INL/CON-25-84516-Revision-0

Parametric and Nonparametric Models of U.S. Cost Overruns for Nuclear Power Plants

April 2025

Changing the World's Energy Future

Nahuel Guaita, Ryan Matthew Spangler

Idaho National Laboratory

INL/CON-25-84516-Revision-0

Parametric and Nonparametric Models of U.S. Cost Overruns for Nuclear Power Plants

Nahuel Guaita, Ryan Matthew Spangler

April 2025

Idaho National Laboratory
Idaho Falls, Idaho 83415

http://www.inl.gov

Prepared for the
U.S. Department of Energy
Under DOE Idaho Operations Office
Contract DE-AC07-05ID14517

Parametric and Nonparametric Models of U.S. Cost Overruns for Nuclear Power Plants

Nahuel Guaita[1], Ryan Spangler[1], Jason Hansen[1]

[1]Idaho National Laboratory, Idaho Falls, ID

ABSTRACT

This study presents new data-driven models to estimate the effect of capacity on the percentage of cost overruns in the United States for nuclear power plant construction projects before and after the Three Mile Island accident. Parametric and nonparametric models have been developed that describe the significant shifts in nuclear energy costs during the dynamic environment. Employing a contemporary descriptive methodology and a quantitative analysis, we furnish a comprehensive overview of the alterations in cost overrun distribution and show the changes observed in other pivotal metrics alongside cost overruns. Our emphasis lies in documenting the fluctuations in cost overruns alongside nuclear reactor capacity levels and the increase of the overnight capital costs to build nuclear reactors.

Our results show that increasing the size of nuclear reactors is not a factor statistically significant to decrease the percentage of cost overruns, and the probit model results provide evidence that an increase in size increases the probability of having cost overruns larger than 100% (double the estimated cost). We also compare our findings to two other regions: Asia and Europe.

Keywords: cost overruns, distribution, regression analysis, probit, non-parametric

1. INTRODUCTION

The expensive characteristic of nuclear power technologies in comparison to other energy sources is one of the reasons used to explain the delay of new nuclear reactor deployments. Given this, despite its low-carbon benefits, nuclear energy struggles to compete with more affordable energy sources [1].

The history of nuclear energy in the United States has been characterized by a combination of achievements and setbacks, particularly visible in the realm of cost overruns and project delays. For instance, there was an initial rapid expansion of nuclear capacity during the 1960s, characterized by significant unit scaling and operational enhancements, which was followed by a period of rising project durations and costs in the 1970s [2] [3]. Public concerns over the ecological impacts of cooling water usage and low-level radiation have further fueled controversy. The Three Mile Island (TMI) accident in 1979 led to a halt in nuclear construction, amplifying construction costs and delays, which have persisted since then. Post-1970 plant surveys reveal an average overnight cost overrun of 241% [4].

Also, recent nuclear projects in the United States and throughout, such as the VC Summer project in South Carolina and the Vogtle project in Georgia, have faced challenges with cost escalations and delays. The Vogtle expansion, for instance, has been going through continuous cost overruns from initial projections [5]. Similar challenges have been encountered in nuclear construction worldwide with projects in Finland, France, and the United Kingdom experiencing cost escalations and schedule delays that generate cost

---

[1]nahuel.guaita@inl.gov

overruns [6], and [7]. However, countries like China, India, Japan, and South Korea have achieved shorter construction durations and less cost overrun since the 1990s [8]. Despite historical cost increases, various stakeholders in the nuclear industry continue to anticipate cost reductions through investments in next-generation reactor designs, banking on the idea that subsequent plants of standardized designs will be more economical [9]. This optimism extends to small modular reactors with the expectation of cost reductions due to learning effects in factory settings [10].

The primary objective of this paper is to enhance the analysis of overnight capital cost (OCC) overrun and overrun risks by examining the entire percent of cost overrun distribution with data from seven countries grouped into three regions: United States and Canada (North America); France, United Kingdom, and Switzerland (Europe); and India and Japan (Asia). The aim is to provide a more comprehensive understanding of cost overrun shifts in the United States, which might be overlooked when solely focusing on cost overrun and cost trends.

Our approach is a combination of a linear regression and probit model, and kernel density distributions. Our primary aim is to identify key patterns in the changes of the entire cost overrun distribution, extending beyond its central tendency and dispersion. We do not intend to discover all the causes behind these changes but strive to identify similarities and differences in the evolution of various aspects of the cost overrun distribution. This should provide a solid foundation for further research aimed at exploring the causal factors behind these phenomena and, ideally, proposing pertinent cost strategies to avoid overruns.

This paper is organized as follows. Section 2 starts with a non-parametric method (useful when there is no assumption on the distribution) and description of the data examining the entire spectrum of cost overruns distribution without attempting to distill it into summary measures. Some insights, primarily centered on cost overruns and costs, are presented. In Section 3, we develop a parametric approach presenting a panel model to measure the effect of capacity on cost overrun considering the effects of the different regions created and the TMI accident. Section 4 presents a parametric model to estimate the probability of cost overrun percentage. Finally, Section 5 provides the concluding remarks.

2. A NONPARAMETRIC PERSPECTIVE OF COST OVERRUNS IN THE UNITED STATES

Nonparametric methods prove to be powerful instruments when the objective is to unveil the distributions of pertinent variables, such as percentage of cost overruns [11]. Information regarding nuclear power plant costs indicates that certain plants encountered cost escalations following the TMI accident, and some plants that were already under construction experienced prolonged construction periods [12] [13]. However, [14]found that these patterns were already underway prior to the accident, with plant costs increasing and completion rates decelerating during the latter half of the 1970s [2]. [14]suggested that TMI accident might have intensified these pre-existing trends.

2.1. Data

The database contains the Overnight Capital Cost of 175 reactors for the said seven countries. From [15], the descriptive statistics show that the “% of cost overrun” has a mean around of 119.87% with a minimum of -7.9% and a maximum of 1,279.7%. The reactor capacity in megawatt electric shows a mean of 990 MWe with a minimum of 100 MWe and a maximum of 3,512 MWe.

Table I. Summary of descriptive statistics for the selected variables. [15]

Variable	Number of Observations	Mean	Standard Deviation	Min	Max
Capacity (MWe)	175	990.34	344.11	100	3512
% of cost overrun	175	119.87	153.43	-7.9	1279.7

The data sets, includes 75 U.S. light-water reactors ranging in size from 476 MWe to 1,335 MWe between 1971 and 1988. In Table II the descriptive statistics for the created region are shown. Note that North America is the region with the highest cost overrun percentage.

Table II. Summary of descriptive statistics for the selected variables by Region. [15]

	Number of Observations	Mean	Standard Deviation	Min	Max
North America					
Capacity (Mwe)	76	1022.12	339.98	476.00	3512.00
% of cost overrun	76	217.09	178.30	3.50	1279.70
Asia					
Capacity (Mwe)	39	700.85	332.74	100.00	1175.00
% of cost overrun	39	63.98	95.76	-7.90	435.60
Europe					
Capacity (Mwe)	60	1138.25	228.09	917.00	1561.00
% of cost overrun	60	33.06	34.74	0.00	236.20

The present paper does not seek to find the real causes of OCC overruns, but to show that in 1979, after the partially melted core on March 28, 1979, near the TMI Unit 2 reactor in Middletown, PA, the cost overrun percentage of the nuclear reactors in the United States increased, which can be seen in the distributions in Figure 1.

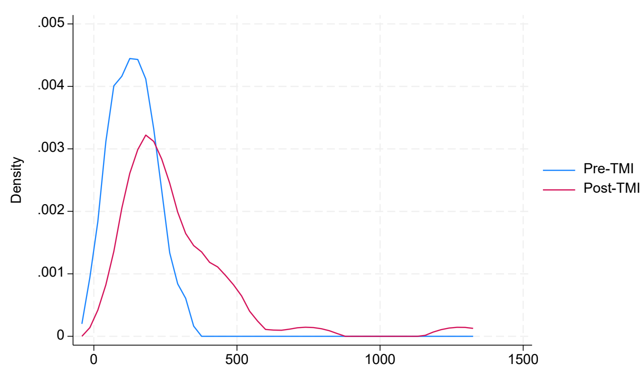

Figure 1. Standard nonparametric kernel density of U.S. percent cost overrun distribution pre- and post-TMI.

The average of percent cost overrun goes from 136.5% pre-TMI to 285.8% post-TMI. Given the increase in the percentage of cost overrun during this period, it prompts consideration of whether other elements of the distribution underwent similarly drastic changes in dispersion and central tendencies. Furthermore, how this sudden change in cost overruns impacted the whole distribution becomes an interesting discussion point.

The percentage of cost overrun over the initial budget for three regions is estimated separately using standard non-parametric kernel methods. Bandwidths are chosen using a data-driven algorithm. Figure 2 shows how the distribution of North America is centered to the right of Europe and Asia, and furthermore, the percentage of cost overruns are much higher. When comparing these three regions, the left tail of the distribution remained unchanged for all, while a significant portion of the center of North America's region distribution shifted to the right.

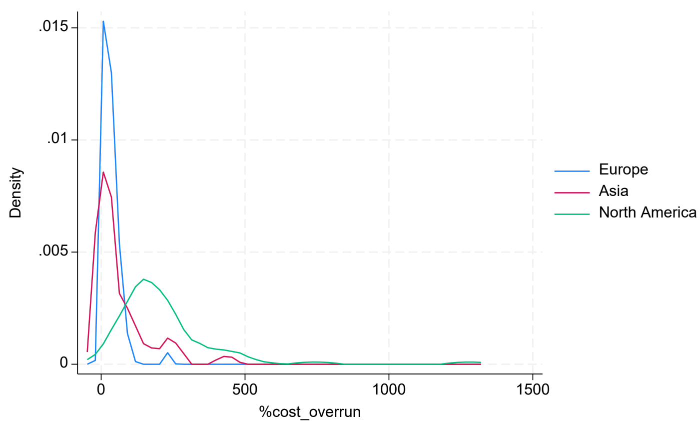

Figure 2. Kernel density distribution of percentage of cost overruns by region.

More accumulated density on the right means that more reactors have a high percentage of cost overruns. Finally, even though the estimated densities clearly point toward a much more peaked distribution for Europe and Asia, with a stronger prevalence of lower percentage cost overruns, the results do not suggest obvious unambiguous clustering based on percentage of cost overruns.

Also, the kernel cumulative distribution function (CDF) can show a nonparametric representation of the probability density function of the percent of cost overrun. In

Figure **3**, the CDF shows how the probability density for the North American region is shifted to the right, which means that the probability of having a cost overrun is higher in North America than in any other region.

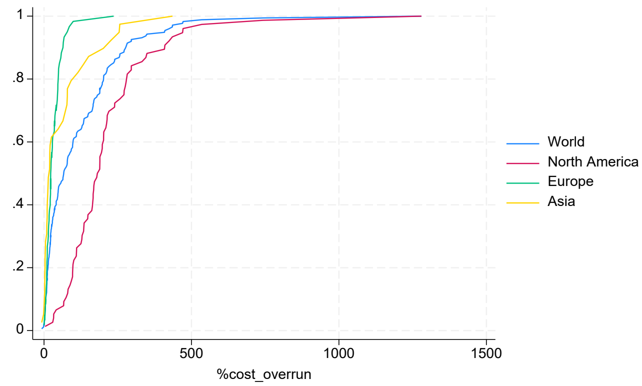

Figure 3. Cumulative distribution function of percent of cost overrun by region.

Note that visual examination does not uncover significant multimodal patterns that would allow the categorization of nuclear reactors into different groups. There are specific country characteristics such as institutional design, social, political, and technological context that are not detailed within this analysis. While a more thorough investigation into the potential presence of country-specific patterns is not within the scope of this article, it remains a potential path for future work. Addressing it needs a more meticulous approach to shed light on the asymmetric nature of percentage of cost overrun distributions and the application of more formal tests.

3. PREVIOUS EMPIRICAL RESULTS

Previous econometric analyses commonly aim to analyze the fluctuations in nuclear power plant construction costs through regression analysis, where overnight construction costs serve as the dependent variable and a range of explanatory variables are considered [16] [17] [12] [18] [19]. From a policy and decision-making standpoint, this discovery is significant because, all else being equal, it favors the development of large power plants over smaller ones.

A functional form commonly used is

$$\begin{aligned}&\ln\text{(Overnight Capital Costs)}\\&\qquad= \beta_0 + \beta_1 * \text{Size} + \beta_2 * \text{Tower} + \beta_3 * \text{First} + \beta_4 * \text{Learn} + \beta_5 * \text{Regulation} + u\end{aligned}$$

where:

- ln is natural logarithm
- Size is the capacity in megawatt electric (Mwe)
- Tower indicates for the presence of a cooling tower
- First indicates if the reactor is the initial one at a site
- Learn is a managerial metric representing industry experience
- Regulation indicates regulatory factors such as the safety regulation rigor
- u is the error term.

We estimate the following equation utilizing the percentage of cost overrun as the dependent variable

$$\ln(\text{\% of Cost Overruns}) = \beta_0 + \beta_1 * \ln\text{(CapacityMWe)} + \beta_2 * TMI + u$$

where:

- CapacityMWe is the nameplate capacity in Megawatt electric
- TMI is a binary variable equal to 1 if the reactor was built after the Three Mile Island accident
- u is the error term.

The estimations and t-statistics rely on heteroskedastic consistent covariance matrices as proposed by White (1980).

3.1. Results

Based on the estimated linear regression model and the results presented in **Error! Reference source not found.**, with a coefficient of determination (R^2) equal to 0.23, it can be stated that 23% of the variation in the dependent variable in question (logarithm of the cost overrun percentage over the initial budget) is explained by the independent or predictor variable, one dummy variable that is equal to 1 if the reactor was constructed after the TMI accident. The coefficients of the TMI dummy variable represent the differences in the percentage of cost overrun compared to the base category (omitted dummy) for reactors that were built after TMI.

Table III. Linear regression results.

Dependent Variable: natural logarithm of Cost Overruns Model 1				
	Coefficient	p-value	t-value	[95% confidence interval]
Intercept	5.30* (3.14)	0.09	1.69	[-0.96 to 11.55]
ln(Capacity-Mwe)	-0.09 (0.46)	0.84	-0.19	[-1.01 to 0.83]
Dummy (TMI)	0.80*** (.19)	0.00	4.26	[0.43 to 1.18]
R-squared	0.23			
Adjusted R-squared	0.21			
Number of observations	75			

Standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1

In the first place, given the obtained p-values, with the null hypothesis stating the non-significance of the parameters and the alternative hypothesis suggesting their significance, it was found that:

- The coefficient accompanying the variable "TMI" is significant in explaining the variation of the percentage of cost overruns. This means that if the reactor was built after TMI, the percentage of cost overruns increases by 80 percentage points.
- The coefficient accompanying the variable "Capacity-Mwe" is non-significant; with a low t-statistic observed, it is inferred that its standard deviation is high. With all other variables constant, it is expected that each increase of 1% in capacity increases the cost overruns by 0.09 percentage point.

The result is in line with [20] that found when all other variables considered in the analysis remained constant, larger plants typically exhibited lower costs per unit of capacity. However, the larger plants also had longer construction times compared to the smaller ones. Consequently, the increased costs from the additional construction time needed for the larger units outweighed the direct cost savings associated with economies of scale It is critical to note that the term, economies of scale, in this paper is not related to the economic idea of economies of scale meaning the more units produced will decrease costs because of the learning-by-doing concept, but this concept does not necessarily apply to bigger units. Instead of using economies of scale, the term, economies of size, should be more consistent; however, we prefer [20] definition where scale refers to more units produced and not to bigger size.

4. PREDICTING THE PERCENTAGE OF COST OVERRUNS

One characteristic of linear probability models is the former assumes constant marginal effects of explanatory variables. In a linear model, the partial derivative coincides with the coefficient associated with the explanatory variable in question. Additionally, linear models can be estimated by ordinary least square (OLS). Estimating the linear probability model by OLS does not guarantee that the estimated values of Yi are between 0 and 1; therefore, it cannot be given a probabilistic interpretation which yields values within said interval.

Given we need a probabilistic interpretation, a nonlinear model known as Probit and Logit is necessary. Those approaches use a nonlinear function to model the conditional probability function of a binary dependent variable. Specifically, this study applies a probit model that can be estimated by maximum likelihood since it contains desirable properties for large samples. Note that using a probit implies diminishing marginal effects of partial effects, a big difference in relation to OLS.

The disadvantage of using a probit is that, being nonlinear, the interpretation of the coefficients is not straightforward unlike in a linear model, and we need to do an additional step known as margins estimation to interpret the coefficient as in the OLS model [21]. On the other hand, it can be shown that this estimation does not imply efficiency of the estimators since we are dealing with heteroskedasticity of errors [22].

A probit model is estimated to analyze the impact on cost overruns exceeding 100%. The variable to be explained is Y (equal to 1 if it has cost overruns exceeding 100%), while the regressors are region (if the reactor was built in the North America, Europe, or Asia), TMI (if the reactor was built after the TMI accident), and the natural logarithm of capacity (nameplate capacity in MW electric):

$$P(Y=1|X) = \beta_0 + \beta_1 * \ln(Capacity) + \beta_2 * Dummy_{Asia} + \beta_3 * Dummy_{Europe} + \beta_4 * TMI + u$$

As can be observed in **Error! Reference source not found.**, all variables are statistically significant at 0.01% level when explaining percentage of cost overrun over 100%, except for the logarithm of capacity, which is not significant at any level.

Table IV. Probit model results.

Dependent Variable	Independent Variables	Model (1)
% Cost Overruns > 100% =1	CapacityMWe	-0.000
		(.0003)
	Asia	-1.92***
		(.2885)
	Europe	-3.44***
		(.4010)
	TMI	1.03***
		(0.3050)
	Observations	175

Standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1

In a binary probit-style model, it becomes relevant to calculate the marginal effects at different points, as the influence of explanatory variables on the dependent variable is not solely dependent on the coefficients but also on the specific values that these variables take. By increasing the level of capacity in MWe from its minimum value to its maximum value, ceteris paribus, a significant variation in the coefficients of all explanatory variables can be observed.

Given that we cannot interpret the previous coefficients values, we run estimations for specific values of capacity (500, 1000, and 1500 MWe) and obtained the partial derivative for each region. The results for the United States are presented in Figure 4.

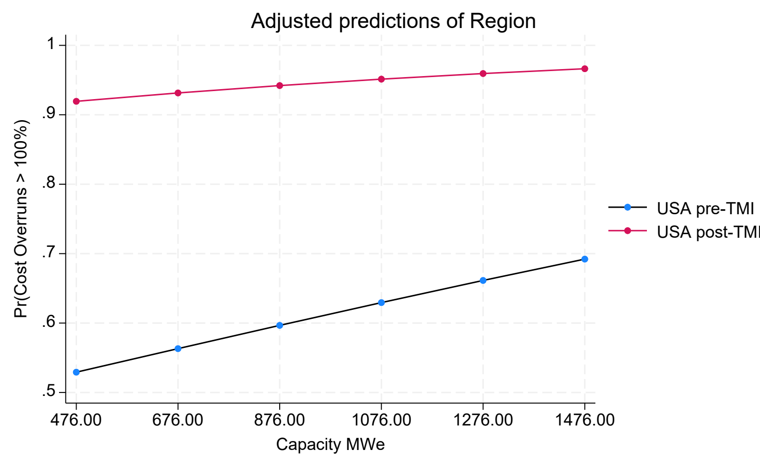

Figure 4. Probability of cost overruns > 100% predictive margins for the U.S. pre- and post-TMI.

Figure 4 shows that the probability of cost overruns increased significatively after the TMI accident, going from a minimum around of 53% to 92% when the capacity is around 500MWe.

In a similar way, as a point of comparison, the predictive margins were calculated using all the datasets for the three regions. The result in Figure 5 shows a higher probability of having more than 100% of cost overruns for the North American reactors in relation to European and Asian reactors. When the capacity is increased, the probability of cost overruns bigger than 100% start to get closer for the three regions.

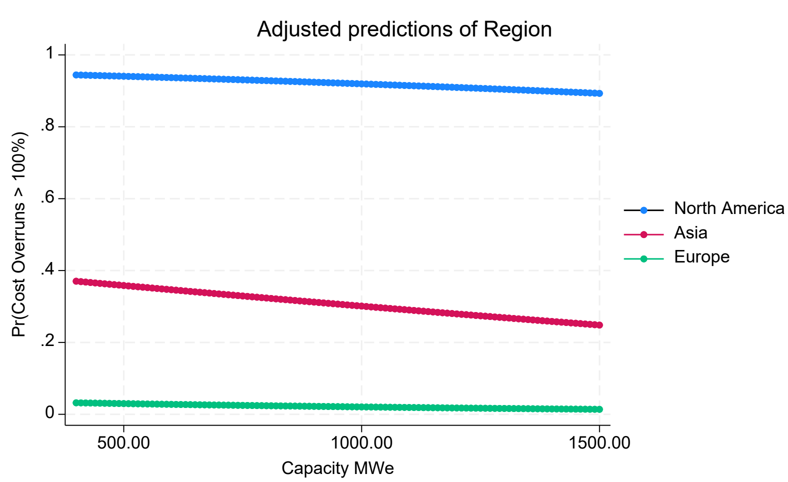

Figure 5. Probability of cost overruns > 100% predictive margins by region post-TMI.

5. CONCLUSION

Since 1970, nuclear reactors have been through continuous cost overruns of different magnitudes. After the TMI accident, the levels of cost overrun became higher. Delving into the dynamics of the cost overrun requires concentrated exploration. This paper's results show that it is important to trace out the whole distribution of cost overrun and use a mix of different approaches to understand the dynamics better. The cost overrun percentage was studied using parametric and nonparametric models. Dynamic panel is a parametric approach that permitted leveraging cross-sectional variations to compensate for limited continuous temporal dimensions. On the other side, the nonparametric approach allowed us to apply a data-driven model without imposing a functional form on the data.

The panel model shows that the effect of capacity on the cost overrun percentage is not statistically significant. Also, the North American region has a higher probability of having cost overruns at different levels of capacity, while Asia and Europe have lower ones.

Another strong result of this paper is that after the TMI, the distribution of cost overrun shifted to the right generating higher percents of cost overruns in North America than in other regions.

Numerous aspects of this research deserve deeper examination. The recurring patterns of cost overrun, a long-standing concern in technoeconomic and financial literature of nuclear reactors, need a carful investigation. However, the use of time series analysis or bigger data sets is challenging due to the scarcity of extensive and consistent data.

6. REFERENCES

[1] T. Roulstone, "The importance of know-how," *Nucl Eng Int,* no. https://www.neimagazine.com/features/featurethe-importance-of-know-how-7194172/., 2019.

[2] J. Koomey and N. Hultman, "A reactor-level analysis of busbar costs for us nuclear plants, 1970–2005," *Energy Policy,* vol. 35, p. 5630–5642, 2007.

[3] J. M. Marshall and P. Navarro, "Costs of nuclear power plant construction: theory and new evidence," *The RAND Corporation,* vol. 22, p. 148–154, 1991.

[4] P. Eash-Gates, M. M. Klemun, G. Kavlak, J. McNerney, J. Buongiorno and J. E. Trancik, "Sources of cost overrun in nuclear power plant construction call for a new approach to engineering design," *Joule,* vol. 4(11), pp. 2348-2373, 2020.

[5] S. Company, "Vogtle Cost Monitor: 29th Semi-Annual Report," 2023.

[6] G. Locatelli and M. Mancini, "Looking back to see the future: building nuclear power plants in Europe," *Construction Management and Economics,* vol. 30, p. 623–637, 2012.

[7] R. Horgan, "Hinkley Point C at risk of delay as cost soars by £2.9bn," *New Civil Engineer,* p. 3, 25 September 2019.

[8] J. Buongiorno, M. Corradini, J. Parsons and D. Petti, "The future of nuclear energy in a carbon-constrained world, an MIT interdisciplinary study," *MIT Energy Initiative,* no. http://energy.mit.edu/wp-content/uploads/ 2018/09/The-Future-of-Nuclear-E, 2018.

[9] The Economic Modeling Working Group Of the Generation IV International Forum, "Cost estimating guidelines for generation IV nuclear energy systems," Generation IV International Forum, https://www.gen-4.org/ gif/upload/docs/application/pdf/2013-09/ e, 2007.

[10] C. A. Lloyd, T. Roulstone and R. E. Lyons, "Transport, constructability, and economic advantages of SMR modularization," *Progress in Nuclear Energy,* vol. 134, no. 103672., 2021.

[11] Q. Li and J. Racine, "Nonparametric Econometrics: Theory and Practice," *Princeton University Press,* 2006.

[12] C. Komanoff, Power Plant Cost Escalation, New York: Van Nostrand Reinhold Company, 1981.

[13] I. Bupp and J. Derian, The Failed Promise of Nuclear Power: The Story of Light Water, New York: Basic Books, 1981.

[14] N. E. Hultman and J. G. Koomey, "Three Mile Island: the driver of US nuclear power's decline?," *Sage Journals,* vol. 69, no. (3), p. 63–70, 2013.

[15] B. K. Sovacool, A. Gilbert and D. Nugent, "Risk, innovation, electricity infrastructure and construction cost overruns: Testing six hypotheses," *Energy,* vol. 74, pp. 906-917, 2014.

[16] W. Mooz, "Cost Analysis of Light Water Reactor Power Plants," Santa Monica: The RAND Corporation, 1978.

[17] S. Paik and W. Schriver, "The Effect of Increased Regulation on Capital Costs and Manual Labor Requirements of Nuclear Power," *The Engineering Economist,* vol. 26, pp. 223-244, 1981.

[18] M. Zimmerman, "Learning Effects and the Commercialization of New Energy Technologies," *Bell Journal of Economics,* vol. 13, pp. 297-310, 1982.

[19] R. Cantor and J. Hewlett, "The Economics of Nuclear Power: Further Evidence on Learning, Economies of Scale, and Cost Estimation," 1987.

[20] Energy Information Administration, "Analysis of nuclear power plant construction costs," no. https://doi.org/10.2172/6071600, 1986.

[21] R. Williams, "Using the margins command to estimate and interpret adjusted predictions and marginal effects," *Stata Journal 12,* p. 308–331, 2012.

[22] H. White, "A Heteroskedasticity-consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity," *Econometrica,* vol. 48, pp. 817-838, 1980.