Optimal Preventive Maintenance: Balancing Reliability and Costs in Power Systems

Seon-Yong Kim[1], M. Choi[2], Dae-Wook Kim[3] and Man-Keun Kim[4*]

[1] Department of Economics, University of Missouri, Columbia, MO, USA. skxbb@umsystem.edu
[2] KEPCO
[3] Department of Economics, Soongsil University, Seoul, Korea. daekim@ssu.ac.kr
[4*] *Corresponding author*, Department of Applied Economics, Utah State University, Logan UT, USA.
mk.kim@usu.edu and 1-435-797-2359

Abstract

This study addresses the challenge of achieving optimal preventive maintenance within power systems, aiming to strike a harmonious balance between reliability and costs. The primary focus is on unraveling the intricate relationship between preventive maintenance expenditures and the failure rate of essential transmission components, with a specific emphasis on transformers in substations. We contribute to the existing literature by offering two key insights. Firstly, we establish a theoretical foundation for determining the optimal level of preventive maintenance, which can be extended across various electricity facilities. By scrutinizing the relationship between preventive maintenance costs and failure rates, our goal is to identify the investment level that guarantees a dependable power system while minimizing financial burdens. Secondly, we explore the functional form that accurately characterizes the link between preventive maintenance costs and failure rates. The diminishing marginal rate of failure, contingent on various functional forms, highlights the variability of optimal preventive maintenance expenses. Utilizing Lasso regression, we identify the functional form that optimally characterizes the relationship between outage occurrences and maintenance costs. These insights extend well beyond the confines of the Korean power industry, offering a sustainable approach to managing transmission facilities and enhancing the overall stability and efficiency of global electricity networks.

Key words: Preventive maintenance, Substations, Survival analysis

JEL codes: C41, D24, Q40

1. Introduction

A stable and reliable power supply plays a pivotal role in sustaining economic activities. It is widely recognized as a fundamental element for ensuring energy security and fostering economic growth (Ateba et al., 2019; Neelawela et al., 2019). The significance of a stable power supply is particularly important for the Republic of Korea (hereafter referred to as Korea) due to the unique characteristic of its electric grid being isolated (see Figure 1). Unlike some countries with cross-border power grid interconnections, Korea's geographical location on the Korean Peninsula in Northeast Asia places it between China (to the west with the Yellow Sea) and Japan (to the southeast with East Sea and Korea Strait), with North Korea located to its north (Kim et al., 2020). This geopolitical configuration results in the absence of interconnections with neighboring countries, making the maintenance of a self-reliant and robust power supply infrastructure of paramount importance to meet the nation's energy demands and ensure energy security (Kim et al., 2020).

Achieving a stable power system heavily relies on the reliability of generators, transmission, and distribution facilities (U.S. Department of Energy, 2016). However, ensuring such reliability demands substantial investments in preventive maintenance costs, which may have adverse effects on the economic performance of power companies and efficient allocation of resources (Espiritu et al., 2007). For example, the Korea Electric Power Corporation (KEPCO) exemplifies this scenario, making significant investments in preventive maintenance costs to uphold a certain level of reliability. A study by Kim et al. (2020) emphasized that power plants in Korea might be incurring unnecessarily high preventive maintenance expenses, even though they already achieve the world's lowest forced outage rates for their generators.

This study aims to explore the intricate relationship between preventive maintenance costs and the failure rate of transmission facilities, with a particular emphasis on transformers in substations[1]. Substation equipment, utilized for extended periods, undergoes gradual deterioration due to aging and continuous operation, eventually impacting the overall system reliability. The significance of equipment deterioration is particularly pronounced in substations, being the pivotal

[1]A transformer is an electrical device that transfers electrical energy from one electrical circuit to another circuit or multiple circuits. A substation is a vital part of an electrical generation, transmission, and distribution system. Substations are responsible for transforming voltage from high to low or vice versa, and they serve various other essential functions. Electric power may flow through several substations at different voltage levels between the generating station and the end consumers (Atwa, 2019). A substation may include transformers used to change voltage levels between high transmission voltages and lower distribution voltages or at the interconnection of two different transmission voltages (Atwa, 2019).

junction point between generation, transmission, and distribution systems. Therefore, maintenance is imperative to prolong the availability of substation equipment (Sudket and Chaitusaney, 2014). By employing a Cox proportional hazard rate model (Cox, 1972) and considering various functional forms ranging from linear to cubic, we thoroughly investigate the association between maintenance costs and the failure of transformers. Our objective is to determine the optimal level of preventive maintenance required to uphold a reliable substation system while taking into account the trade-offs between the economic implications of maintenance investments and the reliability of the transmission and distribution infrastructure.

This study makes three significant contributions to the existing literature. Firstly, we establish a theoretical framework for determining the optimal level of preventive maintenance for substations, with the potential for extension to various types of electricity facilities. By investigating the correlation between preventive maintenance costs and substation failure rates, our goal is to pinpoint the most effective investment level that guarantees a dependable power system while minimizing financial burdens. Secondly, we delve into identifying the appropriate functional form that accurately reflects the relationship between preventive maintenance costs and failure rates. The diminishing marginal rate of failure can differ based on various functional forms, leading to variable optimal preventive maintenance costs. Therefore, comprehending the underlying nature of this relationship is essential for estimating the most suitable preventive maintenance costs to ensure the reliability of transmission and distribution facilities. Our research casts light on these vital aspects, offering valuable insights into power system maintenance and reliability. To determine a more fitting functional form that explains the connection between outage occurrences and maintenance costs, we employed least absolute shrinkage and selection operator (LASSO) regression (Tibshirani, 1997). This technique helps us discern the model that best elucidates the interplay between these factors, addressing concerns such as variable selection and overfitting. These findings hold significance not only for the Korean power industry but also for power systems globally, facilitating the sustainable management of transmission facilities and contributing to the overall stability and efficiency of electricity networks. Lastly, our study is instrumental in discussing the optimal level of preventive maintenance spending and evaluating whether allocation towards preventive maintenance costs is appropriate or excessive. These analyses demonstrate the practical application of the theoretical framework developed, emphasizing its potential relevance for diverse countries, and represent another contribution of this paper.

This paper is organized into seven sections. In Section 2, we present an overview of the industry background, focusing on the power system in Korea to provide readers with the necessary context. In Section 3, we introduce the conceptual model aimed at maximizing the net benefit from investments in preventive maintenance spending on transmission facilities. The methodology used to estimate the hazard model based on the data discussed in Section 4.3 is presented in Section 4. We describe the empirical model, which builds upon the conceptual framework introduced in Sections 3 and 4. This model allows us to estimate the optimal level of maintenance costs for transmission facilities in Korea and assess whether the current spending is unnecessarily high. Section 5 presents the results of our empirical analysis, offering valuable insights into the optimal preventive maintenance level for transmission facilities in Korea. Furthermore, we discuss the policy implications of our findings and provide limitations of the study in Section 6.

2. Industry Background

As per the Electric Utility Act of Korea (`https://elaw.klri.re.kr/`), the electricity industry is categorized into distinct sectors, including power generation, transmission, distribution, electricity sales, and district electricity businesses. The transmission business is specifically defined as the operation primarily focused on the installation and management of electrical facilities essential for transmitting electricity generated in power plants to distribution operators. Korea Power Exchange (KPX), which operates under the umbrella of the Ministry of Trade, Industry and Energy (MOTIE), is the sole transmission system operator (TSO) for electricity supply in Korea (International Energy Agency, 2023). Korea Electric Power Corporation (KEPCO), also majority-owned by the government, is responsible for transmission facility management as the asset owner. Figure 1 presents Korea's electric infrastructure in 2022, which was presented in International Energy Agency (2023).

As of the end of 2020, the Korea Electric Power Corporation (KEPCO) operated a total of 2,868 power transformers in 877 substations (Electric Power Statistics Information System). In the context of large-scale transmission substations responsible for transmitting power from power plants to high-voltage transmission lines (765kV or 345kV), voltage conversion is a critical step. This process involves the use of a 765kV/345kV transformer to lower the voltage and a subsequent

Figure 1: Electric infrastructure in Korea

Note: Image is taken from International Energy Agency (2023)

345kV/154kV transformer to further adjust it. Near the demand points, a distribution transformer (154kV/22.9kV) is employed to decrease the voltage even further, allowing for effective distribution to consumers. To supply power at the desired levels (380V or 220V), small-scale transformers installed on utility poles in close proximity to consumers play a vital role. These power transformers installed within substations specifically refer to the 765kV/345kV, 345kV/154kV, and 154kV/22.9kV transformer configurations.

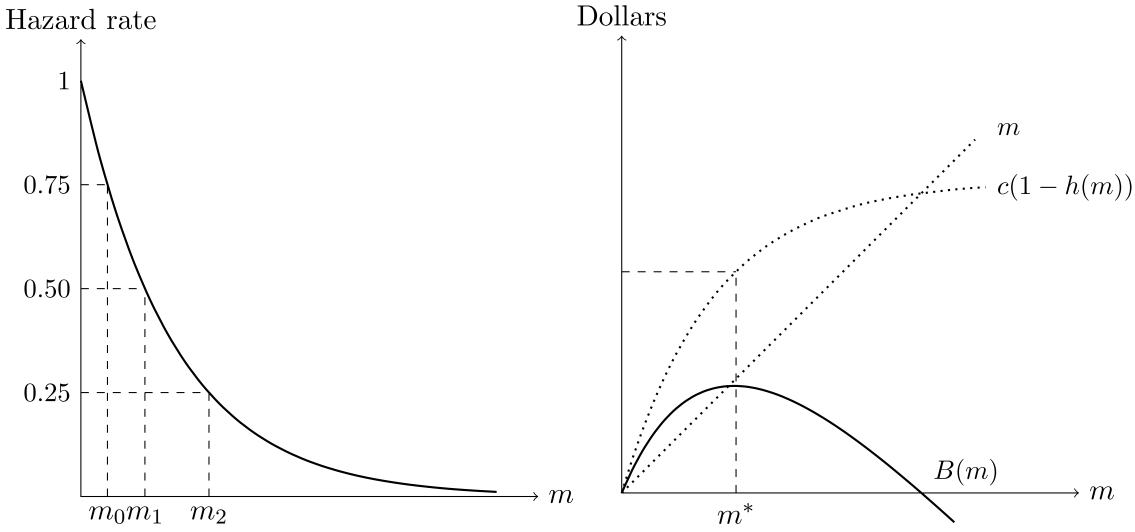
Transmission and substation facilities consist of numerous devices, including overhead transmission lines, underground transmission cables, transformers, circuit breakers, disconnectors, lightning arresters, and more (Atwa, 2019). Each device possesses unique characteristics, and the costs associated with individual facilities are not separately managed. This makes it impractical to conduct comprehensive analyses for each individual facility. Therefore, there is a need to limit the scope of facilities and review them in order to measure the degree of quality improvement in relation to the costs invested in transmission and substation facilities. In this study, to enable efficient analysis, the power transformer, considered the most crucial facility within transmission and substation systems, was selected as the primary subject of analysis. The survival analysis method was adopted to examine the factors influencing the failure rate. Failure in transmission facilities is predominantly caused by natural disasters (24 out of 54 cases in 2021), making it challenging to limit the targets of failure due to their wide distribution. Additionally, the allocation of costs becomes difficult. Taking these factors into consideration, transmission facilities were excluded from the analysis, and power transformers, which are situated within confined spaces called substations and allow for easy cost distribution, were chosen as the subject of analysis.

3. Conceptual Model

Suppose that investing in preventive maintenance leads to a reduction in the hazard rate (failure rate) of transformers in substations. The hazard rate is denoted by the function $h(m)$, where m represents the level of preventive maintenance spending. We assume that $h'(m) < 0$ and $h''(m) > 0$, indicating that as the level of preventive maintenance increases, the hazard rate of transformers in a substation decreases. However, with each incremental increase in preventive maintenance, the marginal reduction in the hazard rate diminishes. In other words, as the level of maintenance spending increases, the additional decrease in the hazard rate achieved by investing more in preventive maintenance becomes less significant. Panel A in Figure 2 represents the function $h(m)$ illustrating the relationship between preventive maintenance spending and the hazard rate. When $m = 0$, $h(m) = 1$, implying that a transformer is certain to fail. In contrast, when $m = m_2$, the transformers only has a 25% chance of failure. As m approaches a very large value, the failure probability converges to zero, i.e., $\lim_{m\to\infty} h(m) = 0$.

Suppose that the outage loss (explained below) remains constant at \$$c$, which implies that the expected gain from preventive maintenance with a cost of m will be given by the following expression:

$$B(m) = \underbrace{c(1 - h(m))}_{\text{avoided outage loss}} - m \tag{1}$$



Panel A. Hazard rate of substation Panel B. Gain from prevented maintenance

Figure 2: Preventive Maintenance Spending and Benefits

where $B(m)$ represents the net benefit from reducing the hazard rate of the substation by investing m in preventive maintenance costs. Since $h(m)$ in equation (1) denotes the hazard rate, $1 - h(m)$ represents the survival probability or *reliability* of the substation, that is, the probability of avoiding outages. Therefore, $c(1 - h(m))$ corresponds to the (expected) avoided outage loss in monetary terms. First order condition to maximize $B(m)$ is

$$B'(m) = -ch'(m) - 1 = 0 \quad \rightarrow \quad h'(m) = -\frac{1}{c} \tag{2}$$

Simply put, the optimal preventive maintenance spending is determined when the marginal reduction in hazard rate is equal to the inverse of the outage loss. Moving to Panel B of Figure 2, this concept is visually depicted through the interplay of avoided outage cost, $c(1-h(m))$, maintenance cost, m, and their differences, $B(m)$ as in equation (1). The optimal maintenance cost is denoted as m^*, which equates to $h'(m) = -\frac{1}{c}$ (Panel B in Figure 2).

Outage loss in electricity transmission refers to the costs and negative impacts resulting from disruptions or failures in the power supply, leading to temporary electricity loss for consumers and critical infrastructure (Centolella, 2010). The literature on electricity outage costs is extensive, with various studies conducted in different contexts. Woo et al. (2021b) specifically addresses outage costs in residential areas, while Woo et al. (2021a) provides the estimation of outage costs in non-residential sectors. Both studies include a comprehensive literature review. It's worth noting that the present study, while not centered around estimating outage costs directly, will leverage a simplified approach, elaborated further in the subsequent empirical sections, to approximate outage costs based on previous research for example Woo et al. (2021b) for residential sector. These estimates will be instrumental in determining outage cost, c in equation (1), and the optimal preventive maintenance spending in the empirical sections of this research.

4. Empirical Strategies and Data

This section outlines the empirical methodologies employed for estimating the hazard function, denoted as $h(m)$ in equation (1), and for computing outage costs. These estimations are instrumental in discussing the optimal level of preventive maintenance spending and evaluating whether KEPCO's allocation towards preventive maintenance costs is appropriate or excessive. These analyses demonstrate the practical application of the theoretical framework developed earlier, emphasizing its potential relevance for diverse countries. This practical application represents the primary contribution of this paper.

4.1. Hazard Function

The core objective of this study is to explore the intricate connection between maintenance costs and the incidence of failures in power transformers through the lens of survival data analysis. Survival data analysis proves highly apt for scrutinizing time-to-event data, which pertains to the duration leading up to an event, while accounting for multiple covariates that might impact the event's likelihood. In survival data analysis, the dependent variable reflects the time until the event occurs, often measured in discrete and non-negative terms (Klienbaum and Klein, 2012).

Conventional regression analysis, predicated on a continuous dependent variable following a normal distribution, proves unsuitable for survival data analysis due to the discrete and non-negative attributes inherent to the dependent variable. Consequently, to effectively model and estimate the intricate link between maintenance costs and the probability of failure cessation in power transformers, we turn to survival data analysis techniques. Notably, we employ methods like the Kaplan-Meier estimator (Kaplan and Meier, 1958) and the Cox proportional hazard model (Cox, 1972). Through the lens of survival data analysis, we can discern how maintenance costs influence the survival probabilities of transformers. This approach offers a nuanced perspective on the optimal level of preventive maintenance and its economic ramifications in ensuring a dependable and efficient power supply system.

Incorporating survival analysis into this study entails addressing the concept of right-censored data, a fundamental consideration in the field (Klienbaum and Klein, 2012). Right-censored data pertains to observations that do not experience the event of interest within the designated analysis period (Klienbaum and Klein, 2012). While we possess knowledge that these observations endured throughout the study's duration, the exact survival time eludes our grasp. In instances like these, the utilization of specialized methodologies within survival data analysis becomes imperative to accurately estimate and validate models during analysis. It is noteworthy that the dataset employed in our study exclusively centers on transformers that encountered failures, thereby excluding any instances of right-censored observations. However, given the dataset's structure, where the survival time functions as the dependent variable, as previously elucidated, it becomes a natural progression to apply survival analysis models. Thus, the conduct of this analysis hinged on the adoption of survival analysis models to navigate the research landscape effectively.

As previously mentioned, survival analysis is a methodology utilized for the examination of survival data. This field commonly employs two approaches: the nonparametric Kaplan-Meier estimator and the semi-parametric Cox proportional hazard model (Klienbaum and Klein, 2012; Cleves et al., 2016). The Kaplan-Meier estimator is a nonparametric technique that avoids the need for distributional assumptions, making it particularly useful for estimating survival probabilities when continuous variables are not essential. However, its utility in elucidating survival probabilities based on continuous variables, aside from categorical variables, is limited. In contrast, the Cox proportional hazard model, a semi-parametric approach, assumes the proportional hazards principle, implying that the hazard ratio remains consistent over time. While the model requires the assumption of proportional hazards, it avoids distributional assumptions and is capable of integrating both categorical and continuous variables (Klienbaum and Klein, 2012; Cleves et al., 2016). Consequently, this study employed survival analysis techniques, encompassing the Kaplan-Meier estimator and the Cox proportional hazard model, to facilitate its analysis.

Upon examining the current status of the transmission and substation operation, it is evident that there were no significant structural changes during the analysis period. Therefore, it is reasonable to assume that the effects of the variables used in the analysis on failure mitigation or cessation remain constant throughout the analysis period, making the use of the Cox proportional hazard model appropriate. Consequently, this study aims to investigate the impact of maintenance costs on transformer failure cessation by complementary utilizing the Kaplan-Meier estimator and the Cox proportional hazard model. The Kaplan-Meier analysis will examine various variables that may influence failure cessation through graphical representations. Based on this analysis, a model will be constructed to estimate the survival probability of transformers. The Cox proportional hazard model will then be employed as a foundation to estimate this model.

Suppose that the hazard function represents the instantaneous probability of transformers in a substation experiencing failure at time t, given that it has survived without failure until the current time. Suppose that failures of a transformer are random events drawn from a particular probability distribution function (PDF), $f(t)$, where t is a duration of the transformer. The cumulative distribution function (CDF), $F(t) = \Pr(T \leq t)$, gives the probability that the transformer failure has occurred by duration t. Note that, generally, the probability of failure will increase over time.

It is convenient to work with the survival function such as

$$S(t) = \Pr(T \geq t) = 1 - F(t) = \int_t^\infty f(u)du \tag{3}$$

which gives the probability that the transformer failure has not occurred by duration t or the transformer is still running at time t or, equivalently the probability of failing after t (Cleves et al., 2016). If $S(1200) = 0.75$, it means that there are 75% of the transformers that are running at $t = 1200$ (days), for example. Basically, $S(0) = 1$ (no transfer has failured), $S(t)$ is a non-increasing function, and $S(\infty) = 0$ (eventually all the transformers fail). For a dataset with observed failure times, $t_1, \cdots, t_k$, where k is the number of distinct failure times observed in the data, the Kaplan-Meier estimate at any time t is give by

$$S(t) = \prod_{j|t_j \leq t} \left(\frac{n_j - d_j}{n_j} \right) \tag{4}$$

where n_j is the number of individuals at risk at time t_j and d_j is the number of failures at time t_j. The product is over all observed failure times less than or equal to t (Cleves et al., 2016).

In the Cox proportional hazard (PH) model (Cox, 1972), an alternative characterization of the distribution in equation (3) is given by the hazard function, $h(t)$, defined as

$$h(t) = \lim_{\Delta t \to 0} \frac{\Pr\left(t < T < t + \Delta t | T > t\right)}{\Delta t} = \frac{f(t)}{S(t)} \tag{5}$$

In words, the hazard function gives the instantaneous potential per unit time for the failure occur given that the individual transformer has survived up to time t. Note that, as indicated in Klienbaum and Klein (2012), in contrast to the survival function in equation (3), which focuses on not failing, the hazard function in equation (5) focuses on failing, i.e., on the forced outage occurring. Equation (5) can be modified to $S(t) = \frac{f(t)}{h(t)}$ and thus specifying one of the three functions specifies the other two functions. The hazard for transformer i at time t can be estimated by the Cox model using the following form as discussed in Cleves et al. (2016):

$$h(t \mid \mathbf{x}) = h_0(t) \exp\left(\mathbf{x}'\beta\right) \tag{6}$$

In this equation, the term $h_0(t)$ is an unspecified baseline hazard and $\exp\left(\mathbf{x}'\beta\right)$ represents the linear form of the model, similar to the estimated coefficients and covariates seen in conventional regression analysis. The coefficient β in equation (6) represents the relative hazard ratio. For example, if β is estimated as 0.1, a one-unit change in x would result in a 10.5% increase in the hazard rate, that is, $\exp(0.1) - 1 = 0.105$. Conversely, if β is estimated as -0.1, the hazard rate would decrease by 9.5%, that is, $\exp(-0.1) - 1 = -0.095$. The regression coefficients are estimated by maximizing the log partial likelihood (Tibshirani, 1996; Cleves et al., 2016):

$$L(\beta) = \sum_{i=1}^{n} \left(\mathbf{x}'_{j(i)}\beta - \log\left(\sum_{j \in R_i} \exp\left(\mathbf{x}'_j \beta\right) \right) \right) \tag{7}$$

where $\mathbf{x}_i$ is the vector of covariates for i transformer, R_i is the set of indices, j, with $t_j \geq \tau_i$ which at risk at time τ_i. The index $j(i)$ denote the transformers which failed at time τ_i.

The covariates included in the model are maintenance costs (in million won), substation type (outdoor or indoor), substation voltage (345kV/154kV), and transformer age. Except for maintenance costs and transformer age, the remaining variables are categorical. To examine the potential nonlinearity on failure cessation, various forms of maintenance costs were incorporated into the model such as quadratic terms. The determination of the best functional form for the Cox model depends on the relationship between the variables being examined. The final choice of functional form should be guided by statistical significance, goodness of fit measures such as pseudo R^2, and the theoretical relevance of the relationship.

Tibshirani (1997) introduce the application of the Lasso (Least Absolute Shrinkage and Selection Operator) method in the context of the Cox proportional hazard model. The study addresses the challenge of selecting relevant variables for inclusion in the Cox model while also dealing with the problem of multicollinearity and overfitting. When including a regularization term in equation (7), the resulting Lasso-Cox model (Tibshirani, 1997) can be represented as:

$$L(\beta) = \sum_{i=1}^{n} \left(\mathbf{x}'_{j(i)}\beta - \log \left(\sum_{j \in R_i} \exp\left(\mathbf{x}'_j \beta\right) \right) \right) - \lambda \sum_{k=1}^{p} |\beta_k| \tag{8}$$

where $\lambda > 0$ is a control parameter. It is particularly useful when dealing with datasets where there are many potential predictor variables, and you want to identify the most influential factors affecting time-to-event outcomes (Tibshirani, 1997).

4.2. Back of Envelope Estimate of Outage Cost

Numerous studies have been conducted to estimate the outage cost in the industrial and residential sectors. These studies can be categorized into two main groups based on the data employed for estimating the economic impact of electricity outages. The first category involves estimating economic losses using publicly available data. The second category employs survey data to estimate the willingness to pay (WTP) for avoiding an outage, as detailed in the economic theory behind WTP approaches, which can be found in Gorman (2022). The comprehensive list of previous research on outage cost estimation is provided and discussed in Kim and Cho (2017), Woo et al. (2021a), and Woo et al. (2021b).

Estimates for residential outage costs vary significantly, ranging from under US$0.5 per kWh unserved, based on the residential marginal electricity rate, to over US$10 per kWh unserved, obtained through customer survey data using the contingent valuation (CV) method (Woo et al., 2021b). Given the wide range of residential outage cost estimates and potential biases in CV survey data, Woo et al. (2021b) proposes a market-based estimation of a household's outage cost per kWh unserved, yielding an estimated range from US$0.12 to US$0.34 per kWh unserved. These estimates will be instrumental in determining outage cost, c in equation (1), and the optimal preventive maintenance spending in the empirical sections of this research.

4.3. Data

The dataset employed in this analysis was compiled from the failure records of all substations managed by the Korea Electric Power Corporation (KEPCO) from 2008 to 2018. As the dataset contains internal information, it is not publicly accessible. However, anonymized data can be made available upon request. These failure records offer valuable information about the substations in which transformer failures took place, including their associated power districts. Furthermore, the dataset encompasses pertinent data about the age of the transformers that experienced failures.

In determining the maintenance costs for substations, the approach involves calculations at the power district level, after which the costs are allocated among the substations within each district. Consequently, the maintenance cost attributed to each substation is obtained by dividing the annual maintenance cost allocated to the corresponding power district by the total number of substations within that district. Moreover, the analysis takes into consideration the non-failure days, which denote the duration in days from the previous failure of a transformer to the current failure event. For the initial failure, the analysis period commences on January 1, 2008, as the starting point of observation. These components of the dataset form the foundation for exploring and understanding the relationship between maintenance costs and the failure rate of transformers in substations.

Based on the data used in the analysis, we can summarize the key statistics of non-failure days, maintenance costs, and the age of failed transformers by substation type and voltage, as presented in Table 1. When examining non-failure days by substation type, it is evident that indoor substations have, on average, approximately 406 days longer non-failure periods compared to outdoor substations. This observation suggests that indoor substations are less susceptible to external factors such as weather conditions, which can contribute to a more reliable operation. Furthermore, in terms of voltage, high-voltage substations tend to experience more frequent failures compared to low-voltage transformers. As Table 1 indicates, the average non-failure period of low-

Table 1: Descriptive Statistics by Substation Type

Variable	Mean	Std. Dev.	Min	Max	Obs.
	Entire sample				
Non-failure period (days)	1,276	1,102	1	3,821	274
Maintenance cost (million KRW/year)	86.6	40.6	4.1	260.9	274
	Indoor substations				
Non-failure period (days)	1,499	1,165	4	3,732	123
Maintenance cost (million KRW/year)	88.7	41.0	4.1	260.9	123
	Outdoor substations				
Non-failure period (days)	1,094	1,016	1	3,821	151
Maintenance cost (million KRW/year)	84.9	40.4	30.1	230.6	151
	154 kV				
Non-failure period (days)	1,373	1,116	4	3,821	200
Maintenance cost (million KRW/year)	86.2	40.8	4.5	260.9	200
	345 kV				
Non-failure period (days)	1,013	1,025	1	3,633	74
Maintenance cost (million KRW/year)	87.8	40.4	4.1	230.6	74

1. KEPCO internal data,
2. Substations with a voltage of 765kV depicted in Figure 1 as well as underground substations were excluded from the analysis due to the limited number of available observations,
3. The average exchange rate in 2018 was 1 million KRW ≈ US$900. Consequently, 86.6 million KRW is approximately US$77,940.

voltage (154kV) substations is 359 days longer than that of high-voltage substations. However, it is important to note that among the total of 123 indoor substations, only 9 of them are high-voltage (345kV) substations, with the majority being outdoor substations. Considering this imbalance, it becomes necessary to investigate the influence of voltage on failure cessation by incorporating interaction terms of the two dummy variables during model estimation.

Upon analyzing the maintenance costs of substations based on their voltage levels, it becomes evident that the overall average maintenance cost (86.6 million KRW, approximately US$77,940) remains consistent, regardless of the voltage. However, when considering the substation type, indoor substations show slightly higher maintenance costs, amounting to approximately 3.8 million KRW (approximately US$3,420) more than outdoor substations. Nevertheless, this difference is relatively small, indicating that the maintenance costs for substations tend to be similar, regardless of their type or voltage level.

5. Results and Discussions

5.1. Hazard Function

Drawing upon the data analyzed earlier, two estimation methods, nonparametric and semiparametric approaches, were employed to explore the connection between substation failures, maintenance costs, and other relevant covariates. To begin with, the nonparametric Kaplan-Meier estimator (Kaplan and Meier, 1958) was utilized to examine the survival probabilities, as depicted in Figure 3. As shown in Panel A of Figure 3, we examined the survival probabilities based on the type of substations. Outdoor substations had a survival probability of approximately 15% at 2,000 days, whereas indoor substations exhibited a notably higher survival probability of around 35%, representing an approximate 20% difference between the two types. Panel B in Figure 3 presents the survival probability by voltage type, 154 kV vs. 345 kV. In general, 154 kV substations have a slightly higher survival rate, but this difference is not significant.

Moving on to Panel C in Figure 3, we explored the survival probabilities based on maintenance costs. Substations with higher maintenance expenses displayed a survival probability of approximately 39% at 2000 days, while substations with lower maintenance costs exhibited a survival probability of about 19%. Note that maintenance costs greater than the mean value of 86.6 million KRW (≈ US$77,940) (Table 1) were classified as *High*, while costs equal to or below the mean were classified as *Low*. It is also worth noting that the disparity in survival probabilities between these two groups varied throughout the non-failure period, making it challenging to draw definitive conclusions.

Moreover, to gain a more comprehensive understanding, we transformed the continuous variables, such as maintenance costs and transformer age, into categorical variables. However, analyzing these variables as continuous variables would provide more valuable insights, allowing us to examine the impact of a one-unit increase in these continuous variables on the survival probability. Therefore, in order to investigate the influence of maintenance costs on substation failures, we intend to employ the PH model, considering the continuous nature of these variables, to conduct a more robust analysis.

The estimated results of the PH model, which includes maintenance costs and other covariates affecting substation failures as described earlier, are presented in Table 2. The models can be categorized into two groups: Model 1, which doesn't consider power district fixed effects, and

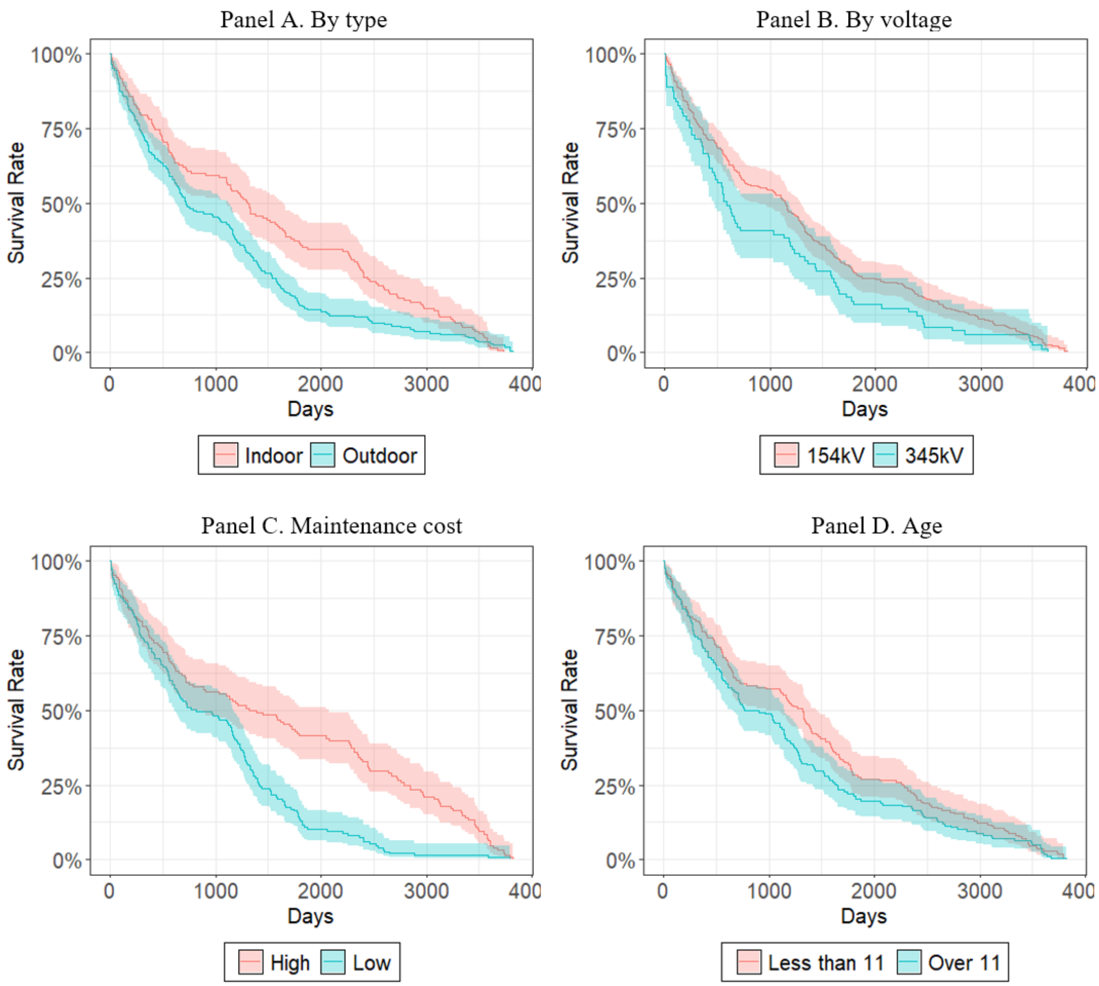


Figure 3: Survival probabilities

Note: Regarding maintenance costs, values greater than the median value of 86.6 million KRW ($\approx$ US\$77,940) were classified as *High*, while values equal to or below the median were classified as *Low*. As for transformer age, it was categorized into two groups using the median age of 11 years.

Models 2 and 4, which incorporate power district fixed effects. Model 4 is the most robust model employing the Cox Lasso technique discussed in equation (8). The estimation outcomes are consistently similar across all models. Focusing on the estimated coefficients of categorical variables related to substation characteristics, it becomes evident that the coefficient for `indoor:154kV`, representing low-voltage indoor substations, is consistently estimated to be approximately -0.6 with a significance level of 1% across all models. This implies that the hazard rate of indoor substations is roughly 45% lower compared to high-voltage outdoor transformers. A similar pattern emerges for `outdoor:154kV` substations, where the coefficient is approximately -0.4 and statistically significant, indicating an around 33% lower hazard rate than high-voltage outdoor transformers. Notably,

Table 2: Hazard Model Estimates[a, b]

	Model 1	Model 2	Model 3	Model 4
cost	-0.009^{***}	-0.016^{***}	-0.034^{***}	-0.025^{***}
	(0.002)	(0.003)	(0.008)	(0.005)
cost^2			0.0001^{***}	
			(0.000)	
cost^3				0.0000002^{**}
				(0.0000001)
age	-0.022^{**}	-0.023^{**}	-0.022^{**}	-0.022^{**}
	(0.010)	(0.010)	(0.010)	(0.010)
indoor:154kV	-0.548^{***}	-0.609^{***}	-0.599^{***}	-0.597^{***}
	(0.159)	(0.198)	(0.199)	(0.199)
outdoor:154kV	-0.474^{***}	-0.421^{**}	-0.408^{*}	-0.405^{*}
	(0.175)	(0.211)	(0.212)	(0.212)
indoor:345kV	-0.401	0.405	0.314	0.366
	(0.358)	(0.531)	(0.533)	(0.531)
(outdoor:345kV)[c]				
Fixed effect	No	Yes	Yes	Yes
Observations	261	261	261	261
Pseudo R^2	0.144	0.379	0.394	0.391
Log lik.	-1,174.7	-1,133.0	-1,129.8	-1,130.2

[a] Numbers in parentheses are standard errors; *, **, and *** indicate the significance at 10%, 5% and 1%, respectively.
[b] The PH model estimates the equation $h(t) = h_0 \exp(\mathbf{x}_i'\boldsymbol{\beta})$, where $h(t)$ is the forced outage at time t, $\mathbf{x}_i'$ is a vector of covariates, and $\boldsymbol{\beta}$ is a vector of regression coefficients. A negative coefficient implies a longer expected duration.
[c] Reference type of substation is presented in parentheses.

the coefficient for high-voltage indoor transformers, `indoor:345kV`, are not statistically significant across all models.

The coefficients estimated for the `age` variable are close to the value of -0.2, indicating that for each additional year of transformer age, there is an associated reduction of roughly 2% in the hazard rate. This finding may appear counterintuitive. Commonly, one would anticipate a positive coefficient for the `age` variable since machinery typically becomes more susceptible to failures over time due to wear and tear, resulting in an increased likelihood of failure. A plausible explanation for this can be found in the bathtub curve theory (Klutkey et al., 2003). According to this theory, during the initial phases of machinery assembly and installation, the occurrence of failures is relatively high due to trial and error. As time progresses, the equipment stabilizes, resulting in a decline in failures. However, after a certain duration, wear and tear begin to take their toll, causing failures to rise due to component replacements and other factors. Examining our data, it's apparent that transformers with an age of up to 10 years constitute approximately 54% of the entire dataset. Even within a narrower 5-year span, they still make up around 35% of the total. This distribution suggests that a considerable portion of the dataset involves relatively recently installed transformers. Consequently, the observed negative sign in the estimated `age` coefficient can be attributed to this context, where recent installations dominate the data.

When examining the estimated coefficients for maintenance costs, their values vary based on the chosen functional form, but they are statistically significant across all models. In models assuming a linear functional form (Model 1 and Model 2 in Table 2), the coefficient for `cost` was estimated as -0.009 in the case of the model without considering fixed effects (Model 1) and -0.016 when fixed effects were incorporated (Model 2). Including fixed effects, which account for regional variations in outage occurrences (such as higher frequencies of outages in coastal areas compared to urban areas), is considered more appropriate, as it aligns better with the data and enhances the model's accuracy, as indicated by the Pseudo R^2 values. However, determining the definitive model to explain the relationship remains challenging. Nevertheless, Model 3 appears to be a promising candidate in terms of Pseudo R^2. To further refine the functional form that better explains the relationship between substation failure and maintenance costs, we explored various functional forms using Lasso regression (Tibshirani, 1997), as discussed in equation (8). Model 4, as presented in Table 2, represents the results of this Lasso regression analysis.

The estimation results clearly indicate that the impact of increasing maintenance costs diminishes, that is, diminishing marginal benefits of reducing hazard rate. This relationship is visualized in Figure 4, which presents the predicted hazard rates and their corresponding maintenance costs based on Models 2, 3, and 4 as outlined in Table 2. With an average maintenance spending of 86.6 KRW, the hazard rate is predicted to be 0.26 (Model 2), 0.09 (Model 3) and 0.13 (Model 4). It's worth noting that the marginal change, the slope of the predicted hazard function, i.e., the approximation of $h'(m)$ in equation (2), around the average spending, is relatively small, as depicted in Figure 4, which are -0.004 (Model 2), -0.002 (Model 3) or -0.003 (Model 4).

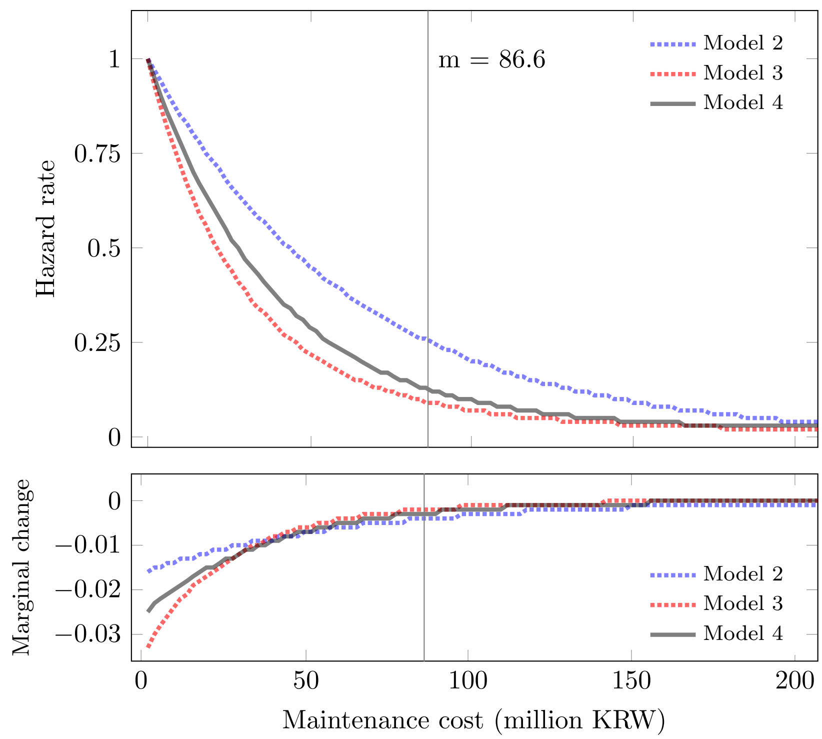


Figure 4: Predicted Hazard Rate and Marginal Changes

5.2. Balancing Reliability and Maintenance Cost

Let's consider that Model 4 represents the typical hazard rate for substations in Figure 4. Additionally, let's assume that the current average maintenance spending, which is 86.6 million KRW (≈ US$77,940), represents the optimal expenditure that maximizes the net benefit derived from avoiding outages, as presented in equation (2). The estimated slope of the hazard function at m = 86.6 is approximately -0.003 (Figure 4). This suggests that the outage cost, denoted as c in equation (2), would amount to 357 million KRW (≈ US$0.321 million) per year.[2] In terms of practical ramifications, an average substation in Korea typically caters to around 28,400 households.[3] Consequently, the outage cost per household per year could be 12,580 KRW per household (≈ US$11.30).

The question arises as to whether the estimated outage cost of 12,580 KRW (≈ US$11.30) per household accurately reflects the actual outage cost. If the true outage cost is lower than this estimate, it suggests that KEPCO is allocating excessive resources to maintain substation

[2] Derived from equation (2), $c = -1/h'(m)$.

[3] In 2018, the total number of 154kV substations in Korea was 753, which supply electricity to residential sector, while the number of households amounted to 205 million. This equates to an average of approximately 28,353 households per substation.

reliability. Conversely, if the actual outage cost is higher, it indicates that KEPCO is investing less than the optimal maintenance cost level. However, due to the absence of a specific study on Korean households to provide a precise estimate of outage costs, making a definitive judgment is challenging.

One potential approach is to rely on estimates from previous studies, such as Woo et al. (2021b) for U.S. households. Nevertheless, it's important to recognize that direct comparisons might yield misleading conclusions, given the variations in household characteristics, climate conditions, and other factors that can influence electricity consumption. Nonetheless, we believe that conducting such a comparison is valuable. It's worth noting that Woo et al. (2021b) estimated the market-based outage cost per kWh for the U.S. residential sector, resulting in a range from US$0.12 to US$0.34 per kWh. To convert the outage cost estimates in Woo et al. (2021b) to an annual basis, we use the U.S. household average electricity consumption (10,632 kWh per year (EPA, 2020)) and outage statistics (7.5 hours in 2021 (EPA, 2023)), which results in approximately 9.1 kWh of lost electricity per year due to outages. With Woo et al. (2021b) estimates, this equates to an outage cost ranging between US$1.10 and US$3.20 per household per year. Clearly, this range is considerably lower than the estimated US$11.30 from above, KEPCO might be allocating excessive maintenance costs or investing in unnecessarily high levels of maintenance on average.

6. Conclusion and Policy Implications

In this study, we have delved into the critical domain of preventive maintenance spending for electrical substations. Our analysis was underpinned by the overarching goal of optimizing the allocation of resources to ensure grid reliability while minimizing costs. We developed a comprehensive conceptual model that elucidates the intricate relationship between preventive maintenance spending, substation hazard rates, and outage costs. Through empirical estimation, we sought to uncover insights that would inform decision-makers at KEPCO regarding the efficiency of their maintenance expenditure.

Our empirical analyses have yielded several noteworthy findings. First, we observed diminishing returns in the reduction of hazard rates as preventive maintenance spending increased. This suggests that while higher maintenance investments enhance substation reliability, the incremental improvements in reliability diminish with each additional unit of expenditure. Second, our estimations suggest that the current average maintenance spending of 86.6 million KRW might be excessive when compared to the outage costs for U.S. households. However, it's crucial to recognize that direct comparisons could lead to misleading conclusions. This is due to variations in household characteristics, climate conditions, and other factors that can significantly influence electricity consumption.

The contributions of this study can be summarized as follows:

- Concetptual framework: This study provides a conceptual framework for balancing reliability of substation and the preventive maintenance expenditure using the hazard rate modeling and outage cost considerations.
- Empirical insights: Through empirical analysis, the study sheds light on the actual impact of preventive maintenance spending on substation reliability with covariates including diminishing returns of the maintenance cost
- Methodological contribution: This research contributes methodologically by employing hazard rate modeling and Cox Lasso regression to estimate the relationship between maintenance spending and reliability. These methods can be applied in similar studies in the utility sector and beyond.

Two significant caveats should be highlighted. Firstly, the findings of this study are derived from a dataset specific to Korea and may not be directly transferable to all regions or utility companies. Substation maintenance requirements and outage costs can vary significantly due to local conditions, infrastructure age, and other contextual factors. Consequently, the results should be generalized cautiously. Secondly, the challenge of estimating outage costs to determine whether KEPCO is overspending on maintenance costs has inherent difficulties and subject to uncertainties. Relying on estimates from other countries introduces potential disparities due to variations in electricity consumption patterns, climate conditions, and consumer behavior. Therefore, further investigation into outage costs for Korean households and a comprehensive evaluation of the trade-off between reliability and maintenance expenses are necessary and should be considered as a key focus for future research.

References

Agency, U. E. I. (2020). Residential Energy Consumption Survey (RECS). Available at https://www.eia.gov/consumption/residential/data/2020/, accessed on Sep 17, 2023.

Agency, U. E. I. (2023). Annual Electric Power Industry Report. Form EIA-861, Available at https://www.eia.gov/electricity/data/eia861/, accessed on Sep 17, 2023.

Ateba, B., Prinsloo, J. and Gawlik, R. (2019). The Significance of Electricity Supply Sustainability to Industrial Growth in South Africa, *Energy Reports* **5**: 1324–1338. https://doi.org/10.1016/j.egyr.2019.09.041.

Atwa, O. S. (2019). Substations, *Practical Power System and Protective Relays Commissioning*, Academic Press, pp. 9–21. https://doi.org/10.1016/B978-0-12-816858-5.00002-2.

Centolella, P. (2010). Estimates of the Value of Uniterrupted Service for the Mid-West Independent System Operator. Harvard Electricity Policy Group, John F. Kennedy School of Government, Harvard University, Available at https://hepg.hks.harvard.edu/publications/estimates-value-uninterrupted-service-mid-west-independent-system-operator, accessed on July 26, 2023.

Cleves, M., Gould, W. and Marchenko, Y. (2016). *An Introduiction to Survival Analysis Using Stata*, Stata Press, College Station, Texas.

Cox, D. (1972). Regression Models and Life-Tables, *Journal of the Royal Statistical Society, Series B* **34**(2): 187–220.

Espiritu, J. F., Coit, D. W. and Prakash, U. (2007). Component Ciriticality Importance Measures for the Power Industry, *Electric Power Systems Research* **77**: 407–420. https://doi.org/10.1016/j.epsr.2006.04.003.

Gorman, W. (2022). The Quest to Quantify the Value of Lost Load: A Critical Review of the Economics of Power Outages, *The Electricity Journal* **35**(8): 107187. https://doi.org/10.1016/j.tej.2022.107187.

International Energy Agency (2023). Korea Electricity Security Policy. Available at https://www.iea.org/articles/korea-electricity-security-policy, Accessed on July 23, 2023.

Kaplan, E. and Meier, P. (1958). Nonparametric Estimation from Incomplete Observations, *Journal of American Statistics Association* **53**(282): 457–481.

Kim, K. and Cho, Y. (2017). Estimation of Power Outage Costs in the Inudstrial Sector of South Korea, *Energy Policy* **101**: 236–245. https://doi.org/10.1016/j.enpol.2016.11.048.

Kim, T.-W., Chang, Y., Kim, D.-W. and Kim, M.-K. (2020). Preventive Maintenance and Forced Outages in Power Plants in Korea, *Energies* **13**(14): 3571. https://doi.org/10.3390/en13143571.

Klienbaum, D. and Klein, M. (2012). *Survival Analysis A Self-Learning Text*, Springer, New York, NY.

Klutkey, G., Kiessler, P. and Wortman, M. (2003). A Critical Look at the Bathtub Curve, *IEEE Transactions on Reliability* **52**(1): 125–129. https://doi.org/10.1109/TR.2002.804492.

Neelawela, U., Selvanathan, E. and Wagner, L. (2019). Global Measure of Electricity Security: A Comprehensive Index Approach, *Energy Economics* **81**: 433–453. https://doi.org/10.1016/j.eneco.2019.04.009.

Sudket, N. and Chaitusaney, S. (2014). Optimal Maintenance of Substation Equipment by Considering Maintenance Cost and Reliability. 11th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTI-CON), https://doi.org/10.1109/ECTICon.2014.6839858.

Tibshirani, R. (1996). Regression Shrinkage and Selection via the Lasso, *Journal of the Royal Statistical Society, Series B* **58**(1): 267–288. https://www.jstor.org/stable/2346178.

Tibshirani, R. (1997). The Lasso Method for Variable Selection in the Cox Model, *Statistics in Medicine* **16**(4): 385–395. https://doi.org/10.1002/(SICI)1097-0258(19970228)16:4<385::AID-SIM380>3.0.CO;2-3.

U.S. Department of Energy (2016). Maintaining Reliability in the Mdoern Power System. Energy Policy and System Analysis (DOE-EPSA), Available at https://www.energy.gov/policy/articles/maintaining-reliability-modern-power-system, Accessed on August 10, 2023.

Woo, C., Tishler, A., Zarnikau, J. and Chen, Y. (2021a). A Vack of the Envelope Estimate of the Average Non-residential Outage Cost in the U.S., *The Electricity Journal* **34**(4): 106930. https://doi.org/10.1016/j.tej.2021.106930.

Woo, C., Tishler, A., Zarnikau, J. and Chen, Y. (2021b). Average Residential Outage Cost Estimates for the Lower 48 States in the U.S., *Energy Economics* **98**: 105270. https://doi.org/10.1016/j.eneco.2021.105270.