2017

Bridge design efforts and cost estimation models for PPCB bridge projects

Tejas Ostwal
Iowa State University

Bridge design efforts and cost estimation models for PPCB bridge projects

by

Tejas Ostwal

A thesis submitted to the graduate faculty

in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

Major: Civil Engineering (Construction Engineering and Management)

Program of Study Committee:
Hyung Seok "David" Jeong, Major Professor
Douglas D. Gransberg
Brent Phares

Iowa State University

Ames, Iowa

2017

TABLE OF CONTENTS

Page

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

Firstly, I would like to thank Dr. David Jeong for the inspiration and support to complete this research. The meetings and chapter reviews were greatly appreciated.

Secondly, many thanks to staff at Iowa DOT, who have supported with data collection and sharing of estimating practices through meetings and correspondence. I would especially like to acknowledge Ronald Meyer and James Nelson.

I would also like to thank my friends for my fantastic experience at Iowa State University.

ABSTRACT

The estimation of design effort and cost plays a vital role in authorizing funds and controlling budget during the project development process. Typically, the design phase consists of various engineering activities that require substantial efforts in delivering final construction documents for bid preparation. Estimating these efforts accurately and efficiently is critical for transportation agencies to properly allocate funds and assign appropriate time and resources.

Previous studies have reported several problems associated with the estimation of design effort such as lack of predictive tools, inaccurate forecasts and misallocation of efforts. Thus, there is a need for a proactive scheme to estimate more accurate and reliable design efforts and costs in order improve the confidence of the design office at the negotiation table with consulting firms and finally enhance the accountability and transparencies of funding decisions.

This study develops advanced design effort and cost estimation models using multivariate linear regression (MLR), multivariate polynomial regression (MPR) of second degree and Artificial Neural Network (ANN) methods. First, the study develops a master database that consolidates various data points of historical pretensioned prestressed concrete beam (PPCB) bridge projects designed by external consultants. The master database includes data attributes such as bridge design attributes, various physical attributes of bridges, consultant's proposed fees and workhours, Iowa Department of Transportation (DOTs) proposed fees and workhours, contracted fee and work hours and actual amount of fee paid after the completion of design. Seven MLR, MPR and ANN models have been developed to estimate design fees and work hours at different negotiation and design stages. Two approaches are used to predict cost and workhours, the first one utilized the data from bridge design attributes and the other one utilizes bridge design attributes along with number of various kinks of design sheets. The MLR, MPR and ANN models are developed using commercial prediction analytics software JMP pro.

The performances of all the models are evaluated by comparing the MAPE (Mean absolute percentage error). The MLR models did not perform well because of their inability to recognize non-linear patterns whereas MPR models performed slightly better due to the ability of recognizing some non-linearity in the data. However, ANN models are able to detect non-linear patterns along with interactions in the training data. The results reveal that ANN models perform significantly better with MAPE range of 5-13% with only bridge data attributes and MAPE range of 4-12% with bridge data attributes and different design sheets as inputs.

CHAPTER 1. INTRODUCTION

The estimation of design effort and cost plays a vital role in authorizing funds and controlling budget during the project development process. Typically, the design phase consists of various engineering activities that require substantial efforts in delivering final construction documents for bid preparation. Estimating these efforts accurately and efficiently is critical for transportation agencies in properly allocating funds and assigning appropriate time and resources. However, due to the relatively small proportion of project cost consumed by design service activities, "there has generally been little effort invested in improving management of those activities in the preconstruction phase" (Persad et al. 1995). Additionally, researchers that have focused on preconstruction management have reported a lack of predictive tools to estimate design costs (Knight and Fayek 2002, and Woldesenbet and Jeong 2012). Inaccurate forecast of design costs or misallocation of these efforts can result in a) increased design errors and low quality work leading to project delays, and increased construction costs, or b) lost opportunities to authorize other projects due to unrecognized over-expenditure.

Today, the rapid increase in the number and complexity of transportation projects has led agencies to accelerated project delivery, variation in workload, and overall changes in legal and policy requirements that necessitate specialized skills (Persad et al., 2010). To meet these changing project management needs, design and engineering activities are being augmented by private consulting firms that the agencies negotiate with before funding authorization for consulting agreements. Figure 1 show a typical process when a consulting company is hired for design and engineering by Department of Transportation (DOT). First, the DOT requests a proposal from the most qualified consulting company, with their selection based on the company's expertise, capabilities and previous records. Then, the consulting company proposes a design fee and anticipated work hours based on the scope of work for the design project. The DOT then compares the consulting company's proposal with their estimated costs and typically makes a counter offer to the consulting company, before both parties finally agree on the scope of work, design fee and work hours. The actual amount of design fee paid may be different from the contracted amount because of scope changes and actual work hours different from estimated work hours in the agreement.

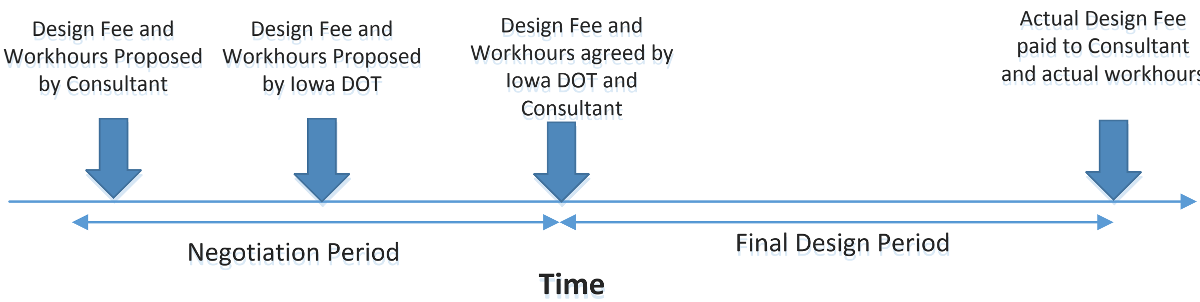

Figure 1. Contracting process of design by consultant

Federal regulation now stipulates that a detailed independent cost estimate must be formed when an engineering work is requested of an external consultant. Such an estimate serves as an important reference point for arbitrations with the Consultant" (UDOT 2013) and must include "an appropriate breakdown of specific types of labor required, workhours, and an estimate of the consultant's fixed fee…for use during negotiations" (23CFR Section 172.7).

However, the efficiency and reliability of outsourcing design services and determining reliable costs associated with the engineering activities have been a huge challenge since most highway agencies do not have a structured and well-defined procedure to reasonably estimate design efforts and costs. A limited number of studies in this area revealed drastic variances in terms of the accuracy of design cost estimates. A study done by the Washington DOT (WSDOT, 2002) reported that the estimated design costs of a sample bridge by ASSHTO Subcommittee members from 25 states ranged from 4% to 20% of construction costs (WSDOT 2002). A recent study of 344 bridge projects constructed by the North Carolina DOT (NCDOT) found that preliminary engineering costs ranged from 0.04% to 138% of construction costs. A similar study by Kuprenas (2003) found design costs ranging from 25% to 50% while Liu et al. (2013) obtained 7% to 16% of the construction cost as the design costs of roadway projects. Lastly, Gransberg et al. (2007) found that using historic construction cost percentages for Oklahoma Turnpike Authority bridge projects caused consultant design fees to be underestimated in almost every case and resulted in statistically significant (R^2 = 0.92) construction cost growth due to change orders required to correct design errors and omissions.

Thus, there is a need for a proactive scheme to estimate more accurate and reliable design efforts and costs in order to meet the federal regulations, improve the confidence of the design office at the negotiation table with consulting firms, and finally enhance the accountability and transparencies of funding decisions.

Motivation

Due several problems associated with the estimation of design effort such as lack of reliable predictive tools and inaccurate forecasts an effort is made to add some intellectual value in this domain. The data from Iowa DOT is used as a part of this study. Pretensioned Prestressed Concrete Bam (PPCB) bridges are the most popular kinds of bridges preferred by Iowa DOT (Nelson and Waterhouse 2016) and on an average 13 PPCB bridge designs are outsourced by Iowa DOT every year. After the financial crisis of 2007-2008, Iowa DOT experienced significant reduction in workforce. However, the workload has not decreased comparatively but in fact has increased for some time because of American Recovery and Reinvestment Act of 2009, a federal stimulus program. A new legislation also requires Iowa DOT to find additional funds for its operations to the tune of $10 million and moreover also reduce the timeframe for project delivery with this increasing workload (Nelson and Waterhouse 2016).

The key motivation behind this study is to develop a need for a proactive scheme to estimate more accurate and reliable design efforts and costs in order to meet the federal regulations, improve the confidence of the design office at the negotiation table with consulting firms, and finally enhance the accountability and transparencies of funding decisions. The current system that the office of Bridges and Structures at Iowa DOT uses is based on percentage of construction cost to estimate the bridge design costs but the office needs to improve and upgrade the system to take advantage of technological advances and historical data that they have accumulated.

Research Approach

Figure 2 shows the overall research approach. First, the agreement files of PPCB design projects conducted by external consultants are studied and important pieces of data points for this study are extracted. In addition, bridge characteristics data that are stored in the bridge information system (BRIS) at Iowa DOT are imported into a spreadsheet file. The master database consists of various data attributes for each PPCB bridge design project including consultant's proposed design fee and workhours, Iowa DOT's proposed design fee and workhours, contracted design fee and workhours, actual amount of design fee paid, and several bridge attributes. The design fees are adjusted to account for inflation and some data cleaning techniques are applied to address missing values and other data issues. Two estimation methods are developed in this research as follows:

- Multivariate regression - Multivariate regression is a technique which estimates a single regression model with more than one output variable. The word multiple is used because there are more than one predictor or input variables. The purpose why regression is chosen for this study is to formulate a relation between several predictor variables and a dependent variable. Furthermore, two kinds of regression multivariate linear regression (MLR) and multivariate polynomial regression (MPR) are used in this study to find if there is any linear relation between the variables in case of MLR and non-linear relation in case of MPR.
- Artificial Neural Network (ANN) – Artificial neural network is a computational approach, which is modeled in the same way as that of a biological brain, which has large clusters of neurons connected by links. Similarly, the neural network is composed of large number of interconnected nodes to solve problems. Neural networks are used in this study to recognize complex patterns in terms of interactions, which is not possible in case of regression.

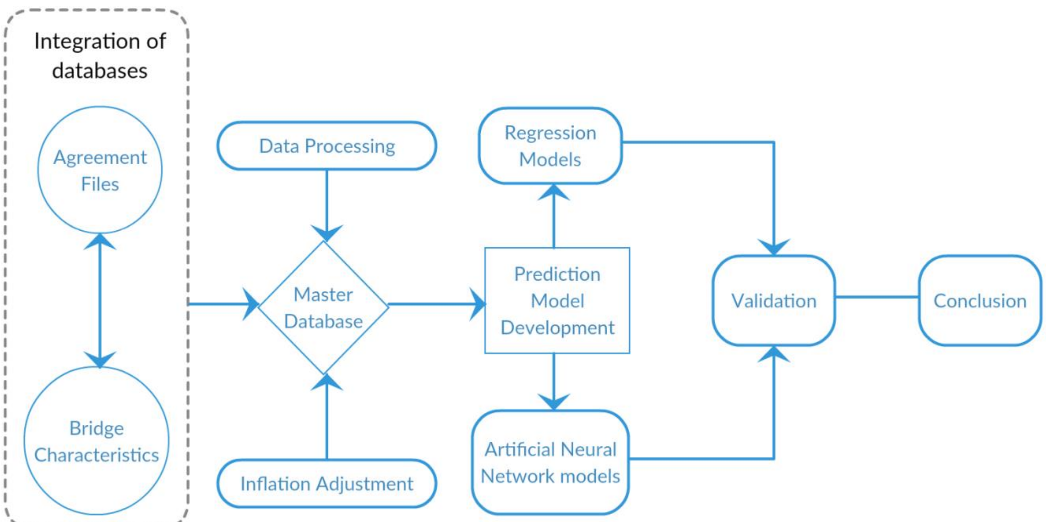

Figure 2. Overall research approach

Problem Statement

The literature shows that statistical tools like regression and artificial intelligence technique can be applied to develop a parametric estimating tool to determine bridge design engineering efforts and costs using historical contract data. Despite the availability of these tools, data-driven techniques are not ben currently used by Iowa DOT or any other agencies for that matter for cost prediction.

The aim of this research is to demonstrate the use of tools that Iowa DOT may utilize in order to predict bridge design cost and efforts by data driven methods. DOT's in other states may also apply the methods demonstrated to develop their own equations for prediction. The detailed objectives are to:

(1) Develop a master database to track bridge design effort and costs.

(2) Develop a parametric bridge design effort and cost estimation models using tools like Regression and Artificial Neural Network.

(3) Validate the developed models and tool, which would assist Iowa Department of Transportation in making reasonable judgements.

Content Organization

Chapter 1 explains the motivation behind this study research, the research approach and furthermore, the problem statement for this study is summarized in this chapter.

Chapter 2 contains research background. To be specific, the current state of predicting design costs is discussed. Previous research in design cost prediction using tools like regression, artificial neural network and random forest will be summarized. Furthermore, the problem statement is also stated.

Chapter 3 highlights the overall approach and assumptions in data extraction. In this chapter the creation of a database is explained. This includes data collection methods, selecting input variables and predicting output variables.

Chapter 4 contains descriptive statistics of all the bridge projects considered as a part of this study. This section discusses the quantitative results of descriptive statistics applied to the master database of 58 PPCB bridge design projects.

Chapter 5 proposes an approach by incorporating bridge characteristics that could be used to develop multivariate regression and neural network models.

Chapter 6 proposes an approach by incorporating bridge characteristics and design sheet types to develop multivariate regression and neural network models

Chapter 7 summarizes the main conclusions from the papers and discusses the limitations of those conclusions.

Chapter 8 outlines the key contributions to the bridge design estimating body of knowledge and areas for future research.

CHAPTER 2. LITERATURE REVIEW

This chapter explains the current state-of-the-practice used to estimate the preliminary engineering cost of projects. It then discusses the application of statistical tools like regression artificial intelligence tools like neural network to the field of PE cost estimation. Finally, the chapter poses the specific research problems investigated in this thesis.

PE as Percentage of Construction Cost

Turochy et al. (2001) conducted a survey of some DOT's in order to know about the procedures they followed in estimating the Preliminary Engineering (PE) costs on highway projects. The survey results revealed that: Delaware DOT considers project size in determining the procedure i.e. in case of large projects; they calculate the percentage of estimated construction cost & in case of smaller projects, work hours are estimated. Florida DOT applies percentages on construction estimates to calculate design costs. Kentucky Transportation Cabinet (KYTC) applies around 10% of the estimated construction cost to determine the design costs & furthermore, the depending on the size of the project the percentage can vary. Minnesota DOT has not included PE costs in calculating the cost estimate but they are planning to change the trend soon. Penn DOT adopts a percentage somewhere between 10% & 20% on estimated construction costs in determining the initial PE costs. Tennessee DOT usually applies 10% of the estimated construction cost to determine the PE costs. Texas DOT typically estimates the PE costs based on the percentage applied on estimated construction costs. Furthermore, they also develop estimates based on estimated ROW width. Washington State DOT also applies percentage on estimated construction cos to estimate PE costs. West Virginia DOT typically adopts 8% on estimated construction costs to determine the PE costs. To conclude, most of the state DOT's usually apply percentages between 5% & 20% based on the scope & size of the project.

To analyze and compare the PE costs with construction costs, information from 25 DOTs was collected by Washington State DOT (WSDOT 2002). PE cost in this study was defined as "the work that goes into preparing a project for construction". The participators were asked to identify project PE cost as a percentage of construction cost and as a result, 10.3 percent was found to be the average PE cost. The range was found between 4 – 20%.

Virginia DOT (VDOT) developed a robust cost estimating system in 2002 (Kyte et al.). Based on the spreadsheet used by Fredericksburg district in Virginia, which produced reliable results for a particular bridge and road projects, the taskforce at VDOT, expanded this tool by incorporating many other project data and improved it further. Fredericksburg district was using this toll since 1999. With large rowing database, minor adjustments were made to refine the values however enhancements on this tool were made only to improve the ease of usage. Originally, this tool just had length, width and complexity option as inputs. In complexity, three sub inputs like simple, moderate and complex were provided. The output was cost per square foot for a bridge with particular complexity. Further enhancements by Fredericksburg district included estimation of PE costs for roads and bridges. Project data reveals that the PE costs ranged from 8% of construction costs for large projects to 20% of construction costs for small ones. In addition to that, inflation actor was also accounted in the database annually.

VDOT later in 2004 (Kyte et al.), did lots of adjustments to account this prototype for state wide usage like accounting for statewide cost variation, inclusion of interstates, accounting for consultant PE costs which are generally higher than in-house costs, bridge component adjustments, additional worksheets for Right-of-way (ROW) and utilities worksheets, etc. To estimate PE costs, a cost curve also called PE curve was derived which was the ratio of PE costs over construction costs. Initially the PE curve was constructed based on 30 roadway projects. Later, the sample was increased to 136 projects to verify the PE curve for all the districts in Virginia. To account for consultant PE costs, a 50% factor is applied to raise the costs over in-house PE work. This 50% factor came from VDOT's previous study on the costs of design consultants. The user can also choose the percentage of total work to be done in-house and to be outsourced to consultants. Similar PE cost curve was developed from a database consisting of 23 completed bridge projects. As we see, VDOT's PE estimating tool is not a single percentage based; the developed PE curve utilizes historical data based on various factors. However, it is a better approach compared to most of the DOTs but still it is not regression based which considers only significant parameters in predicting PE costs.

Use of Statistical Analysis Techniques in PE Cost Estimation

It has been well known that design fees are typically calculated as a percentage of the construction contract value. However, the American Society of Civil Engineers (ASCE) discourages the use of this method to determine the design fees (Carr and Beyor 2005). Carr and

Beyor (2005) investigated the effects of using "outdated percentage of construction fees schedules" to determine design fees. The study found out that design fees were significantly underestimated because of the use of outdated and stagnant fee schedules. Additionally, they reported that there was a significant need to develop a guide for estimating appropriate design fees (Carr and Beyor 2005). The literature shows two cost estimating data driven methods multiple linear regression (MLR), artificial neural network (ANN) been demonstrated to predict PE costs of bridge projects and highway projects (Mahamid 2011). Multiple linear regression is nothing but developing a regression equation to link independent input variables to the dependent output variable.

Multiple Regression Models

Lowe et al. (2006) tried to predict the early stage construction cost of buildings by developing linear regression models. The data was collected from 286 construction projects from the United Kingdom. Forty-one independent variables were identified for developing the model. The input variables were categorized as project strategic variables, site related variables and design related variables. They used regression to predict the log (cost), cost per unit area ($\$/m^2$), and log (cost per unit area) instead of using costs at completion of construction. The researchers recommended reviewing error spread in each model by analyzing the dependent variable versus error illustrated on scatter plots. The model performance was evaluated using R^2 and mean absolute percentage error (MAPE). The best-performed regression model reported an R^2 value of 0.66 and MAPE of 19.3%. They reported that the models underestimated the costs of very expensive projects and overestimated the costs of inexpensive projects when scatter plots of cost/m2 were compared against error.

Woldesenbet and Jeong (2012) used historical data provided by Oklahoma DOT to estimate the preliminary engineering (PE) costs of roadway projects. This study is an example of data driven based prediction model for PE costs. Using data mining techniques, they identified influential factors that affected the PE costs and developed decision tree and regression models to estimate PE costs. Totally 5 major categories like project scope, geographic attributes, design attributes, environmental attributes and external factors were identified. 25 critical factors were sub-identified under these five categories, which affected PE costs. Using the regression model, project length was found to be the most significant factor and in case of decision tree model, projects type, route type and fund type was found to have a profound effect on PE costs.

Furthermore, the potential to predict PE cost as a function of workhours per sheet, number of sheets and cost per workhour was also noticed with R^2 greater than 41.64%. However, the decision tree approach was found to be comparatively better than regression model in predicting PE costs.

Hollar (2012) studied on calculating estimates on PE costs and PE duration for North Carolina DOT's bridge projects. In this research, the data was acquired from ten sources to form a database on 461 bridge projects. Twenty-eight independent variables were identified through correlation analysis and Analysis of Variance (ANOVA). Regression models were developed for PE cost ratio and PE duration. The modeling strategies like Multivariate linear regression (MLR), hierarchical linear models (HLM), Dirichlet process linear models (DPLM) and multilevel Dirichlet process linear models (MDPLM) were included in the study. For a goodness of fit testing, the mean absolute percentage error (MAPE) was used to rank predictive performance when each candidate model was applied to a validation set. The MLR model including eight variables achieved a MAPE of 18.9%.

Jianxiong (2015) developed a bottom up design hour estimation template for California Department of Transportation (Caltrans). This study proposed a work plan template development procedure to simplify the workhour estimation process. The bottom up approach is considered to provide a better estimation of design hours, but it is generally labor intensive. Turner and Miller (2015) developed regression models to estimate PE costs. The researchers used project length, project duration, level of environmental review and whether or not the project is administered by Virginia DOT to develop the regression models. Initially, only construction cost was used to develop regression models. Later on, other project characteristics were added, which reduced the mean percentage error of estimation from 134% to 44%.

New York DOT has developed a MS Access based tool to search projects with given characteristics (Williams et al. 2013). Historical data from past 64 NYSDOT projects were collected for regression analysis. The regression model was developed to predict the design hours based on the following project characteristics:

1. Complexity,
2. Project type,
3. Number of sub-consultants,
4. Construction costs,

5. Number of lanes,
6. Number of plan sheets,
7. State Environmental Quality Review (SEQR) classification,
8. National Environmental Policy Act (NEPA) classification,
9. Predominant bridge type,
10. Number of bridges,
11. Highway classification, and
12. Length of project.

Due to insufficient data, the above project characteristics apart from 3-6 and 12 were not included in the analysis. Separate regression models were developed for in-house project designs by NYSDOT and outsourced project designs by private consultants. For consultant data, the simple linear regression model developed to predict workhours based on construction cost had an R^2 value of 0.689 and a multi linear regression model developed to predict workhours based on no. of plan sheets, no. of lanes and length had an R^2 value of 0.628. Backward regression technique was used to remove the insignificant input variables i.e. variables which had least affected the R^2 value were considered insignificant. In this study, the total number of design sheets were considered in analysis. In reality, the no. of design sheets consists of standard design sheets, modified design sheets and new design sheets. Furthermore, the designer need not completely design new sheet in case of standard sheets and modified sheets and may put less effort in designing them and more hours in designing a complete new sheet. Hence, this issue was not addressed while developing this estimating tool, which could have made a better model.

The Iowa DOT developed a multivariate regression model in order to estimate the hours of a PPCB bridge design project based on a database that included following parameters: Length, Width, Number of spans, No of beams, Horizontal curve, Design Methodology, Span arrangement, Pier type, Expansion joint type, Skew, Construction staging, Abutment type, Abutment foundation type, and Beam type (Nelson and Waterhouse 2016). The regression model was developed based on the historical data of 45 projects constructed between 2000 and 2015. The prediction of design hours used the following parameters as significant variables;
a) Number of spans, b) Standard span arrangement, c) Pier type, d) Expansion joint type,
e) Skew, and f) Construction staging.

Using the above-mentioned parameters, the tool predicted the expected, lower 95%, and upper 95% estimates for the bridge design hours. However, the tool needed to be enhanced by considering other factors such as bridge aesthetics, deep foundation type, accelerated bridge construction, deck area, and barrier rail type (Nelson and Waterhouse 2016).

Artificial Neural Network

Another popular approach is called the artificial neural network (ANN). As the name states the neural network is inspired by that of brain's neuron. This is an approach where human intelligence is incorporated into machine to perform complex analysis. In the case of ANN there is no need of presuming that a link exists between the input and output variables (Kim et al. 2004). By creating layers of arbitrary data, the input variables are transformed into output variables. The data from past projects is used to train the model, which develops relationships within the database to predict the output variables. Finally, the trained model is used to predict output variable by recognizing patterns in the trained data.

Researchers have been applying ANN in wide areas of construction applications since early 80's. In civil engineering, the first paper in neural networks was published in 1989 (Adeli, 2001). (Flood, 2006) stated that, ANNs have proven to be more flexible and precise for academic researches and some practical applications. Furthermore, (Flood, 2006) also challenged researchers to produce a convincing model for prediction in the future.

Louisiana DOT developed artificial neural network (ANN) to predict conceptual highway construction cost based on cost indexes (Wilmot and Cheng 2003). The cost index involved composite measure of material, labor, equipment, contract characteristics and contract environment. The developed ANN model predicted the construction costs with 95% level of accuracy (Wilmot and Cheng 2003).

A framework was developed in 2006 in Egypt to train and test artificial neural network in predicting estimates for concrete activities (Ezeldin and Shahara 2006). An attempt was made to identify the influential parameters in concrete activities and based on input parameters a prediction model was developed. The outcome of this research was that the identified influential factors demonstrated a reasonable improvement in predicting the future values.

Gransberg et al. (2015) developed a holistic framework which included both top-down and bottom-up approaches based on case studies performed in nine state DOTs. This study published a national guidebook on estimation of preconstruction services costs which describes a systematic step-by-step process of developing Preconstruction Services cost estimating models with data driven approaches such as Multivariate regression modeling, decision tree analysis and neural network. The report recommended a top-down approach for estimating Preconstruction costs (PCS) costs during the planning phase when the project information is very limited and a bottom-up approach during the detailed design stage when a work breakdown structure is available at each functional level. Furthermore, the need of maintaining and updating a PCS cost database has been strongly emphasized to continuously improve the estimation results, which would reduce project cost uncertainties.

Many researchers have developed data-driven models with reliable results using both Regression and Neural Network. In 1987, Bell and Ghazanfer developed a cost estimation model for highway maintenance project with data from 174 projects (Bell and Ghazanfer 1987). This cost model had an average error of 17% when tried against testing data. A neural network model was developed by Creese and Li in 1995 to predict the construction cost of timber bridges with an error of 8.24%. Kim et. al (2004) compared case based reasoning, multiple regression and neural network to analyze performance by extracting data from 530 projects. The error for regression model was 7% and 3% for neural network models.

CHAPTER 3. DATA EXTRACTION, PROCESSING AND VALIDATION

Data Extraction

Available Data Sources

The data available for PPCB bridge design effort estimation model development is from three different sources:

1. Bridge Information System (BRIS) data
2. Design Fee Data and Workhours data
3. Agreement files

Bridge Information System (BRIS) is a database of bridge projects in the Iowa DOT's computer system that is accessible from within the Iowa DOT only. This database comprises bridge characteristics of all kinds of bridges, which have been let by Iowa DOT. The categories of bridge data attributes in the BRIS database are shown in Table 1.

Table 1. BRIS Bridge data attribute categories

Bridge General Information	Bridge Misc. Information	Bridge Bearing Description
Bridge Location	Bridge Work Description	Bridge Abutments Description
Bridge Geometrical Features	Bridge Deck Description	Bridge Piers Description
Bridge Design Method	Bridge Joints Description	Bridge Aesthetics Description
Bridge Misc. Features	Bridge Beams Description	

Design Fee and Workhours data includes proposed design fee and workhours by consultant, design fee and workhours by Iowa DOT, contracted design fee and workhours. In this study, the total design fee without contingency has been considered for estimation model development. Another spreadsheet file in this data source comprises billing data and actual amount paid to the consultant. This gives information about the actual design fee received to date by the consultant.

Agreement Files consist of contract documents related to a particular project. These files have data, which includes the scope of work, itemized engineering hours and design fee, which has detailed breakdown of direct labor costs, overhead costs, direct expense, contingency and sub consultant expenses.

Data Extraction Assumptions

Assumptions made during the data extraction from the agreement files are below:

a) Multi-bridge fees – In case of Multi Bridge projects encountered in agreement files or fee data sheets (excel files); the fees are split proportionately based on the available workhours data. For example, individual workhours for each lane are not available as shown in Figure 3.

	Proj Mgt	Engr.	Hours	Avg Direct Labor	Over-head	Misc. costs	Non FF Misc. costs	Fixed Fee	Hourly Cost of Doing Business (w/ FF)	Total w/o contingency	Contin-gency	Total
Final Bridge	6.2%	60%	1,409	32.49	182.60%	2,638		12.0%	104.71	147,531	13,201	160,731
Proposed Negotiation	5.0%	60%	825.00	32.49	182.60%	2,638		12.0%	106.03	87475.76	7,839	95,314

Figure 3. Non-availability of individual workhours

The Fee data for this particular contract is merged together for both lanes in North and South direction. For these kinds of projects with same design characteristics, the design fee and workhours are split into equal ratios (50:50) for both lanes.

$$Workhours = \frac{1409}{2} = 704.5\ hours$$

$$Design\ Fee = \frac{\$147{,}530.61}{2} = \$73{,}765.31$$

Another example with a different case is for those projects, which have itemized workhours for each lane as shown in Figure 4.

Staff Hour Classification				
Design No.	Engineer Hours	Technician Hours	Administrative Hours	Totals
TS&L Verification	16	4	0	20
Railroad Coordination	10	0	0	10
Aesthetics, Lighting	24	48	0	72
I-35 NB Median	218	204	14	436
I-35 NB Mainline	494	436	32	962
I-35 SB Median	164	147	9	320
I-35 SB Mainline	360	306	22	688
Shop Drwg. & RFI review	80	0	4	84
Sub-total Staff Hrs	1,366	1,145	81	2,592

Figure 4. Availability of individual workhours

The workhours and Design Fee for I-35 SB Median are calculated as follows:

$$Workhours = \frac{320}{436+962+320+688} * 2592 = 345\ hours$$

$$Design\ Fee = \frac{345}{2592} * \$270{,}165 = \$35{,}932.36$$

b) Increased fee resulting from the change of scope is not included in the database.

c) Only one record from each BRIS table is used to develop the preliminary database for modeling. For example, Project No. ESIMX-035-1(105)33--1S-20 represents a project that is repeated twice in the BRIS datasheet for east bound and west bound which has the same attributes apart from lane direction. In these cases, only one lane is considered because all the technical specifications are identical and it would just be duplication of data.

d) If the project has missing data for consultant/Iowa DOT's proposed workhours and design fee, the actual fee paid is assumed as consultant/ Iowa DOT's proposed fee. Furthermore, the corresponding workhours have been interpolated based on the contract hours.

e) If the data for actual fee paid is missing, the contracted design fee is assumed to be the same as actual fee paid.

f) If the percentage of design completion is greater than 90% and the actual fee paid to date is less than the contract amount, then the adjusted actual fee paid is assumed to be the same as the contract amount as shown in Table 2. Adjusted actual fee paid is a value assigned in place of actual fee paid for design projects, which have not been completed.

g) If the actual fee paid to date is higher than the contract amount with the design close to completion (> 90%); the adjusted actual fee paid to date is considered the same as actual fee paid as shown in Table 2.

Table 2. Adjusted actual fee paid example

Contract Fee	Actual Fee Paid	Adjusted Actual Fee Paid	Design Percentage Completion
$131,012.30	$125,202.00	$131,012.30	99%
$216,979.93	$235,515.00	$235,515.00	96.90%

Data Processing

Missing Value Imputation

Some data attributes have missing values and most of them are filled using reasonable assumptions. Missing values in one or more fields will make the entire record of a bridge design project unusable when developing estimation models such as regression models or neural networks. As such, it is important to fill those missing values to retain the highest possible number of records. Missing value imputations are conducted using the following rules:

- No construction staging if the number of construction stages is zero.
- Feature crossed field is assumed based on the "Over" field. The "Over" field contains the following attributes: creeks, rivers, roads, multiple and railroad.
- Route type is determined based on the route number (Interstate, US, or Iowa) stored in project number.
- Bridges with a blank "bridge width variable" record are assumed to have a constant bridge width.
- Number of spans is calculated based on the count of span lengths field.
- Number of piers is calculated as number of spans minus one.
- Bridges with a blank "sidewalk" record are assumed to have no sidewalk.
- Bridges with a blank "Aesthetic items" record are assumed to have no aesthetic items.
- Number of aesthetic items is counted and added as a separate field as the number of aesthetic items.

Exclusion of Attributes

Based on the questionnaire survey responses and the availability of the data, the following attributes are excluded:

- Design specification parameter because the majority of PPCB bridges are designed by using LRFD
- Bridge out to out width because of missing data
- Number of abutments because of missing data
- Deck reinforcement type because of missing data
- Beam description because of missing data

- Abutment foundation because the majority of the bridges have piles except one bridge that has a drilled shaft
- Pile type because of missing data
- Outside/inside rail because of missing data
- Work complexity and urgency because almost all PPCB design projects in the database have the same value and differences in values were negligible.

Outlier Detection and Exclusion

A modified z-score, which is calculated, based on the median absolute deviation (MAD) and median of a dataset is used to find outliers of data points (Iglewicz and Hoaglin 1993). The modified z-score is calculated as shows in equation (1):

$$M_i = \frac{0.6745\overline{x_i-(x)}}{MAD} \tag{1}$$

Where X_i is the observation value, $\bar{X}$ is the median, and MAD is the absolute difference between the median of data and observation value. As noted by Iglewicz and Hoaglin (1993), modified z-scores that are greater than 3.5 are considered as outliers. As such, seven projects are removed based on the criteria.

Inflation Adjustment

Available projects for analysis are spread from 2008 to 2016. Thus, design fees need to be adjusted for inflation. In this study, all the design fees are adjusted to 2015 dollars. In order to make that adjustment, the Consumer Price Index (CPI) is used. Table 3 shows the CPI indexes obtained from the Bureau of Labor and Statistics (2016).

Table 3. Consumer price indexes (Bureau of Labor and Statistics 2016)

Year	CPI
2008	215.303
2009	214.537
2010	218.056
2011	224.939
2012	229.594
2013	232.957
2014	236.736
2015	237.017
2016	237.855

Questionnaire Survey Results

A short questionnaire is distributed to three bridge design experts in Iowa DOT to determine the level of importance of the bridge attributes on design efforts and costs. Respondents are first asked to indicate whether the attribute is well known, somewhat known or unknown before receiving a consultant's proposal. The respondents reported that most attributes are quite well known before the final design starts except for the following two attributes; a) Abutment: Abutment Pile Type and b) Pier: Pile Type. Thus, these attributes are excluded from design effort estimation model development.

In addition, respondents are asked to assign a level of importance score on a scale of one to three where one has the lowest effect on design fee and three has the highest effect. The responses of this question are summarized in Table 4. As shown in Table 4, these experts believe that type of work, number of construction stages, bridge width variable, number of piers and pier type are the most influential factors on design fee and workhours, which appear to be in the list along with some other significant attributes as derived from polynomial regression models in Chapter 5.

Table 4. Summary of the questionnaire survey responses

Attribute	Response #1	Response #2	Response #3	Total score
Type of work	3	3	3	9
Number of construction stages	3	3	3	9
Bridge width variable	3	3	3	9
Number of piers	3	3	3	9
Pier type	3	3	3	9
Number of spans	2	3	3	8
Beam spacing variable	3	3	2	8
Bridge length	1	3	3	7
Aesthetic item	3	2	2	7
Aesthetic description	3	2	2	7
Horizontal curve radius	1 or 3	3	3	7
Skew	2	2	2	6
Sidewalk	2	2	2	6
Joint expansion type	2	2	2	6
Abutment type	2	2	2	6
Abutment foundation	2	2	2	6

Table 4. (Continued)

Attribute	Response #1	Response #2	Response #3	Total score
Foundation type	-	3	2	5
Feature crossed	2	2	1	5
Bridge width	1	2	2	5
Bridge out to out width	1	2	2	5
Number of beam lines	1	2	2	5
Beam type	1	1	3	5
Route	2	1	1	4
Span lengths	1	1	2	4
Beam spacing	1	2	1	4
Design load parameters (LRFD, LFD, and WSD)	1	1	1	3
Vertical curve type	1	1	1	3
Number of abutments	1	1	1	3
Rail – inside and outside - left and right	1	1	1	3
Deck reinforcement type	1	1	1	3
Beam description	1	1	1	3
Abutment pile type	1	1	1	3
Pier foundation	2			2
Pile type	1	1	1	3

Data Validation

To validate the models, fifteen percent of the available data, (i.e. eight records), are kept for testing purposes. The performances of the MLR, MPR and ANN models are evaluated by calculating the mean absolute percentage error (MAPE) that is measured with equation given below.

$$MAPE = \left(\frac{100}{n}\right)\sum_{i=1}^{n}\left|\frac{P_i - A_i}{A_i}\right|$$

Where, n is the number of testing data-points, P_i is the predicted design fee or engineering workhours, A_i is the actual design fee or engineering workhours.

CHAPTER 4. DESCRIPTIVE STATISTICS AND DATA TRANSFORMATION

This section discusses the quantitative results of descriptive statistics applied to the master database of 58 PPCB bridge design projects. The descriptive statistics results are based on the assumption that all other attributes related to bridge design remain constant apart from the attributes, which are compared. Another technically important issue is that most of the data attributes used to describe bridge properties are string and text data, which make the use of regression models inapplicable. For example, the abutment type attribute is described as integral abutment or stub abutment. As such, it is necessary that string attributes need to be converted to categorical attributes by comparing the average contracted design fee for each category for regression analysis. For example, the average contracted design fee for projects with integral abutment (category 0), is $93,651.50 while the average contracted design fee for projects with stub abutment (category 1), is $153,299.20. Similarly, string attributes are converted using the same method. In some cases, some categories have very similar average fees and hence combined in one category. However, beam type and construction staging attributes are categorized based on the beam type requirement (PPCB bulb Tee beam vs non-tee beams) and application of construction staging. Furthermore, pier types are categorized based on the average contract value of bridges.

Figures 5 shows the average contract values of bridge design for new structures and replacement structures. The interpretations have been made from 47 replacement structure projects and 11 new structure projects. The results show that type of work does have a small impact on the average contract values as the average design cost of new structures are 12.4% above the average cost of replacement structures. In this case, the replacement structures have been assigned with binary 0 to group them in one category and new structures have been assigned with binary 1 to group them in another category.

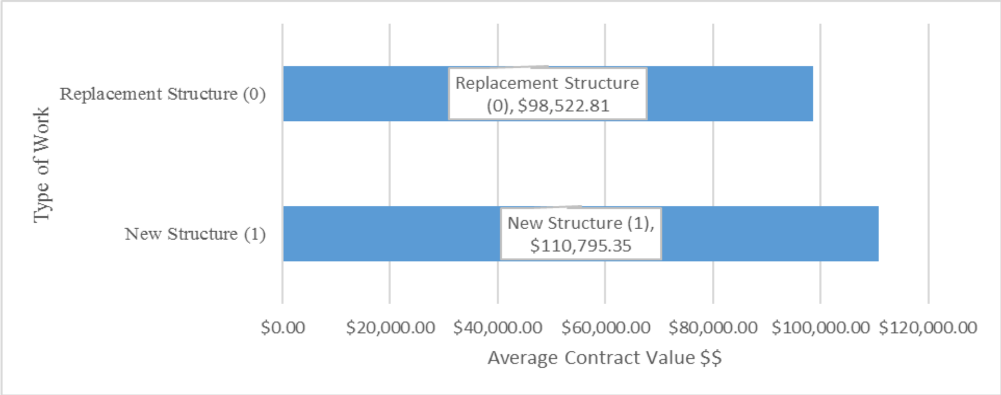

Figure 5. Type of work categories and impact on average contract value

Figures 6 shows the average contract values of bridge design for various route types. Of the 58 projects, 33 have been designed on interstate and 25 have been designed on US highways and Iowa roads. Route type significantly affects the average contract values as bridges designed on Interstate cost comparatively higher by 16%-39% when compared to bridges designed on US Highways and Iowa Roads respectively. Although, the difference between average design fees for bridges built on US Highways and Iowa Roads is found to be relatively high (20%) which is attributed to 2 expensive projects on US Highways and 1 economical project on Iowa road. If these 3 projects are removed, then the difference between the average contract values of bridges is only 4% and hence, they have been assigned the same variable (0).

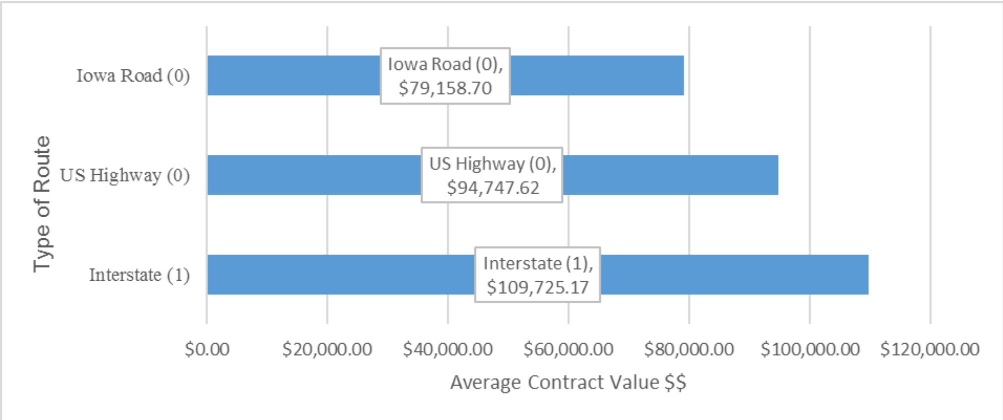

Figure 6. Route type categories and impact on average contract value

Figure 7 shows the average contract values of bridge design with constant and variable widths. Totally, 54 projects have been designed with constant bridge widths and 4 have been designed with variable bridge widths. A bridge with variable width can increase the average design fee up to 56% compared to bridges with constant widths.

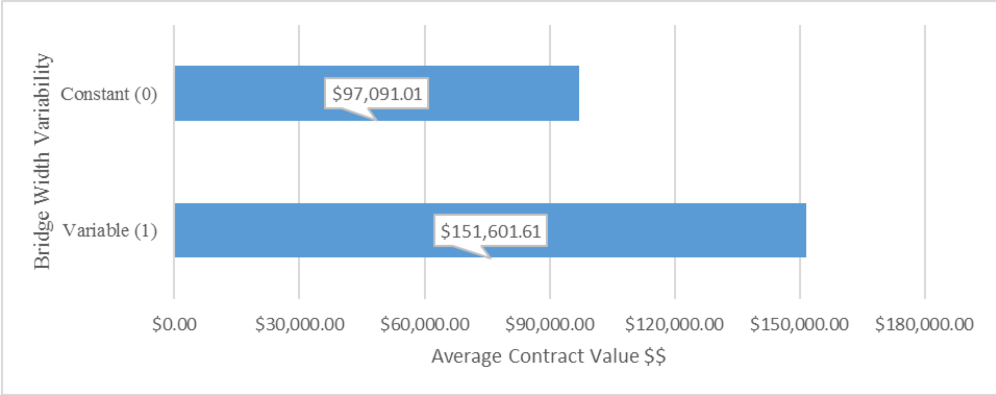

Figure 7. Bridge width variability and impact on average contract value

Figure 8 shows the average contract values of bridge design with and without horizontal curvature. The interpretations have been made from 48 bridge designs with no horizontal curves and 10 bridge designs with horizontal curves. A bridge with horizontal curve would increase design fees by 57% on average compared to bridges with no curves. Hence, width variability and horizontal curves may play an important factor in driving the average bridge design fee.

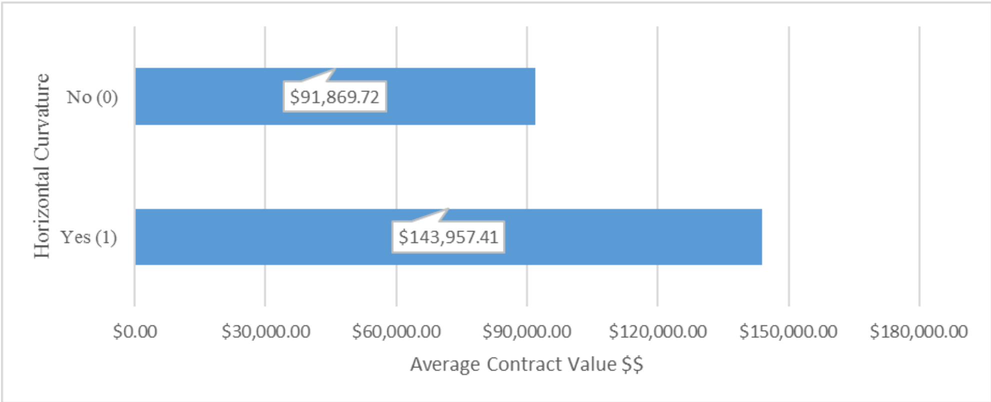

Figure 8. Horizontal curvature requirement and impact on average contract value

Figure 9 shows the average contract values of bridge designs with 55 projects having constant beam spacing and 3 projects with variable beam spacing. Variable beam spacing can pose a significant impact on bridge design fee as bridges with variable beam spacing can increase the design fees by 71% compared to bridge design with constant beam spacing. This deviation can be attributed to the complexity in design as the average workhours for bridges designed with variable beam spacing are 55% higher than bridges designed with constant beam spacing.

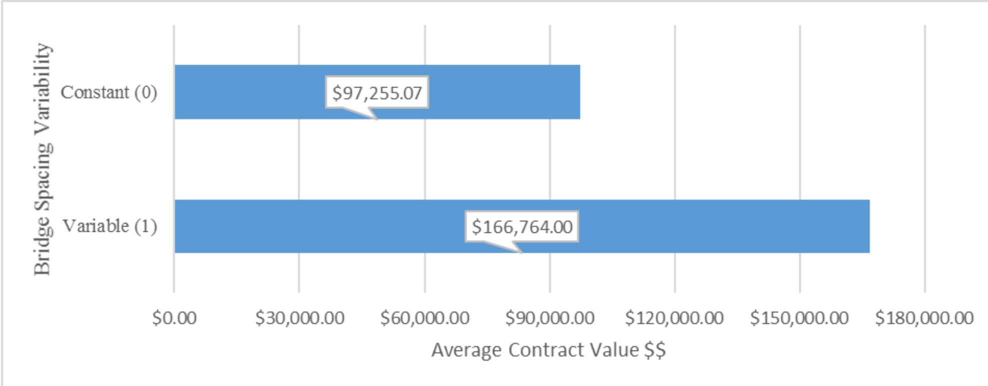

Figure 9. Beam spacing variability and impact on average contract value

Figure 10 shows the average contract values of bridge designs with 50 projects designed without any sidewalks and 8 projects having sidewalks on either sides or both. Having sidewalks on the bridge can increase the design fees by 15% compared to bridges having no sidewalks because the average engineering hours for designing bridges with sidewalks is 1098 workhours compared to 913 workhours for bridges designed without sidewalks. Hence, beam spacing and sidewalks poses a significant impact on bridge design fee.

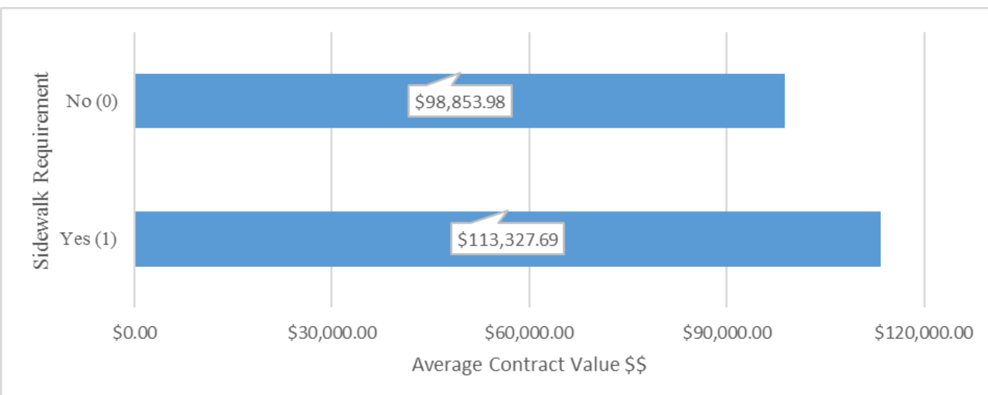

Figure 10. Sidewalk requirement and impact on average contract value

Figure 11 shows the average contract values of bridge design with different kinds of beams. The interpretations have been made by grouping 49 projects under bulb tee Beam design category and 9 projects under non-bulb tee beam design category. The average contract value of bridges designed with bulb-tee beams is found to be higher than 40% when compared with average contract value of bridges designed with non-bulb tee beams. As seen from figure, the average contract value of PPCB bulb tee C beam bridge is less than PPCB D beam bridges. This reduction is attributed to 2 economical PPCB bulb tee C beam bridges and if these two bridges are removed, the average comes around $90,000. Since there isn't much difference between the average contract values of bulb-tee beams, they have been assigned with the same variable (1), and the other beams have been grouped with the variable (0).

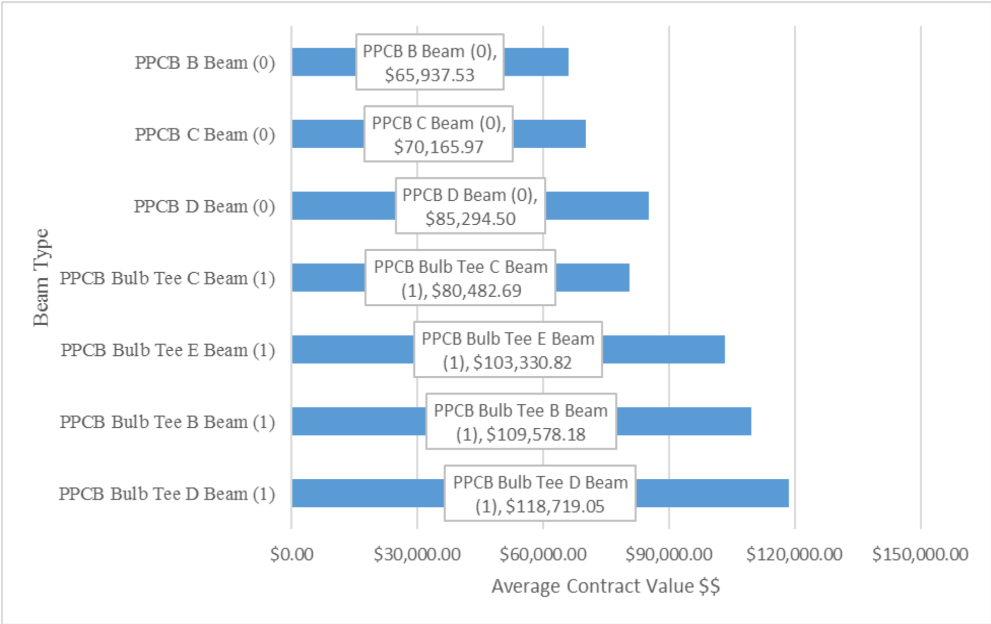

Figure 11. Beam type requirement and impact on average contract value

Figure 12 shows the average contract values of bridge design with different pier types. The inferences have been made from 4 designs with bent pile type, 30 designs with frame pier, 15 designs with T-pier, 6 designs with solid concrete diaphragm and 3 designs with no piers. When compared with bridges designed with bent pile and T pier, bridge designs with frame pier, solid concrete diaphragm and no piers can increase design fee up by 39%. It is found that bridges with no piers have higher design fees when compared to other projects with different pier types. The reason why bridge design with no piers is very high is probably because these three projects with no piers have been designed on interstates with Bulb Tee beams and horizontal curvature. In case of bent pile and T pier, all the bridges have been designed without horizontal curvature, having constant bridge width and constant beam spacing, which might have drastically reduced the average design fee. Thus, pier type poses a large impact on the average bridge design fees. Furthermore, bent pile & T-pier are merged with the same variable (0) due to small difference in average contracted design fee and the others are merged with variable (1).

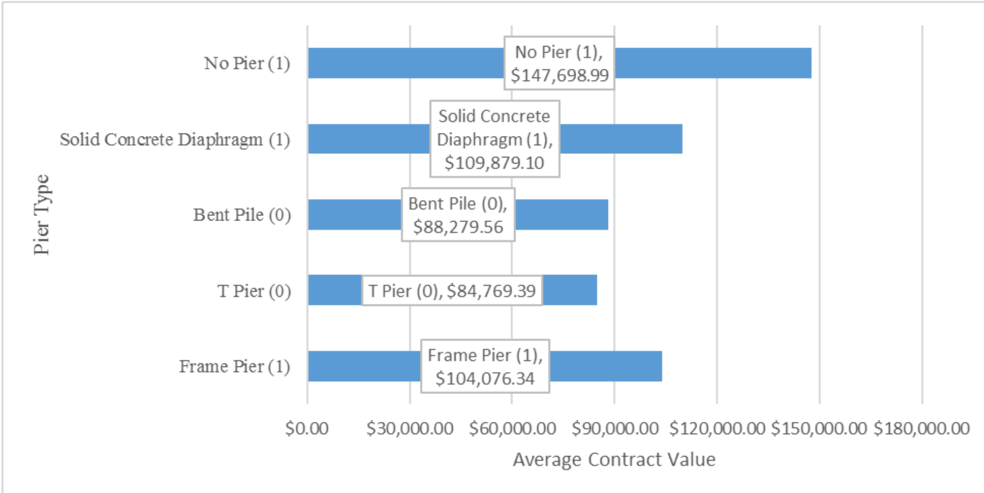

Figure 12. Pier type requirement and impact on average contract value

Figure 13 shows the average contract values of bridge design with 7 projects under stub abutment category and 51 projects under integral abutment categories. Bridges with stub abutments can increase the design fees by 64% compared to bridges with integral abutments. Hence, the type of abutment may have a considerable impact on bridge design fee.

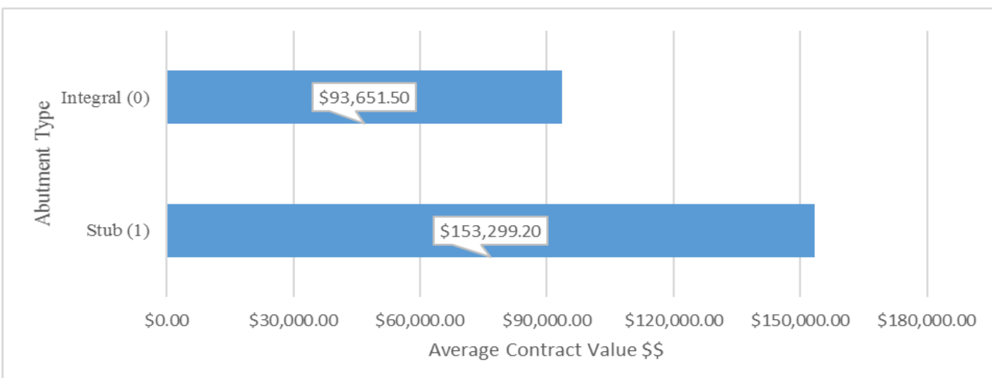

Figure 13. Abutment type requirement and impact on average contract value

Figure 14 shows the average contract values of bridge designs with and without staging. Totally, 49 projects have been designed with no staging at all and 9 projects have been designed with staging. It appears that construction staging has a small impact on average bridge design fees. Having staging in bridge design can increase the costs by 21% compared to bridges with no staging. It is important to note that all the project having construction staging have been designed with PPCB bulb Tee beams and furthermore, 67% of the projects with staging have been designed with either frame pier or solid concrete diaphragm, which may be additional factors, which have increased the average design fee.

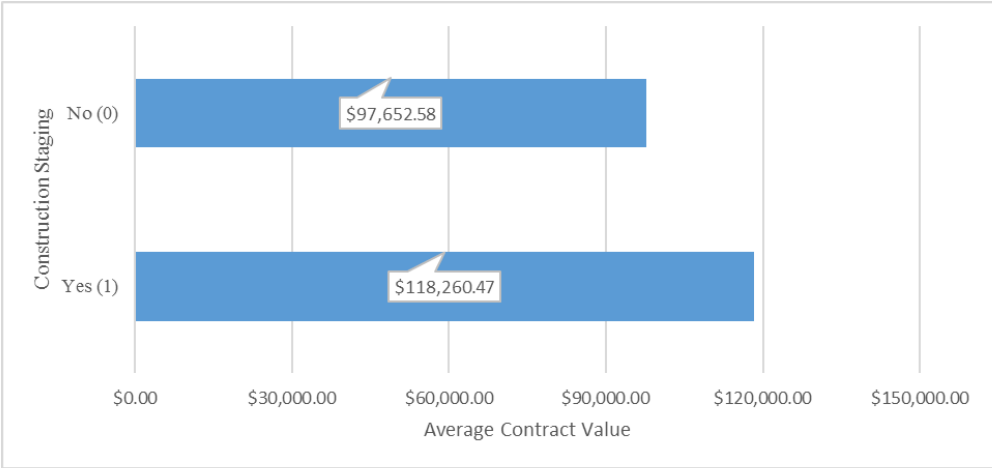

Figure 14. Construction staging requirement and impact on average contract value

Figure 15 shows the average contract values of bridge designs with and without aesthetic items. Totally, 34 projects have been designed with no aesthetic items at all and 24 projects have been designed with aesthetic items. It appears that aesthetic items have a small impact on average bridge design fees. Having aesthetic items in bridge design can increase the average design costs by 8% compared to bridges with no aesthetic items.

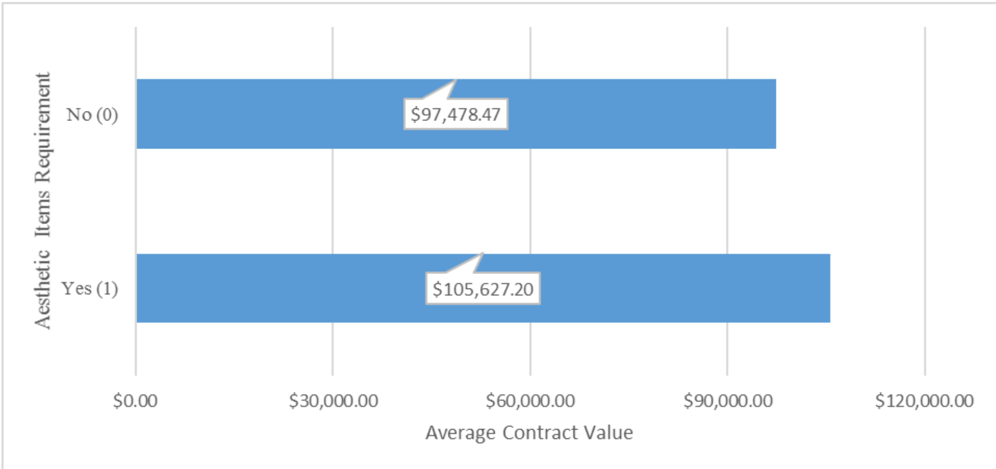

Figure 15. Aesthetic items requirement and impact on average contract value

CHAPTER 5. BRIDGE DESIGN ESTIMATION USING DESIGN ATTRIBUTES

Multivariate Linear Regression (MLR)

Linear Regression is an approach in which a relationship is modeled between a dependable variable and an independent variable. In case of multiple regression, a relationship is modeled between several independent variables and a dependent variable. In design effort estimation, the dependent variables predicted are design fees or workhours. Independent variables are design parameters and bridge characteristics. A linear regression model is represented as shown in equation (2).

$$\text{Estimated Design Fee/workhours} = I + V_1 \text{ x } C_1 + V_2 \text{ x } C_2 + V_3 \text{ x } C_3 + \cdots + V_n \text{ x } C_n \qquad (2)$$

Where;

I = Intercept;
V_i= ith input variable;
C_i= Coefficient with respect to ith input variable;
n= Number of input variables.

If the coefficient associated with a corresponding variable is positive, then the Design Fee/workhours increases with the increase in value of the corresponding variable. Similarly, if the coefficient of an input variable is negative, then the Design Fee/workhours decreases with the increase in value of the variable. In estimating bridge design fee/workhours, the input variables have to be in the numerical form, as multiple regression cannot use categorical variables as input data. Conversion of categorical data into numerical data is described in the previous sections of the report. In total, seven models are developed for four different types of design fees and three different types of workhours as shown below in Table 5.

Table 5. Multivariate linear regression models

Design Fee Estimation Models	Workhour Estimation Models
- Design Fee Proposed by the Consultant	- Workhours Proposed by the Consultant
- Design Fee Proposed by Iowa DOT	- Workhours Proposed by Iowa DOT
- Contracted Design Fee	- Contract Workhours
- Actual Design Fee paid	

Various data mining tools are available for developing regression models. In this study, the Microsoft data-mining client for Excel is used to develop regression models. Once a preliminary model is developed, the model is then optimized by discarding the input/independent variables one by one starting with a variable with the highest p-value. The final model consists of statistically significant independent variables with p-values less than 0.05. A p-value less than 0.05 means that the independent variable has a significant impact on the dependent variable to the tune of 95%. Any p-value less than 0.05 is considered acceptable in the model. The adjusted R-squared value is used to measure the performance of a regression model developed. The adjusted R-squared value ranges from 0% to 100%. If the value gets closer to 1, that indicates the model explains the data better. The adjusted R-squared value increases only if the addition of a new input/independent variable improves the performance of the existing model.

Design Fee Proposed by Consultant

The regression model developed for estimating consultant's proposed fee is provided in Equation (3) below. The significantly variables are shown in Table 6. The adjusted R-squared value of 69% is obtained for the model, which indicates that independent variables in this model can explain variability between the independent variables and the dependent variable up to 69% with an MAPE of 34%.

$$Y = 24{,}791.72 + 98{,}339.90(B) + 34{,}806.64(C) + 4.70(D) + 40{,}635.16(F) - 25{,}976.39(O) \quad (3)$$

Table 6. Significant attributes for consultant's proposed design fee (MLR)

Symbol	Significant Attributes
B	Beam spacing variability
C	Beam type
D	Bridge area
F	Horizontal curve requirement
O	Pier type

Workhours Proposed by Consultant

The regression model developed for estimating work hours proposed by consultant is presented in equation (4) and Table 7 shows the statistically significant variables. The adjusted R-squared value of 69% is obtained for this model with an MAPE of 31%.

$$Y = 523.44 + 323.14(A) + 689.12(B) + 0.02(D) + 374.01(F) + 123.68(G) + 200.34(J) \quad (4)$$

Table 7. Significant attributes for consultant's proposed design workhours (MLR)

Symbol	Significant Attributes
A	Abutment type
B	Beam spacing variability
D	Bridge area
F	Horizontal curve requirement
G	Number of construction stages
J	Sidewalk requirement

Design Fee Proposed by Iowa DOT

The regression model developed for estimating the design fee proposed by Iowa DOT is given in Equation (5). The independent variables shown in Table 8 are found to be significant and the adjusted R-squared value of 72% is obtained for this model with an MAPE of 25%.

$$Y = 38{,}899.70 + 112{,}380.96(B) + 3.15(D) + 11{,}819.08(G) \tag{5}$$

Table 8. Significant attributes for Iowa DOTs proposed design fee (MLR)

Symbol	Significant Attributes
B	Beam spacing variability
D	Bridge area
G	Number of construction stages

Workhours Proposed by Iowa DOT

The regression model developed for estimating the design fee proposed by Iowa DOT is given in Equation (6). The independent variables shown in Table 9 are found to be significant. By comparing Table 8 and 9, the significant variables for both estimation models are found to be the same. The adjusted R-squared value of 67% is obtained for this model with an MAPE of 21%.

$$Y = 442.24 + 923.30(B) + 0.02(D) + 112.36(G) \tag{6}$$

Table 9. Significant attributes for Iowa DOTs proposed design workhours (MLR)

Symbol	Significant Attributes
B	Beam spacing variability
D	Bridge area
G	Number of construction stages

Contracted Design Fee

The regression model developed for estimating the contracted amount of design fee is provided in Equation (7) with its statistically significant variables in Table 10. The adjusted R-squared value of 71% is obtained for this model with an MAPE of 31%.

$$Y= 10{,}096.86 + 60{,}480.42(B) + 23{,}615.73(C) + 2.60(D) + 39{,}996.58(F) + 9{,}386.75(I) \qquad (7)$$

Table 10. Significant attributes for contracted fee (MLR)

Symbol	Significant Attributes
B	Beam spacing variability
C	Beam type
D	Bridge area
F	Horizontal curve requirement
I	Number of piers

Contracted Workhours

The regression model developed for estimating the contracted amount of designed fee is provided in Equation (8) with its statistically significant variables in Table 11. The adjusted R-squared value of 70% is obtained for the model with an MAPE of 32%. In addition to the attributes in Table 10, number of different spans is an additional attribute, which is found to be significant for contracted workhours.

$$Y= 231.39 + 446.79(B) + 261.06(C) + 0.017(D) + 371.82(F) - 83.80(H) + 149.66(I) \qquad (8)$$

Table 11. Significant attributes for contracted workhours (MLR)

Symbol	Significant Attributes
B	Beam spacing variability
C	Beam type
D	Bridge area

Table 11. (Continued)

Symbol	Significant Attributes
F	Horizontal curve requirement
H	Number of different spans
I	Number of piers

Actual Fee Paid

The regression model for estimating the actual amount of design fee paid to consultant is provided in Equation (9) with its significant variables explained in Table 12. The adjusted R-squared value of 72% is obtained for this model with an MAPE of 32%.

$$Y= 5{,}164.36 +41{,}033.15(B) +29{,}953.34(C) +2.77(D) +28{,}144.07(F) +11{,}125.43(I) \qquad (9)$$

Table 12. Significant attributes for actual fee paid (MLR)

Symbol	Significant Attributes
B	Beam spacing variability
C	Beam type
D	Bridge area
F	Horizontal curve requirement
I	Number of piers

Summary

Table 13 shows the summary of significant attributes of linear regression models developed in this study. In this study, only 58 projects are considered in model development. It is inferred that beam spacing variability and bridge area are the most common significant attributes derived from these linear regression models. Furthermore, the adjusted R-squared values of all models are between 67-72%, which indicates that this method does not explain the variability of remaining 28-33% of the available data. This can be attributed to the lack of ability to track non-linear relation in the data by linear regression.

Table 13. Significant attributes for MLR models

		Consultant		Iowa DOT		Contract		Actual
Sl No.	Significant Attributes & Symbol	Fee	Work Hours	Fee	Work Hours	Fee	Work Hours	Fee
1	Abutment type (A)		✓					
2	Beam spacing variability (B)	✓	✓	✓	✓	✓	✓	✓
3	Beam type requirement (C)	✓				✓	✓	✓
4	Bridge Area(D)	✓	✓	✓	✓	✓	✓	✓
5	Bridge width variability (E)							
6	Horizontal curve requirement(F)	✓	✓			✓	✓	✓
7	Number of construction stages(G)		✓	✓	✓			
8	Number of different spans(H)						✓	
9	Number of piers(I)					✓	✓	✓
10	Pier type requirement (O)	✓						
11	Sidewalk requirement(J)		✓					
	Adj. R-squared Values	69%	69%	72%	67%	71%	70%	72%
	Mean absolute percentage error	34%	31%	25%	21%	31%	32%	32%
	Standard Error	$30,769	262	$24,107	221	$23,924	206	$23,116

Multivariate Polynomial Regression (MPR)

A polynomial term i.e. quadratic (squared) or cubic (cubed) transforms a linear regression model into a curve. The choice of degree and the evaluation of quality of fit solely depend on the judgement of the analyst. The powers of input variables are added to the equation in order to see if it increases R-square significantly. In this study 2nd degree, term was only considered in the regression equation. Before going ahead with polynomial regression, plots of cost/workhours vs bridge length and width were analyzed and a need of switching towards second degree was felt. Hence, polynomial regression models were developed. A polynomial regression model of second degree is represented as shown in equation (10).

$$\text{Estimated Design Fee/workhours} = I + V_1 \text{ x } C_1 + {V_1}^2 \text{ x } C_2 + V_2 \text{ x } C_3 + {V_2}^2 \text{ x } C_4 \ldots + {V_n}^k \text{ x } C_n \quad (10)$$

Where;

I = Intercept;
V_i= ith input variable;
C_i= Coefficient with respect to ith input variable;
K= order of the equation;
n= Number of input variables

Design Fee Proposed by Consultant

The regression model developed for estimating consultant's proposed fee is provided in Equation (11) below. The significantly variables are shown in Table 14. The adjusted R-squared value of 75% is obtained for the model, which indicates that independent variables in this model can explain variability between the independent variables and the dependent variable up to 75% with an MAPE of 38%.

$$Y = 29{,}889.65 - 96{,}177.49(B) + 24{,}982.49(C) + 31{,}438.13(F) + 5{,}885.34(G-0.3)(G-0.3) - 8{,}175.95(H-2.38)(H-2.38) + 11{,}502.96(I-2.24)(I-2.24) + 209.64(K) - 22{,}808.44(O) \quad (11)$$

Table 14. Significant attributes for consultant's proposed design fee (MPR)

Symbol	Significant Attributes
B	Beam spacing variability
C	Beam type
F	Horizontal curve requirement
G	Number of construction stages
H	Number of different spans
I	Number of piers
K	Bridge length
O	Pier type

Workhours Proposed by Consultant

The regression model developed for estimating work hours proposed by consultant is presented in equation (12) and Table 15 shows the statistically significant variables. The adjusted R-squared value of 78% is obtained for this model with an MAPE of 27%. By comparing Table 14 and Table 15, the significant variables for both estimation models are found to be the same.

$$Y = 321.85 + 700.06(B) + 195.90(C) + 335.33(F) - 397.73(G) + 240.46(G-0.3)(G-0.3) - 77.05(H-2.38)(H-2.38) + 110.08(I-2.24)(I-2.24) + 1.75(K) - 165.37(O) \quad (12)$$

Table 15. Significant attributes for consultant's proposed design workhours (MPR)

Symbol	Significant Attributes
B	Beam spacing variability
C	Beam type
F	Horizontal curve requirement
G	Number of construction stages
H	Number of different spans
I	Number of piers
K	Bridge length
O	Pier type

Design Fee Proposed by Iowa DOT

The regression model developed for estimating the design fee proposed by Iowa DOT is given in Equation (13). The independent variables shown in Table 16 are found to be significant and the adjusted R-squared value of 77% is obtained for this model with an MAPE of 25%.

$$Y = 39{,}392.2 + 111{,}767.02(B) + 6{,}670.81(G-0.3)(G-0.3) - 5{,}131.30(H-2.38)(H-2.38) + 8{,}829.01(K-300.44)(K-300.44) + 8{,}829.01(I-2.24)(I-2.24) + 132.99(K) \quad (13)$$

Table 16. Significant attributes for Iowa DOTs proposed design fee (MPR)

Symbol	Significant Attributes
B	Beam spacing variability
G	Number of construction stages
H	Number of different spans
I	Number of piers
K	Bridge length

Workhours Proposed by Iowa DOT

The regression model developed for estimating the design fee proposed by Iowa DOT is given in Equation (14). The independent variables shown in Table 17 are found to be significant. The adjusted R-squared value of 72% is obtained for this model with an MAPE of 22%.

$$Y = 457.50 + 748.05(B) + 195.34(F) + 68.93(G-0.3)(G-0.3) - 47.41(H-2.38)(H-2.38) + 1.03(K) + 0.0033(K-300.44)(K-300.44) \quad (14)$$

Table 17. Significant attributes for Iowa DOTs proposed design workhours (MPR)

Symbol	Significant Attributes
B	Beam spacing variability
F	Horizontal curve requirement
G	Number of construction stages
H	Number of different spans
K	Bridge length

Contracted Design Fee

The regression model developed for estimating the contracted amount of design fee is given in Equation (15) with its statistically significant variables in Table 18. The adjusted R-squared value of 87% is obtained for this model with an MAPE of 17%.

$$Y = -1957.32 - 31{,}548.10(A) + 56{,}608.87(B) + 26{,}359.64(C) - 21{,}946.80(E) + 44{,}483.82(F) - 36{,}708.10(G) + 17{,}752.73(G-0.3)(G-0.3) - 7{,}284.43(H) - 14{,}204.30(H-2.38)(H-2.38) + 191.44(K) + 0.96(K-300.44)(K-300.44) + 686.59(L) - 40.05(L-49)(L-49) + 15{,}761.55(M) \quad (15)$$

Table 18. Significant attributes for contracted fee (MPR)

Symbol	Significant Attributes
A	Abutment type
B	Beam spacing variability
C	Beam type
E	Bridge width variability
F	Horizontal curve requirement
G	Number of construction stages
H	Number of different spans
K	Bridge length
L	Bridge width
M	Type of work

Contracted Workhours

The regression model developed for estimating the contracted amount of designed fee is provided in Equation (16) with its statistically significant variables in Table 19. This model is found to be the best performing one as the adjusted R-squared value of 82% is obtained for the model with an MAPE of 16%.

$$Y = 337.77 - 262.64(A) + 394.51(B) + 196.17(C) + 367.78(F) - 363.59(G) + 182.99(G-0.3)(G-0.3) - 82.57(H) - 107.85(H-2.38)(H-2.38) + 1.78(K) + 0.0073(K-300.44)(K-300.44) \quad (16)$$

Table 19. Significant attributes for contracted workhours (MPR)

Symbol	Significant Attributes
A	Abutment type
B	Beam spacing variability
C	Beam type
F	Horizontal curve requirement
G	Number of construction stages
H	Number of different spans
K	Bridge length

Actual Fee Paid

The regression model for estimating the actual amount of design fee paid to consultant is provided in Equation (17) with its significant variables explained in Table 20. The adjusted R-squared value of 81% is obtained for this model with an MAPE of 24%.

$$Y = -2{,}316.94 + 45{,}660.39(B) + 26{,}827.17(C) + 27{,}203.83(F) - 8{,}175.42(H-2.38)(H-2.38) + 152.88(K) + 0.57(K-300.44)(K-300.44) + 677.09(L) - 33.51(L-49)(L-49) \quad (17)$$

Table 20. Significant attributes for actual fee paid (MPR)

Symbol	Significant Attributes
B	Beam spacing variability
C	Beam type
F	Horizontal curve requirement
H	Number of different spans
K	Bridge length
L	Bridge width

Summary

Table 21 shows the summary of significant attributes of polynomial regression models developed in this study. In this study, only 58 projects are considered in model development. It is inferred that beam spacing variability, bridge length and number of different spans are the most common significant attributes derived from these polynomial regression models. Furthermore, the adjusted R-squared values of all models are between 72-87%, which indicates that this model does not explain the variability of remaining 13-27% of the available data. The improvement in adj. R squared vales is due to the non-linear relationship between input and output variables and polynomial regression can account for this relationship, which cannot be possible by linear regression.

Table 21. Significant attributes for MPR models

		Consultant		Iowa DOT		Contract		Actual
Sl No.	Significant Attributes & Symbol	Fee	Work Hours	Fee	Work Hours	Fee	Work Hours	Fee
1	Abutment type (A)					✓	✓	
2	Beam spacing variability (B)	✓	✓	✓	✓	✓	✓	✓
3	Beam type requirement (C)	✓	✓			✓	✓	✓
4	Bridge width variability (E)					✓		
5	Horizontal curve requirement (F)	✓	✓		✓	✓	✓	✓
6	Number of construction stages (G)		✓			✓	✓	
7	Number of construction stages square (GG)	✓	✓	✓	✓	✓	✓	
8	Number of different spans (H)					✓	✓	
9	Number of different spans square (HH)	✓	✓	✓	✓	✓	✓	✓
10	Number of piers square (II)	✓	✓	✓				
11	Bridge Length (K)	✓	✓	✓	✓	✓	✓	✓
12	Bridge Length square (KK)				✓	✓	✓	✓
13	Bridge width (L)					✓		✓
14	Bridge width square (LL)					✓		✓
15	Type of work (M)					✓		
16	Pier type requirement (O)	✓	✓					
	Adj. R-squared Values	75%	78%	77%	72%	87%	82%	81%
	Mean absolute percentage error	38%	27%	25%	22%	17%	16%	24%
	Standard Error	$27,537	220	$21,755	206	$16,170	158	$19,147

Artificial Neural Network

Artificial Neural Network (ANN) is defined as a heuristic learning technique that aims at finding non-linear patterns and relationship between inputs and outputs by using a training dataset (Hsu et al. 1995). The technique has been used in the construction industry for the past two decades. For example, Hegazy and Ayed (1998) as well as Adeli and Wu (1998) used the ANN to estimate highway projects costs. Berry and Linoff (1997) defined neural network as a powerful tool for the purpose of prediction, classification and clustering sometimes termed it as mysterious “black boxes” for internal workings.

Figure 16 shows the main components of ANN. Typically; ANN consists of an input layer, a hidden layer(s), and an output layer. Additionally, each layer consists of a number of neurons that are assigned with different values during the training phase. As the number of neurons increase, the trained ANN is expected to perform better by reducing the prediction error. However, increasing the number of neurons may cause an overfitting problem. This means that even when the percentage error within the training dataset is very small, the percentage errors will be very large when new data is introduced to make predictions. As such, it is important to hold a portion of the data and use them as a testing dataset to measure the percentage error in prediction and ensure that the trained ANN is not over fitted. In this research, 15% of data points are used for testing while the other 85% are used for training.

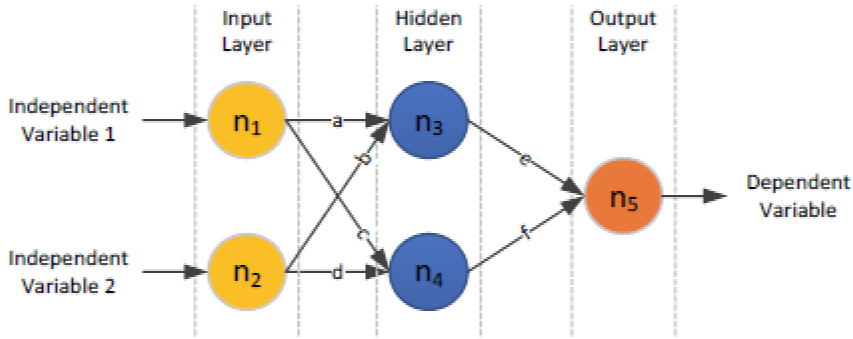

Figure 16. Typical neural network configuration

Just like MLR models, seven ANN models are developed to estimate the design fees and work hours. The following sixteen input variables are used for developing ANN models.

- Bridge length
- number of different spans
- type of work
- route type
- skew
- sidewalk requirement
- abutment type
- pier type
- bridge width
- number of piers
- number of construction stages
- bridge width variability
- beam spacing variable
- horizontal curvature
- beam type
- number of aesthetic items

It is worth noting that ANN can deal with any type of variables and hence no data transformation is needed. The ANN models are capable of showing complex input and output non-linear relationships (Young II et al., 2008). Initially, the input nodes capture the information. The captured information then is converted to a form, which could be numerically communicated. The captured information is expressed in terms of activation values where in each node is given a number; higher the number greater would be the activation. Later, this information is passed on through the network and based on connection strengths (weights), transfer functions (linear, sigmoid, Gaussian), the activation value is passed on from node to node. Every single node then sums up the activation values and modifies the value based on the transfer function. Through the hidden layers, the activation value flows through the whole network until it reaches the output node. The error i.e. difference between actual and predicted output will be promulgated back to each node based on the weights assigned to each node, which it would be liable for (Saed Sayad 2016). The neural networks are classified into fed-forward and feedback networks.

The hidden node ration or the number of neurons in the hidden layer is an important parameter that should be tuned in order to improve the performance of the neural network. In this study, ANN models are trained by using a hidden node ratio of eight since it achieved better performance than models with other ratios. The activation functions used were linear, TanH (Hyperbolic) and Gaussian while tuning the hidden layer structure. When training a new ANN, users are recommended to tune the activation functions until the best performance is reached. The commercial software tool used is predictive analytics software called JMP Pro developed by SAS institute. More information regarding JMP Pro is found through this link <http://www.jmp.com/support/help/Neural_Networks.shtml>.

Design Fee Proposed by Consultant

The results of ANN model developed for estimating consultant's proposed fee are shown below in Table 22. In the case of consultant's proposed fee, the R^2 value is found to be 60% for the validation data with MAPE error of 13%.

Table 22. ANN model results: Consultant's proposed fee

	Consultant's proposed fee	
Measures	Training data	Validation data
Rsquare	0.90	0.60
Standard Error	$17,993.37	$16,438.85
MAPE Error		13%

Workhours Proposed by Consultant

The results of ANN model developed for estimating consultant's proposed fee are shown below in Table 23. In the case of consultant proposed workhours, the R^2 value for the validation data is found to be 54% with MAPE error of 11%.

Table 23. ANN model results: Consultant's proposed workhours

	Consultant's proposed workhours	
Measures	Training data	Validation data
Rsquare	0.65	0.54
Standard Error	283.04	127.81
MAPE Error		11%

Design Fee Proposed by Iowa DOT

The results of ANN model developed for estimating Iowa DOTs counter-proposed fee are shown below in Table 24. In the case of Iowa DOTs counter-proposed fee, the R^2 value for validation data is found to be 68% with MAPE error of 6%.

Table 24. ANN model results: Iowa DOTs counter proposed fee

	Iowa DOTs counter proposed fee	
Measures	Training data	Validation data
Rsquare	0.97	0.68
Standard Error	$8,762.48	$7,189.49
MAPE Error		6%

Workhours Proposed by Iowa DOT

The results of ANN model developed for estimating Iowa DOTs counter-proposed workhours are shown below in Table 25. In the case of Iowa DOTs counter-proposed workhours, the R^2 value for validation data is found to be 48% with MAPE error of 6%.

Table 25. ANN model results: Iowa DOTs counter proposed workhours

	Iowa DOTs counter proposed workhours	
Measures	Training data	Validation data
Rsquare	0.99	0.48
Standard Error	34.24	60.35
MAPE Error		6%

Contracted Design Fee

The results of ANN model developed for estimating contracted fee are shown below in Table 26. In the case of contracted fee, the R^2 value for validation data is found to be 66% with MAPE error of 9%.

Table 26. ANN model results: Contracted fee

	Contracted fee	
Measures	Training data	Validation data
Rsquare	0.87	0.66
Standard Error	$16,137.81	$9,882.95
MAPE Error		9%

Contracted Design Workhours

The results of ANN model developed for estimating contracted workhours are shown below in Table 27. In the case of contracted workhours, the R^2 value for validation data is found to be 61% with MAPE error of 7%.

Table 27. ANN model results: Contracted workhours

	Contracted workhours	
Measures	Training data	Validation data
Rsquare	0.95	0.61
Standard Error	82.90	74.79
MAPE Error		7%

Actual Fee Paid

The results of ANN model developed for estimating actual fee paid are shown below in Table 28. In the case of actual fee paid, the R^2 value for validation data is found to be 67% with MAPE error of 5%.

Table 28. ANN model results: Actual fee paid

	Actual fee paid	
Measures	Training data	Validation data
Rsquare	0.44	0.67
Standard Error	$32,638.52	$5,236.31
MAPE Error		5%

Summary

The results from the developed ANN model are shown in Table 29 below. As seen below, the MAPE errors from all the seven models are within 14% with R^2 vales greater than 74%

Table 29. ANN model results using bridge design attributes

		Consultant		Iowa DOT		Contract		Actual
Sl No.	Parameters	Fee	Work Hours	Fee	Work Hours	Fee	Work Hours	Fee
1	R-squared Values - Validation	90%	65%	97%	99%	87%	95%	44%
2	R-squared Values - Training	60%	54%	68%	48%	66%	61%	67%
	Mean absolute percentage error	13%	11%	6%	6%	9%	7%	5%

Conclusion

As seen from Table 30, for estimation using bridge design attributes, the MPR regression models performs better in estimating the consultant's proposed workhours, contracted design fee, contracted design workhours and actual paid fee. In estimating the consultant's proposed fee, there is no significant difference between the two regression models. However, the MLR model performs better in estimating the Iowa DOTs proposed fee and workhours. Furthermore, if the standard error values and adjusted R-squared values are compared, then MPR outperforms MLR in all the seven prediction models. ANN models, on the other hand perform significantly better than MLR and MPR approaches by a huge margin because ANNs can model nonlinear relationships as well as interactions between the variables.

Table 30. MAPE for all MLR, MPR and ANN models

	MLR	**MPR**	**ANN**
Consultant's proposed fee	34%	38%	13%
Consultant's proposed workhours	31%	27%	11%
Iowa DOTs proposed fee	25%	25%	6%
Iowa DOTs proposed workhours	21%	22%	6%
Contracted design fee	31%	17%	9%
Contracted workhours	32%	16%	7%
Actual paid design fee	32%	24%	5%

CHAPTER 6. BRIDGE DESIGN ESTIMATION USING DESIGN ATTRIBUTES AND DESIGN SHEETS

Another approach to predict bridge design cost and workhours is carried out which had additional input variables in terms of design sheets. The design sheets are sub divided into standard design sheets, modified design sheets and new design sheets. The standard design sheets contain pre-engineered details for many structural components like sound walls, underground structures, earth retaining systems etc. Hence, the usage of such standard design sheets directly reduces the efforts in detaining for specific types of bridges i.e. PPCB bridges in this study. When any kind of modification is made to the standard design sheet, it becomes a modified sheet. Usually, the number of modifications on a design sheet is clearly indicated in the design sheet. Example for some modified sheets is abutment details, superstructure details, beam details etc. This modification would require more effort when compared to standard design sheet. The last one is the new design sheet in which the detailer would put in maximum effort as it requires starting from scratch like slab elevations, situation plans, slab reinforcing layout, pier details etc. Hence, new design sheets consume maximum efforts in bridge design. Not much is done in the past regarding these differences in design sheets hence, an effort was made to see if this difference in design sheets would help in improving the performance of models in a better way.

Initially, the trials are carried out using MLR but it is found that none of the number of standard sheets, modified sheets and new sheets is statistically significant in all the models apart from consultant workhours model and actual cost model. Furthermore, the MAPE error did not improve significantly as desired for these two models. Later, the trials are performed using multivariate polynomial regression with second degree. The MPR models did not perform better and had high MAPE error compared to the previous approach in all the models apart from contract cost model and contract workhours model. The MAPE error in these two models again did not improve significantly as desired. Hence, additional data of number of standard, modified and new sheets did not reveal better results using MLR and MPR approach. Finally, the trails are carried out using artificial neural network and the results found are comparatively better than the previous approach of just using bridge characteristics as input variables. The analysis is carried out using a commercial tool called JMP Pro and number of layers, hidden node ratio, learning rate and other activation functions are kept same as previously explained in chapter 6.

Design Fee Proposed by Consultant

The results of ANN model developed for estimating consultant's proposed fee are shown below in Table 31. In the case of consultant's proposed fee, the R^2 value is found to be 79% for the validation data with MAPE error of 12%.

Table 31. ANN model results: Consultant's proposed fee

	Consultant's proposed fee	
Measures	Training data	Validation data
Rsquare	0.90	0.79
Standard Error	$17,901.05	$12,811.31
MAPE Error		12%

Workhours Proposed by Consultant

The results of ANN model developed for estimating consultant's proposed fee are shown below in Table 32. In the case of consultant proposed workhours, the R^2 value for the validation data is found to be 71% with MAPE error of 10%.

Table 32. ANN model results: Consultant's proposed workhours

	Consultant's proposed workhours	
Measures	Training data	Validation data
Rsquare	0.87	0.71
Standard Error	169.59	112.42
MAPE Error		10%

Design Fee Proposed by Iowa DOT

The results of ANN model developed for estimating Iowa DOTs counter-proposed fee are shown below in Table 33. In the case of Iowa DOTs counter-proposed fee, the R^2 value for validation data is found to be 85% with MAPE error of 5%.

Table 33. ANN model results: Iowa DOTs counter proposed fee

	Iowa DOTs counter proposed fee	
Measures	Training data	Validation data
Rsquare	0.95	0.85
Standard Error	$11,054.85	$5,427.10
MAPE Error		5%

Workhours Proposed by Iowa DOT

The results of ANN model developed for estimating Iowa DOTs counter-proposed workhours are shown below in Table 34. In the case of Iowa DOTs counter-proposed workhours, the R^2 value for validation data is found to be 84% with MAPE error of 4%.

Table 34. ANN model results: Iowa DOTs counter proposed workhours

	Iowa DOTs counter proposed workhours	
Measures	Training data	Validation data
Rsquare	0.75	0.84
Standard Error	198.37	37.53
MAPE Error		4%

Contracted Design Fee

The results of ANN model developed for estimating contracted fee are shown below in Table 35. In the case of contracted fee, the R^2 value for validation data is found to be 84% with MAPE error of 7%.

Table 35. ANN model results: Contracted fee

	Contracted fee	
Measures	Training data	Validation data
Rsquare	0.93	0.84
Standard Error	$11,861.87	$8,523.91
MAPE Error		7%

Contracted Design Workhours

The results of ANN model developed for estimating contracted workhours are shown below in Table 36. In the case of contracted workhours, the R^2 value for validation data is found to be 75% with MAPE error of 6%.

Table 36. ANN model results: Contracted workhours

	Contracted workhours	
Measures	Training data	Validation data
Rsquare	0.97	0.75
Standard Error	69.85	76.40
MAPE Error		6%

Actual Fee Paid

The results of ANN model developed for estimating actual fee paid are shown below in Table 37. In the case of actual fee paid, the R^2 value for validation data is found to be 77% with MAPE error of 5%.

Table 37. ANN model results: Actual fee paid

	Actual fee paid	
Measures	Training data	Validation data
Rsquare	0.78	0.77
Standard Error	$20,240.72	$7,048.79
MAPE Error		5%

Summary

The results from the developed ANN model are shown in Table 38. As seen below, the MAPE errors from all the seven models are within 12% with R^2 vales greater than 71%.

Table 38. ANN model results using bridge design attributes and design sheets

		Consultant		Iowa DOT		Contract		Actual
Sl No.	Parameters	Fee	Work Hours	Fee	Work Hours	Fee	Work Hours	Fee
1	R-squared Values - Validation	90%	87%	85%	84%	93%	75%	77%
2	R-squared Values - Training	79%	71%	95%	75%	84%	97%	78%
	Mean absolute percentage error	12%	10%	5%	4%	7%	6%	5%

Conclusion

In this approach by adding three types of design sheets along with bridge data attributes, MLR and MPR models does not have design sheets as significant variables in most of the models. This only means that there is no significant linear or even a non-linear relation between different kinds of design sheets when plotted against design cost and workhours. The ANN models on the other hand perform slightly better when compared to ANN models in the previous approach by just using bridge data attributes. The improved performance in ANN can be attributed to complex interactions between number of design sheets and bridge attributes when plotted against design cost and workhours. However, the only impediment in this approach is the prediction of standard, modified and new sheets along with bridge data attributes. This impediment can be circumvented by comparing similar projects in the past and professional judgments in order to predict the number of different kinds of design sheets to be produced for a new bridge project.

CHAPTER 7. CONSOLIDATED CONCLUSIONS AND LIMITATIONS

Conclusions

This section presents the main findings from Chapters 5 and 6. Chapter 5 presents a method to estimate bridge design efforts and workhours by using only design attributes as input variables.

- For estimation using bridge design attributes, the MPR regression models performs comparatively better in estimating design costs and workhours when compared with MLR regression models. Furthermore, if the standard error values and adjusted R-squared values are compared, MPR outperforms MLR in all the seven prediction models.
- ANN models, on the other hand performs significantly better than MLR and MPR approaches by a huge margin due to the fact that ANNs can model nonlinear relationships as well as interactions between the variables which is not possible by MLR and MPR.

The MAPE values reported for validation projects range from 21-34% in MLR models, 16-38% in MPR models and 5-13% in case of ANN models.

Chapter 6 presents a method to estimate bridge design efforts and workhours by using design attributes and the number of various types of design sheets as input variables.

- Adding three types of design sheets (standard, modifies and new sheets) along with bridge data attributes, MLR and MPR models does not have design sheets as significant variables in most of the models which means that there is no significant linear or even a non-linear relation between different kinds of design sheets when plotted against design cost and workhours.
- The ANN models on the other hand perform slightly better when compared to ANN models in the previous approach by just using bridge data attributes. The MAPE values range from 4-12% for ANN models. The improved performance in ANN can be attributed to complex interactions between number of design sheets and bridge attributes when plotted against design cost and workhours.

In this study, only 50 data points are used for training regression and ANN models. The small size of the data points might be the main reason for the poor performance of regression models. The reason why the models did not perform well in case of linear regression is because in case of regression, the data is modeled by a straight-line function, which is not the case very often, and hence it does not recognize any non-linearity between the attributes. However, the polynomial regression models perform better than linear regression models because polynomial regression extends the linear regression to fit some general non-linear functions but it does not fit categorical variables with interactions. Overall it appears that neural network perform significantly better in predicting bridge design efforts and cost as it can model nonlinear relationships as well as interactions between the variables.

It is expected that the performance of all the models will continue to improve if the number of data points for training increases as the training algorithm will be able to detect and fine-tune the relationships between the inputs and outputs more accurately in the future. This leads to the conclusion that ANN models are well suited for use in model development for bridge design efforts and cost estimation as they can be trained easily without any conversion from categorical inputs to numerical inputs which can be laborious for any new data to be entered in the future and hence, it is recommended for use in bridge design efforts and cost prediction.

Limitations

Every research study has some limitations. It must be stressed that the developed models are valid only within the data in which they were trained. To add a few more limitations, if Iowa DOT plans to use this model in the future, the models can only be applicable to bridge projects designed by the firms/consultants considered in the training data.

Furthermore, the models developed are sensitive to the design techniques. The models cannot be used to predict bridge design costs if there is any change in the method of design.

To make the predictions more reliable, the models have to be updated frequently so that the database expands and performance of the model improves which is also supported by literature studies (Setyawati et al. 2002, Tatari and Kucukvar 2010, Gunduz et al. 2011).

CHAPTER 8. CONTRIBUTIONS AND RECOMMENDATIONS FOR FUTURE RESEARCH

Contributions

Chapter 2 provides the foundation for this research as it summarizes the current practices in various agencies in estimating PE costs. Chapter 3 determines the bridge characteristics by a questionnaire survey, which assisted in predetermining attributes before detailed design process starts. Sixteen bridge attributes are selected based on score of importance, which serves as input for prediction modeling. This research presents a case for estimating bridge design efforts and work hours for PPCB bridge projects. Chapter 5 and 6 presents popular predicting tools like regression and neural networks, which are utilized as a part of this research. Two approaches like bridge design attributes and design attributes with number of different types design sheets are applied to develop and compare the performance of predictions. The results indicate that the performance of ANN models is found to be better in all the approaches with fewer hassles in data transformation.

In this study, a major attempt is made to include different types of design sheets and number of various types of design sheets like standard sheets, modified sheets and new sheets. Most of the bridge designed by consultants in this study have work break down hours for each design sheets submitted by them. When the workhours of a new bridge project are to be estimated, the DOT can compare similar projects from the past data and estimate the number of standard, modified and new sheets, which will be prepared by the consultant for the new project. Later, during the process of negotiation DOT can have a good idea as to how many new sheets would be prepared by the consultant and based on the complexity of the project and the workhours allotted for each type of design sheet, the new contract workhours can be well negotiated as standard and modified sheets would not require the same number of workhours when compared to designing new sheets.

Recommendations for Future Research

The major observation with the historic bridge workhours data is that most of this data, which is derived from consultants, which is used in training the model is not uniform in terms of designation and expertise of engineers involved in designing the project. In this study, the models were prepared by collecting the bridge data from 14 different consultants. The problem with this method is of assumption that all 14 firms use the same software's, same approach and same experience of engineers for all the projects, which may not be true. The effort should be made in standardizing the data as much as possible, which can be possible by segregating the projects, by individual firms to bring some uniformity in terms of team set up because no matter how sophisticated the tools are but finally, the designs are analyzed by humans. The models by this method would be expected to perform a lot better as the training data is more standardized and can be utilized to deal with the particular firm.

An approach recommended by Craigie (2015) is to estimate bridge design costs and workhours using a "bottom-up" approach in other words a CSBS (Cost and scope breakdown structure). The CSBS is basically a spreadsheet, listing preconstruction tasks specific to different departments to which work hours are assigned. In this bottom up approach also called as functional level estimate the design tasks are divided into work breakdown structure (WBS) which are then grouped into various packages. This approach provides a methodological classification of work tasks which can be clearly identified and efficiently managed.

When this same CSBS approach be applied for an individual firm taking its historical data into consideration, the performance of the model would probably be better. The only issue with this is the availability of enough data of an individual firm for a specific kind of bridge within a DOT. The only way this issue can be tackled is by having collaboration with various DOTs in collecting data for a particular firm and for a specific kind of bridge as firms work for many DOTs nationwide. This way plenty of projects can be included in the data base for prediction modeling.

BIBLIOGRAPHY

Adeli, H., & Wu, M. (1998). Regularization neural network for construction cost estimation. Journal of Construction Engineering and Management, 124(1), 18-24.

Adeli, H. (2001). Neural networks in civil engineering: 1989-2000. Computer-Aided Civil and Infrastructure Engineering, 16(2), 126-142.

Bell, L. C., and Ghazanfer, A. B. (1987). "Preliminary Cost Estimating For Highway Construction Projects." *AACE Transactions*, C6.1-C6.4.

Berry, M. J., & Linoff, G. (1997). Data mining techniques: for marketing, sales, and customer support. John Wiley & Sons, Inc.

Carr, P. G., & Beyor, P. S. (2005). Design fees, the state of the profession, and a time for corrective action. Journal of Management in Engineering,21(3), 110-117.

Creese, R. C., and Li, L. (1995). "Cost Estimation of Timber Bridges Using Neural Networks." *Cost Engineering*, 34(5), 17-22.

Douglas D. Gransberg, H. David Jeong, Emily K. Craigie, Jorge A. Rueda-Benavides, and K. Joseph Shrestha. (2015) "Preconstruction Services Cost Estimating Guidebook", National Cooperative Highway Research Program (NCHRP 15-51), Transportation Research Board, Washington DC.

Ezeldin, A. S., & Sharara, L. M. (2006). Neural networks for estimating the productivity of concreting activities. Journal of construction engineering and management, 132(6), 650-656.

Flood, I. (2006). Next generation artificial neural networks for civil engineering. Journal of Computing in Civil Engineering, 20(5), 305-307.

General Services Administration (2015). Code of Federal Regulations, Part 172-Administration of Engineering and Design Related Service Contracts. 23 CFR Section 172.7.

Gransberg, D. D., Lopez del Puerto, C., & Humphrey, D. (2007). Relating cost growth from the initial estimate to design fee for transportation projects. Journal of construction engineering and management, 133(6), 404-408.

Gunduz, M., Ugur, L. O., and Ozturk, E., (2011). "Parametric cost estimation system for light rail transit and metro trackworks." *Expert Systems with Application*, 38, 2873-2877.

Hegazy, T., and Ayed, A., (1998). "Neural Network Model for Parametric Cost Estimation of Highway Projects." *Journal of Construction Engineering and Management*, 124(3), 210-218.

Hollar, D. A., Rasdorf, W., Liu, M., Hummer, J. E., Arocho, I., & Hsiang, S. M. (2012). Preliminary Engineering Cost Estimation Model for Bridge Projects. Journal of Construction Engineering and Management, 139(9), 1259-1267.

Hsu, K. L., Gupta, H. V., & Sorooshian, S. (1995). Artificial neural network modeling of the rainfall-runoff process. Water resources research, 31(10), 2517-2530.

Iglewicz, B., & Hoaglin, D. C. (1993). How to detect and handle outliers (Vol. 16). Asq Press.

Iowa Department of Transportation, I-JOBS Transportation Funding. (2014) Retrieved from: http://www.iowadot.gov/IJOBS/ijobsfunding.html

Kim, G. H., Yoon, H. E., An, S. H., Cho, H. H., & Kang, K. I. (2004). Neural network model incorporating a genetic algorithm in estimating construction costs. Building and Environment, 39(11), 1333-1340.

Knight, K., & Robinson Fayek, A. (2002). Use of fuzzy logic for predicting design cost overruns on building projects. Journal of Construction Engineering and Management, 128(6), 503-512.

Kuprenas, J. A. (2003). Project management actions to improve design phase cost performance. Journal of Management in Engineering, 19(1), 25-32.

Kyte, C. A., Perfater, M. A., Haynes, S., and Lee, H. W. (2004a). “Developing and validating a highway construction project cost estimation tool.” Rep. 05-R1, Virginia Transportation Research Council, Charlottesville, VA.

Kyte, C. A., Perfater, M. A., Haynes, S., and Lee, H. W. (2004b). “Developing and validating a tool to estimate highway construction project costs.” Transportation Research Record 1885, Transportation Research Board, Washington, DC, 35–41.

Lowe, D. J., Emsley, M. W., and Harding, A. (2006). “Predicting Construction Cost Using Multiple Regression Techniques.” *Journal of Construction Engineering and Management*, 132(7), 750-758.

Mahamid, I. (2011). “Early Cost Estimating for Road Construction Projects Using Multiple Regression Techniques.” *Australasian Journal of Construction Economics and Building*, 11(4), 87-101.

Nelson, J., Waterhouse, E. (2016). The Development of a Parametric Design Time Estimating Tool for Pretensioned Prestressed Concrete Beam Bridges in Iowa. *PCI Convention and National Bridge Conference: Nashville, Tennessee*, Proceedings [USB Flash Drive]

Persad, K. R., Oconnor, J. T., & Varghese, K. (1995). Forecsting engineering manpower requirements for highway preconstruction activities. Journal of Management in Engineering, 11(3), 41-47.

Setyawati, B. R., Sahirman, S., and Creese. R. C. (2002). “Neural Networks for Cost Estimation.” *AACE International Transactions*, 13.1-13.9.

Svicarovich, K., Wang, Z., Zhang, G., Walton, C. M., Conway, A. J., & Persad, K. (2010). Arterial Intelligent Transportation Systems Implementation and Integration: A Simulation-Based Case Study. In ITS America 20th Annual Meeting & Exposition.

Saed Sayad. 2016. An Introduction to Data Mining. [ONLINE] Available at: http://www.saedsayad.com/artificial_neural_network.htm. [Accessed 6 December 2016].

Tatari, O., and Kucukvar, M. (2011). “Cost premium prediction of certified green buildings: A neural network approach. Building and Environment.” *Building and Environment*, 46, 1081-1086.

Turner, B. L., & Miller, J. S. (2015). Forecasting Project Preliminary Engineering Costs. In Transportation Research Board 94th Annual Meeting (No. 15-0975).

Turochy, R. E., Hoel, L. A., & Doty, R. S. (2001). Highway project cost estimating methods used in the planning stage of project development (pp. 1-290). Virginia Transportation Research Council.

Utah Department of Transportation (UDOT) (2013). Consultant Services Manual of Instruction. Salt Lake City, UT.

Wilmot, C. G., & Cheng, G. (2003). Estimating future highway construction costs. Journal of Construction Engineering and Management-Asce, 129(3), 272-279.

Williams, Trefor P., Neville Parker, and James Klotz. 2013. “Development of an Estimation Tool for Transportation Project Design Effort.” In Transportation Research Board 92nd Annual Meeting

Woldesenbet, A., & Jeong, D. H. S. (2012). Historical data driven and component based prediction models for predicting preliminary engineering costs of roadway projects. In Construction Research Congress 2012: Construction Challenges in a Flat World (pp. 417-426). ASCE.

WSDOT (2002), “Highway Construction Cost Comparison Survey”, Washington State Department of Transportation, April 2002, <http:/www.wsdot.wa.gov/biz/construction/pdf/IC_Const_Cost.pdf> (July 7, 2008) [WSDOT 2002]

Young II, W. A., Holland, W. S., & Weckman, G. R. (2008). Determining hall of fame status for major league baseball using an artificial neural network. Journal of Quantitative Analysis in Sports, 4(4), 1-44.

Yu, J. (2015). Development of Baseline Work Plan Templates for Transportation Projects. In Transportation Research Board 94th Annual Meeting (No. 15-0085).