energies

MDPI

Article

Construction Time Estimation Function for Canadian Utility Scale Power Plants

Herve Kabanda †, Alex Romard †, Fuze Yurtsever †, Anjali Wadhera, Joshua Andrews and Craig Merrett *

Department of Mechanical and Aerospace Engineering, Carleton University, Ottawa, ON K1S 5B6, Canada; hervekabanda@gmail.com (H.K.); aromard@gmail.com (A.R.); dejanmb456@gmail.com (F.Y.); anjali.wadhera@gmail.com (A.W.); joshwgandrews@gmail.com (J.A.)

* Correspondence: craig.merrett@carleton.ca; Tel.: +1-315-268-6580

† These authors contributed equally to this work.

Abstract: Construction time and time overruns for infrastructure projects have been frequently studied; however, the construction time of power plants has not been studied. This lack of study is problematic, as more renewable energy power plants, such as wind and solar, are planned for many jurisdictions. Accurately estimating the construction time of a power plant will assist construction planning, budget estimates, and policy development encouraging the use of more renewable sources. The construction times of utility scale power plants in Canada were studied using publicly available data. Multiple linear regression analysis techniques were applied to the data to generate construction time estimation functions for all power plants together, and for individual technologies. The analyses reveal that construction time is sensitive to jurisdiction and the decade of construction, indicating that decisions made by individual Canadian provincial governments at different times had statistically significant impacts on construction time. The analyses also indicated that construction time is a strong function of installed capacity, independent of technology. This finding suggests that large solar or wind energy facilities will encounter longer construction times similar to large hydroelectric facilities.

Keywords: construction planning; delays in power plant construction; grid integration of renewable energy; nuclear power; hydroelectric power

check for updates

Citation: Kabanda, H.; Romard, A.; Yurtsever, F.; Wadhera, A.; Andrews, J.; Merrett, C. Construction Time Estimation Function for Canadian Utility Scale Power Plants. *Energies* **2021**, *14*, 5421. https://doi.org/10.3390/en14175421

Academic Editor: Georgios Christoforidis

Received: 6 July 2021
Accepted: 11 August 2021
Published: 31 August 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

The focus of this paper is the development of an estimation function for the construction time of utility scale power plants, and the associated statistical analysis of the results from the estimation function. Construction cost of power plants has been often studied in the literature; however, a statistical analysis of the construction time for power plants has not been executed previously. The development of this function is novel and aims to increase technical accuracy and to inform policy makers. The technical motivation is to provide a function for use in the preliminary project planning phase to accurately estimate the required construction time, logistics, and budget. The policy motivation is to inform policy makers about the time required to implement renewable energy sources and integrate the sources into the grid because a concern of many governments is energy security. This paper determines the construction time for power plants using stepwise, multiple linear regression techniques with interaction terms. The paper is limited in geographical scope to Canadian power plants because of the limited number of provinces. Data was collected for the provinces of Alberta, British Columbia, Manitoba, New Brunswick, Nova Scotia, Ontario, Prince Edward Island, Quebec, and Saskatchewan.

Prior to a discussion of the investigation, the statistical tools and approach are explained and justified. The investigation begins with existing construction time data for hydroelectric, nuclear, and wind power plants across Canada. These data are explored for sensitivities to year of construction, technology, location, and size. A regression is

performed, leading to estimation functions that are subsequently compared to the construction time of the power plants. These estimation functions will answer the following research objectives:

1. Determination of construction time dependency on jurisdiction of construction;
2. Assessment of the impact of power plant technology on construction time;
3. Influence of power plant installed capacity on construction time;
4. Impact of historical factors on construction time.

2. Literature Review

The majority of previous research on construction time has focused on transportation and civic infrastructure rather than power systems. The main exception is nuclear power plants, which have been subject to a number of economic and construction analyses. The value of these analyses is to provide improved estimates of the necessary budgets and timelines to complete large projects so that subsequent projects may be completed more efficiently.

Estimates for construction time and costs are part of the larger field of project and operations management. Models of project management have developed over many centuries, with early models using trial and error approaches without conception of the final products [1]. Standardized management tools and practices did not appear until the 1950s. These tools and practices were developed in response to military and aerospace programs of the Cold War, with one example being the critical path model. Standard project monitoring tools were not enforced until the 1980s by American federal agencies along with the development of concurrent engineering practices [1].

The development of project management tools have aided large construction projects; however, the effectiveness of the tools is subject to the users. Several researchers have identified that social biases exist in project planning such as future-perfect strategizing [2], strategic misrepresentation [3–6], and escalation of commitment [7]. The biases have been studied in the context of power plant construction. Strategic misrepresentation was investigated for offshore wind farms for the United Kingdom coupled with reference class forecasting [6]. The authors considered 10 UK wind power plants and applied Flyvbjerg's reference class forecast (RCF) approach. Flyvbjerg's approach identified cost overruns as a result of optimism bias and strategic misrepresentation. Nine of the 10 wind power plants considered exhibited a cost overrun, and all had time overruns exceeding 10% of the original duration. Strategic misrepresentation was found to extend to the benefits of the wind farms as capacity factors were overestimated. Koch and their colleagues applied the RCF to the London Array and suggested extending the construction time by 30%.

A similar study was completed by Locatelli for the Flamanville, France and Olkiluoto, Finland nuclear power plants [7]. Both power plants are facing cost overruns and construction delays, with optimism bias and unfamiliarity with the design identified as contributing factors. The study included detailed analyses of the two projects using ATLAS coding of qualitative data to identify collaborations, network management, procurement, etc. An earlier study revealed that scaling up nuclear power led to significant increases in cost and time [8]. Subsequent work by Locatelli predicted the completion of Flamanville for 2019–2020 [9], and associates the delays to a lack of standardization. Locatelli contrasts the French construction experience with the South Korean construction of nuclear plants, which occur on time and on budget because of a standardized construction and supply chain.

Another study applied the Construction Industry Institute Engineering Procurement Construction (EPC) project life cycle to nuclear power plants to determine the origins of delays [10]. The EPC life cycle identifies six phases; however, each phase may be sensitive to the type of project such as nuclear. This result confirms an earlier study completed for the Ford Amendment on nuclear that concluded delays in the 1970s and 1980s resulted from a lack of experienced personnel, poor project management, and regulatory approval of permits in a dynamic regulatory and political environment.

The majority of research on construction times has focused on waterworks and sewage [11], buildings [12], transportation infrastructure [13], and petrochemical projects [14]. Six categories of delay were identified: contractor performance, owner administration, planning and design, government regulations, environmental assessments, and supervision. Of these, contractor performance is considered the major delay cause. Upward of 70% of projects experience delays with an average delay of 39% of the intended construction duration [11]. Contractor performance was echoed as a cause of delay by Alhajri et al. with specific examples cited of poor site management and conflict with the contractor from interviews conducted with project managers, engineers, and supervisors [14].

Al-Momani considered construction delays and causes for 130 public projects in Jordan such as residential, medical, and schools. The sources of delays varied based on the group surveyed. Engineers identified cash flow from the project owners as the main delay, while owners identified design errors by the engineers and labor availability of the contractors as the main sources of delay. The survey groups also identified that the early stages of a project were the origins of most delays, which coincides with the findings by Wright. The analysis used Excel to develop a simple linear regression for the construction delays [12].

A more recent study of power transmission projects attempted to determine the origins of the delays that afflict that sector. A survey of 311 stakeholders identified 63 delay factors that can be organized into ten groups: sector-specific, general, administrative, employer, contractor, consultant, materials, equipment, labor, and unavoidable. The sector-specific and general factors are novel from this study as they identified factors that are particular to power transmission projects such as access roads, right-of-way, poor site management, and poor coordination between different types of work [15]. The study did not include any regression techniques and was solely empirical; however, the identification of factors impacting delays is useful for power plant related analyses.

Love and their colleagues examined cost overrun probability distributions for 276 Australian construction projects with the objective to see if the distribution is normal [13]. Normality was assessed as applying the incorrect distribution would affect the predicted results. This assessment is important since Flyvbjerg assumes Gaussian distributions in their technique. The authors used Kolmogorov–Smirnov, Anderson–Darling, and Chi-squared tests to compare distributions and found that the distributions are non-Gaussian and are best described by a three parameter Frechet function for cost overruns. This result suggests that Flyvbjerg's normality assumption for cost overruns is incorrect.

Further, there is debate whether the project type and geographical location affects the cost overrun prediction. Flyvbjerg predicts independence of location but dependence of project type [16]. Dependence upon project type was also identified by Bhargava and their colleagues for highway projects, and project type influenced both cost and construction time [17]. Meanwhile, Odeck predicts independence of project type [18]. Last, Flyvbjerg's approach is that projects of the same class are assumed to have the same optimism bias [19]; however, earlier work by Love and Odeck show that the bias is not uniform for a given class [18,20,21]. The present work assesses the dependence of location and project type, and applies non-Gaussian distributions, thus contributing to the literature. The novelty is the consideration of construction time instead of cost.

The technical approach considered here is the use of multiple linear regression (MLR) techniques, as described in Section 4. Regression analyses have an established presence in predicting cost and time overruns of infrastructure projects. Samarghandi et al. used regression techniques to quantify construction delay factors in Iran for residential projects and educational facilities [22]. Educational facilities have been a particular interest for assessing cost and time overruns, with a more advanced regression analysis executed by Asiedu et al. [23]. Samarghandi et al. used linear regression techniques where a single dependent variable was correlated with a single independent variable. The Asiedu et al. team used multiple linear regression (MLR), where they identified 10 predictive variables that could influence cost overruns. The MLR analysis determined that five of the variables influenced the overruns [23]. A study of 911 building projects in Ghana also applied MLR

analysis in the statistical software *R* and considered seven predictive variables [24]. A total of three variables were determined to influence the completion cost of the projects.

Other construction projects, such as buildings and drainage, have been analyzed using regression techniques. Senouci et al. used linear regression models to predict cost overruns in Qatari construction projects. These models were developed in Excel and determined the correlation between contract price and cost overrun [25]. Croatian water infrastructure projects have also been modeled using MLR techniques. Ninety-three projects were considered with data collected via survey of project managers. These surveys identified 108 variables that were evaluated using MLR, and a set of 5 variables were determined to have the majority of the impact on the success of a water project [26].

Electrical infrastructure projects have been studied less often; however, an extensive analysis of cost overruns was done in 2017 of 401 projects across 57 countries. These projects were constructed between 1936 and 2014, and the team applied a linear regression analysis to the construction costs. They determined that large projects such as hydroelectric and nuclear facilities have a high correlation with cost overruns while decentralized facilities such as wind and solar have a negative correlation [27]. This study did not include a time trend, as the data was clustered in certain decades. For example, most of the nuclear power plants used in this study were constructed during the 1970s. A subsequent study examined hydroelectric power plants and their time overruns by considering 57 hydroelectric dams installed between 1975 and 2015 [28]. This study only included hydroelectric facilities financed by the World Bank Group and assessed the uncertainties in the costs and benefits of the technology. The data analyzed indicated that 80% of the 57 facilities considered experienced a time overrun, which highlights the need to accurately predict the construction of hydroelectic facilities [28].

3. Materials

A list of power plants across Canada was assembled, and the construction start and end dates compiled by searches through utility documents, press releases, and newspaper reports. The construction dates are subject to variability because the definitions of a "start" and "end" are non-uniform. Some "start" dates referred to when a license or regulatory approval was granted for a project, and other "start" dates were when the design process began, or when construction began. Similarly, the "end" date may be when the power plant is grid connected, or when the balance of equipment is installed. The variability of the dates affects the final results and is discussed in Section 6.

The data collected is a combination of continuous variables and categorical variables, as shown in Appendix A—Table A1. The continuous variables are the power plants rated capacities, and construction times. The categorical variables are power plant type, province where they were constructed, and decade of construction. These are categorical variables because they can only have discrete values. The categorical variables are necessary to include because each province in Canada may have different regulations and procedures for the approval of a power plant. Further, nuclear power plants are subject to federal approval; therefore, power plant type becomes a relevant category. The decade of construction is an additional categorical variable, and is intended to capture larger effects such as the economic or global conditions that may impact the construction of a large project.

4. Model

The theoretical framework is composed of early diagnostics, regression analysis, and residual analysis, and was completed using R, a free programming language and software for statistical analysis and graphics. A couple of early diagnostics were performed prior to the regression analysis. First, the dependent variable (Time) was plotted against each independent variable to observe any non-linearities. A correlation matrix was computed to see if there is any indication of multicollinearity among the independent variables. In regression analysis, a function must be established that relates the dependent variable to the independent variable. The function may use the dependent and independent variables directly, known as a level–

level, or use the logarithmic of one or both variables. Additionally, the function may include interaction terms which are other variables that may influence the dependent variable. Further, the function may be linear or a polynomial. For the regression analysis considered here, several models were considered of the following functional forms. The objective was to determine the functional form that captured the relationship between the construction time and the independent variables the best. Each of the following functional forms was applied to the data, and a stepwise regression applied within each form. The stepwise regression added parameters to the model such as technology type, decade of construction, or midpoint of construction. At each step and form, the residuals of the model were calculated and recorded. The form that had the lowest residual was selected as the best model. Below are the forms considered:

- Level–level without interaction terms;
- Log–level without interaction terms;
- Level–log without interaction terms;
- Log–log without interaction terms;
- Level–level with interaction terms;
- Log–level with interaction terms;
- Level–log with interaction terms;
- Log–log with interaction terms;
- Polynomial models with linear dependent variable;
- Polynomial models with logarithmic dependent variable.

The level–level form without interaction terms is the simplest model and uses the dependent and independent variables directly. For example, construction time as a function of power plant installed capacity. The log–level form without interaction terms is similar to the level–level form except the logarithm of the dependent variable is used. In this case, that would be the logarithm of construction time as a function of installed capacity. A logarithm may be necessary, as some technologies have long construction periods. A related form is the level–log form without interaction terms. This model would be the construction time, as a function of the logarithm of installed capacity. The power plants considered here range in capacity from a few MWs to the GW scale; therefore, the logarithm of installed capacity may provide a better fit. Lastly, the log–log form without interaction terms uses the logarithm of both the dependent and independent variables.

Interaction terms describe any relationships between the independent variables that may have an effect on the construction. Models with interaction terms are more sophisticated than the basic models and contribute to the stepwise regression approach used here. The simplest models were attempted first, then interaction terms added to determine if a more accurate model in terms of predicting construction time was possible. An example of interaction terms in the context of power plant construction, is the jurisdiction where the power plants are constructed can have an interaction with power plant size. A jurisdiction with large hydroelectric resources may become more efficient at construction of large installed capacities; therefore, the construction time would be shorter compared to other jurisdictions. The model would require a term to account for the interaction of jurisdiction with power plant size.

The level–level, log–level, level–log, and log–log models all attempt to construct a linear relationship between the independent variables and the dependent variables. A line of best fit is the result of these models. With some data, a linear relationship is not feasible and a nonlinear relationship is required. The polynomials describe nonlinear relationships such as quadratic, cubic, or quartic relationships between the independent and dependent variables. These models are more sophisticated than the linear models and may also use logarithms of the variables.

Each model is estimated by the ordinary least squares (OLS) method [29]. Initial model development and selection of the functional form was performed using a stepwise regression procedure. For each functional form, stepwise regression selects a group of independent variables by employing forward selection, backward elimination, or bidirectional elimination

methods. Stepwise regression adds or removes independent variables to a regression model by optimizing an objective function/criterion such as Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC), and R^2 value. The set of models suggested by the stepwise regression was then reduced by applying several selection criteria: standard error of regression, overall significance of a model, individual significance of independent variables, and adjusted-R^2 value. The number of models was further reduced by a multicollinearity test and a residual analysis. The ideal regression model will determine the function that relates the dependent variable to a single independent variable. Actual data may involve the dependent variable being related to several independent variables or the independent variables may be related to each other, which is multicollinearity. The existence of these relationships must be checked and the extent of the relationships evaluated.

Residual analysis was implemented to see if the surviving models satisfied the six multiple linear regression (MLR) Gauss–Markov assumptions. For this purpose, first residuals were plotted against fitted values to detect if there was a serious violation of the linear parameters, zero conditional mean, and homoskedasticity assumptions. Similarly, a histogram and a normal probability plot of residuals were created to check the normality assumption. Finally, outliers and influential observations were determined using outlier tests, standardized residuals, and Cook's Distance. After the residual analysis, model selection was completed by assessing the remaining models in terms of their predictive power.

5. Results

The statistical techniques described in Section 4 were applied to several data samples of Canadian power plants constructed over the past 100 years. The total sample size was 87 power plants with 25 wind, 23 nuclear, and 39 hydroelectric power plants. The total sample size was first assessed to distinguish if any power plant technologies appeared to be outliers, as shown in Section 5.1. The subsequent subsections consider models for hydroelectric power plants, and nuclear power plants. In these models, time is returned in months while the power plant size is provided in MW.

The wind power plants were modeled in isolation, but none of the models were significant compared to the arithmetic average construction time because of the very small variation in construction times of wind power plants. The available data for the wind plants were to the nearest year of construction unlike the other power plants where the data were available to the nearest month of construction. Construction time should be a continuous variable, which is possible for monthly data; however, yearly data creates a discrete variable. In the specific case of the wind power plant data, the recorded construction times were either 12 months or 24 months; therefore, a function was not generated from this data.

5.1. Modeling All Power Plants Together

When considering all power plants together, the stepwise regression procedure determined that a log–log with interaction terms model passed all model selection criteria and gave the lowest residual error. All functional forms were considered, and a subset of the their resulting R^2 values are shown in Table 1.

Table 1. Stepwise regression results for modeling all power plants.

Model	n	s	R^2	Adjusted-R^2	F-Statistic	p-Value
Log-level with province and plant type indicators	87	0.3896	0.7653	0.7568	90.21	2.2×10^{-16}
Log-log with province and plant type indicators	87	0.3714	0.7867	0.7790	102.0	2.2×10^{-16}
Log-log with mid-point of construction	87	0.3523	0.8127	0.8012	70.3	2.2×10^{-16}

Only a subset of the stepwise regression is shown in Table 1 because the other forms and combinations of indicators gave results that were not statistically significant. The three models reported in the table are all statistically valid; however, the third model indicates the best fit via the R^2 value and is reported in the following.

The selection process identified that the key plant-type indicators are for hydro and wind, and that the construction time is dependent upon the size of the power plant. The process also identified that the location of the power plant does not affect the construction time. Lastly, the process identified another key indicator, mPoint, that represents the year at the mid-point of the construction time of a given power plant. For example, a power plant constructed between 2001 and 2005 would have 2003 as the mid-point of the construction time. The purpose of including the mid-point was to detect any sensitivity of the model to the historic and economic time period such as the 2008 economic recession. Capital projects under-construction during 2008 may have been delayed because of larger financial pressures [30]. The estimated model is

$$ln(Time) = \beta_0 + \beta_1 Hydro + \beta_2 Wind + \beta_3 mPoint + \beta_4 ln(MW) + \beta_5 ln(mPoint) + \epsilon \quad (1)$$

where β_0 through to β_5 are the regression coefficients, while ϵ is the residual error. The regression process determined the values and significance of the coefficients as shown in Table 2.

Table 2. Dependent variable: ln(Time)—coefficients for all power plants.

Model	$\hat{\beta}$	$se(\hat{\beta})$	t-Stat	*p*-Value
Constant	14,030	4293	3.267	0.00159
Hydro	−0.323	0.1015	−3.181	0.00208
Wind	−1.719	0.1905	−9.022	7.19×10^{-12}
mPoint	1.087	0.3319	3.274	0.00156
ln(MW)	0.216	0.04049	5.338	8.37×10^{-7}
ln(mPoint)	−2131	652.2	−3.268	0.00159

Substituting the parameters from Table 2 into Equation (1) gives

$$\widehat{ln(Time)} = 14{,}030 - 0.323 Hydro - 1.719 Wind + 1.087 mPoint + 0.216 ln(MW) - 2131 ln(mPoint) \quad (2)$$

where the $\widehat{\ }$ denotes the estimated nature of the equation. The summary statistics of the model are provided in Table 3, where N indicates the number of data points used in the model, s is the standard error and is the average distance of the data points from the regression line.

Table 3. All power plants model summary.

n	*s*	R^2	Adjusted-R^2	F-Statistic	*p*-Value
87	0.3523	0.8127	0.80120	70.3	2.20×10^{-16}

In terms of the units of the model, a standard error of 0.3523 corresponds to a difference of 1.4 months. The summary statistics indicate that the model is statistically significant. The use of technology indicators in the model shown in Section 5.1 reveals that a single estimation model for the construction time is not feasible. The inclusion of technology indicators in Equation (2) means that the type of technology will alter the predicted construction time. For example, if a hydroelectric power plant is considered, the construction time will be reduced given the −0.323 before the hydro indicator. If construction time was insensitive to technology, then Equation (2) should not include "hydro" or "wind". The subsequent step is to consider regression models for individual power plant technologies.

5.2. Modeling Hydro Power Plants

The form of this estimation function for hydroelectric power is different than the form used for all power plant technologies combined; indicating a different relationship between construction time and size of the power plant. Applying the functional forms shown in Section 4, a level-log with interaction terms model was found to provide the best fit according to the regression procedure. The results from other possible, statistically valid, model forms are shown in Table 4.

Table 4. Stepwise regression results for modeling hydroelectric power plants.

Model	n	s	R^2	Adjusted-R^2	F-Statistic	p-Value
Level–level with province and decade indicators	39	1.804	0.5193	0.4781	12.6	9.64×10^{-6}
Level–log with province and decade indicators	39	1.733	0.5565	0.5184	14.64	2.43×10^{-6}
Level–log with province and multiple decade indicators	39	1.681	0.5945	0.5468	12.46	2.4×10^{-6}

The hydro power model introduced interaction terms for province and decade of construction. The inclusion of additional decade indicators was necessary because large hydro plants may require 10 or more years to complete construction. The longer duration means that the construction process is sensitive to economic conditions at the time. A logarithm for the power plant size was also introduced as the data for hydro power plants included a large range of sizes, thus the logarithm may linearize the relationship between size and construction time. The resulting model is

$$Time = \beta_0 + \beta_1 QC + \beta_2 D_{1910} + \beta_3 D_{2000} + \beta_4 ln(MW) + \epsilon \quad (3)$$

where D_{1910} and D_{2000} indicate decade of construction, and QC indicates the province of construction. All are categorical data that take discrete values of 0 or 1. The model, with coefficients, is shown in Equation (4)

$$\widehat{Time} = 0.714 + 2.574QC + 4.229D_{1910} + 1.872D_{2000} + 0.6502ln(MW) \quad (4)$$

where the coefficients originate from the regression results shown in Table 5.

Table 5. Dependent variable: Time—coefficients for hydroelectric power plant model.

Model	$\hat{\beta}$	$se(\hat{\beta})$	t-Stat	p-Value
Constant	0.7143	1.1738	0.609	0.547
QC	2.5744	0.8063	3.193	0.00303
D_{1910}	4.2293	1.2345	3.426	0.00162
D_{2000}	1.8718	1.0476	1.787	0.08289
ln(MW)	0.6502	0.2154	3.019	0.00478

With the summary statistics provided in Table 6.

Table 6. Hydro power model summary

n	s	R^2	Adjusted-R^2	F-Statistic	p-Value
39	1.681	0.5945	0.5468	12.46	2.404×10^{-6}

The summary statistics of the model, shown in Table 6, indicate a higher standard error for the estimated construction time compared to the standard error achieved with the model for all power plants. The standard error is equivalent to 1.7 months compared to the earlier error of 1.4 months.

5.3. Modeling Nuclear Power Plants

Assessment of the construction time for nuclear power plants generated a log–level with interaction parameters model. The logarithm is applied to the construction time, which is significant because the nuclear power plants took disproportionally long to construct for their nameplate capacity. Compared to the previous models, the construction time of nuclear power plants is more influenced by geographic location, and the decade of construction if a nuclear accident occurred. Similar to the hydro power plants, nuclear power plants may take 10 or more years to construct, thus the economic and socio-political environment will influence the continued construction. Unlike nuclear power plants, hydro power plants have not encountered suspended construction because of accidents thus the hydro power plant model did not require a logarithmic function for time.

The nuclear power plant model is a function of mPoint. The parameter mPoint is the year corresponding to the midpoint of the construction period for a given power plant. This parameter is a continuous variable and is necessary to capture major shifts in public acceptance for the continued construction of a nuclear power plant. Accidents such as Three-Mile Island, Chernobyl, and Fukushima occurred in short periods of time; however, the resulting shift in public acceptance had lingering effects. For example, after the Three-Mile Island accident, no new nuclear reactors for power generation were constructed in the United States [31]. The results from the leading model forms are shown in Table 7.

Table 7. Stepwise regression results for modeling nuclear power plants.

Model	n	s	R^2	Adjusted-R^2	F-Statistic	p-Value
Level–log with province and site indicators	23	0.6896	0.8608	0.8289	27.82	1.72×10^{-7}
Level–log with province, site and mid-point indicators	23	0.5909	0.9034	0.875	31.81	4.69×10^{-8}
Log–log with site, and mid-point indicators	23	0.09267	0.8729	0.8446	30.9	7.68×10^{-8}
Log–log with province, site, and mid-point indicators	23	0.08198	0.906	0.8784	32.78	3.74×10^{-8}

This model is unique in that cross-products of interaction terms and independent variables are present, as shown in Equation (5).

$$ln(Time) = \beta_0 + \beta_1 QC + \beta_2 Darlington + \beta_3 MW + \beta_4 mPoint + \beta_5 MW \cdot mPoint + \epsilon \quad (5)$$

where *Darlington* indicates a particular nuclear power plant. The Darlington power plant appears as an indicator because the plant was under construction when the Chernobyl accident occurred. The ramifications of this accident on the construction time are elaborated in Section 6. The regression process applied to this model generated the coefficients tabulated in Table 8.

Table 8. Dependent variable: ln(Time)—coefficients for nuclear power plant model.

Model	$\hat{\beta}$	$se(\hat{\beta})$	t-Stat	p-Value
Constant	−235.4	30.75	−7.654	6.63×10^{-7}
QC	0.2106	0.07243	2.908	0.00979
Darlington	0.2884	0.09473	3.044	0.00733
MW	0.2529	0.04973	5.085	9.17×10^{-5}
mPoint	0.1203	0.01557	7.725	5.86×10^{-7}
MW·mPoint	−0.0001282	0.00002516	−5.097	8.95×10^{-5}

Substitution of the parameters in Table 8 into Equation (5) gives

$$\widehat{ln(Time)} = -235.4 + 0.2106QC + 0.2884Darlington + 0.2529MW + 0.1203mPoint - 0.0001282MW \cdot mPoint \quad (6)$$

The resulting model has a standard error of 0.082, shown in Table 9, which is equivalent to a 1 month discrepancy in construction time.

Table 9. Nuclear power plant model summary.

n	*s*	R^2	Adjusted-R^2	F-Statistic	*p*-Value
23	0.08198	0.906	0.8784	32.78	3.737×10^{-8}

6. Discussion

The discussion of the results is organized into three subsections that consider all the power plants together, hydro plants in isolation, and nuclear plants in isolation. Finally, the limitations of the models and the data are discussed and assessed against the hypothesis.

6.1. Modeling All Power Plants Together

The model shown in Equation (1) indicates that the coefficients for the Hydro and Wind variables are statistically significant, as their *p*-values are less than 0.05. The presence of the Hydro and Wind variables means that the construction times of wind and hydro power plants are significantly different from that of nuclear power plants. This result agrees with Flyvbjerg's assertion that project type is a factor for cost overrun predictions [16], rather than Odeck's later assertion of type independence [18]. This difference in type required the derivation of separate estimation functions for each type of power plant. Despite the need for project type indicators, the model has a small standard error of 1.4 months, which is substantially less than the 30% increase in construction time proposed by Koch [6] for wind turbines and the 39% determined by Al-Khalil [11]. Thus, this model provides a more accurate estimate for construction time.

The model also demonstrates a statistically significant correlation with the power plant size because the MW variable has a *p*-value less than 0.05. The model includes the mid-point of construction to achieve a more accurate prediction. The mid-point of construction is important because for longer construction times, such as those experienced by hydro and nuclear power plants, there may be a change in economic conditions during the construction that influence the balance of the project.

6.2. Modeling of Hydro Power Plants

The hydro power model indicates that hydro power construction time is consistent across Canadian provinces and decades except for construction in Quebec, and construction during the 1910 and 2000 decades. The model, Equation (4), indicates that the time to construction is dependent upon the size of the power plant because the MW variable appears. Further, the construction time exhibits a linear relationship with power plant size. The exceptions for Quebec and the two decades indicate an increased construction time.

The longer construction time in Quebec may be a function of the interactions of Hydro Quebec with the First Nations whose traditional lands were affected by the hydroelectric installations and reservoirs [32,33]. For example, the La Grande project began construction in 1971, but a court order brought by the Quebec Association of Indians stopped construction in 1975, and construction was not completed until 1985 [32]. Since the La Grande project, with its associated long construction time, was included in the data set used for the regression analysis, the Quebec parameter appeared. The appearance of the Quebec parameter also contradicts Flyvbjerg's assertion of independence of location for cost overrun predictions [16].

The extended construction time for plants built during the 1910s pertains to Great Falls Generating Station in Manitoba and Sir Adam Beck I Generating Station in Ontario. The Great Falls Generating Station's construction commenced in 1919 and was completed in 1928. Construction was delayed by World War 1 and subsequent "material shortages and uncertain financing", caused additional delays [34]. The first generating station came online in 1923, and in 1928 there were a total of six generators in operation [34].

Two factors that influenced the duration of construction projects were the accessibility of transportation and available technologies. A lack of road infrastructure during the 1910s made transporting materials and heavy machinery more difficult in comparison to later decades. A period report states that heavy equipment was moved using skids, horses, and steam hoists. In the warmer months, ferries could be used, but when the water froze over, equipment had to be moved over the ice [34].

The Sir Adam Beck (SAB) I generating station project located near Niagara Falls, Ontario commenced construction in May 1917. By 5 December 1925, it had nine generators in operation, with a tenth generator to be finished and in service by July of 1930. The 437 MW project was unparalleled to anything the Ontario government had done before, and made it the world's largest hydroelectric power station for its time [35]. Being the first of its kind, SAB encountered a number of novel construction challenges. The power station required a 20 km hydro canal to be excavated across uphill terrain to exploit a larger hydraulic head, starting in Chippawa and ending in Queenston [35].

The construction delay during the 2000s is attributable to the economic recession in 2008, which impacted the budgets associated with the necessary construction logistics [30].

6.3. Modeling of Nuclear Power Plants

The nuclear power plant model (Equation (6)) demonstrates that the construction time for nuclear power plants is consistent across provinces and dependent upon the size of the power plant, and the mid-point of the construction period. The consistency across provinces is demonstrated by the absence of provincial variables except for Quebec, and the Darlington power plant variable. The construction time for nuclear power plants in Quebec is longer because of the data for the Gentilly 1 facility. This facility introduced many new technologies that encountered delays, and is only one of two nuclear power plants constructed in Quebec. The impact of the single power plant on the province is higher than for provinces with more nuclear power plants. The models also demonstrate a dependence on the mid-point of the construction period, which is expected because of the number of phases that nuclear construction must complete in Canada [36].

The Darlington nuclear power plant was constructed in Ontario between 1982 and 1993, and there are several factors to consider when investigating this lengthy construction time. The Chernobyl accident occurred during construction, which led to the provincial government requesting further analysis of the power plant's safety. The Canadian Deuterium Uranium (CANDU) reactor is technologically very different from Chernobyl's RKMB reactor design; however, a construction halt was imposed during the review. A change of scope for Darlington caused the most amount of delays along with uncontrolled economic factors [37]. A decrease in forecasted electricity load was the main reason for major project delays for all four units. Units 3 and 4 suffered further economic impact as policies stipulated a reduction in the borrowing for capital expenditures [38]. This sudden change in access to capital negatively impacted the construction time. In 1986, there was a strike by the electrical workers on site causing six month delays on both Unit 1 and Unit 2 [37].

While these unexpected factors caused the majority of the delay for Darlington, other technical and mismanaged resource issues created further delays. The first major observed issue was the damage to the fuel bundles caused by vibrations from the main heat transport pumps followed by cracks observed in the main generator shaft which required a redesigned rotor shaft [39]. The turnover of trained operating staff cascaded from Units 1 and 2 delays [37]. The technical issues were observed early on in the project for Units 1 and 2, allowing Units 3 and 4 to be commissioned with the optimally designed parts.

6.4. Limitations and Evaluation of Hypothesis

The estimation functions generated for all the power plants combined, hydro power plants in isolation, and nuclear power plants in isolation are limited by several factors. The first factor is that the regression functions were developed using Canadian data, thus the specific functions may not be applied to other countries. The function forms reveal the independent variables and indicators that should be included in the construction time estimation functions. These parameters are transferable to other countries. The second factor is the limited data set used to generate the regression functions. The wind power plant data included in the models for all construction plants lack a continuous variable for construction time as the reported completion of these power plants is to the nearest year rather than nearest month. The third factor is that the data was collected through public resources. More precise analyses are possible if the utilities and contracting companies provide their schedules from each facilities' construction so that construction start and end times may be narrowed to a uniform definition and narrowed to the day rather than the year or month. More accurate analyses are possible by increasing the size of the data set, which is possible if the same utilities provide the required data on all their facilities.

The hypothesis is that the construction time for a power plant is a strong function of installed capacity. The presence of installed capacity in all the models confirms the hypothesis. The models also reveal that the construction time is dependent on the technology, thus a universal model for predicting construction time is limited.

7. Conclusions

The construction times of utility scale power plants in Canada were examined using multiple linear regression analysis techniques to assist in identifying the factors that may contribute to time overruns. Construction time overruns for power plants is an area where there has been little study, while other infrastructure projects such as educational facilities, water systems, and highways have been thoroughly studied. This lack of study has implications for construction planning, policy development to adopt more renewable power, and larger grid level planning. If the construction time is poorly estimated, the logistics of certain supplies and workers to arrive on site will be incorrect, policy windows may close that are favorable to certain technologies, and the retirement of older, inefficient power plants may be delayed.

The MLR analysis of Canadian power plants revealed that the construction time is strongly dependent on the installed capacity of the power plant, shown by all the models (Equations (2)–(6)) requiring the inclusion of the power plant size through the MW variable. The construction time is also a function of the type of technology, as the analyses demonstrated that hydro and nuclear power plants required separate models (Equations (4)–(6)) to accurately estimate their construction time compared to a model developed from hydro, nuclear, and wind (Equation (2)). Lastly, the construction time is a function of location, as both the hydro and nuclear models required location indicators.

The inclusion of technology indicators on construction time means that an accurate estimate of the construction time for novel power plants may be difficult in jurisdictions where no power plants of that type currently exist. Construction planners are recommended to use the worst case, i.e., longest construction time, from existing technologies. In the Canadian context, the regression model developed here for hydroelectric facilities is recommended. These regression models are also of use to policy developers because there is typically a policy window within which a policy should be enacted. Policy analysts aiming to encourage the use of more utility-scale renewable energy are recommended to use the regression models to estimate how long the power plants will take to construct. This duration will indicate how long a particular construction may remain at the forefront of the public's perception. The duration will also help with planning to meet environmental targets. Lastly, the inclusion of jurisdictions in the models suggests that policy planners should examine the policies implemented by certain Canadian provinces that may have eased the implementation of power plant technologies.

The development of regression models for the construction time of Canadian utility-scale power plants fills a gap in the knowledge of construction time and time overruns for energy infrastructure. These models have several limitations, the major limitation being that the models were developed with publicly available data. As a result, not every power plant in Canada was used in the regression analyses as many power plants have not released the necessary data. For this research to continue further, power plant construction companies and energy ministries are encouraged to make time line data publicly available. A second limitation is that the end times of the construction may not be uniformly defined across the entire data set. Some reported end of construction times were when a power plant was grid connected, while others used when a generator became operational. The last limitation is that the start times of the construction varied where some facilities used the issuance of government licenses as the start, while others used the breaking of ground as the start.

Author Contributions: Conceptualization, H.K., A.R. and C.M.; methodology, F.Y.; software, F.Y.; validation, H.K., A.R., A.W. and J.A.; formal analysis, F.Y.; investigation, H.K., A.R., A.W. and J.A.; resources, C.M.; data curation, H.K., A.R., A.W. and J.A.; writing—original draft preparation, H.K., A.R., F.Y., A.W., J.A. and C.M.; writing—review and editing, C.M.; supervision, C.M. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data are available by request to the authors.

Acknowledgments: The authors would like to thank Y. Han for her editorial assistance with the paper. The authors would also like to thank A. Boehm for her research assistance. This work was not funded by any organization.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A. Definition of Model Variables

Table A1. Definition of variables for models of power plants (Hydro, Nuclear, Wind).

Time	Duration of construction of power plant (from release of the RFQ to connection the grid system)	Dependent
ln(Time)	Natural log of Time variable	Dependent
MW	Installed capacity in Megawatt (MW)	Independent
ln(MW)	Natural log of MW variable	Independent
AB	Provincial indicator, AB = 1 if the power plant is in Alberta, AB = 0 otherwise	Independent
BC	Provincial indicator, BC = 1 if the power plant is in British Columbia, BC = 0 otherwise	Independent
MB	Provincial indicator, MB = 1 if the power plant is in Manitoba, MB = 0 otherwise	Independent
NB	Provincial indicator, NB = 1 if the power plant is in New Brunswick, NB = 0 otherwise	Independent
NS	Provincial indicator, NS = 1 if the power plant is in Nova Scotia, NS = 0 otherwise	Independent
ON	Provincial indicator, ON = 1 if the power plant is in Ontario, ON = 0 otherwise	Independent
PEI	Provincial indicator, PEI = 1 if the power plant is in Prince Edward Island, PEI = 0 otherwise	Independent

Table A1. *Cont.*

QC	Provincial indicator, QC = 1 if the power plant is in Québec, QC = 0 otherwise	Independent
SK	Provincial indicator, SK = 1 if the power plant is in Saskatchewan, SK = 0 otherwise	Independent
NL	Provincial indicator, NL = 1 if the power plant is in Newfoundland and Labrador, NL = 0 otherwise	Independent
Hydro	Plant type indicator, Hydro = 1 if power plant is a hydroelectric facility, Hydro = 0 otherwise	Independent
Nuclear	Plant type indicator, Nuclear = 1 if power plant is a hydroelectric facility, Nuclear = 0 otherwise	Independent
Wind	Plant type indicator, Wind = 1 if power plant is a hydroelectric facility, Wind = 0 otherwise	Independent
mPoint	The year that corresponds to midpoint of the construction time. (i.e., for a power plant started in 2004 and completed in 2008, mPoint = 2006)	Independent
ln(mPoint)	Natural log of mPoint variable	Independent
Decades	Defines the decade of the beginning of construction (example: 1910 = from 1910 to 1919)	Independent
ln(Decades)	Natural log of Decades, variable	Independent
D_1910	Decade indicator, D_1910 = 1 if power plant was built between 1910 and 1919, D_1910 = 0 otherwise	Independent
D_1920	Decade indicator, D_1920 = 1 if power plant was built between 1920 and 1929, D_1920 = 0 otherwise	Independent
D_1940	Decade indicator, D_1940 = 1 if power plant was built between 1940 and 1949, D_1940 = 0 otherwise	Independent
D_1950	Decade indicator, D_1950 = 1 if power plant was built between 1950 and 1959, D_1950 = 0 otherwise	Independent
D_1960	Decade indicator, D_1960 = 1 if power plant was built between 1960 and 1969, D_1960 = 0 otherwise	Independent
D_1970	Decade indicator, D_1970 = 1 if power plant was built between 1970 and 1979, D_1970 = 0 otherwise	Independent
D_1980	Decade indicator, D_1980 = 1 if power plant was built between 1980 and 1989, D_1980 = 0 otherwise	Independent
D_2000	Decade indicator, D_2000 = 1 if power plant was built between 2000 and 2009, D_2000 = 0 otherwise	Independent
Pickering	Site indicator for nuclear power plant, Pickering = 1 if the power plant is at Pickering, Pickering = 0 otherwise	Independent
Bruce	Site indicator for nuclear power plant, Bruce = 1 if the power plant is Bruce, Bruce = 0 otherwise	Independent
Darlington	Site indicator for nuclear power plant, Darlington = 1 if the power plant is at Darlington, Darlington = 0 otherwise	Independent
Gentilly	Site indicator for nuclear power plant, Gentilly = 1 if the power plant is at Gentilly, Gentilly = 0 otherwise	Independent
PLP	Site indicator for nuclear power plant, PLP = 1 if the power plant is at Point LePreau, PLP = 0 otherwise	Independent

References

1. Garel, G. A history of project management models: From pre-models to the standard models. *Int. J. Proj. Manag.* **2013**, *31*, 663–669. [CrossRef]
2. Winch, G. Escalation in major projects: Lessons from the Channel Fixed Link. *Int. J. Proj. Manag.* **2013**, *31*, 724–734. [CrossRef]

3. Ehrenfeucht, R. Megaprojects and risk: A conversation with Bent Flyvbjerg. *Crit. Plan.* **2004**, *11*, 53–62.
4. Flyvbjerg, B. From Nobel Prize to project management: Getting risks right. *Proj. Manag. J.* **2006**, *37*, 5–15. [CrossRef]
5. Flyvbjerg, B. Truth and lies about megaprojects. In *Inaugural Speech*; Delft University of Technology: Delft, The Netherlands, September 2007.
6. Koch, C. Contested overruns and performance of offshore wind power plants. *Constr. Manag. Econ.* **2012**, *30*, 609–622. [CrossRef]
7. Locatelli, G.; Mancini, M. Looking back to see the future: Building nuclear power plants in Europe. *Constr. Manag. Econ.* **2012**, *30*, 623–637. [CrossRef]
8. Grubler, A. The costs of the French nuclear scale-up: A case of negative learning by doing. *Energy Policy* **2010**, *38*, 5174–5188. [CrossRef]
9. Locatelli, G. *Why Are Megaprojects, Including Nuclear Power Plants, Delivered Overbudget and Late? Reasons and Remedies*; Technical Report; MIT CANES: Cambridge, MA, USA, 2018.
10. Wright, E.; Cho, K.; Hastak, M. Assessment of critical construction engineering and management aspects of nuclear power projects. *J. Manag. Eng.* **2014**, *30*, 04014016. [CrossRef]
11. Al-Khalil, M.; Al-Ghafly, M. Important causes of delay in public utility projects in Saudi Arabia. *Constr. Manag. Econ.* **1999**, *17*, 647–655. [CrossRef]
12. Al-Momani, A. Construction delay: A quantitative analysis. *Int. J. Proj. Manag.* **2000**, *18*, 51–59. [CrossRef]
13. Love, P.; Wang, X.; Sing, C.; Tiong, R. Determining the probability of a project cost overruns. *J. Constr. Eng. Manag.* **2013**, *139*, 321–330. [CrossRef]
14. Alhajri, A.R.; Alshibani, A. Critical factors behind construction delays in petrochemical projects in Saudi Arabia. *Energies* **2018**, *11*, 1652. [CrossRef]
15. Pall, G.K.; Bridge, A.J.; Gray, J.; Skitmore, M. Causes of delay in power transmission projects: An empirical study. *Energies* **2020**, *13*, 17. [CrossRef]
16. Flyvbjerg, B.; Mette, S.; Buhl, S. Underestimating costs in public works projects: Error or lie? *J. Am. Plan. Assoc.* **2002**, *68*, 279–295. [CrossRef]
17. Bhargava, A.; Anastasopoulos, P.; Labi, S.; Sinha, K.; Mannering, F. Three-stage least-squares analysis of time and cost overruns in construction contracts. *J. Constr. Eng. Manag.* **2010**, *136*, 1207–1218. [CrossRef]
18. Odeck, J. Cost overruns in road construction—What are their sizes and determinants? *Transp. Policy* **2004**, *11*, 43–53. [CrossRef]
19. Flyvbjerg, B. Curbing Optimism Bias and Strategic Misrepresentation in Planning: Reference Class Forecasting in Practice. *Eur. Plan. Stud.* **2008**, *16*, 3–21. [CrossRef]
20. Love, P. Influence of project type and procurement method on rework costs in building construction projects. *J. Constr. Eng. Manag.* **2002**, *128*, 18–29. [CrossRef]
21. Love, P.; Edwards, D.; Smith, J.; Walker, D. Congruence or divergence? A path model of rework in building and civil engineering projects. *J. Perform. Constr. Facil.* **2009**, *23*, 480–488. [CrossRef]
22. Samarghandi, H.; Tabatabaei, S.M.M.; Taabayan, P.; Hashemi, A.M.; Willoughby, K. Studying the reasons for delay and cost overrun in construction projects: The case of Iran. *J. Constr. Dev. Ctries.* **2016**, *21*, 51–84. [CrossRef]
23. Asiedu, R.O.; Frempong, N.K.; Alfen, H.W. Predicting likelihood of cost overrun in educational projects. *Eng. Constr. Archit. Manag.* **2017**, *24*, 21–39. [CrossRef]
24. Coffie, G.H.; Aigbavboa, C.O.; Thwala, W.D. Modelling construction completion cost in Ghana public sector building projects. *Asian J. Civ. Eng.* **2019**, *20*, 1063–1070. [CrossRef]
25. Senouci, A.; Ismail, A.; Eldin, N. Time delay and cost overrun in Qatari public construction projects. *Procedia Eng.* **2016**, *64*, 368–375. [CrossRef]
26. Klepo, M.S.; Radujkovic, M. Early warning system in managing water infrastructure projects. *J. Civ. Eng. Manag.* **2019**, *25*, 531–550. [CrossRef]
27. Sovacool, B.K.; Proudlove, A.; Green, N. Scale, risk, and construction cost overruns for electricity infrastructure. In *The Governance of Infrastructure*; OUP Oxford: Oxford, UK, 2017; pp. 1–23.
28. Baurzhan, S.; Jenkins, G.P.; Olasehinde-Williams, G.O. The economic performance of hydropower dams supported by the World Bank Group, 1975–2015. *Energies* **2021**, *14*, 2673. [CrossRef]
29. Osborne, J. *Regression and Linear Modeling: Best Practices and Modern Methods*; SAGE Publications: Los Angeles, CA, USA, 2017.
30. Deutch, J.; Forsberg, C.; Kadak, A.; Kazimi, M.; Moniz, E.; Parsons, J.; Du, Y.; Pierpoint, L. *Update of the MIT 2003 Future of Nuclear Power*; Technical Report; MIT: Cambridge, MA, USA, 2009.
31. Levy, M. Three Mile Island Plant, site of nuclear accident, to close. *Associated Press News*, 8 May 2019.
32. Ma, J.; Hipel, K.; De, M. Strategic analysis of the James Bay hydro-electric dispute in Canada. *Can. J. Civ. Eng.* **2005**, *32*, 868–880. [CrossRef]
33. Maxwell, J.; Lee, J.; Briscoe, F.; Stewart, A.; Suzuki, T. Locked on course: Hydro-Quebec's commitment to mega-projects. *Environ. Impact Assess. Rev.* **1997**, *17*, 19–38. [CrossRef]
34. Affairs, P. *A History of Electric Power in Manitoba*; Technical Report; Manitoba Hydro: Winnipeg, MB, Canada, 2011.
35. OPG. Sir Adam Beck I Generating Station. Available online: http://www.opg.com/generating-power/hydro/southwest-ontario/Pages/sir-adam-beck-i.aspx (accessed on 18 July 2016).

36. Canadian Nuclear Safety Commission (CNSC). *Licence Application Guide: Licence to Construct a Nuclear Power Plant*; Regulatory Document REGDOC-1.1.2; CNSC: Ottawa, ON, Canada, August 2019.
37. Canadian Nuclear Association. *Nuclear Canada Yearbook*; Staff Report Darlington Cost Increases; Canadian Nuclear Association: Ottawa, ON, Canada, 1992; pp. 14–16.
38. Ontario. Power Corporation Act. Available online: https://www.ontario.ca/laws/statute/90p18 (accessed on 18 July 2016).
39. Nixon, A. *The Canadian Nuclear Power Industry*; Background Paper BP-365E; Library of Parliament: Ottawa, ON, Canada, December 1993.