.2023.102376.
- [2] S. Mitra, "Discrete choice model of agricultural shipper's mode choice," *Transportation Journal*, vol. 52, no. 1, pp. 6-25, Jan. 2013. DOI: 10.5325/transportationj.52.1.0006.
- [3] R. A. Waraich and K. W. Axhausen, "Agent-based parking choice model," *Transportation Research Record*, vol. 2319, no. 1, pp. 39-46, Jan. 2012. DOI: 10.3141/2319-05.
- [4] L. Ding and N. Zhang, "A travel mode choice model using individual grouping based on cluster analysis," *Procedia Engineering*, vol. 137, pp. 786-795, 2016. DOI: 10.1016/j.proeng.2016.01.317.
- [5] J. Wang, G. Z. Dane, and H. J. P. Timmermans, "Carsharing-facilitating neighbourhood choice: a mixed logit model," *Journal of Housing and the Built Environment*, vol. 36, no. 3, pp. 1033-1054, Sep. 2021. DOI: 10.1007/s10901-020-09791-z.
- [6] L. Carolevschi, "Monetary policy choice in developing countries: a multinomial probit model," *The Journal of Developing Areas*, vol. 52, no. 3, pp. 125-138, 2018. DOI: 10.1353/jda.2018.0041.
- [7] X. Xie, Y. Wang, and X. Li, "The usage analysis and policy choice of CNG taxis based on a multi-stage dynamic game model," *Computational Economics*, vol. 54, no. 4, pp. 1379-1390, Dec. 2019. DOI: 10.1007/s10614-016-9645-5.
- [8] D. McFadden and K. Train, "Mixed MNL models for discrete response," *Journal of Applied Econometrics*, vol. 15, no. 5, pp. 447-470, Sep. 2000. DOI: 10.1002/1099-1255(200009/10)15:5<447::aid-jae570>3.0.co;2-1.
- [9] D. A. Hensher and L. W. Johnson, "Applied discrete choice modelling," in *Croom Helm and Wiley*, 1st ed. London, UK, 2018.



Hyungwook Shim

received his PhD degree in City Planning from Seoul National University 2021. He is senior researcher in Korea Institute of Science and Technology Information.



Jaegyoon Hahm

received his MS degree in Computer Science from Korea Advanced Institute of Science and Technology 2002. He is principal researcher in Korea Institute of Science and Technology Information.