International Journal of Applied Industrial Engineering
Volume 7 • Issue 1 • January-June 2020

Application of Multiple Regression and Artificial Neural Networks as Tools for Estimating Duration and Life Cycle Cost of Projects

Brian J. Galli, Hofstra University, USA
https://orcid.org/0000-0001-9392-244X

ABSTRACT

Project managers face complex challenges when planning project stages because contract durations and project costs are difficult to predict accurately. The purpose of this study is to investigate statistical tools and concepts that can be integrated in the second phase of the project life cycle: the planning stage. Furthermore, this study aims to compare the accuracy of multiple regression and artificial neural network models, as well as the application of simulation in construction models used in predicting project duration and cost. This paper will also discuss the industry's current estimation methods, the use of statistical approaches, simulation, and the relationship between the application statistical tools and project success. Thus, this review identifies the trending statistical tools used by scholars to develop regression and neural models to solve the complexity of cost and duration estimation. The findings indicate that although the industry needs more accurate predictions and estimating tools, and regardless of the investigations and advancements made with integrating statistical tools, implementing these statistical approaches is faced with barriers.

KEYWORDS

Duration Estimation, Life Cycle Cost, Multiple Regression, Neural Networks, Project Management, Simulation

INTRODUCTION

In the field of project management, the life cycle of a project is composed of the following phases or stages: initiation, planning, execution-monitoring, and project closure. The focus of my investigation is concerned with the planning stage, namely cost estimating and duration estimation. This paper intends to provide an understating of the application of statistical methods and concepts used in the forecasting of cost and durations in construction projects.

The project life cycle, as described previously, lacks the phase of project implementation or post-occupancy planning. Construction projects, such as power plants, treatment plants, buildings, dams, factories and other types of structures, undergo a period of implementation. This paper explores the application of statistical tools for forecasting the running cost for operating and maintaining completed projects. In the context of this paper, operation and maintenance is treated as the last phase of the project life cycle prior to retiring the asset or terminating an initiative.

Furthermore, this paper reports on existing literature, concerning the application of statistical models by industry professionals, to determine duration of contracts, manage cost, and estimate

DOI: 10.4018/IJAIE.2020010101

cycle costs. Additionally, we will discuss the steps involved in developing regression and the neural network model. We compare the view of researches and results of regression and neural networks, as well as the advantages and disadvantages of each method.

The literature appears to share the same concept of significant items. Cost significant items refer to elements of the work that influence cost and duration when compared to others. The identification of these items is essential for developing models of high accuracy. Researchers have developed models for various project types, but implementing statistical principles to determine project cost and duration during the planning phase remains low among professionals.

The difficulty of estimating project cost and durations accurately results in a dilated schedule; longer durations will require additional resources, and project cash flows may be impacted. Nonetheless, computer technology advancement and the integration of applications and software have made statistics helpful in estimating project success factors, such as cost estimating, duration, and scheduling. The wide implementation of statistical approaches to estimating is hindered by cost of technology, skill level, and awareness.

Scholars agree that cost is a life cycle constituent of great importance. A poor estimate can make the difference between success and failure, so the accuracy of cost estimates and monitoring is essential to avoid cost overruns (Chew, 2017; Burnes, 2014; Galli, Kaviani, Bottani, & Murino, 2017). As a response to this problem, the industry is adopting modern simulation tools that are gaining ground, as computing power can handle large amounts of data and can perform advanced mathematical and statistical analysis.

A typical construction contract is coupled to a construction schedule that is legally binding. The duration of projects is determined by the owner and consultants. Once the contractor is selected and the contract is awarded, then the contractor must finish on or before the stipulated date. This research will report on the quantitative data processing methods and techniques, such as stepwise regression, multiple regressing, and artificial neural networks. Researchers agree that regression and neural networks are viable methods for estimating project cost and project duration (Adjei-Kumi, 2017; Marcelino-Sádaba Pérez-Ezcurdia, Lazcano, & Villanueva, 2014; Schwedes, Riedel, & Dziekan, 2017).

The current method of estimating durations represents a risk to the contractor, as there is a high probability that the project duration determined by the owner is not reasonable, which makes the contractor liable for penalties and liquidated damages (Jin, 2016; Zwikael, & Smyrk, 2012; Todorović, Petrović, Mihić, Obradović, & Bushuyev, 2015; Medina & Medina, 2015). Furthermore, unplanned durations increase litigation and compromises construction quality. Similarly, if duration is overestimated, then the client will incur damages, so the difficulties of projects are unique and require high-level customization.

Deviation from planned duration in developing parts of Africa ranges from 51% to 92%. This affects building and transportation projects that are reported to deviate from the stipulated completion date by 20%. Stakeholders, project owners, the government, and construction professionals will benefit from more accurate project durations (Mensah, 2016; Galli & Hernandez-Lopez, 2018; Easton & Rosenzweig, 2012; Brown & Eisenhardt, 1995). Overall, there is consensus in the literature that the duration of projects is strongly related to the quantities of work, rather than to the project cost. Popular scheduling computer programs are based on the Gant Chart, the critical path method. The Gant Chart illustrates the project schedule and the dependency relationship between items of work, and cost is not influential. Developed countries, such as the U.S Japan, Hong Kong, and countries in Europe, have developed models for estimating the construction duration of bridges, roads, and buildings (Ahmadu, 2015; David, David, & David, 2017; Galli & Kaviani, 2018; Hartono, FN Wijaya, & M. Arini, 2014; Parast, 2011).

Moreover, the literature improves the analysis of the quantities of work for non-significant items to be identified. Only the most important items are considered for study. Also, this paper will report on statistical significant work items that impact the duration of bridge construction projects. The completed activities will be the independent variables, and the goal is to understand significant work activities that greatly influence contract duration. This study aims to understand the development of models using regression and artificial neural network methods.

In addition to discussing the application of regression and neural networks for estimating duration of projects, we also explain the application of these statistical tools for the estimation of the life cycle buildings (Whyte, 2014; Sharon, Weck, & Dori, 2013). The literature emphasizes the common notion that the initial capital cost of projects represents a fraction of the operating and maintenance cost. A cradle to grave or whole-life approach is required to capture future expenditures accurately.

The concept of life cycle cost is important because decisions regarding elements of construction during design development will be selected based on operation and maintenance cost, rather than on the initial cost. Moreover, cost savings can be identified and applied before the project goes into the execution phase (when the implementation of design changes is costly).

According to PMBOK, the project life cycle is broken up into phases to make the project more manageable. Therefore, the design or planning phase creates the roadmap to a successful project completion and implementation. The estimation accuracy of a project's life cycle cost is heavily dependent on the amount of information that is available during the design phase.

The planning phase provides the opportunity to identify areas of waste, to enhance the design by improving specifications, to foresee challenging construction tasks, and to plan for efficient operation and maintenance once the project is implemented. Researches attribute the increased importance of life cycle cost to the development of simulation software, which allows the inclusion of estimating and managing facilities into the model.

The unique nature of construction projects makes them good candidates to benefit from the enhanced accuracy of statistical estimating tools. Primarily, the cost estimation of projects is reliant on the accuracy of the estimation of labor and material, and the cost estimate is subjected to the specification provided by the design or planning team (Whyte, 2014; Parker, Parsons, & Isharyanto, 2015; Nagel, 2015). Inaccuracy in cost estimates continues to affect the industry; estimators fail to recognize factors that affect accuracy and continue to rely on the aggregate value of estimates by generating estimators for various elements of the project. Effective methods for cost forecasting must be developed for the design intent to be met with reduced delays, and life cycle cost is estimated with greater confidence (Yamn, 2009; Papke-Shields & Boyer-Wright, 2017; Milner, 2016).

Clearly, the significance of the planning phase in project management is supported by research, primarily the cost estimation component. Authors state that cost estimation provides significant amounts of useful information in decision-making, regarding planning, resource allocation, and monitoring. The sophistication of construction systems and the request to reduce running cost drives the market to look for innovation. This is often found in digital estimating tools that can handle the large range of variables found in construction projects.

Ultimately, the goal of developing such estimating tools is to help solve the challenges of cost prediction. Researchers argue that simulation and statistical tools can be used to develop models that close this gap (Cheng, 2010; Svejvig & Andersen, 2015). Regression models are gaining popularity for determining the impact of significant factors on cost estimation. To select the most appropriate regression equation, the relationship between significant factors and cost estimation must be established. The literature indicates that modeling methods can address prediction problems in construction cost estimating (Whyte, 2014; Xue, Baron, & Esteban, 2016; Zhang, Bao, Wang, & Skitmore, 2016).

Artificial neural networks propose to solve estimating challenges by formulating the relationship between variables, but formulating or mapping the relationship between data points is challenging. Therefore, methods for identifying activities that are strongly related to cost factor must be applied. Interestingly enough, the Pareto rule is present in construction cost estimating, which contains about 80% of total construction cost that can be contained in 20% of construction activities. In developing the models, significant activities and non-cost activities will be applied.

This literature review focuses on the application of regression analysis and artificial neural models to predict the construction duration and life cycle cost of construction projects. Also, this investigation seeks to demonstrate the use of statistical models during the planning phase of the project to predict durations and cost. The selection of appropriate variables and simulations models in cost management will be discussed. The research papers reviewed for this report relate to the use of statistical tools in public works and commercial construction, but topics related to project life cycle apply to all project environments.

Background

Typically, the duration of construction projects is estimated poorly, with methods that are antiquated and unable to handle the complexity of modern projects. The critical path method is the prevailing tool for construction scheduling and activity planning. Unfortunately, the development of estimation models is delegated to teams that do not account for all factors. The type of equipment used, weather conditions, labor force, cash flows, the availability of materials, issuance of required permits, and other unforeseen conditions affect duration. Similarly, cost estimating is performed without accounting for significant cost factors, leading to profit loss for investors. In the traditional method, an estimator performs a quantity take-off to which a unit cost is assigned. The product will constitute a bill of quantity item, and the sum of the line items is the project cost. In this method, the project is subdivided into multiple sections, and each section is estimated individually by individual estimators. Then, all of the estimates are added to arrive at a project cost. Both durations and cost are difficult to estimate because the condition of each project is unique, and projects are customized.

Previously, research has been done to predict the lowest bidder for public schools. The studies concluded that for accurate results to be achieved, the appropriate number of significant factor must be used (Whyte, 2014; Xue, Baron, & Esteban, 2017; Winter, Andersen, Elvin, & Levene, 2006a; Von Thiele Schwarz, 2017). Also, the building sector applied the multiple regression method to develop a model to estimate the duration of housing projects (Jin, 2016; Yun, Choi, Oliveira, Mulva, & Kang, 2016). The research indicates that neural networks can minimize uncertainties when estimating building systems and some authors have conducted studies that demonstrate that the neural networks are more accurate than regression techniques (Elkassas, 2011; Xiong, Zhao, Yuan, & Luo, 2017).

Additionally, scholars in Nigeria and Kuwait have developed regression models to predict the duration of building construction. Duration models were constructed using buildings characteristics or elements of construction as the major determinants (Jarkas, 2015; Shenhar & Levy, 2007). Nonetheless, neural networks cannot show the relationship between the predictor with the outcome, and regression cannot assist in selecting a cost model that fits the variables to an established level of accuracy (Wheaton, 2009; Usman Tariq, 2013; Sutherland, 2004).

However, scholars in Bosnia, Korea, Vietnam, have developed artificial network and regression models to predict construction duration (Petruseva, 2013). Probabilistic models have also developed models to predict risk and its influence in public works. The sample size used for developing models was greater than 30 in all cases. If the sample size is sufficiently large ($n>30$) and the sample data is approximate the normal distribution, then sample measures and data description will closely approximate that of the population.

In the United States, the Department of Transportation of North Carolina analyzed over 400 bridge projects to determine the duration of the design development phase. The researcher determined a number of factors that had an influence in the duration of the engineering phase (Liu, 2012; Ahern, Leavy, & Byrne, 2014). Of the factors identified, four were significant: 1) geographical location of project, 2) environmental requirements, 3) scope of work, and 4) team in charge of assembling environmental documents. The study was validated using a sub-sample drawn from the larger sample. The result was a set of regression equations that the DOT can use to estimate the duration of the engineering and design phase.

Governments and administrations can benefit from statistical models to more accurately estimate duration during the second phase of project life cycle, the planning stage. An accurate prediction of project durations minimizes completion delays and maximizes the use of project resources. Other research points to Neuronal networks as a more accurate method, and it is argued that neural models can predict construction cost without detailed drawings. This notion deviates from current practices; projects must reach a high development level for an accurate estimate to be generated (Arafa, 2013; Al-Kadeem, Backar, Eldardiry, & Haddad, 2017a; Andersen, 2014).

Overall, this research review contributes to the understanding of effectively integrating statistical concepts into construction project planning. The body of this paper is framed to study the integration of statistical models in the construction industry against the background of current estimating methods.

Problem Statement

Project managers in the construction industry are faced with challenges in the second phase of the project life cycle: planning. The planning phase includes three vital parts: 1) estimating the durations of projects, 2) project cost, and 3) the cost of running buildings after completion. Traditional methods prevail in the industry, but these methods are being displaced by statistical models that can make predictions more accurately. This is an area of concern to investors and managers because it directly impacts profits. Also, the planning stage of a project is vital because it sets procedures for monitoring the execution of the project. Hence, the ability to accurately estimate duration, project, cost, and operating cost is vital for project success. This paper seeks to shed light on the benefits of statistical tools, as well as to provide a background and practical examples of regression and neural networks.

Research Hypothesis

Statistical models can be effectively integrated into a project life cycle. Artificial networks, regression, and modeling will be explored to confirm the effectiveness of statistical models applied to project planning. Regardless of the poor performance, traditional methods continue to be widely used. If the findings reveal that statistical models are more reliable in estimating durations and life cycle cost, then the industry should consider investing in resources that would allow the adoption of statistical models.

Research Objective and Research Gap

For the most part, literature addresses how these variables, concepts, and models are necessary in project management and performance, but certain information is not addressed. There can be more literature on why these variables, their concepts, and models ease the progression of project management and performance. Thus, a research gap has developed. The variables, concepts, and models will be assessed to find what they share and what they do not. As a result, this will allow for a universal framework that features their best aspects. In this study, there are evidence-based answers for primary questions about these variables, their concepts, and models, such as how to maximize on them for project management and performance objectives. This study can act as a reference for future research, as well.

Originality

The originality of this literature review is expressed to explore the reach of statistical estimating methods in the construction industry. In addition, the existing literature does not mention the traditional method and anecdotal approaches to estimating. This paper heavily focuses on the application of statistical models, but it also touches on traditional estimating methods and drawbacks. Furthermore, this literature review aims to synthesize research performed on the topics of 1) statistical models, 2) traditional methods, and 3) simulation techniques.

This study will contribute more literature to expand on the effectiveness of these variables, their concepts, and models and their likenesses and differences. Studies that have tested this paper's hypotheses contribute information, as well. This study uses a design-science-investigate strategy, and it then approves a valuable growth reveal to apply reasonably and hypothetically. In the end, this study provides an assessment model for these variables, their concepts, and models. The evaluation instrument is emphasized as responses to the examination question, and the instrument is reviewed with an explanation of the approach to the outline and results of the meetings. The conclusion features primary findings and ideas to arrange investigative limitations and future studies.

Contribution to the IE/EM/PM

Clearly, this paper contributes to the field of project management and engineering management because it explores non-traditional methods for estimating cost and duration of construction projects. Also, this study introduces the combination of statistical tools and simulation techniques for better duration and cost estimation. These research findings show the benefits of these variables, their concepts, and models. However, this study depicts the detriments of not addressing performance and sustainability. This study also highlights the need for real-life examples, as it features examples to apply these theories to the real world.

Managerial Relevance

Engineering managers must always make decisions, which will become even more important in the future of project management and engineering. This study addresses such a future and the role of the engineering manager and the engineering management field. Furthermore, the implications are addressed within different organizational levels, such as corporate level, managerial level, and project team level. Also, an engineering management practitioner can utilize the conclusions to capitalizing on these variables, concepts, models, and their relationship within project environments and operations. The variables, their concepts, and models will be assessed to propose a framework that will fill a research void in pre-existing literature. Many different business subjects can be enhanced by this study, as it contributes to each body of knowledge with novel approaches for future research.

Paper Organization

This paper begins with section two, which features the literature review of literature in these fields of research. The research methodology is presented in section three, and the findings and analysis are in section four. Lastly, the implications are outlined in section five for practitioners, but it also features ideas for future research, limitations, and general conclusions.

LITERATURE REVIEW

Multiple Regression Cost Prediction Application

Construction projects present complex estimating challenges. There are many variables or factors that can affect the outcome of a project, as conditions are unique and products are personalized. The diversity of challenges in construction projects makes it difficult to predict durations and costs. Hence, to find a practical solution to the estimation of project duration and running costs, an equation that can handle multiple variables is required. In this case, we are interested in the variables that influence the mean duration and running cost of projects.

Before we introduce research information on multiple regression, basic definitions and concepts should be discussed. The multiple regression model is an extension of the simple linear regression analysis. The linear regression only uses two variables: a dependent variable, such as cost, and an independent variable, such as building type. We say a regression is linear when the relationship between the independent and dependent variable is linear. With this technique, we try to explain the variation in the dependent variable.

Now that we have explained the basic concept of linear regression, we can discuss multiple regression. We will begin with introducing an experiment performed by a researcher to determine the operation and maintenance cost using 20 construction projects. The data was taken from three separate sources. The significant cost factor affecting the cost of operation and maintenance had been established from previous research (Whyte, 2014; Arumugam, 2016; Badi & Pryke, 2016). Furthermore, the parameters of the experiment indicate that 11 cost significant items were identified as most influential, and these items applied to all buildings.

We will also introduce the definition of neural network, as the following example will compare the accuracy of both methods. Artificial neuron network models are nonlinear and find the complex relationship between inputs and outputs. Artificial networks differ from regression models, in that they employ the concept of machine learning to reach the desired outcome (Ontepeli, 2012; Besner & Hobbs, 2012; Cova & Salle, 2005; Detert, 2000).

Existing publications indicate that the running cost of building is approximately 70% of the building's life cycle cost. This measure is the result of an 18-year study done on the operation and maintenance cost of buildings. Also, 7 non-cost factors that are reported as having an influence on the estimation cost are also established by an existing study.

Overall, eight factors will be used as the input for the multiple regression model. These inputs are the nodes that will be linked to the hidden layers via weight connection. We will later perform the same experiment, but by using the neural network method. The multiple regression model equation for the mentioned outputs will be constructed as follows:

$$Y=a+b_1x_1+b_2x_2+b_3x_3+b_4x_4+\dots b_nx_n + u$$

Where:
Y= The variable that we are trying to predict (DV)
X= The variable that we are using to predict Y (IV)
a= The intercept
b= The slope (Coefficient of X1)
u= The regression residual error

In our equation, *a* is the intercept b_1-b_n. The R-squared value is a measure of how close a data point is to the fitted regression line. R-squared value of 0% indicates that the model cannot explain the variability of the data around the mean. Similarly, p-values are key for the evaluation of the model. A p value $< .05$ means you can reject the null hypothesis. Also, if a predictor has a low p-value, then it is most probably a significant addition to the model.

Resuming our model, the R-squared values indicate the percentage cost variability. If our R-squared value is close to or equal to 1, then there is a strong correlation between the model output and the actual value. Moreover, p-values were used to determine if changes in the predictor resulted in changes in the response variable. Those variables that were not deemed significant were eliminated. Thus, the research argues that variables with a p-value less than .05 should be included in the model (Ontepeli, 2012; Eskerod & Blichfeldt, 2005; Galli, 2018c).

In regression analysis, we have the option to bring all possible independent variables into the model in one step using full regression, but the case selected for illustration utilizes a stepwise method to identify significant variables. In this method, we determine the p-value for each of the eight variables originally identified. Any variable with a p-value greater than .05 is eliminated until only the significant variables are remaining.

At this stage of the experiment, we move on to regression validation to decide if the results are acceptable descriptions of the data. Validation can be done by determining the aptness of the model (Sosmez, 2010; Galli, 2018a; Galli, 2018b). However, our research utilizes the cross-validation method. From the 20 projects, 17 are used to develop the model and 3 are used to validate the model. The model performance will be measured using the mean percentage error. This value is the calculated average of all percentage errors by which the predictions differ from the actual value. For this calculation, we use the difference of the actual output and the predicted/model output.

Neural Networks Cost Prediction Application

As discussed earlier, neural networks utilize a hidden layer called the black box, as the relationship between the variables is determined. The inputs are selected and are processed through the back box; the model value is then generated. In this section, we will consider the same scenario presented in the previous regression model, where 20 construction projects are being studied to estimate the operation and maintenance cost.

In our model of 20 projects, the goal is to optimize the model and to reduce the neural network weighted error to zero. To reduce the weighted error, the weights will be adjusted from input to black box and from black box to outputs. A neural model can be trained to predict outcomes. Optimizing one node at a time can turn into a daunting task, so the process of training begins by guessing or selecting a random weight. The process is then allowed to run, and the resulting deviation is the analyzed and then weights adjusted. Finally, the process is repeated.

The number of hidden layers can be calculated using traditional parametric methods. As previously mentioned, during the training process, the weights and hidden number in the black box are adjusted to find a model that will result in the lowest value of mean percentage error and root mean square. The sigmoid curve is used to analyze the neural model.

Similar to the regression model, the 20 projects are divided in two sets, while 17 projects are used for model development and the remaining set for testing the procedure. It is worth emphasizing the importance of the training the model for the best arrangement of the neurons to be identified. One must keep in mind that acceptance of the model is judged based on the root

mean square and mean percentage error values.

In this section, we will talk about the connection weight method that is used to determine the importance of the input. Our inputs will be the eight variables, which are also called predictors, and our outputs are the running cost (Olden, 2014; Gimenez-Espin, 2013; Hoon Kwak, & Dixon, 2008). In this step, we calculate the sum of the products from the output to the hidden node and then from the hidden node to the output. We determine the importance of a node based on how large the sum of the products of the weights. Also, a larger number means the more influence on the corresponding input value.

Considering both the regression model and the neural network models are the same, a paired test was conducted to compare the accuracy of the two estimation methods. The rules were set up as follows:

H_0: No accuracy difference between the two methods
H_1: There is difference in accuracy between the two methods

The comparison indicates that neural network models are more accurate than regression models. The neural network model is argued to predict the cost of operation and maintenance with high accuracy, while the regression model is less accurate. The difference may seem insignificant, but we must pay attention to the minimum number of variables required to develop a model. The lowest number of variables required mans that the technique would be more useful for practitioners during the planning stage of a project. Accurate estimates generated early during planning will transcend and have positive impacts on the project outcome.

Regression and Duration Estimation

In this section, we will demonstrate the application of multiple linear regression and artificial neural network for the estimation of project durations. For the development of this model, a sample of 30 completed bridges was taken. From these projects, the bill of quantities was collected. Information on the bill on quantities will be used to determine the items that can potentially affect the duration of projects (Czarnigowska, 2014; Labedz & Gray, 2013; Lee Lapira, Bagheri, & Kao, 2013).

The work items that were selected as inputs are elements of construction and characteristics of the bridges, and these elements of construction represent work activities that have been billed. Overall, 11 activities were selected and analyzed statistically, while the items selected from the bill of quantities represent the independent variables and are analyzed. Then, descriptive statistics of mean and standard deviation are calculated. For the purpose of the analysis, the scheduled and actual completion data were secured. These parameters will be used to compare calculated predictions.

Multiple regression analysis is gaining popularity among practitioners, but scholars use it to advance models to predict project durations. The formula for the duration estimation is identical to the regression formula used in cost prediction in the previous example. In this case, the cost of maintenance will be replaced by the "T" for time duration: $T=a_0+a_1x_1+a_2x_2+a_3x_3+a_4x_4+\ldots a_nx_n + u$.

"X" represents the independent variables, and "a" is the regression parameter.

The literature explains the effects of multicollinearity and how to determine if the model appears reasonable. Multicollinearity can be a problem when two independent variables provide duplicated information to the model. When highly correlated independent variables are part of the model inputs, then the regression results can be negatively impacted. The standard error of the regression coefficients is increased and impacts the accuracy of the regression model (Jarkas, 2015; Loyd, 2016).

Although there are disagreements, some researchers argue that if the inflation factor is greater than 10 between two pairs of independent variables, then these are regarded as high correlation or multilinearity. In the analysis under study, a variable inflation value exceeding 2 is enough evidence of multicollinearity. The model under examination consists of 30 completed projects to develop project duration models; the regression model designated 80% of the projects for model development and 20% to validate the model (Irfan, 2013).

For the 30 projects, four models are developed using stepwise regression to identify the significant variables. The process identified four significant work items from the bill of items. Also, the model selected will be the one with an R-squared value closest to 1, which indicates the most accuracy in predicting duration (Nani, 2017). This result, if less than zero, would suggest that our best model has a percentage of variance between predicted value and actual value. The R-squared is the proportion of total variation in the significant variables.

Conversely, the R-squared value would indicate the percentage of variation of the dependent values that are accounted for: 1) material transportation and handling, 2) cast in place concrete, 3) weight of structure, and 4) underlayment subbase. For the standardized coefficients of the model to be considered significant, their respective p-values must be lower than .05. The variance inflation factor in good models, according to scholars, should be lower than 2. This process will confirm that multicollinearity between variables was removed during the stepwise regression process. Also, the R value of the model selected must be checked, as it indicates the relationship significance between the predicting variables and the duration of construction. The result is a mathematical expression to estimate time, "t":

T= Standard Coefficient + Bata Coefficient X (bill of quantity)

Neural Networks and Duration Estimation

Artificial neural networks, as mentioned earlier, are mathematical models that use nodes as processing units in the form of layers. A simple model would contain three layers: the input layer, the hidden layer and the output layer. Also, all layers of nodes are connected with weights (Afrifa, 2013). The node value and the weights are adjusted until the desired outcome is achieved. The significant quantities observed during the regression analysis will be entered as the input layer, the hidden layer, and internally will represent the issue at hand. Lastly, the output layer will provide the time duration prediction.

The artificial neural network described its termed multilayered perceptron, and it is characterized by comprising three layers of nodes. Furthermore, all nodes, except the input layer, are neurons that employ a non-linear activation function. In this perceptron, the quantity of hidden layer is determined by trial and error during the model training phase. The neurons in the hidden layer, also called the black box, are activated by a predetermined function and generate an output (Ahiaga-Dagbui, 2013).

The artificial neural network is sensitive to the amount of data used for training. Researchers agree that the ability of the model to accurately make predictions is affected by training data. The accuracy of the model can be assessed by calculating the mean absolute percentage error, MAPE. The proportions for training and testing are consistent with proportions used by previous scholars in developing neural models. In the case illustrated in this paper, 22 projects will be used for training and 8 for validating the model. This represents 75% and 25% respectively.

The hypotheses developed for comparing the regression and artificial neural model are as follows:

1).
H_1: There is no significant difference between actual mean and predicted duration
$H_{11:}$ There is significant difference between actual mean and predicted mean of duration
2).
H_2: There is no significant difference between the planned and actual duration means
$H_{22:}$ There is significant difference between the planned and actual duration means

The artificial neural network was developed using 22 projects, and 8 projects were used for validation. As in the regression model, four independent variables were used: material transportation, cast in place concrete, weigh of structure, and underlayment subbase material. A total of five models are developed, while the R, R-squared, MAPE, and average accuracy are calculated to compare the accuracy of the artificial neural model (Nani, 2017). R-value indicates the strength of the relationship between the variable and the outcome, and the R-squared value the determination. The selected model should have the highest accuracy level and the lowest mean absolute percentage error (MAPE) (Asiedu, 2017).

Simulation Techniques

Simulation methods involve mathematical and logic models that rely on 1) equations with known and specific output values and 2) random variables. The result is a graphical representation of construction activities (Chew, 2017). Simulation is used to study the influence that variables or uncertainties have on project planning. Also, simulation can be a powerful tool during the planning stages of a project, namely duration estimation, scheduling, cost estimating, risk assessment, and resource loading (Jahangirian, 2011).

Currently, the construction industry relies on commercial computer software, such as Cristal Ball, Innovaya, Monte Carlo, Revit, and Bentley (Liozou, 2013). These programs are used to model uncertainties and to help construction professionals make decisions. The Monte Carlo simulation technique is one of the pioneering simulation tools applied in the construction industry. The simulation is based on probability distribution, and statistical sampling.

Furthermore, this method uses uncertainties as inputs and will then predict different scenarios of risk by randomly selecting inputs to estimate outputs that represent the desired solutions (Grinstead, 2013). In the planning phase of construction, this method can be useful to estimate construction cost and/or durations. Monte Carlo relies on random sampling to calculate values that are plotted to form a uniform distribution. In our study, construction project variables are used to generate values for the variable, and a computer will repeat this process multiple times to determine the distribution of the project (Potts, 2011).

Another simulation method used in construction is 4-diamentional. The 4th dimension in construction modeling is the schedule, allowing the construction to be simulated before actual project execution. The literature states that although cost values cannot be linked to the model, the simulation of time provides further insight to managers during the planning phase of the project. The modern computer aided drawing allows the creation of multi-layered models that display the different building systems, such as plumbing, electrical, and mechanical (Latiffi, 2013).

RESEARCH METHODOLOGY

Literature Review Research Approach

Two steps went into the literature review. Step one involved searching for relevant information, which included inputs from keywords. Step two was more structured, as it involved the use of databases and search strong for the review process. The tables of contents were also searched through, as two journals were applicable.

Part 1: Explorative and Unstructured Literature Review

This study aims to reconsider certain keywords, so publications that reflected the keywords were assessed. This led to 31 applicable journal articles and 7 books. With the 38 publications, the keywords were studied to be search terms in the structured review.

Part 2: Structured Literature Review

In general, this step involved a structured and systematic approach to conduct reviews, which was taken from other literature. Four phases went into this step, as phase one entailed preparation and scoping. Phase two was the review planning, and phase three was the search, evaluation, and selection of literature. Phase four entailed evaluating the literature.

The phase (1) review scope was based on key concepts on projects, marketing, and strategic planning. This phase promised to result in adequate data from journals that would contribute to the study.

To gain more information, phase (2) involved connecting other concepts to the keywords. Other concepts were the keywords and their relationship and interaction. Some concepts were too vague, such as success, evaluation, and impact because their results were not practical.

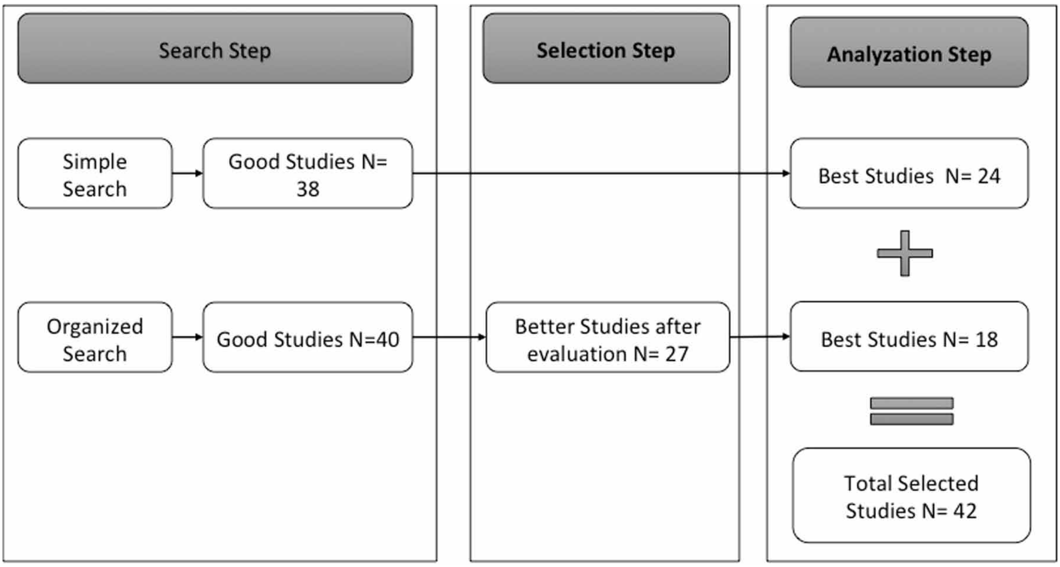
Then, phase (3) began with a successful compilation of applicable results from many databases. Some databases included ProQuest, Business Source Complete, EBSCO, and ScienceDirect. As a result, 15 conference papers and 25 results were found that reflected the journals. Overall, 40 results were compiled, which featured conference papers and the journal entries.

The search ended with looking through the tables of contents for tier 1 and tier 2 journals. These journals were academic and practitioner-based, as any pertinent article was applied that may not have matched the keywords. Below, Figure 1 illustrates that there were three streams to the search and selection phase: the explorative and unstructured search, the structured search with search strings, and the tables of contents search.

Applying the three streams condensed the results to 42 publications. With the selection process, between 24 and 18 results were collected with a concentration on results from academic journal articles, literature reviews, conference papers and proceedings, and books. Triangulation methods were utilized, while the first selection should indicate a link between the resulting publications and the keywords to the project research. Also, the evaluation was executed with inclusion and exclusion criteria that highlighted the abstract, as some publications used the introduction or the entire paper.

For phase (4), the collected information was organized into an inductive and deductive analysis. This was then documented with a software package, as the deductive analysis involved documenting the author's university and country. To indicate the research genre, empirical research, theory development, research essays and literature reviews, or the category of "other" were used. With proof that the publications utilized theoretical frameworks, the deductive coding was added. Such frameworks included a research-based perspective and contingency theory. Lastly, there was an indication if the publication had a model.

Figure 1. Research approach for literature review



A grounded theory approach was utilized for the inductive analysis to code publications with open and selective codes. Since the annual number of citations was the basis of most of the selected publications, older aged publications were balanced out. Vital literature reviews were applied, as well as some current publications that contribute to the keywords research.

Furthermore, phase (4) involved the creation of certain key themes by studying the list of open codes. These would then be collected into axial and selective codes. Throughout April and August of 2018, parts 1 and 2 of the literature review took place. Also, there was a final evaluation performed during this time for pertinent data and how they overlap.

The collection of these papers has revealed that there are some key themes shared between the variables, concepts, and models. By performing a statistical analysis and investigation of other variables or factors, this study's research conclusions were given more weight. The following section includes Table 1, which includes the 42 studies and key themes.

Assessing the 42 studies showed that the literature evaluated the keywords with many statistical methods that took relational and causal perspectives. This added significance to the research conclusions, as well. In Table 2, the statistical methods for the 42 studies were summarized. Additionally, Table 3 summarizes the factors or variables that were assessed in the journals.

The subsequent section features the findings of the research methods. These findings are rooted in the themes and topics that are featured later within this study.

FINDINGS

The findings section will be formatted in the corresponding order with the sections and topics covered in the literature review section of this paper. Multiple regression and neural networks modeling techniques were applied in construction cost estimation; regression and neural models were to estimate project durations, and then there was simulation for cost management.

Before discussing the results of the literature review, the data collection methods and the instruments for collecting data will be discussed. The experiments and studies reviewed for this report collected sample data from 1) bill of quantities and 2) costs index data. There are also two inherent

Table 1. Identified studies from research approach by theme

Theme #1	Theme #2
Adjei-Kumi (2017) Ahern, Leavy, & Byrne, (2014) Arumugam, (2016) Cova & Salle (2005) David, David, & David, (2017) Eskerod & Blichfeldt, (2005) Galli & Kaviani, (2018) Galli et al., (2017) Hartono, FN Wijaya, & Arini, (2014) Jarkas, (2015) Xue, Baron, & Esteban, (2016) Xue, Baron, & Esteban, (2017) Andersen, (2014) Schwedes, Riedel, & Dziekan, (2017) Petruseva, (2013)	Afrifa, (2013) Al-Kadeem et al., (2017a) Badi & Pryke, (2016) Irfan, (2013) Liu, (2012) Medina & Medina, (2015) Milner, (2016) Parast, (2011) Parker, Parsons, & Isharyanto, (2015) Sosmez, (2010) Sharon, Weck, & Dori, (2013) Shenhar & Levy, (2007) Yun, et al. (2016) Gimenez-Espin, (2013) Kwak, & Dixon, (2008) Sutherland, (2004) Wheaton, (2009)
Theme #3	**Theme #4**
Ahiaga-Dagbui (2013) Arafa, (2013) Cheng, (2010) Detert, (2000) Easton & Rosenzweig, (2012) Elkassas, (2011) Galli, (2018c) Galli & Hernandez-Lopez, (2018) Grinstead, (2013) Jin, (2016) Liozou, (2013) Svejvig & Andersen, (2015) Todorović et al., (2015) Labedz, & Gray, (2013) Olden, (2014) Ontepeli, (2012) Lee et al., (2013) Potts, (2011) Usman Tariq, (2013) Von Thiele Schwarz, (2017) Yamn, (2009) Zwikael & Smyrk, (2012)	Ahmadu, (2015) Asiedu, (2017) Besner & Hobbs, (2012) Brown & Eisenhardt, (1995) Burnes, (2014) Chew, (2017) Czarnigowska, (2014) Galli, (2018a) Galli, (2018b) Jahangirian, (2011) Latiffi, (2013) Xiong et al., (2017) Winter et al., (2006a) Loyd, (2016) Marcelino-Sádaba et al., (2014) Mensah, (2016) Nani, (2017) Nagel, (2015) Papke-Shields & Boyer-Wright, (2017) Whyte, (2014) Zhang et al., (2016)

problems with the source of the data: 1) inaccuracy of quantity initial estimate and 2) price indexes estimated probabilistically.

Construction estimates are not accurate, and the quantities recorded on a bill of quantities differ from the actual values. The final cost of projects can be predicted with more accuracy than the components of the total cost. Literature indicates that construction estimates are incorrect 17% of the time (Jarkas, 2015). Thus, collecting and using poor data quality affects the variables and inputs that are selected for the regression and the artificial neural network.

Estimates in construction and other project-based industries continue to be estimates and not exact predictions of cost and duration. Traditionally, estimators add a safety factor to the estimate as contingency, but this contingency is based on the estimators experienced and not on actual project cost records. This precisely is the problem that statistical tools and concepts aim to solve for the construction industry.

Table 2. Systematic analysis results by statistical analysis method

Statistical Method	Number of Articles (Frequency)	Author(s)
Regression	17 (22.97% of total articles)	Adjei-Kumi (2017) Asiedu, (2017) Chew, (2017) Cova & Salle, (2005) David, David, & David, (2017) Detert, (2000) Easton & Rosenzweig, (2012) Elkassas, (2011) Galli et al., (2017) Gimenez-Espin, (2013) Irfan, (2013) Loyd, (2016) Nani, (2017) Petruseva, (2013) Sutherland (2004) Xue, Baron,& Esteban, (2017) Zwikael & Smyrk, (2012)
ANOVA	13 (17.57% of total articles)	Afrifa, (2013) Ahern, Leavy, & Byrne, (2014) Ahiaga-Dagbui (2013) Brown & Eisenhardt, (1995) Cheng, (2010) Galli, (2018b) Galli, (2018c) Latiffi, (2013) Nagel, (2015) Papke-Shields & Boyer-Wright, (2017) Potts, (2011) Xiong et al., (2017) Yun, et al., (2016)
Q-Test	13 (17.57% of total articles)	Arumugam, (2016) Badi & Pryke, (2016) Czarnigowska, (2014) Grinstead, (2013) Kwak & Dixon, (2008) Jarkas, (2015) Jin, (2016) Labedz & Gray, (2013) Olden, (2014) Parker, Parsons, & Isharyanto, (2015) Schwedes, Riedel, & Dziekan, (2017) Usman Tariq, (2013) Von Thiele Schwarz, (2017)
t-Test	15 (20.27% of total articles)	Ahmadu, (2015) Andersen, (2014) Arafa, (2013) Besner & Hobbs, (2012) Burnes, (2014) Eskerod & Blichfeldt, (2005) Galli, (2018a) Winter et al., (2006a) Hartono, FN Wijaya, & Arini, (2014) Lee et al., (2013) Sharon, Weck, & Dori, (2013) Shenhar & Levy, (2007) Sosmez, (2010) Yamn, (2009) Zhang et al., (2016)
Chi-Square Test	17 (22.97% of total articles)	Al-Kadeem et al. (2017a) Galli & Kaviani, (2018) Galli & Hernandez-Lopez, (2018) Jahangirian, (2011) Liozou, (2013) Liu, (2012) Marcelino-Sádaba et al., (2014) Medina & Medina, (2015) Mensah, (2016) Milner, (2016) Ontepeli, (2012) Parast, (2011) Svejvig & Andersen, (2015) Wheaton, (2009) Whyte, (2014) Todorović et al., (2015) Xue, Baron, & Esteban, (2016)

Similarly, cost data indexes are probabilistically determined. Costs are determined based on representative samples. Sampling errors are inherent in probabilistic methods of estimation, so there is a measurable level of uncertainty in the variables selected for model development. Since the same

Table 3. Systematic analysis results by number of variables studied

No. Factors Studied	Number of Articles (Frequency)	Author(s)
1	14 (18.92% of total articles)	Andersen, (2014) Arafa, (2013) Besner & Hobbs, (2012) David, David, & David, (2017) Galli, (2018b) Grinstead, (2013) Latiffi, (2013) Lee et al., (2013) Medina & Medina, (2015) Nagel, (2015) Papke-Shields & Boyer-Wright, (2017) Sutherland, (2004) Von Thiele Schwarz, (2017) Yun et al., (2016)
2	14 (18.92% of total articles)	Afrifa, (2013) Ahiaga-Dagbui (2013) Ahmadu, (2015) Al-Kadeem et al., (2017a) Brown & Eisenhardt, (1995) Cheng, (2010) Elkassas, (2011) Galli & Hernandez-Lopez, (2018) Jahangirian, (2011) Jarkas, (2015) Shenhar & Levy, (2007) Sosmez, (2010) Petruseva, (2013) Xue, Baron, & Esteban, (2016)
3	17 (22.97% of total articles)	Arumugam, (2016) Badi & Pryke, (2016) Chew, (2017) Czarnigowska, (2014) Eskerod, & Blichfeldt, (2005) Galli et al., (2017) Gimenez-Espin, (2013) Irfan, (2013) Liu, (2012) Loyd, (2016) Marcelino-Sádaba et al., (2014) Mensah, (2016) Nani, (2017) Svejvig & Andersen, (2015) Usman Tariq, (2013) Winter et al. (2006a) Zhang et al. (2016)
4	11 (14.86% of total articles)	Adjei-Kumi (2017) Ahern, Leavy, & Byrne, (2014) Detert, (2000) Easton & Rosenzweig, (2012) Galli, (2018a) Hoon Kwak & Dixon, (2008) Labedz & Gray, (2013) Parast, (2011) Potts, (2011) Todorović et al., (2015) Zwikael & Smyrk, (2012)
5	9 (12.16% of total articles)	Asiedu, (2017) Burnes, (2014) Cova & Salle, (2005) Galli, (2018c) Jin, (2016) Liozou, (2013) Sharon, Weck, & Dori, (2013) Wheaton, (2009) Xue, Baron, & Esteban, (2017)
6	10 (13.51% of total articles)	Galli & Kaviani, (2018) Hartono, FN Wijaya, & Arini, (2014) Milner, (2016) Olden, (2014) Ontepeli, (2012) Parker, Parsons, & Isharyanto, (2015) Schwedes, Riedel, & Dziekan, (2017) Whyte, (2014) Xiong et al., (2017) Yamn, (2009)

data samples were used to develop the regression and neural network models, the effects of standard error will impact the predictive accuracy of both models.

The main results of the literature review, related to the hypothesis statement, indicate that statistical analysis tools can be successfully integrated into the planning life cycle stage of construction. There are disadvantages found in random sampling, such as sampling bias, in which some data points are less likely to be selected during sampling. Nevertheless, statistical tools can be applied to engineering management and project management fields.

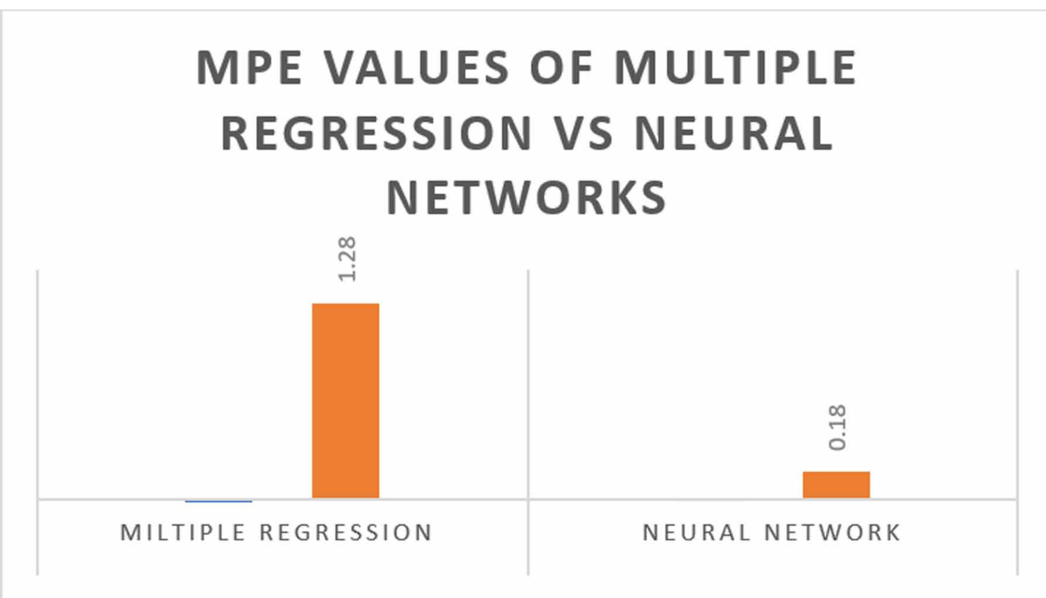
Furthermore, in the literature review, the authors demonstrate the leading statistical methods used in construction for cost and duration estimation. Moreover, the authors demonstrate the performance of each modeling technique and rank the predictive accuracy of each modeling techniques. The literature argues that artificial neural models have superior predictive accuracy over multiple regression.

The inability to explain or to trace the steps of the hidden layer creates skepticism among practitioners, which affects the acceptance of findings, its implications, and the practical application in the field. The hidden layer can approximate the statistical algorithm or function, but the link between the weight and the node, or the axon and the neuron, cannot be studied. Thus, no insight is gained concerning the relationship between the input variables and the outputs. Two distinct artificial neuron models can produce the same result, but there are different connection weights. The inability to produce a formula or equation that will always produce the same result is problematic. Below, Table 4 and Figure 2 show the calculated accuracy of the cost estimating models developed:

Table 4. Accuracy of cost estimating models

		MPE	Standard Deviation
Multiple Regression	Training	-0.02	1.67
	Testing	1.28	1.77
Neural Network	Training	-0.01	0.38
	Testing	**0.18**	0.31

Figure 2. Multiple regression vs. neural networks



In our cost prediction studies, the mean percentage error in the testing stage for the neural network was significantly lower than that of multiple regression. The mean percentage error for regression is 1.28 compared to 0.18 of the neural model. It is demonstrated by this literature review that artificial neural models not only work well in predicting cost, but they also display higher predictive accuracy. In our study, the neural network can predict cost with an accuracy of 99.8% (Whyte, 2014).

In the literature review, we discussed the importance of ranking the independent variables, such that only cost significant items are selected as inputs. Given the accuracy of the developed model, the literature demonstrates cost items selected as significant are influencing the output values. In other words, presumed cost significant items influence the life cycle cost of buildings.

In the regression model, the stepwise method was applied to the input variables to select those influencing the model. Also, the aim of the stepwise method is to optimize the model. The p-value for each variable is determined, and the value of coefficient of correlation is calculated to determine the relationship strength between variables. A low value means a low association. Also, a value that is greater than zero means that the variable is both significant and positively associated to the outcome. As the value of the independent value increases, the value of the output is also affected.

Consequently, the best regression model is identified. The literature indicates that the most accurate regression model has a mean percentage error of 1.28, compared to the same statistic for the neural network model, 0.18. This result tells us that the regression model can predict the life cycle cost of buildings with an accuracy of 98%. Despite the evidence that artificial neural models are more accurate than regression model, the results confirm the adequacy of statistical models in the prediction of project costs. Therefore, the null hypothesis is rejected, as there is a difference in the predictive accuracy of the two models.

H_0: No accuracy difference between the two methods - REJECTED
H_1: There is a difference in accuracy between the two methods

Overall, our study constituted four models for the multiple regression analysis and five models for the artificial neural network analysis. The stepwise method was used to determine the accuracy of the regression models. After the process, regression model four was selected, as it has a calculated R-square value of 0.71. This value is higher than models 1, 2, and 3, and it represents the percentage of variance that is explained by the model. The selected model also has a calculated R of 0.84, while the calculated R-value confirms the significance of the predictors used to build the model. Thus, the input variables do influence the duration of projects. Table 5, below, summarizes the R-squared values for each of the models.

The four regression models under consideration have a mean and median R-squared value 0f 0.53, which means that on average, the level of accuracy of the four models considered is approximately 50%. The standard deviation is 0.15, which represents the extent of deviation for the group of models. The model with the least accuracy has an R square value of 0.36, and the highest was the selected model, model four.

The difference between the actual value and the predicted value show no indication of autocorrelation, as the study shows that the variation inflation factor for all models was less than 2. Variation inflation models with values less than 2 are assumed to have no collinearity, and that predictor is not influencing each other. A model where collinearity between variables exists will exhibit a high variance. Variables found to have a linear relationship were eliminated from the model during the stepwise process. Table 6 highlights the R-squared descriptive statistics for this model:

In the case of neural network models, five models were developed. The aim is to select the model with the MAPE closest to zero and the highest percentage of accuracy. The MAPE score was as follows: 6.6, **4**, 4.6, 5.8, and 7.9 for models 1, 2, 3, 4, 5, respectively. Also, the validation test was 26.8, 26.0, **27.7**, **27.7**, and 26.1 for models 1, 2, 3, 4, 5, respectively.

Table 5. Summary of R-squared values for each model

Model	R-Square
1	0.36
2	0.47
3	0.58
4	**0.71**

Table 6. R-squared descriptive statistics

R-Square Descriptive	
Standard Deviation	0.15
Mean	0.53
Standard Error	0.08
Median	0.53

From the results, we can see that model two has the lowest MAPE score, so it was selected. However, models three and four indicate a higher percentage of accuracy, while model two is selected based on MAPE criteria. The selected model has an R value lower than the other four models considered. This indicates that variables in models not selected have a stronger relationship with the outcome value.

DISCUSSION

Implications to the Field of Engineering Management and Project Management

Cost and duration estimating has always been a challenging task in the planning phase of construction projects. Poor estimates lead to cost overruns, and the dilation of project schedules impacts project success and investor's profits. Planning is the roadmap to project success, but if planning is not done properly, then the roadmap can lead to project failure. Also, resource management is particularly impacted by poor duration and cost estimation.

The industry has systems and methods in place for scheduling and estimating project costs. Large amounts of time and resources are allocated to scheduling and estimating projects. However, the results obtained through such methods lack the accuracy needed for capital budgeting and investment decision-making. Investors, owners, end users, and managing teams are affected by the inaccuracy of project schedule and cost estimates.

This literature review demonstrates that the industry can benefit from utilizing statistical tools and concepts in estimating cost and duration of projects. Multiple regression, artificial neural networks, and construction simulation are models capable of predicting construction cost and duration with high levels of accuracy. The literature demonstrates that regardless of neural networks superior predictive accuracy, both methods perform well and provide high-level confidence estimates and accurate predictions.

Furthermore, the literature provides step-by-step instructions on how to build working models for managers and administrators. The literature presents the two specific methods for collecting input variables, namely bill of quantities and cost significant items. The bill of quantities can be obtained from cost records of completed projects, and surveying construction professionals and project teams can procure from published cost index data or the significant cost items.

One issue that construction companies face is poor data collection and record keeping. Companies typically dispose of project records after completion, and the records kept during the project are not always representative of real conditions and costs. Prices are typically set based on negotiations or the estimator's experience, rather than quantitative mathematical approaches. This culture, coupled with antiquated estimating methods, has a negative impact on the accuracy of duration and project cost.

Furthermore, the accurate estimation of quantities has its set of challenges. An accurate estimation of quantities is dependent on the quality and level of development of the design documents. Although there are commercial computer programs dedicated to construction estimation, the implementation of these programs remains low, due to cost and level of skill required and culture.

In the literature, the authors illustrate methods and techniques for designing a predictive model. The stepwise concept and Durbin Watson's statistic to address issues with correlation and multicollinearity are covered. The advantages of developing an equation that can be used for estimating cost and project duration can be of strategic value. Predicting events accurately during the planning phase is vital for projects to achieve milestones and targeted budgets, which will optimize resource allocation and investment planning.

The models studied in this paper can be used by project managers for predicting the duration and cost of construction projects. The literature demonstrates that project durations stipulated in contract documents deviate significantly from the actual duration. The scholars recommend using the estimated quantities once the project is ready for bidding. However, the problem with inaccurate quantity take-offs and unit prices can affect the inputs used for model development. Nevertheless, the research demonstrates that statistical methods are effective and can be integrated in the planning phase of project life cycle.

The multiple regression model appears to be the most user-friendly choice, since once the model is developed, then only the values for significant items are needed for a prediction to be made. The application of neural networks requires the use of a computer and an operator to enter the significant items and to run the simulations until the required project duration is created. Multiple regressing is a more accessible tool, and the results are very similar.

Scholars suggest that regression models be replaced by artificial neural network models because for higher accuracy. Also, neural models can handle a larger number of inputs, whereas the regression model equation may become too complex with an increasing number of cost factors. Moreover, the relationship between variables must be analyzed prior to designing the model, adding to the complexity of the problem if a large number of variables are being handled.

The study demonstrates that both models are capable of identifying the significant variables affecting cost and duration. The important factors identified by the authors in the papers reviewed match those identified by other researches in previous studies. The literature also points out the opportunity estimate of a project's life cycle, taking into account taxes, inflation, and other non-cost factor that influence the total life cycle cost of the asset. An accurate selection and specification of materials during planning stage can translate into future savings accrued during the life of the asset.

Additionally, the implications of this study in project management are profound. Effectively integrating statistical tools and concepts into the planning phase of the project life cycle can change current practices. The implementation of statistical models in cost and duration estimation can help in the optimization of project management and delivery. Thus, statistical models will increase the capacity of managers and firms to handle more complex projects. Good planning translates in the effective use of resources, namely time and materials.

Organizational Implications

Through studying the acquired skill and management strategies from these variables, concepts, and models, it is seen that they are useful tools for business projects and project management. Furthermore, they encourage teamwork skills to better achieve company goals, so they are more valuable than current technology for projects or management. Within the results, it is emphasized that strategic planning is important. This includes a top-down and bottom-up approach from leadership, especially with project management, operations management, and process improvement elements.

Also, this study indicates that current organizational problems are derived from poor leadership skills, as only having a bottom line focus can be damaging. Long-term problems can arise by only focusing on profits and costs. Thus, the necessary tools and information is needed for supervising project management and operational performance. Having only a financial focus is not as beneficial as managing multiple business elements, such as operations, project management, financials, performance, strategy, and human resources. As a result, multiple organizational elements will improve, including performance, profits, and costs.

Managerial and Team Implications

The primary implication from this study is that the results evaluate the variables, concepts, models, and their relationship in a novel way. This new approach aims to fill a knowledge gap in research by evaluating how the variables and their relationship affect each other and outside elements. A business' performance and effectiveness can be impacted, so using these variables in the best way is important.

Secondly, an outline can be derived from this study for business projects and performances. Having a better understanding of the relationship between these variables can yield better management performances. Business leaders can then generate more comprehensive constructs to mentor teams to find their shortcomings and performance gaps. Thus, teams can find ways to avoid these shortcomings for better performances.

Thirdly, this study reveals how businesses can benefit from a more comprehensive training program to improve projects, performances, and overall effectiveness, especially for project teams, project leadership, and organizational leadership. In this study, they can find ways to evaluate the performance of a team, project, or business, so as to measure them against standard and industry accepted models. Also, leaders can find ways to better manage teams and projects to reveal how teams and projects influence overall performance and effectiveness.

Applications to the Field of Engineering Management and Project Management

The techniques and methods discussed in this paper have many applications to the field of project management, particularly in the planning phase. Of the five phases in the project life cycle, planning and monitoring are of great importance. Two components of the planning phase that were investigated in this literature review were cost and duration estimation. These topics were selected because they play an essential role in the success of projects.

Project managers face challenges when developing project duration schedules and cost estimates, so they typically rely on reports and advice from consultants. However, these scheduling and estimating consultants are not contractually bound to the accuracy of their estimates. Estimates are reported to be 30% off target the majority of times. To account for this variation, contingency money is added to the estimate. In the case of duration estimates, consultants arbitrarily determine the duration of projects.

Furthermore, project managers and other practitioners accurately estimate the cost and duration of projects with multiple regression and artificial neural models. They can do this by developing equations and models for each project type. Managers have the necessary tools to identify significant cost factors that can be used as inputs for preliminary experimentation.

Independent variables can be obtained from the project's schedule of values; this schedule provides a breakdown of construction cost by category. Also, these categories are taken from the construction specifications institute and are related to construction components of a project, such as metals, thermal protection, conveying equipment, and earthwork. These sections can be used as starting points for selection dependent variables.

Once variables of interest are identified, managers can begin experimenting with cost and duration models. The stepwise process variables that are not significant contributors to either cost or duration can be eliminated. The literature offers techniques, such as the stepwise method, to determine how much the input variable has on the output variable. To perform this step, managers need to calculate the p-value for each variable. If the value is higher than .05, then the variable is considered non-significant and is eliminated. Overall, the resulting model should be a working model that can be used by project managers to predict cost and duration.

For the traditional construction firms, I suggest the use of the multiple regression modeling technique because it can be computed and programmed using excel or a basic calculator. Once an equation is developed, then adjustments can be made as variable change over time. For a more sophisticated firm, I suggest using neural network model. Also, for the implementation of these models, computers and a program are needed to perform the iterations produces an output.

Construction simulation is a technique reviewed in the literature. There are commercial computer programs available for construction simulation. This can be a powerful tool for designers, managers, and builders, as it allows teams to execute the project virtually and to gain insight about constructability issues. The tool is currently used for estimating, logistics planning, and risk analysis.

The simulation approach can be useful in communications management because it allows stakeholders to visualize the construction process. This visualization can provide stakeholders with the information required to make project decisions. Improved decision-making during the design phase can reduce the cost of construction errors and rework. The simulation of construction can have a positive impact on construction and site management. Also, the benefits of simulation can be shared by designers, owners, and builders, as simulation creates a platform for collaboration.

Engineers require more attention, as well. The job of engineers used to involve problem-solving with math and technology, but they now provide stakeholders with economically viable solutions. These variables, their concepts, and models are undoubtedly vital for engineering decisions, as a product must be created on economically sound manufacturing to succeed. Business management and maturity models can help engineers to find technical knowledge that can help their investors.

According to research, these models can identify certain project elements from a business perspective. This study takes an engineering perspective to also address pure engineering filed techniques like budgeting, equipment, and purchasing material. The IE/EM profession and research field heavily relies on project management and operational performance, as lean thinking is not always the answer. As a result, these variables, their concepts, and models are best used to create different environments in this profession. However, the structural orientation of a scope can make those within IE/EM create the required scopes of interest at each level. A strategy can only come from applying the concentrations needed for every interest level.

Also, stakeholders, such as system engineers, project managers, and others within industrial engineering and engineering management, can find information on applying maturity to project management. Stakeholders will also be guided to capitalize on the roles of system engineering and project management, so their success rate will increase.

CONCLUSION

Recommendation for Future Research

These research findings should be done to implement predictive models in construction. Also, research should investigate the degree of variance in construction quantity estimations. This is important because estimated quantities and cost will then become inputs for predictive modeling, so the quality of the independent variables is essential. For future research, it can be studied how these variables, concepts, and models relate within other environments, such as other industries and managerial settings. This can reveal any benefits, detriments, impacts, outside influences, and outside perspectives. Future research can also assess the organizational, strategic, or cultural perspectives on this topic to find how culture, strategy, human resources, and operations influence these variables.

Limitations

The primary limitation of this research is the quality of the data used for developing the models. The errors in the data collected are not accounted for in the experiments. Further, the findings of this study pertain to commercial construction, primarily in the public sector. A secondary limitation is the limited availability of relevant literature. Although the findings can be applied to any project-based company, the literature review applies to construction projects and to the life cycle cost of completed projects. Hence, this study is somewhat limited to the prediction of cost and duration of commercial construction projects.

Additionally, this study is limited by a small sample size that only addresses key factors within them. Thus, there is bias and validity that could be avoided with a larger sample size. Since the study only assesses certain key factors and their relationship from a project environment, then the conclusions and analysis can only apply to project environments. The findings are also not applicable to supply chain management, operations management, strategic management, and more. As a result, the findings cannot be deployed to other industries or managerial settings.

Conclusion of Research

Decision-making becomes difficult when there is limited information. Predicting the operating cost of a building will impact the choice of materials and specifications for the project during the planning phase. The first step towards developing a model is the identification of significant variables. Also, regression models cannot select the most accurate design model and can become complex with an increasing number of input variables. On the other hand, neural networks can handle many variables. Although the results identify the artificial neural network as the best model, multiple regression can be more practical for project management.

Furthermore, the stepwise method is critical in developing a regression model because it helps the estimator to rank the input variables in order of importance. The goal is to identify the predictors that will be used to develop the model. The implementation of the models discussed in this literature review is recommended to project managers, designers, and owners. Also, the resources referenced in this study can develop methodologies for real-world scenarios. Statistical tools may require a level of understanding that has not been observed in construction firms. Construction planning can benefit from applying statistical tools, as information can be gained in cost estimation, project duration, constructability, and risk analysis.

Overall, project management professionals are ambassadors for the implementing new technologies. The best way to promote innovative estimating techniques is to use them within forums and professional organizations. Thus, stakeholders can greatly benefit from statistical tools and simulation during the planning and monitoring phases of the project life cycle.

REFERENCES

Adjei-Kumi, G. N. (2017). Duration estimation models for bridge construction. *Journal of Engineering, Design and Technology*, 754-777.

Afrifa, R. A.-A.-A. (2013). Artificial neural network model for low estrength RC beam shear capacity. *Journal of Science and Technology*, 119–132.

Ahern, T., Leavy, B., & Byrne, P. J. (2014). Complex project management as complex problem solving: A distributed knowledge management perspective. *International Journal of Project Management*, *32*(8), 1371–1381. doi:10.1016/j.ijproman.2013.06.007

Ahiaga-Dagbui, D. T. (2013). A neuro fzzy hybrid model for predicting final cost of water infrastructure projects. In *29th Annual ARCOM Conference*, (pp. 181-190). Association of research in construction management.

Ahmadu, H. I. (2015). Modeling building construction durations. *Journal of Financial Management of Property and Construction*, 65-84.

Al-Kadeem, R., Backar, S., Eldardiry, M., & Haddad, H. (2017a). Review on using system dynamics in designing work systems of project organizations: Product development process case study. *International Journal of System Dynamics Applications*, *6*(2), 52–70. doi:10.4018/IJSDA.2017040103

Andersen, E. S. (2014). Value creation using the mission breakdown structure. *International Journal of Project Management*, *32*(5), 885–892. doi:10.1016/j.ijproman.2013.11.003

Arafa, M. a. (2013). Early stage cost estimation of buildings construction projects using artificial neural networks. *Journal of Artificial Intelligence*, 63-75.

Arumugam, V. A., Antony, J., & Linderman, K. (2016). The influence of challenging goals and structured method on six sigma project performance: A mediated moderation analysis. *European Journal of Operational Research*, *254*(1), 202–213. doi:10.1016/j.ejor.2016.03.022

Asiedu, R. F., Frempong, N. K., & Alfen, H. W. (2017). Predicting likelyhood of cost overrun in educational projects. *Engineering, Construction, and Architectural Management*, *24*(1), 21–39. doi:10.1108/ECAM-06-2015-0103

Badi, S. M., & Pryke, S. (2016). Assessing the impact of risk allocation on sustainable energy innovation (SEI): The case of private finance initiative (PFI) school projects. *International Journal of Managing Projects in Business*, *9*(2), 259–281. doi:10.1108/IJMPB-10-2015-0103

Besner, C., & Hobbs, B. (2012). The paradox of risk management; A project management practice perspective. *International Journal of Managing Projects in Business*, *5*(2), 230–247. doi:10.1108/17538371211214923

Brown, S. L., & Eisenhardt, K. M. (1995). Product development: Past research, present findings, and future directions. *Academy of Management Review*, *20*(2), 343–378. doi:10.5465/amr.1995.9507312922

Burnes, B. (2014). Kurt Lewin and the planned approach to change: A re-appraisal. *Journal of Management Studies*, *41*(6), 977–1002. doi:10.1111/j.1467-6486.2004.00463.x

Cheng, M.-Y. T.-C., Tsai, H.-C., & Sudjono, E. (2010). Conceptual cost estimates using evolutionary fuzzy hybrid neural networks for projects in construction industry. *Expert Systems with Applications*, *37*(6), 4224–4231. doi:10.1016/j.eswa.2009.11.080

Chew, A. M. (2017). Simulation techniques for cost management and performance in construction projects. *Built Environment Project and Assessment Management*, 534-545.

Cova, B., & Salle, R. (2005). Six key points to merge project marketing into project management. *International Journal of Project Management*, *23*(5), 354–359. doi:10.1016/j.ijproman.2005.01.006

Czarnigowska, A. a. (2014). Time-cost relatinship for predicting construction duration. *Archives of Civil and Mechanical Engineering*, 518–526.

David, M. E., David, F. R., & David, F. R. (2017). The quantitative strategic planning matrix: A new marketing tool. *Journal of Strategic Marketing*, *25*(4), 342–352. doi:10.1080/0965254X.2016.1148763

Detert, J. R., Schroeder, R. G., & Mauriel, J. J. (2000). A framework for linking culture and improvement initiatives in organizations. *Academy of Management Review*, *25*(4), 850–863. doi:10.5465/amr.2000.3707740

Easton, G. S., & Rosenzweig, E. D. (2012). The role of experience in six sigma project success: An empirical analysis of improvement projects. *Journal of Operations Management*, *30*(7), 481–493. doi:10.1016/j.jom.2012.08.002

Elkassas, E. M. (2011). The neural network model for predicting the financing cost of construction projects. *International Journal of Project Organization and Management*, 321-334.

Eskerod, P., & Blichfeldt, B. S. (2005). Managing team entrees and withdrawals during the project life cycle. *International Journal of Project Management*, *23*(7), 495–503. doi:10.1016/j.ijproman.2004.12.005

Galli, B. (2018a). Application of system engineering to project management–How to view their relationship. *International Journal of System Dynamics Applications*, *7*(4), 76–97. doi:10.4018/IJSDA.2018100105

Galli, B. (2018b). Can project management help improve lean six sigma? *IEEE Engineering Management Review*, *46*(2), 55–64. doi:10.1109/EMR.2018.2810146

Galli, B. (2018c). Risks related to lean six sigma deployment and sustainment risks: How project management can help. *International Journal of Service Science, Management, Engineering, and Technology*, *9*(3), 82–105. doi:10.4018/IJSSMET.2018070106

Galli, B., & Hernandez-Lopez, P. (2018). Risks management in agile new product development project environments–A review of literature. *International Journal of Risk and Contingency Management*, *7*(4), 37–67. doi:10.4018/IJRCM.2018100103

Galli, B., & Kaviani, M. A. (2018). The impacts of risk on deploying and sustaining lean six sigma initiatives. *International Journal of Risk and Contingency Management*, *7*(1), 46–70. doi:10.4018/IJRCM.2018010104

Galli, B., Kaviani, M. A., Bottani, E., & Murino, T. (2017). An investigation of shared leadership & key performance indicators in six sigma projects. *International Journal of Strategic Decision Sciences*, *8*(4), 1–45. doi:10.4018/IJSDS.2017100101

Gimenez-Espin, J. A.-J.-C., Jiménez-Jiménez, D., & Martínez-Costa, M. (2013). Organizational culture for total quality management. *Total Quality Management & Business Excellence*, *24*(5-6), 678–692. doi:10.1080/14783363.2012.707409

Grinstead, C. a. (2013). Introduction to probability. American Mathematical Society, 1554-1602.

Hartono, , & Wijaya, , & Arini. (2014). An empirically verified project risk maturity model: Evidence from Indonesian construction industry. *International Journal of Managing Projects in Business*, *7*(2), 263–284. doi:10.1108/IJMPB-03-2013-0015

Hoon Kwak, Y., & Dixon, C. K. (2008). Risk management framework for pharmaceutical research and development projects. *International Journal of Managing Projects in Business*, *1*(4), 552–565. doi:10.1108/17538370810906255

Irfan, M. K. (2013). Planning-statge estimation of highway project duration on the basis of anticipated project cost, projext type and contract type. *International Journal of Project Management*, 78–92.

Jahangirian, M. E. (2011). Simulation in manufacturing and business. *European Journal of Operational Research*, 1–13.

Jarkas, A. (2015). Predicting contract duration for building construction: Is Bromilow's time- cost model panacea? *International Journal of Project Management*, 1–8.

Jin, R. H. (2016). Application of case base reasoning for estimating duration of building projects. *Journal of Construction Engineering and Management*.

Labedz, C. S., & Gray, J. R. (2013). Accounting for lean implementation in government enterprise: Intended and unintended consequences. *International Journal of System Dynamics Applications*, *2*(1), 14–36. doi:10.4018/ijsda.2013010102

Latiffi, A. M. (2013). Building information modeling applications in construction. *International Journal of Construction Engineering and Management*, 1-6.

Lee, J., Lapira, E., Bagheri, B., & Kao, H. (2013). Recent advances and trends in predictive manufacturing systems in big data environment. *Journal of Cleaner Production*, *3*(10), 45–55.

Liozou, P. a. (2013). Risk and uncertainty in development: A critical evaluation of using monte carlo simulation method as a decision tool in real estate development projects. *Journal of Property Investment and Finance*, 298-210.

Liu, M. R. (2012). Predicting the preliminary engineering duration of bridges. *Construction Research Congress*, 505-514. doi:10.1061/9780784412329.051

Loyd, N. (2016). Implementation of a plan-do-check-act pedagogy in industrial engineering education. *International Journal of Engineering Education*, *32*(3), 1260–1267.

Marcelino-Sádaba, S., Pérez-Ezcurdia, A., Lazcano, A. M. E., & Villanueva, P. (2014). Project risk management methodology for small firms. *International Journal of Project Management*, *32*(2), 327–340. doi:10.1016/j.ijproman.2013.05.009

Medina, R., & Medina, A. (2015). The competence loop: Competence management in knowledge-intensive, project-intensive organizations. *International Journal of Managing Projects in Business*, *8*(2), 279–299. doi:10.1108/IJMPB-09-2014-0061

Mensah, I. A.-K., Adjei-Kumi, T., & Nani, G. (2016). Duration determination for rural roads using the principal component analysis and artifical neural networks. *Engineering, Construction, and Architectural Management*, *23*(5), 638–656. doi:10.1108/ECAM-09-2015-0148

Milner, C. D., & Savage, B. M. (2016). Modeling continuous improvement evolution in the service sector: A comparative case study. *International Journal of Quality and Service Sciences*, *8*(3), 438–460. doi:10.1108/IJQSS-07-2016-0052

Nagel, R. (2015). Operational optimization: A lean six sigma approach to sustainability. *Proceedings of the Water Environment Federation*, *3*(4), 1–12. doi:10.2175/193864715819556688

Nani, G. (2017). Duration estimation models for construction projects. *Journal of Engineering, Design and Technology*, 754-777.

Olden, J. D. (2014). An accurate comparison of methods for quantifying variable importance in artificial neural networks using simulated data. *Ecological Modelling*, 389–397.

Ontepeli, M. (2012). Conceptual cost estimating of urban railway system projects. *Journal of Construction Engineering and Management*, 1017–1028.

Papke-Shields, K. E., & Boyer-Wright, K. M. (2017). Strategic planning characteristics applied to project management. *International Journal of Project Management*, *35*(2), 169–179. doi:10.1016/j.ijproman.2016.10.015

Parast, M. M. (2011). The effect of six sigma projects on innovation and firm performance. *International Journal of Project Management*, *29*(1), 45–55. doi:10.1016/j.ijproman.2010.01.006

Parker, D. W., Parsons, N., & Isharyanto, F. (2015). The inclusion of strategic management theories to project management. *International Journal of Managing Projects in Business*, *8*(3), 552–573. doi:10.1108/IJMPB-11-2014-0079

Petruseva, S. Z.-P. (2013). Neural network prediction model for construction project duration. *International Journal of Engineering Research & Technology (Ahmedabad)*, 1646–1654.

Potts, K. a. (2011). Construction cost management. *Engineering, Construction, and Architectural Management*, 224–237.

Schwedes, O., Riedel, V., & Dziekan, K. (2017). Project planning vs. strategic planning: Promoting a different perspective for sustainable transport policy in European R&D projects. *Case Studies on Transport Policy*, *5*(1), 31–37. doi:10.1016/j.cstp.2016.08.006

Sharon, A., Weck, O. L., & Dori, D. (2013). Improving project–product life cycle management with model–based design structure matrix: A joint project management and systems engineering approach. *Systems Engineering*, *16*(4), 413–426. doi:10.1002/sys.21240

Shenhar, A. J., & Levy, O. (2007). Mapping the dimensions of project success. *Project Management Journal*, *28*, 5–13.

Sosmez, R. (2010). Parametric range estimating of building cost using regression models and bootstrap. *Journal of Construction Engineering and Management*, 1011-1016.

Sutherland, S. (2004). Creating a culture of data use for continuous improvement: A case study of an Edison project school. *The American Journal of Evaluation*, *25*(3), 277–293. doi:10.1177/109821400402500302

Svejvig, P., & Andersen, P. (2015). Rethinking project management: A structured literature review with a critical look at the brave new world. *International Journal of Project Management*, *33*(2), 278–290. doi:10.1016/j.ijproman.2014.06.004

Todorović, M. L., Petrović, D. Č., Mihić, M. M., Obradović, V. L., & Bushuyev, S. D. (2015). Project success analysis framework: A knowledge-based approach in project management. *International Journal of Project Management*, *33*(4), 772–783. doi:10.1016/j.ijproman.2014.10.009

Usman Tariq, M. (2013). A six sigma based risk management framework for handling undesired effects associated with delays in project completion. *International Journal of Lean Six Sigma*, *4*(3), 265–279. doi:10.1108/IJLSS-05-2013-0028

Von Thiele Schwarz, U. N.-H., Nielsen, K. M., Stenfors-Hayes, T., & Hasson, H. (2017). Using kaizen to improve employee well-being: Results from two organizational intervention studies. *Human Relations*, *70*(8), 966–993. doi:10.1177/0018726716677071 PMID:28736455

Wheaton, W. a. (2009). The secular and cyclic behavior of true construction cost. *The Journal of Real State Research*, 1-25.

Whyte, A. A. (2014). Estimation of life cycle cost of building: Regression vs artificial neural network. *Built Environment Project and Asset Management*, 30–43.

Winter, M., Andersen, E. S., Elvin, R., & Levene, R. (2006a). Focusing on business projects as an area for future research: An exploratory discussion of four different perspectives. *International Journal of Project Management*, *24*(8), 699–709. doi:10.1016/j.ijproman.2006.08.005

Xiong, W., Zhao, X., Yuan, J.-F., & Luo, S. (2017). Ex post Risk Management in public-private partnership infrastructure projects. *Project Management Journal*, *48*(3), 76–89. doi:10.1177/875697281704800305

Xue, R., Baron, C., & Esteban, P. (2016). Improving cooperation between systems engineers and project managers in engineering projects-towards the alignment of systems engineering and project management standards and guides. *Proceedings of Joint Conference on Mechanical, Design Engineering & Advanced Manufacturing*, *24*(2), 23-40.

Xue, R., Baron, C., & Esteban, P. (2017). Optimizing product development in industry by alignment of the ISO/IEC 15288 systems engineering standard and the PMBoK guide. *International Journal of Product Development*, *22*(1), 65–80. doi:10.1504/IJPD.2017.085278

Yamn, H. a. (2009). A building cost estimation model based on functional elements. *ZITU Journal of Faculty of Architecture*, 73-87.

Yun, S., Choi, J., Oliveira, D. P., Mulva, S. P., & Kang, Y. (2016). Measuring project management inputs throughout capital project delivery. *International Journal of Project Management*, *34*(7), 1167–1182. doi:10.1016/j.ijproman.2016.06.004

Zhang, X., Bao, H., Wang, H., & Skitmore, M. (2016). A model for determining the optimal project life span and concession period of BOT projects. *International Journal of Project Management*, *34*(3), 523–532. doi:10.1016/j.ijproman.2016.01.005

Zwikael, O., & Smyrk, J. (2012). A general framework for gauging the performance of initiatives to enhance organizational value. *British Journal of Management*, *23*, S6–S22. doi:10.1111/j.1467-8551.2012.00823.x

Brian J. Galli (PhD) works as an Assistant Professor of Industrial Engineering & Engineering Management in the Fred DeMatteis School of Engineering & Applied Science at Hofstra University. He also currently serves as the Graduate Director for the Engineering Management Master degree program in the school. He has previously served as faculty at other academic institutions, including: Long Island University, New York Institute of Technology (NYIT), American Public University (APUS), SUNY Binghamton, and SUNY Stony Brook. Prior to joining academia, he worked in industry for over 12 years in which he applied industrial engineering, project management and continuous improvement in a wide variety of arenas, including healthcare, manufacturing, transactional, and service environments. He also provides project management support and training in the areas of process improvement, project management, and analytics.