Estimating Time Lag Effect of Nuclear National R&D Programs using Almon Polynomial Distributed-lag Model

Jihwan Lim
Korea Atomic Energy Research Institute, Daedeok-daero 989-111, Yuseong-gu, Daejeon, Korea
limjh@kaeri.re.kr

1. Introduction

Big science refers to a scientific field that carries out large-scale projects that include nuclear, space, and marine development programs. Due to its gigantic projects and importance for national development, the government funds these big science projects. One example of government funding is through the national research and development (R&D) program, where research is funded by the National Research Foundation of Korea (NRF).

It is empirically known that a time lag exists between the input of budgets and research outputs. Under a limited budget, it is crucial to analyze and evaluate the budget execution's efficiency of the funded programs. Quantitative evaluations on the scientific and technological outputs are measured through the numbers of SCIE papers and patent applications. Previous research analyzing national R&D programs has shown a time lag of one to four years for basic research, one to three years for applied research, and one to two years for development research [3]. Regardless, for the Korean nuclear sector, time lag effects are not yet considered for evaluation. Furthermore, there is no previous literature on estimating the time lag effects of the Korean nuclear national R&D programs. Thus, this study intends to estimate the time lag effect of nuclear national R&D programs by using an econometrics distributed-lag model.

2. Methods

A distributed-lag model refers to a regression model that includes the explanatory variables of the current and the lagged values. If the length of the lag is defined, it is called the finite distributed-lag model, and likewise, if the length of the lag is undefined, it is called the infinite distributed-lag model. Although the frequently used ordinary least squares (OLS) method can be adopted for these models, multicollinearity will likely be the problem as the values of variables are highly likely to be correlated. In addition, more degrees of freedom will be lost as more variables are included in the equation, which may cause an erroneous statistical implication. Econometrics models have been developed to overcome the problems mentioned above, such as the Koyck or Almon model. However, the Koyck model has based on the premise that β coefficients drop as the lag lengthens. In other words, regardless of the time or amount of R&D funding, the output will always decline as time goes by. Such cannot efficiently capture the properties of R&D funding and its outputs [2]. Therefore, this study intends to use the Almon model.

2.1. Almon Polynomial Distributed-lag Model

Almon suggests estimating the coefficients through a polynomial and the length of the lag [1]. The following equations illustrate the Almon polynomial distributed-lag model. A finite distributed-lag model can be written as below:

$$Y_t = \alpha + \beta_0 X_t + \beta_1 X_{t-1} + \beta_2 X_{t-2} + \cdots + \beta_k X_{t-k} + \mu_t \quad (1)$$

Using the Weierstrass' theorem of mathematics, Almon depicts that β can be approximated through a different degree of polynomial and length of lag. Eq. (2) below shows an mth-degree polynomial of i.

$$\beta_i = a_0 + a_1 i + a_2 i^2 + \cdots + a_m i^m \quad (2)$$

Substituting Eq. (2) into Eq. (1) will provide the following equation:

$$Y_t = \alpha + \sum_{i=0}^{k} (a_0 + a_1 i + a_2 i^2 + \cdots + a_m i^m) X_{t-i} + \mu_t$$

$$= \alpha + a_0 \sum_{i=0}^{k} X_{t-i} + a_1 \sum_{i=0}^{k} i X_{t-i} + a_2 \sum_{i=0}^{k} i^2 X_{t-i} + \cdots + a_m \sum_{i=0}^{k} i^m X_{t-i} + \mu_t \quad (3)$$

Consequently, by constructing variables Z as below,

$$Z_{0t} = \sum_{i=0}^{k} X_{t-i},\ Z_{1t} = \sum_{i=0}^{k} i X_{t-i},\ Z_{2t} = \sum_{i=0}^{k} i^2 X_{t-i}, \ldots, Z_{mt} = \sum_{i=0}^{k} i^m X_{t-i} \quad (4)$$

then Eq. (3) can be rewritten as:

$$Y_t = \alpha + a_0 Z_{0t} + a_1 Z_{1t} + a_2 Z_{2t} + \cdots + a_m Z_{mt} + \mu_t \quad (5)$$

The regression coefficients of Z can be obtained through the OLS method, and once a's have been obtained, the βs can be estimated in order by using Eq. (2).

2.2. Defining the lag length and degree of a polynomial

The equations show that the maximum length of lag k and mth degree of the polynomial must be specified before applying any calculations. According to Gujarati, such is at the discretion of the researcher. Nevertheless, choosing the wrong length of lags may cause omission or inclusion of irrelevant variables. This study used the statistical method of the Akaike information criterion (AIC) on the explanatory variable to find the optimal lag length [2]. The degree of the polynomial was selected based on the insights of previous research [4,5]. As a result, the length of lag k of four and second-degree of polynomial were chosen to be used for the model.

2.3. Data collection

In order to perform this study, all information of national R&D programs performed in Korea Atomic Energy Research Institute, including published SCIE papers by year, patent applications by year, and budget by year from 2010 to 2020, were collected. Subsequently, all programs without any SCIE paper or patent output for ten years were removed from the dataset, as their objectives were not to produce such scientific or technological outputs. Also, understanding that most of the national R&D programs are over five years period, all programs less than four years of the budget were removed. As a result, our dataset was confirmed to have 197 programs with eleven years (n=2,167). All budgets were recalculated by rounding up to one million Korean won, and the GDP deflator of the reference year 2015 was used.

3. Results

Having research budget as our explanatory variable for output of SCIE papers and patent application, the regression results of Z variables are shown below:

Table I: Regression Results

	SCIE Papers	Patent Applications
Intercept	79 (218)	406 (179)
$Z0$	.000812 (.000498)	-.000404 (.000409)
$Z1$	-.000404 (.000614)	.000821 (.000503)
$Z2$	.000032 (.000164)	-.000261 (.000134)
Adj. R-squared	.911	.825
F-value	21.5	10.4
d-value	2.13	2.92

Standard errors are in parenthesis.

Both our models are statistically significant as the F-value results reject the null hypothesis (p<.05). Z variables show large standard errors because they have been constructed from the explanatory variables X, but this does not inevitably lead to βs being statistically insignificant [2].

Figure 1 shows the lag structure of calculated βs by using Eq. (2) from the results of Table 1. From 2010 to 2020, SCIE papers have shown a constant decrease from the first year of research to the fifth year. Meanwhile, patent applications tend to show their output from the second year and peak in the third year.

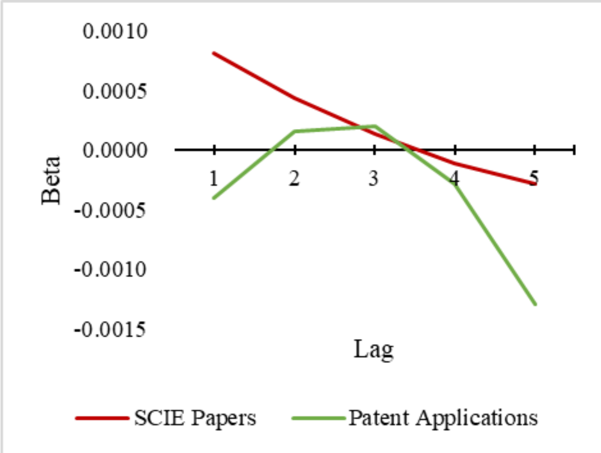

Fig. 1. Lag Structure of Nuclear National R&D Programs.

4. Discussion and Conclusion

Analyzing and evaluating programs is essential, considering that national R&D budgets are limited. Using the Almon polynomial distributed-lag model, our results show that the lag effect of SCIE papers decreases over lags, while patent applications show output from the second year and peak at the third year. This result may be used as a basic note or material to consider when evaluating nuclear national R&D programs.

Nevertheless, there are some limitations to this study. First, the data mining could be improved for the programs. Although the data only included programs with at least five years of research, programs had different starting and ending points. There can be ways of controlling this difference by imposing endpoint restrictions or conducting a preliminary review so programs can be logically unified. Second, characteristics of national R&D programs have not been specified but were collectively analyzed. There may be differences among research stages, such as basic research and applied research. Third, the minimum unit of the data was in years and therefore cannot capture any effects under a year. Precise results may be given if we could obtain quarterly or monthly data. Future research may consider such limitations and provide precise results for further insights.

Acknowledgments

This work was supported by the Korea Atomic Energy Research Institute (KAERI) Institutional Program (Project No. 525341-22).

REFERENCES

[1] S. Almon, The Distributed Lag Between Capital Appropriations and Expenditures, Econometrica, Vo. 33, No. 1, pp.178-196, 1965.
[2] D. Gujarati and D. Porter, Basic Econometrics, New York: McGraw-Hill, 2009.
[3] C.M. Park and B.C. Ku, An Analysis of Distributed Lag Effects of Expenditure by Type of R&D on Scientific Production: Focusing on the National Research Development Program, Journal of Korea Technology Innovation Society, Vol. 19, No. 4, pp. 687-710, 2016.
[4] B.H. Jeong, K.M. Cheon, and J.K. Yang, An Empirical Study on the Time Lag Effect of the Academic Performance of a National R&D Program, Journal of the Society of Korea Industrial and Systems Engineering, Vol. 35, No. 1, pp. 87-92, 2012.
[5] H.J. Lee, C.W. Baek, and J.D. Lee, Analysis on Time Lag Effect of Firm's R&D Investment, Journal of Technology Innovation Research, Vol. 22, No. 1, pp. 1-20, 2014.