OKLAHOMA TRANSPORTATION CENTER

ECONOMIC ENHANCEMENT THROUGH INFRASTRUCTURE STEWARDSHIP

PROCEDURES AND MODELS FOR ESTIMATING PRECONSTRUCTION COSTS OF HIGHWAY PROJECTS

"DAVID" HYUNG SEOK JEONG, PH.D.
ASREGEDEW WOLDESENBET

OTCREOS10.1-19-F

Oklahoma Transportation Center
2601 Liberty Parkway, Suite 110
Midwest City, Oklahoma 73110

Phone: 405.732.6580
Fax: 405.732.6586
www.oktc.org

DISCLAIMER

The contents of this report reflect the views of the authors, who are responsible for the facts and accuracy of the information presented herein. This document is disseminated under the sponsorship of Department University Transportation Center Program in the interest of information exchange. The U.S. Government assumes no liability for the contents or use thereof.

TECHNICAL REPORT DOCUMENTATION PAGE

1. Report No. OTCREOS10.1-19-F	Government Accession No.	Recipient's Catalog No.

4. Title and Subtitle Procedures and Models for Estimating Preconstruction Costs of Highway Projects	5. Report Date July 2012 6. Performing Organization Code
7. Authors "David" Hyung Seok Jeong and Asregedew Woldesenbet	8. Performing Organization Report
9. Performing Organization Name and Address Oklahoma State University Civil & Environmental Engineering 207 Engineering South Stillwater, OK 74078	10. Work Unit No. 11. Contract or Grant No. DTRT06-G-0016
12. Sponsoring Agency Name and Address The Oklahoma Transportation Center (Fiscal) 201 ATRC, Stillwater OK 74078 (Technical) 2601 Liberty Parkway, Suite 110 Midwest City, OK 73110	13. Type of Report and Period Covered Final Report June 2010-July 2012 14. Sponsoring Agency Code

15. Supplementary Notes
University Transportation Center

16. Abstract

This study presents data driven and component based PE cost prediction models by utilizing critical factors retrieved from ten years of historical project data obtained from ODOT roadway division. The study used factor analysis of covariance and correlation matrix to investigate the significance of factors affecting preliminary engineering cost and identify the correlation among them. A combination of three models; regression models, decision tree models and neural network models developed to obtain the optimum roadway PE cost model is presented as a function of three entities a) engineering hours required per number of sheets, b) number of sheets and c) cost per engineering hours for various plan development task outputs. An Excel-based program called Roadway PE Cost Estimator which allows ODOT engineers to estimate a reasonable PE cost is also presented in the study.

The results of this project is expected to significantly influence how efficiently and economically highway projects are planned, executed and managed in the early stages of a project. The developed system allows the engineer to estimate the values for the respective entities; number of sheets, work effort (engineering hours) required per number of sheets, cost per engineering hour, and total cost for a project based on factors such as project length, location, route type and project type. In addition, this system not only allows engineers to easily manipulate the requirements for a specific plan development task, but also help them configure contingencies, as to whether any of the entities are either under/over-estimated, or if there exists misallocation of resources (engineering hours assigned to the respective skilled manpower or number of sheets assigned at a specific level especially when negotiating PE costs with consulting firms.

17. Key Words	18. Distribution Statement
Preliminary engineering, cost estimate, roadway project, preconstruction cost estimate	No restrictions. This publication is available at www.oktc.org and from NTIS.

19. Security Classification. (of this report)	20. Security Classification. (of this page)	21. No. of Pages	22. Price
Unclassified	Unclassified	92 + covers	

SI (METRIC) CONVERSION FACTORS

Approximate Conversions to SI Units				
Symbol	When you know	Multiply by	To Find	Symbol
LENGTH				
in	inches	25.40	millimeters	mm
ft	feet	0.3048	meters	m
yd	yards	0.9144	meters	m
mi	miles	1.609	kilometers	km
AREA				
in^2	square inches	645.2	square millimeters	mm
ft^2	square feet	0.0929	square meters	m^2
yd^2	square yards	0.8361	square meters	m^2
ac	acres	0.4047	hectares	ha
mi^2	square miles	2.590	square kilometers	km^2
VOLUME				
fl oz	fluid ounces	29.57	milliliters	mL
gal	gallons	3.785	liters	L
ft^3	cubic feet	0.0283	cubic meters	m^3
yd^3	cubic yards	0.7645	cubic meters	m^3
MASS				
oz	ounces	28.35	grams	g
lb	pounds	0.4536	kilograms	kg
T	short tons (2000 lb)	0.907	megagrams	Mg
TEMPERATURE (exact)				
°F	degrees Fahrenheit	(°F-32)/1.8	degrees Celsius	°C
FORCE and PRESSURE or STRESS				
lbf	poundforce	4.448	Newtons	N
lbf/in^2	poundforce per square inch	6.895	kilopascals	kPa

Approximate Conversions from SI Units				
Symbol	When you know	Multiply by	To Find	Symbol
LENGTH				
mm	millimeters	0.0394	inches	in
m	meters	3.281	feet	ft
m	meters	1.094	yards	yd
km	kilometers	0.6214	miles	mi
AREA				
mm^2	square millimeters	0.00155	square inches	in^2
m^2	square meters	10.764	square feet	ft^2
m^2	square meters	1.196	square yards	yd^2
ha	hectares	2.471	acres	ac
km^2	square kilometers	0.3861	square miles	mi^2
VOLUME				
mL	milliliters	0.0338	fluid ounces	fl oz
L	liters	0.2642	gallons	gal
m^3	cubic meters	35.315	cubic feet	ft^3
m^3	cubic meters	1.308	cubic yards	yd^3
MASS				
g	grams	0.0353	ounces	oz
kg	kilograms	2.205	pounds	lb
Mg	megagrams	1.1023	short tons (2000 lb)	T
TEMPERATURE (exact)				
°C	degrees Celsius	9/5+32	degrees Fahrenheit	°F
FORCE and PRESSURE or STRESS				
N	Newtons	0.2248	poundforce	lbf
kPa	kilopascals	0.1450	poundforce per square inch	lbf/in^2

ACKNOWLEDGEMENTS

The authors would like to thank ODOT Roadway Design Division: Assistant Division Engineer - Project engineer, Mr. Kirk Goins for sharing his experience, support in extracting ODOT's engineering contract data and continuous effort in completion of this research project. The authors would like to extend their appreciation to various ODOT division representatives including Roadway Division Geometric Design, Mr. Reza Amini, Capital Programs Information System Liaison, Mr. Randy Jones, Survey Division Chief, Mr. Larry D. Reser, Environmental Division, Ms. Siv Sundaram, and Local Government, Mr. Mark Scott.

PROCEDURES AND MODELS FOR ESTIMATING PRECONSTRUCTION COSTS OF HIGHWAY PROJECTS

Final Report

July 31, 2012

Hyung Seok "David" Jeong, Ph.D.
Associate Professor

Asregedew Woldesenbet
Graduate Research Assistant

School of Civil & Environmental Engineering
Oklahoma State University
Stillwater, Oklahoma

Technical Advisor:
Kirk Goins
ODOT Roadway Design Division: Assistant Division Engineer-Project Engineer

Contents

List of Tables

List of Figures

Executive Summary

The purpose of this report is to present data-driven and component-based preliminary engineering (PE) cost prediction models that allow Oklahoma Department of Transportation (ODOT) engineers to reasonably estimate PE costs by utilizing potential factors retrieved from historical roadway project data. The report shows the effect of project-level factors such as project length, route type, location and project type affect PE costs by breaking PE cost into 11 plan development task outputs. A total of 353 roadway preliminary engineering contract data were acquired through ten years of engineering contract documents and interviews and meetings with engineers from the ODOT. In addition, 26 potential factors affecting PE costs were identified and classified into five major categories of project-level factors: project scope, geographical attributes, design attributes, environmental attributes, and external factors.

The report summarizes a comprehensive review of literature of previously conducted studies and current practice of estimating preliminary engineering costs. The report presents the use of a data-driven and component-based estimation system which consists of three functions or entities: a) engineering hours required per number of sheets, b) number of sheets and c) cost per engineering hours for selected 11 major plan development task outputs. The study also discusses the data collection process, data preparation and cleanup process, the significance and correlation of factors, and the development of three prediction models (decision tree models, regression models, and neural network models). A comparison of the three models is conducted and documented in this report to obtain the optimum models along with the development and validation of a standalone Microsoft Excel program to reasonably estimate roadway PE costs.

The developed program allows the engineer to determine a component-based PE cost prediction for the various plan development task outputs with respect to project-level factors. The program also allows engineers to easily manipulate the requirements for a specific task and helps to configure discrepancies or if there exists misallocation of resources (engineering hours assigned to the respective skilled manpower or number of sheets assigned to each task) at a specific level especially when negotiating PE costs with consulting firms. The results of this project are expected to significantly influence how efficiently and economically highway projects are planned, executed, and managed in the early stages of a project. The developed model will not only allow ODOT to be equipped with a streamlined procedure for estimating PE costs but also facilitate consistent practices and a structured format of PE cost estimating.

1. Introduction

1.1 Overview

Preliminary engineering (PE) begins when a federal or state agency receives funding authorization for planning and/or design activities. The delivery of these construction documents for project bid preparation marks the end of PE. Preliminary engineering of projects is the foundation of highway construction to meet the three basic components of project management: scope (requirements), schedule (project completion on time), and budget or cost. Therefore, the management of construction highway projects is established primarily with a reliable PE or preconstruction cost estimate of construction projects.

Estimating PE costs accurately and efficiently is critical for highway agencies due to the fact that PE costs are a large portion of overall project costs and are difficult to estimate in the early project stages. Inaccurate forecast of preliminary cost estimates of highway construction projects will create problems in the financial operations of highway agencies due to marginal budgets which affect these agencies from getting future transportation funds, create inconvenience to the public and additional cost and time for the contractor (Oberlender & Trost, 2001, and Chou et al., 2009). Therefore the accuracy of preliminary engineering cost estimates is not only beneficial to the client, but also creates a win-win situation to all parties involved in the project including the public.

1.2 Problem Statement

The Oklahoma Department of Transportation (ODOT) outsources approximately 50% of their preliminary engineering (design) tasks to consulting firms. Highly complex and high-cost urban projects are typically outsourced, which results in much higher percentage in terms of dollar values. PE costs represent approximately 6 % to 20 % of the overall project cost depending on the type and complexity of a project. However, only one engineer with two technicians in the roadway division is in charge of estimating project construction costs and PE costs. In addition, ODOT does not have any specific guidelines or tools to estimate these costs and highly depends on the estimators' experience, skills, and judgments. This practice does not allow ODOT to estimate reliable PE costs. Despite the significant role of preconstruction studies directed at PE cost estimation, it is minimal in contrast to the amount of research aimed at improving construction estimates. Researchers who focused on preconstruction management have also noted the lack of predictive tools to estimate design costs (Knight and Fayek 2002). Therefore, there is a need to develop a system to accurately estimate PE costs required to plan and design state roadway projects.

1.3 Research Objectives

This project is to develop practical tools to help ODOT engineers in estimating PE costs of roadway projects in a more consistent and reliable manner. The ultimate goal of this study is to develop a framework for estimating preliminary engineering costs of roadway projects. The objectives of this study include:

1) Develop a comprehensive list of factors affecting PE costs in the planning, design, and preconstruction stages of roadway projects in ODOT.
2) Determine uncorrelated significant factors using a factor analysis technique.
3) Determine optimum PE cost prediction models based on comparison of 3 models (decision tree models, regression models and neural network models).
4) Develop a spreadsheet-based software program to estimate PE costs of roadway projects.
5) Validate the software program, and train ODOT engineers.

1.4 Research Methodology

This study is divided into three major phases to address the objectives of this research as shown in Figure 1-1. In the first phase, the researchers employed three tasks: a) review of literature, b) a series of meetings and interviews with ODOT engineers, and c) highway project data collection. These three tasks are conducted to identify the current practices of state agencies and studies from the academia and industry, determine potential factors affecting PE costs, and acquire historically stored project data. In the second phase of the study, d) data preparation and classification, e) data analysis, and f) model development are performed to categorize factors affecting PE costs, identify the relationship and significance of factors, and develop PE cost prediction models, respectively. In the third phase, g) comparison of the developed models is conducted to find the optimum PE cost model, and h) a spreadsheet-based software program is developed and validated to allow ODOT engineers estimate reliable PE costs of roadway projects.

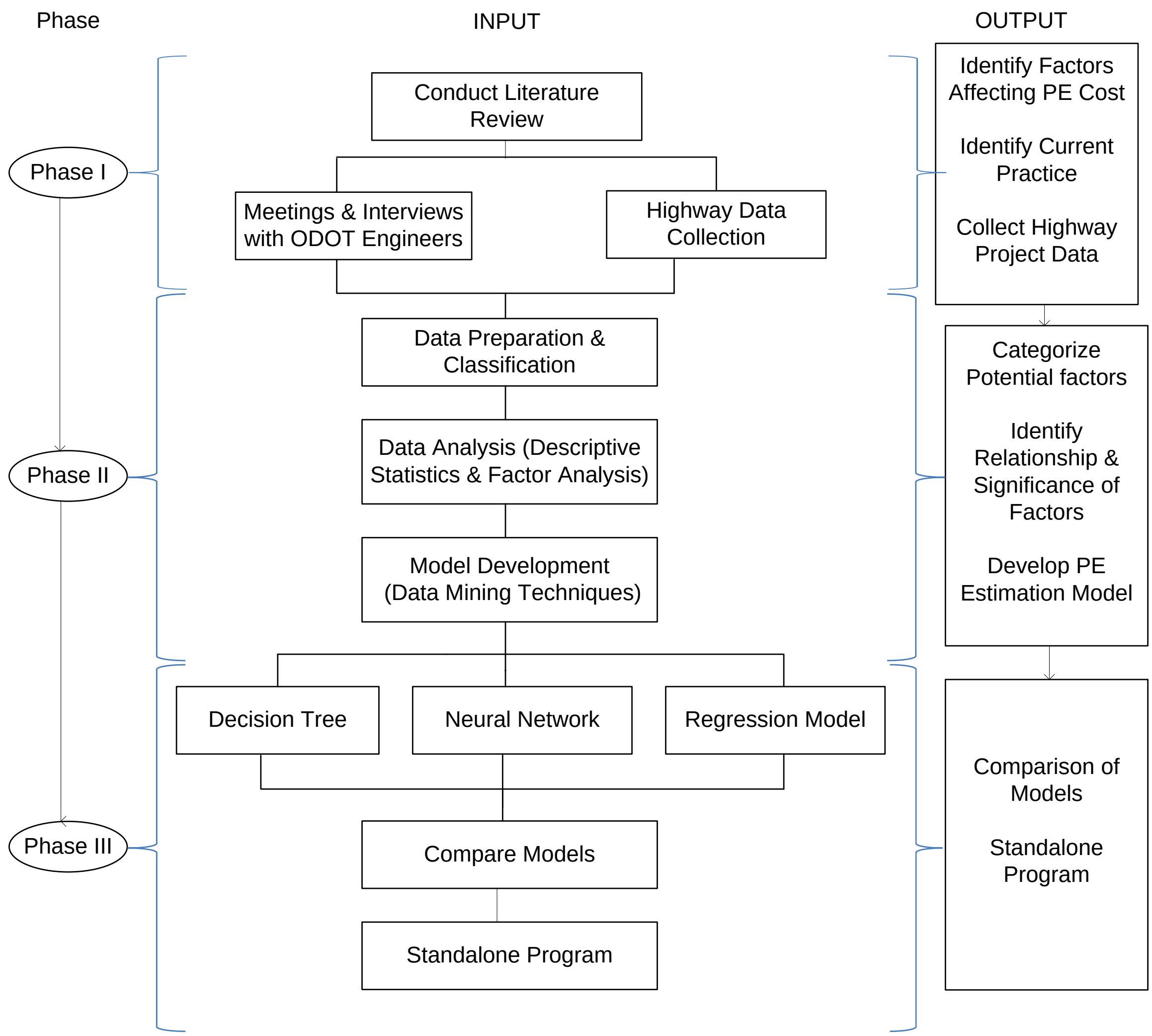


Figure 1-1 Research Methodology Flow Chart

1.4.1 Literature Review

The literature review focuses on prior research and relative studies on identifying current approaches used by various researchers and DOTs in the determination of PE costs of roadway projects. In addition, the study briefly investigates potential factors affecting the preconstruction process and identifies guidelines for estimating PE costs.

1.4.2 Meetings and Interviews

A series of meetings and interviews is conducted with ODOT's central office highway divisions which are responsible for the planning and managing highway projects from the early phase of conceptual estimation to construction document development and bid preparation. These meetings and interviews are primarily aimed at identifying current practice of ODOT's preliminary engineering cost estimation practice, cost estimating management, and contract negotiation process during the planning and preconstruction phase of highway projects.

1.4.3 Highway Data Collection

In this stage, the researchers investigate the existing Oklahoma roadway project database and collect 10 years of previously completed highway engineering contract documents along with PE cost data from the ODOT Roadway division. A data collection Excel spreadsheet is prepared to collect PE costs and potential factors that affect PE costs.

1.4.4 Data Preparation and Classification

Once data collection is completed, data cleanup and preparation is performed for conducting data analysis and developing reliable PE cost prediction models. It includes removing outliers, transforming functions, and replacing missing data points to achieve normal distribution and reduce bias in developing reliable PE cost prediction models. In addition, highway projects are classified into categories based on potential factors such as location, project type, highway type, contract year, and route type.

1.4.5 Data Analysis

Based on the collected data, principal component factor analysis is conducted to measure the relationship between the potential factors and their strengths. Use of factor analysis allows a large number of inter-correlated variables to be reduced into a smaller number of uncorrelated factors. In addition, a descriptive statistics analysis is performed to visualize the effect and trend of factors on PE costs.

1.4.6 Model Development

PE cost prediction models are developed using knowledge discovery in database methods or data mining techniques. For this study, a) decision tree models, b) regression models, and c) neural network models are used to estimate PE costs. Three types of components are used to predict the PE costs of roadway projects: a) engineering hours (man-hours), b) aggregate cost/engineering hours, and c) PE cost. These component–based prediction models will allow ODOT engineers to easily determine the PE cost and work effort required for the various plan development tasks and negotiating purposes with consulting firms.

1.4.7 Model Comparison

Once all models are developed, the next step is to evaluate these three models and obtain the optimum model which will reliably and accurately predict preliminary engineering costs of roadway projects. For this study, two fit statistics, average square error (ASE), and the root average squared error (RASE) are selected for determining the optimum models based on the measure of the variance or square difference between the actual and the predicted value.

1.4.8 Standalone Software Program & Validation

Based on the models developed, a component-based estimation tool for a reliable estimation of roadway PE costs and facilitating the negotiation process with consulting firms will be presented. The tool or system is developed using Microsoft Excel Spreadsheet by utilizing the regression models. The system, Roadway PE Cost Estimator, consists primarily of three sections or tabs: component-based PE cost estimator based on engineering hours, project-level factors-based PE cost estimator, and number of sheets estimator.

1.5 Report Organization

This report is organized into six chapters. Chapter 2 summarizes prior studies conducted in estimating PE costs and current practices by state DOTs. Chapter 3 presents the process of PE cost data acquisition from roadway engineering contract documents and ODOT's PE cost estimation procedure. Data cleanup, preparation, and results of principal component factor analysis are shown in Chapter 4. Chapter 5 illustrates the development and comparisons of component-based PE cost prediction models. Spreadsheet-based software program and the validation of the program are demonstrated in Chapter 6. The final chapter summarizes the findings, contribution, and future studies of this research. In addition, statistical analysis results and models developed in this study are provided in the appendices at the end of the report.

2. Literature Review

This chapter summarizes prior research work and relative studies conducted on identifying current approaches used by various researchers and DOTs in the determination of PE costs. Potential factors affecting the preconstruction process and guidelines for estimating preliminary engineering costs are also investigated in this chapter.

2.1 Major Phases of Highway Construction Projects

Prior to the commencement of a highway construction project, there are various phases or stages that need to be implemented for an effective completion of the project. One of these phases is the preliminary engineering, or preconstruction, phase. The preliminary engineering (design) phase is initiated once the government allocates a budget to transportation highway agencies for the construction of new highway projects or rehabilitation projects and ends with the delivery of construction documents for project bid preparation. During this phase highway agencies perform various types of tasks starting from preparation of preliminary drawings and preliminary estimate of construction cost to final design (detailed estimate) and preparation of contract documents to achieve the goal of solving transportation needs. Figure 2-1 describes the major phases of a typical highway construction project.

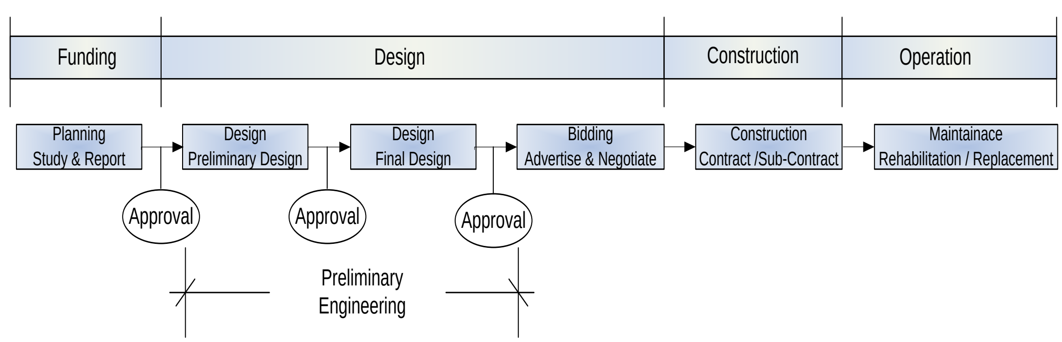


Figure 2-1 Phases of Highway Construction Project

Based on the availability of personnel in highway agencies, the preliminary engineering tasks are either conducted by an in-house team or outsourced to consulting firms or engineers. ASCE (2003) describes the major phases and tasks that need to be accomplished in highway construction projects. The manual classifies the involvement of highway agencies and consulting engineers in a construction project into six major phases.

a) Study and report phase – In this phase, the client's needs are identified; the project scope is defined; and the economic and technical evaluations of feasible alternatives are performed. Services performed by the consulting engineers during this phase may include clarifying project requirements, preliminary data collection (geotechnical investigation, environmental assessment, traffic studies, etc.), identifying government regulations, preparing conceptual designs, and probable cost of the project.
b) Preliminary design phase – this phase involves preparation of preliminary drawings, outlining specifications and preliminary estimate of construction cost. This phase includes services such as reviewing reports and available data, preparing preliminary design documents, determining right-of-way, and preparing revised probable project cost.
c) Final design phase – this phase includes preparation of construction drawings, estimates of probable construction cost, and preparation of other contract documents for initiating the bidding phase. Once the preliminary design phase is approved, a revised total project cost estimate is prepared; applications for regular permits from all authorities are obtained and completed; and construction contract documents (general condition, supplemental conditions, invitation to bid, etc.) are prepared.
d) Bidding/negotiating phase – this phase involves the bidding or negotiating process for the construction of the project. Services performed in this phase include assisting the client in evaluating bids, negotiating on differences between client and the bidder, attending pre-bid conferences, and interpreting and clarifying bid documents.
e) Construction phase – During this phase, the consulting engineers represent the client during the construction of the project and inspect the construction work. Services include reviewing the contactor's submitted shop drawings, inspecting and reviewing tests on materials and equipment, site visit and reporting progress, etc.
f) Operation phase – In this phase, the team assists the client in startup and operation of the project, with the inclusion of periodic inspections. The consultant may prepare a manual for operation and maintenance, observe and report on project operations, identify problems, and assist in obtaining solutions.

The National Cooperative Highway Research Program (NCHRP) has conducted a study to analyze cost estimation and management for the planning, programming, and preconstruction stages of highway projects. Table 2-1 summarizes the development phases of a construction project adopted from NCHRP Synthesis of Highway Practice 331: *Statewide Highway Letting Program Management* (Anderson and Blaschke, 2004).

Table 2-1 Development Phases & Activities of Construction Project (Anderson and Blaschke, 2004)

Development phase	Typical Activities
Planning	Determine purpose and need, determine whether it's an improvement or requirement study, consider environmental factors, facilitate public involvement/participation, and consider interagency conditions
Programming and Preliminary Design	Conduct environmental analysis, conduct schematic development, hold public hearings, determine right-of-way impact, determine project economic feasibility, obtain funding authorization, develop right-of-way, obtain environmental clearance, determine design criteria and parameters, survey utility locations and drainage, make preliminary plans such as alternative selections, assign geometry, and create bridge layouts
Final Design	Acquire right-of-way; develop plans, specifications, and estimates (PS&E); and finalize pavement and bridge design, traffic control plans, utility drawings, hydraulics studies/drainage design, and cost estimates
Advertise and Bid	Prepare contract documents, advertise for bid, hold a pre-bid conference, and receive and analyze bids
Construction	Determine the lowest responsive bidder; initiate contract; mobilize; conduct inspection and materials testing; administer contract; control traffic; and construct bridge, pavement, and drainage

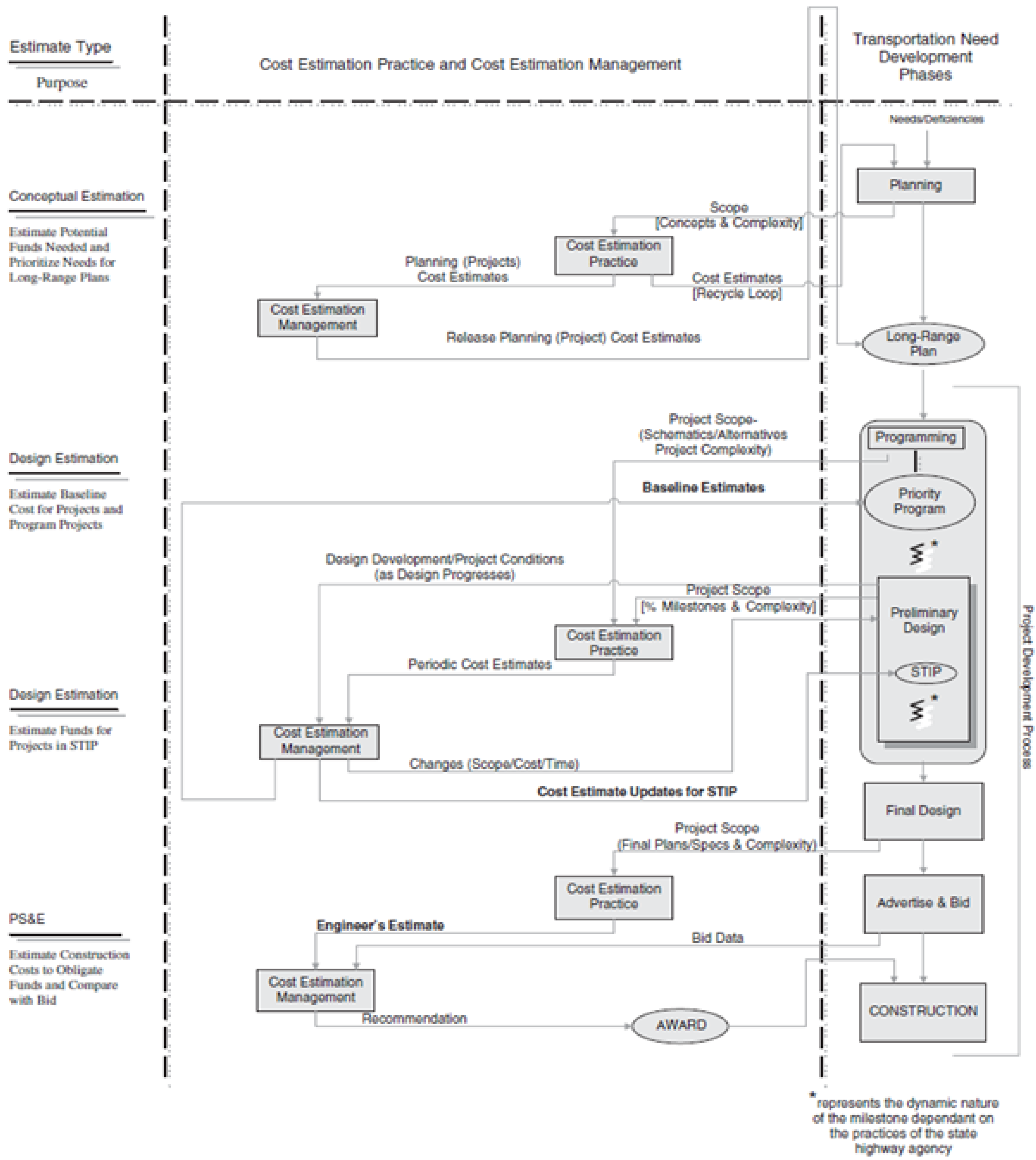
Some studies combine the planning (study and report phase), programming and preliminary design, and the final design phases in cases such as small projects. It should be noted that there is an overlap of these phases due to the cyclic nature of tasks as there are changes to the scope, change orders, and/or errors starting from the conceptual estimate until the completion of the detailed estimate (Anderson et al., 2007).

2.2 NCHRP Guideline for Cost Estimate

The guidebook presents procedures and approaches in preparing cost estimates and managing project cost for state highway agencies to be implemented at organization, program, and project levels (Anderson et al., 2007). "Cost estimates are prepared to support funding decisions as planning documents, program documents, and specific projects are developed. Cost estimation management is performed to support the work of preparing estimates and to ensure that program funding levels are in line with planned funding levels and project budgets. When cost estimation practice and cost estimation management processes are integrated, the transportation agency should have the capability to effectively manage its overall capital program as well as individual project budgets" (Anderson et al., 2007).

The Federal law requires highway agencies to develop a statewide transportation plan (STP) and metropolitans to develop a regional transportation plan (RTP). The purpose of planning for both statewide areas and metropolitan areas is to identify the most relevant and cost effective projects. The long range plan typically includes planning of 25 years. The STP usually establishes strategic directions for state investment in its transportation system, while RTP identifies projects that are to be implemented over the next 25 years, defined in short-, medium-, and long-term implementation stages (Anderson et al., 2007). Figure 2-2 shows a flow chart of cost estimation practice and cost estimation management based on a highway agency's perspective. The chart illustrates the relationship of cost estimation practice and cost estimation management. In addition, it describes the typical estimate types and key purposes of the cost estimates as related to each development phase.

Based on NCHRP guideline, the transportation need development phases show the planning and project development process associated with the type of estimates performed at various stages. Primarily, a preliminary scope of work based on concepts and complexity is defined to release a planning cost estimate. During the planning phase, cost estimation practice and cost estimation management are conducted iteratively to estimate potential funds needed and to prioritize needs for long-range plans (conceptual estimate). This estimation is usually based on a cost per mile tool. The project development phase starts with programming of specific projects that are developed for short-range plans of bidding highway construction projects. The project scope is better defined in terms of schematics and design alternatives to compare cost estimates. During this stage, projects with high priority are selected for preliminary design. Then a design estimation of the baseline cost for program projects and estimate updates of funds for the statewide transportation improvement plan (STIP) is conducted for the selected projects. Once the scope is well defined, with final plans and specifications developed, the detailed estimate is prepared and ready for advertising and bidding the project. During the bidding phase, the contractor's estimate and the engineer's estimate are compared or negotiated depending on the type of delivery method (PS&E). Finally, based on the cost estimation management process using engineer's estimate and bid data, the project is awarded to contractor for construction.



PS&E = plans, specifications, and estimates / STIP = statewide transportation improvement plan

Figure 2-2 Cost Estimation Management (Anderson et al., 2007)

2.3 Current Practice of Estimating Preliminary Engineering Costs

Previous studies reveal that various State Departments of Transportations (DOTs) and researchers utilize a wide variety of methods and approaches in estimating PE costs of highway projects. Chou et al. (2009) summarized approaches currently practiced by the industry and previous studies into eight classifications: a) parametric estimating, b) stochastic cost simulation, c) case based reasoning, d) artificial neural networks, e) activity based costing, f) feature based costing, g) lane mile estimating, and h) in-house spreadsheets. A summary of various studies conducted by different researchers and current methodologies used by DOTs is provided in the following sub-sections.

2.3.1 State DOTs' Practice

State DOTs usually follow the general PE estimating guidelines stated by NCHRP. Based on a report by Schexnayder et al. (2003), it was identified that more than thirty DOTs generate conceptual estimates solely based upon historic lane-mile cost averages for similar projects. However, different studies have indicated that the use of various techniques and approaches resulted in a variety of cost estimates. A summary of some DOTs' estimation techniques is shown in Table 2-2.

Table 2-2 Summary of Methodologies Utilized by DOTs

Agency	Kansas DOT	Nebraska DOR	Florida DOT	Washington DOT
Approach	% of Construction Cost (CC)	% of Construction Cost (CC)	In-house Spreadsheets % of Construction Cost	In-house Spreadsheets % of Construction Cost
% of In-house Design	< 30%	-	18%	-
Parameter	-Small Projects (< $1M) 15% of total CC -Medium Projects ($1M <X< $10M), 10-12% of total CC -Large Projects (>$10M), 8% of total CC	High Level Estimates - 0.5% of CC for Resurfacing (In-House) -4.4% of CC for New Construction (In-house) -8.0% of CC for New Construction (Consultant)	-	15% of total project cost Adjustments - project type, location & total dollar amount
		Mid-Level Estimates - 0.5% of CC for Resurfacing (In-House) -4.4% of CC for New Construction (In-House) -Consultant Contract Amount Plus 5-7%	-	Consulting firm - Use a factor of 1.8 - 2.8 based on type, number and scope of consultants involvement

Based on the study, most DOTs implement a percentage of construction costs as the primary cost estimating method; although, some have developed in-house spreadsheets for estimating their PE costs. When estimators use this method, they try to incorporate factors such as location and type of project and make adjustments to the total dollar amount. However, this sliding percentage of construction cost might not offer a reliable estimate of PE cost as it is difficult to quantify the project attributes and potential factors affecting PE cost and is solely based on the engineers' experience and judgment. In addition, most DOTs outsource their design tasks to consulting firms due to low resources and time constraints. A study conducted by Lafer (2009) on the Oregon DOT program indicated that shrinkage budgets, staff cuts, and a trend towards privatization are the reasons behind contract-out design works. Therefore, a system or tool which allows DOTs to reliably predict PE costs either for in-house design or for negotiating costs with consulting firms should be developed.

2.3.2 Studies by Researchers

Prior academic studies have shown the use of detailed unit cost and parametric approaches in estimating preconstruction costs. Trost and Oberlender (2003) established a parametric model to predict the accuracy of early estimates based on the estimate score for capital projects in the process industry. The required contingency is decided by a rating score for the classified factors based on the experienced estimators. A highway construction cost index was implemented by Wilmot and Cheng (2003) to develop a cost model to estimate future overall highway construction costs in Louisiana. The authors divided the overall model into five sub-models, each of which included one dominant construction item as a predictor. Based on the statistical analysis of more than 2,500 highway and bridge contracts over three years, they determined an average growth rate of 3.3% per year in construction costs in Louisiana.

Molennar (2005) showed that preliminary cost estimates should be represented by a range of estimates by using the Monte Carlo Simulation in developing a probabilistic cost estimation system under various scenarios. The study presented a Cost Estimating Validation Process (CEVP) based on nine case studies. The process was intended to better understand the risks associated with mega highway projects for a more transparent assessment of uncertainty. A statistical approach was utilized by Saito et al. (1991) in developing cost estimation models for bridge replacements using six years of 280 historical project data obtained from Indiana DOT

(INDOT). The study concluded that adding component cost models is better than a total bridge cost model. The model addressed eight explanatory variables including region, bridge type, deck area, substructure area, age, functional class, component condition index, and completed work to explain the response variables, component cost, subtotal costs, and unit cost, respectively. Nassar et al. (2005) applied regression model to estimate design costs of consulting firms based on 59 highway projects obtained from Illinois Department of Transportation (IDOT). However, the study classified project into low, medium, and high complexity, which is difficult to evaluate the level of projects quantitatively.

2.4 Factors Affecting Preliminary Engineering Costs

Actual construction project costs usually exceed the preliminary cost estimate and the amount budgeted for construction. A wide variety of reasons is given for the poor estimate of preliminary cost of highway projects. The NCHRP Guide (Anderson et al., 2007) for cost estimation and management for highway projects during planning, programming, and preconstruction divides factors into internal and external factors. Some of these factors include delivery/procurement approach, project schedule changes, engineering and construction complexities, scope changes, scope creep, poor estimation, inconsistent application of contingencies, and faulty execution. Based on meetings, interviews, and a review of the literature, the study classifies the factors into three major categories:

a) Organizational-level factors,
b) Team-level factors, and
c) Project-level factors.

A. Organizational-Level Factors – include factors such as organizational structure, scope definition (project information), change orders, market conditions (inflation), and standard estimation procedures implemented (Oberlender & Trost 2001, Anderson et al. 2007, and Chou et al. 2009). The organizational structure of highway agencies and/or consulting firms have a big impact on the estimation of PE costs. It affects PE costs in terms of information flow, composition of departments, sets of rules and guidelines, and transparency which shows the strengths of an organization.

   i. Organization Structure –A well-developed organizational structure leads to an effective communication and smooth flow of information that results in minimizing errors, project delays, and lower cost. Since the planning and design of preconstruction projects is the mixture of different divisions interrelated throughout the project phase, a well-planned organization structure along with project development is essential for a reliable estimate of the preconstruction estimate.

ii. Scope Definition – Scope definition during the project phases is another major factor that affects the preliminary engineering (PE) cost. The limited information and project description in the initial stage of a project design leads to a change in the project scope throughout the design phase of a project. "Poor scope definition at the estimation stage and loss of control of project scope results in cost overruns" (Dysert, 1997). For instance, the widening of a shoulder width, or an increase of the number of lanes results in additional work efforts and increase of PE cost estimate during the planning and designing of the project (Chang 2002, and Oberlender & Trost 2001). Therefore as the scope definition increases, the accuracy and reliability of a PE cost estimate also increases.

iii. Design Fees/Quality – Highway agencies nowadays outsource their design work to consultants due to a limited number of in-house designers and engineers. Some highway agencies utilize ASCE's fee curves to estimate design fees based on the cost of construction. Although these curves can be used as a baseline estimate for budget allocation, a design fee based on factors such as scope, project type, etc. should be maintained to obtain quality work and a reasonable cost estimate (Carr and Beyor 2005). A study by Washington Department of Transportation, (WSDOT, 2002) compared a sample bridge project cost estimate between 25 states as part of AASHTO's Subcommittee members which resulted in PE cost that ranges from 4% to 20%. A study by Gransberg et al. (2007) on projects from the Oklahoma Turnpike Authority showed that the percentage of construction cost growth from the engineer's early estimates increases as design fees decrease. In some cases, limited funds available from the government may influence the quality of work. Therefore, there should be a balance in the design fees between highway agencies and consultants.

B. Team-Level Factors– includes factors such as time constraints, design errors, inadequate cost data, and inexperienced team members which associate with the project team.

    i. Time Constraints – Time is another major factor that influences the accuracy of preliminary engineering cost estimates. Highway agencies might not have the sufficient time to produce a feasible PE cost estimate as to obtain project approval and funding from the government as a result of political pressures (Molenaar 2005). On the other hand, consulting firms might take risks to acquire projects that result in cost overruns and delay of project planning and design.
    ii. Inexperienced Team – The availability of skilled personnel (in-house design team or consulting firms) has a significant effect in estimating reliable PE costs. It is obvious that an engineer with 15 years of cost estimation experience might give a better and closer estimate compared to one with 5 years of estimation experience. In addition, the composition of the design and estimation team, type of considerations (factors) taken into account, the collection of up-to-date cost data, and amount of errors (accuracy) encountered leads to a huge gap in cost estimations performed by various personnel.

C. Project-Level Factors– project complexity, project characteristics, or project type can be used interchangeably to explain the nature and difficulty status of a project. These project characteristics or project complexities can be further explained by various attributes including geographical location, size of project, annual average daily traffic (AADT), number of lanes, project length, soil type, etc.

    i. Project Type (Project Characteristic) – The planning and design of construction greatly depends on the type or size of a project. Generally highway agencies utilize project types as a major factor in estimating preconstruction (preliminary engineering) costs of highway projects. Project types usually define the scope of work, describe the characteristics of the project and explain the complexity of the project. The characteristics and complexity of a project type may lead to design errors and changes that result in time delays and fluctuation of preliminary engineering costs (Persad et al. 1995, and Anderson et al., 2007).
    ii. Region/Location – The geographical location or region of a project is an important factor that planners and engineers consider in the estimation of

preconstruction costs. The region of a project identifies whether the project is in the proximity of an urban area or rural location. It determines the daily traffic, availability of materials and skilled labor, design of roadway and route selections, planning the right of way, lane closure tactics, and overall preconstruction estimation and management. Figure 2-3 illustrates the classification of these factors.

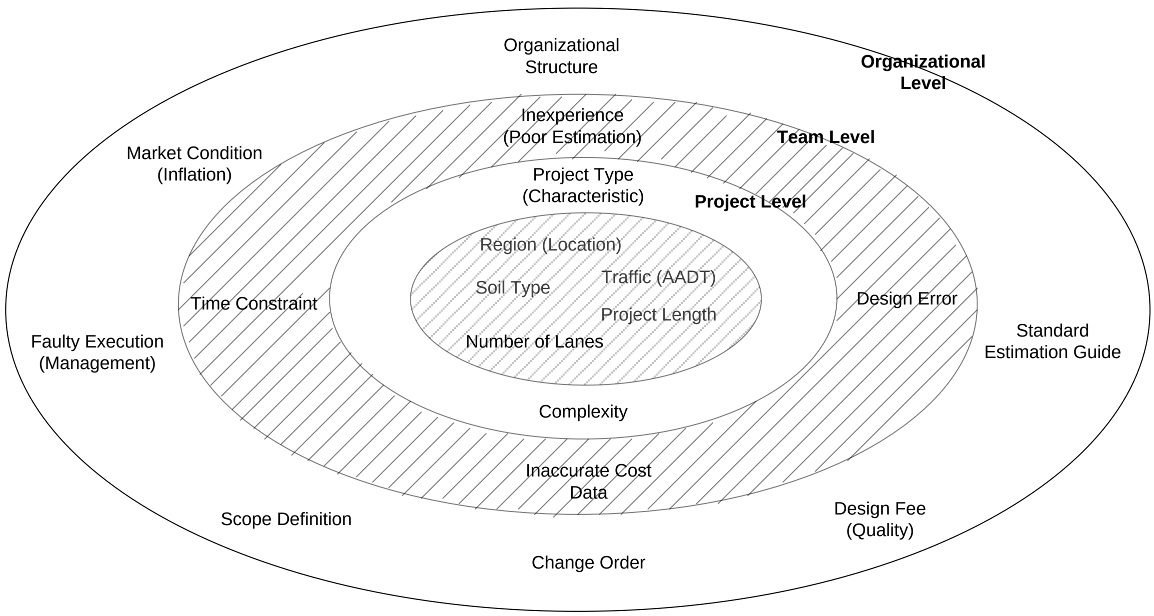


Figure 2-3 Potential Factors Affecting PE Cost

2.5 Costs Associated with Design Fees

The traditional method of charging for engineering services is the use of the percentage of construction cost. Although this method is still being used by some highway agencies in estimating design fees during the early project planning, the use of percentage of construction cost has gradually declined over the years due to the variations in the scope of work, the complexity of assignments, and review submittals from project to project. Based on the ASCE Manual (1972 and 1981), the five methods usually used in charging engineering and consulting services include: i) multiplier, ii) hourly, iii) per diem, iv) cost plus fixed fee (CPFF), and v) lump sum or fixed price.

The first four methods are referred to as variable methods as they are based on the consultant's cost to perform services. Based on ASCE's survey conducted in 2000, Tables 2-3 to 2-5 show the percentage use of compensation methods implemented by consulting firms from the previous year. Based on the study, the usage of the "lump sum" and "hourly" methods of compensation are the most widely used in the various ranges of construction firm sizes. Costs associated with consultant's fee for engineering services fall into five major categories: a) salary cost, b) payroll burden, c) other direct costs, d) general overhead and e) profit as a percentage of the aforementioned methods.

Table 2-3 Compensation Method for Small Consulting Firms (ASCE, 2003)

Method	Small (1-10)
Multiplier with not-to-exceed amount	8.90%
Multiplier without not-to-exceed amount	4.50%
CPFF with not-to-exceed amount	6.60%
CPFF without not-to-exceed amount	1.70%
Lump sum	32.70%
Hourly	37.60%
Other	8.00%

Table 2-4 Compensation Method for Medium Consulting Firms (ASCE, 2003)

Method	Medium (11-100)
Multiplier with not-to-exceed amount	14.80%
Multiplier without not-to-exceed amount	5.20%
CPFF with not-to-exceed amount	11.10%
CPFF without not-to-exceed amount	1.60%
Lump sum	36.80%
Hourly	24.60%
Other	5.90%

Table 2-5 Compensation Methods for Large Consulting Firms (ASCE, 2003)

Method	Large (>100)
Multiplier with not-to-exceed amount	26.70%
Multiplier without not-to-exceed amount	6.80%
CPFF with not-to-exceed amount	15.90%
CPFF without not-to-exceed amount	2.00%
Lump sum	27.90%
Hourly	12.70%
Other	8.00%

A. Multiplier - Salary cost times multiplier plus direct non-salary expense

This method of compensation for engineering services is used based on direct salary times an agreed multiplier which includes the above five costs except other direct costs. Other direct costs are reimbursed based on actual invoice cost plus an administration charge associated with it. The client and the consultant agree on salary ranges on each service and salary escalation clauses are written for any changes. Then the multiplier is chosen based on the type of service, nature, size and experience of the consulting firm, and geographic location. Based on ASCE's survey, an average multiplier for the different consulting firms, percentage markups for sub-contractors and reimbursable items and fixed fees based on the consultant's staff service during construction is shown in Table 2-6 and 2-7 respectively.

Table 2-6 Multiplier and Mark-ups (ASCE, 2003)

Size of firm	Multiplier Office	Multiplier Field	Mark-ups (avg.) %	
			Subs	Reimbursable
Small	2.85	2.58	9.85	9.86
Medium	2.87	2.77	7.78	8.19
Large	3.21	2.65	7.25	8.00
All	2.91	2.61	8.56	8.86

Table 2-7 Multiplier and Fixed Fees (ASCE, 2003)

Size of firm	Multiplier Office	Multiplier Field	CPFF Fixed Fee %	
			Low	High
Small	2.85	2.58	7.72	13.06
Medium	2.87	2.77	8.62	13.43
Large	3.21	2.65	8.69	15.51
All	2.91	2.61	8.30	13.66

B. Hourly Billing Rate

This method is similar to the above method in terms of the costs associated in estimating the hourly rate. It is preferred by consultants where the projects scope is not well defined or to simplify accounting and record keeping.

C. Per Diem

This method is suited for short term engagements, such as expert witness or similar services. It is based on an 8-hr standard day. The consultant is paid for all out-of pocket expenses including travel and standby time. Per-diem rates vary depending on the complexity, risk, expertise, employee classification, location, and period of service.

D. Cost Plus Fixed Fee

This method is usually applicable when the scope is well defined. In this method the consultant is reimbursed for all costs incurred and fixed fee (profit) that is agreed upon by the client and the consultant based on the scope of services to be performed.

E. Lump Sum

This method is applicable for basic services on design-type projects where the scope and complexity of the assignment are clearly and fully defined. A summary of consideration for choosing a method of compensation is shown in Table 2-8.

Table 2-8 Methods for Compensation (ASCE, 2003)

	Reimbursable	Hourly	Per diem	CPFF	lump sum
Scope not well defined	*	*	*	*	
Scope well defined	*	*	*	*	*
Simplified accounting					*
Very short duration assignment		*	*		*
Very complex job	*			*	
On-site construction management services		*	*		

3. Data Collection

The research team collected PE cost data by utilizing two methods a) meetings and interviews and b) historical project data (from engineering contract documents). Based on the data collected, ODOT's current PE cost estimating practice is identified. In addition, PE cost related data such as the work effort (engineering hours) and number of sheets required for preparation of major plan development phases (tasks) along with preliminary engineering cost has been collected.

3.1 Meetings and Interviews

The ODOT is comprised of various highway divisions which are engaged in planning, executing, and managing highway projects from the early phase of conceptual estimation to construction document development and bid preparation. The research team has conducted recurrent meetings and interviews with Mr. Raza Amini and Mr. Kirk Goins (Roadway Division), Mr. Jack Schmidel (Bridge Division), Mr. Ray Sanders (Project Management Division), Mr. Randy Jones (Capital Programs), and Mr. Larry D. Reser (Survey Division). In addition, regular e-mail and phone contacts were made with Mr. Kurt Harms (Right of Way), Mr. Mark Scott (Local Government), Ms. Siv Sundaram (Environmental Division), and Mr. Rob C. Williams (Planning & Research Division of Road Inventory).

3.1.1 ODOT Current Estimating Guideline

The meetings and interviews were primarily aimed at identifying the current practice of ODOT's preliminary engineering cost estimation practice, cost estimating management, and contract negotiation process during the planning and design phase of highway projects. Based on the meetings and interviews, the research team identified guidelines developed by the Project Management Division (PM) for the project development and implementation process (Figure 3-1).

a. First, the Planning Division initiates a project by defining a project scope and conducting a conceptual estimate. The PM Division will then perform the project charter process to obtain approval. Once the project is approved, surveying and environmental studies are conducted by the Survey Division and the Planning Division respectively. Then, the

Roadway Division performs a preliminary roadway plan and passes it to the Bridge Division for hydraulic analysis and setting bridge grade requirements. The Bridge Division then sends it back to the Roadway Division for setting the finished grade.

b. Following the preparation of the preliminary bridge and roadway plans, the PM Division facilitates preliminary field plan review meetings with the preconstruction divisions. The PM Division compiles the logistic information and sends it to the Right-of-Way (ROW) and Utility division. The PM Division gives a four week notification prior to the preliminary field meeting. Then, the ROW and Utility Division will compile and send the data to the Roadway Division and other divisions two weeks before the meeting. The preliminary field review meeting is intended to check environmental concerns, ROW needs, check alignments, verify project scope, etc. In the preliminary meeting, a total of 14 sets of plans are expected from the Roadway Division, Bridge Division, Panning Division, Traffic Division, ROW division, Field Division, and PM Division. A preliminary estimate of earthwork and survey data sheets should also be presented in the meeting. If there is a need for new ROW, a separate meeting will also be conducted with the cost estimate. During this process, while the Material Division performs geotechnical study, the Roadway Division designs the pavement.

c. After the preliminary meeting, a draft agenda will be prepared by the PM Division for review and comments. It will then distribute a final agenda for a final plan field review meeting two weeks prior to the meeting. The meeting is intended to verify plan changes from previous meetings, discuss constructability issues, erosion control, construction sequence, etc. The final plan review meeting will be utilized by ODOT staff and consultants when all information is acquired.

d. The preliminary engineering cost is estimated based on the amount of work effort put or engineering hours spent to conduct these tasks. Each division of ODOT performs its own study to meet project requirements and prepares an Excel sheet of contract fee proposal based on total hours required to develop the set of plans from preliminary stage to the final plan preparation.

Based on the meetings, it was explained that with the exception of the Survey and Geotechnical Divisions, the design services for highway contracts will compensate on an hourly basis. In other words, the consultant is reimbursed for the actual hours worked. It was also noted that the roadway design fees are not negotiated as rigorously as lump sum type contract for this reason. With regard to in-house design, a project template developed by the Project Management Division is utilized by following the above project development guideline.

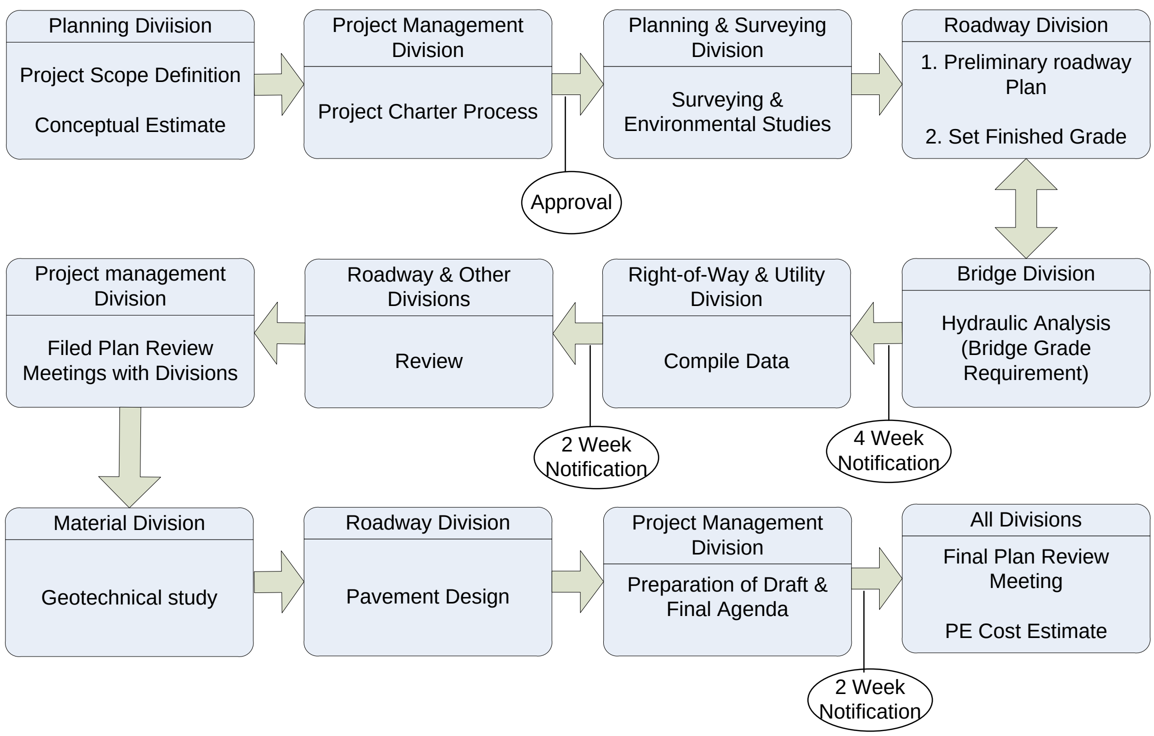


Figure 3-1 ODOT Estimating Guideline

Furthermore, the research team has learned that ODOT follows the Uniform Audit & Accounting Guide developed by the American Association of State Highway and Transportation Officials (AASHTO) for procedures in examining, auditing, and reporting costs that are incurred by Architect/Engineering (A/E) firms for engineering and design services to comply with Federal Regulations. The guide includes cost principles, labor charging systems, techniques in auditing costs associated with general overhead, direct labor cost, and additional costs (payroll additive, indirect costs, direct-non payroll costs, and profit) that need to be included in the total estimated fee for negotiation purpose. ODOT uses an indirect salary cost (payroll additive) factor which varies from 1.35 to 1.375 of the direct salary cost. A factor of 1.5 to 2.0 as a general overhead cost and an approximately 10% project fee (profit) is used as additional cost when negotiating PE cost with consulting firms.

In addition to meetings, e-mail and phone contacts were conducted to extract previous highway project data to develop a database which is comprised of project characteristics along with PE cost. The research team learned that not all divisions outsource their design tasks to consulting firms even for the same project. For instance, the Environmental Division performs approximately 80% of their environmental studies by an in-house design team based on the type of project and type of environmental clearance. Thus, it might be difficult and time consuming to track highway project costs and collect data from each division. Therefore, this study is solely based on highway project data collected from the Roadway Division.

3.1.2 Roadway Contract Fee Proposal

The Roadway Division has an in-house engineering contract cost proposal Excel spreadsheet developed by ODOT engineers based on the amount of work effort (total engineeering hours) required to develop the set of plans from the preliminary stage to the final plan preparation for the purpose of negotiating contracts with consulting firms. The spreadsheet consists of a cross tab of seven main plan development activites, a detailed list of tasks and sub-tasks along with a skilled labor category (Table 3-1). ODOT engineers use this spreadsheet to estimate and match the work efforts required by each engineer for each task based on the amount of sheets required for each task and project length by comparing it with similar, previous highway projects. A sample screenshot of roadway PE cost proposal Excel spreadsheet is shown in Figure 3-2.

Table 3-1 List of Plan Development Activities & Tasks

Plan Development Activities	Typical Tasks
Preliminary Roadway Plans	Develop plan & profile, preliminary finished grade line; design preliminary drainage structure, super-elevation; preliminary construction sequence; create title sheet
Right-of-Way Requirements	Finalize horizontal & vertical alignments; develop right-of-way submission plans; prepare comparative estimates for pavement design
Final Roadway Cross Sections	Develop cross-sections; perform earth work calculations
Final Traffic Plans	Develop final construction traffic control plans, signing & striping plans, signal plans, lighting plans, and summarize traffic plans

Plan Development Activities	Typical Tasks
Final Roadway Plan	Develop final sequence of construction sheet; calculate quantities; generate removal sheets & details, site specific erosion control plans, storm water pollution prevention plan
Estimation & Reporting	Prepare preliminary and final plan field review report & cost estimate, special provisions, and drainage studies & report
Meetings	Confirmation of scope and fee proposal, right-of-way & utility review, final plan filed review & pre-bid

For instance, the tasks associated with preparing a preliminary roadway plan includes creating title sheet and location map, drafting a typical section, developing plan and profile sheets, designing drainage structures, developing finished grade line, designing super-elevation, and developing the preliminary construction sequence. These tasks are further broken down into sub-tasks to estimate the amount of engineering hours required by the skilled laborer (project manager, project engineer, senior engineer, design technician, Computer Aided Design [CAD] technician, and clerk). The sub-tasks in developing the plan and profile sheets include preparing survey files, generating horizontal alignment, generating existing ground and profile, and generating and drafting plan and profile (P&P) sheets. Once the engineering hours are calculated, they are multiplied by the respective labor rate to obtain PE cost. The spreadsheet is well established; however, ODOT does not have any specific tools to estimate the engineering hours and costs required to perform the tasks and would put a lot of burden on the engineer as it highly depends on his/her experience, skills, and judgments.

OKLAHOMA DEPARTMENT OF TRANSPORTATION
ROADWAY ENGINEERING CONTRACT COST PROPOSAL

8/28/2006

E.C. NO. ______ SWO NO. ______ JOB PIECE NO. ______
COUNTY ______ CITY ______ PROJECT NO. ______
HIGHWAY NO. ______ ORGANIZATION SUBMITTING PROPOSAL ______
PROJECT DESCRIPTION :

		Labor Catagory									
11–12	Labor Description	Project Manager	Senior Engineer	Project Engineer	Design Technician	CAD Technician	Clerical				
13	Labor Rate										
15	ROADWAY PLAN DEVELOPMENT	HOURS						Sub-Task	Task	Activity	
16	3.2 Prepare Preliminary Roadway Plans	104	406	682	680	1002	3			2877	$ 77,282
17	3.2.2 Create Title Sheet And Generate Location Map	2	8	16		38			64		
18	3.2.5 Draft Typical Section	4	22	60		108	2		196		
19	3.2.6 Develop Plan & Profile Sheets	92	244	386	284	520	1		1527		
20	3.2.6.2 Prepare Survey Files For Design	2	2	2	4	4	1	15			
21	3.2.6.3 Generate Horizontal Alignment	42	80	148	24	32		326			
22	3.2.6.5 Generate Existing Ground			4	16			20			
23	3.2.6.6 Generate Profile	44	140	184	48			416			
24	3.2.6.7 Generate P&P Sheets	2	4	20	64	164		254			
25	3.2.6.8 Draft P&P Sheets	2	18	28	128	320		496			
26	3.2.7 Design Drainage Structures	0	64	96	56	80	0		296		
27	3.2.7.1 Analyze Existing Drainage System		32	16				48			
28	3.2.7.2 Determine Size of Cross-Drain Structures		16	40	16			72			
29	3.2.7.3 Generate Drainage Map		16	40	40	80		176			
30	3.2.8 Develop Finished Grade Line	2	0	120	308	144	0		574		
31	3.2.8.1 Design Vertical Alignment				88	32		120			
32	3.2.8.2 Develop Templates			56	64			120			
33	3.2.8.3 Develop Preliminary Cross Sections	2		64	116	88		270			
34	3.2.8.5 Develop Preliminary End Areas And Volumes				40	24		64			
35	3.5.3 Design Superelevation	2	28	4	32	32			98		
36	3.7.1 Develop Preliminary Construction Sequence	2	40			80			122		
38	## Establish Right-of-Way Requirements	62	426	755	142	775	2			2162	$ 59,390
39	3.15.1 Finalize Horizonal and Vertical Alignments	12	72	72	72	72			300		
40	3.15.6 Develop R/W Submission Plans	44	324	653	40	697	2		1760		
41	3.15.6.4 Design Driveways	20	120	128	16	152	2	438			
42	3.15.6.5 Design Side Drains	6	24	149	8	77		264			
43	3.15.6.6 Finalize Roadway Drainage Structure Design	16	172	360		452		1000			
44	3.15.6.7 Design Special Drainage Structures	2	8	16	16	16		58			
45	3.18.2 Prepare Comparative Estimates for Pavement Des	6	30	30	30	6		102	102		
47	## Prepare Roadway Final Cross Sections	6	4	112	560	328	0			1010	$ 24,818
48	3.15.4 Develop Cross Sections	4	0	96	344	272	0		716		
49	3.15.4.1 Generate Cross Sections			20	104			124			

Figure 3-2 Screenshot of Roadway PE Cost Proposal Spreadsheet

3.2 Data Collection

The main objective of the data collection stage is to investigate the existing Oklahoma roadway project database and collect PE cost data for roadway projects. ODOT consists of various highway divisions responsible for planning, design, and construction of roadways and bridges. Some of these divisions include Right-Of-Way, Geotechnical, Utilities, Surveying, Environmental, Roadway and Bridge divisions. For this study, the Roadway Division is selected for estimating the PE costs of roadways since it comprises the major cost of a highway project. The Roadway Division is responsible for conducting studies and developing roadway plans. For Oklahoma roadway projects, information on preliminary engineering cost data is partially stored electronically on the Roadway Division engineer's personal computer, but the majority of the project data are stored as hard copies (paper format) as part of the engineering contract data.

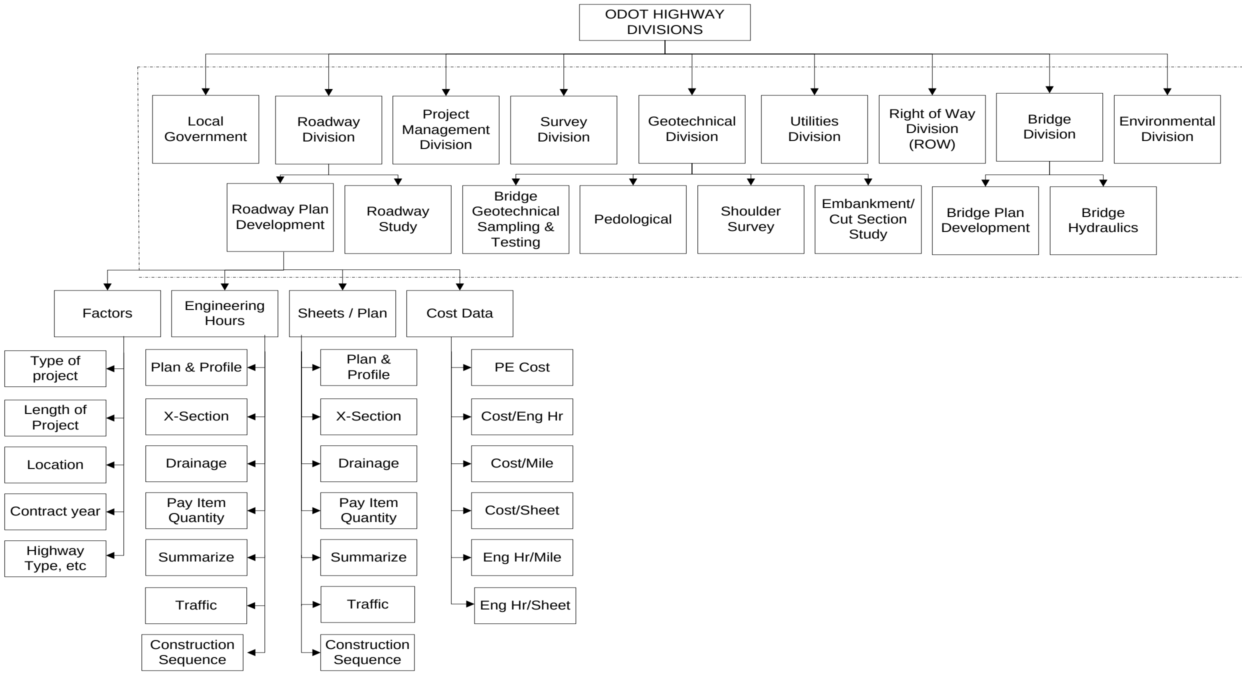


Figure 3-3 Data Collection Process

Based on the engineering contract documents, an Excel spreadsheet was prepared for data collection and cleanup. The data was then prepared for subsequent analysis and model development. The data collection spreadsheet consists of four basic components:

a) Project-level factors or attributes,
b) Engineering hours required to develop the number of sheets or plans,
c) Number of sheets (plan) required for design works, and
d) PE cost data negotiated with consultants.

These four component data are collected for 11 major plan development outputs (plan and profile, drainage, etc.) which will be discussed in the next chapter. For this study, the research team acquired a total of 353 contracted engineering projects along with 25 data attributes or potential factors (length of project, AADT, location, etc.) from ten years of highway project data (projects' contract let year that ranges between January 1, 2001 and July, 31 2011) from the Roadway Division and capital programs inventory. Figure 3-3 illustrates the overall data collection process. It should be noted that the data collection and preparation process was a very time consuming process as the PE cost data is stored manually (paper format), and for each project, data was collected by going through the engineering contract documents (project scope, fee estimate sheet, and the negotiated PE cost data) and inserting it into a data collection Excel sheet. A sample data collection sheet for 5 projects is shown in Tables 3-2 – 3-4.

Table 3-1 Project Attributes and Cost Data

EC No.	Year	Route Type	Consulting Firm	Project length	Number of Sheets	Engineering Hours	PE Cost	Cost/Hr.	Cost/Mile	Cost/Sheet	Hr. / Sheet
57	2001	I	CK	1.495	278	7413	482,000	65.02	96,400	1,734	27
6 Ph 1	-	I	BENHAM	0.56	363	15023	1,035,700	68.94	692,776	2,853	41
6 Ph 2	-	I	BENHAM	0.76	148	4681	358,000	76.48	639,286	2,419	32
6 Ph 3	-	I	BENHAM	0.69	106	3193	244,000	76.42	321,053	2,302	30
					166	4782	360,000	75.28	521,739	2,169	29

Table 3-2 Engineering Hours/Sheet

EC No.	P & P	X-Section	Summarize	Mass Diagrams	PQ	Drainage	Storm Control	Const. Sequence	Traffic	Details
56	3740	1064	360	64	100	762	231	180	258	318
57	3840	1250	1600	0	1120	80	0	2100	800	3040
6 Ph 1	1220	680	270	32	180	520	104	180	746	362
6 Ph 2	840	430	180	32	160	300	104	180	408	300
6 Ph 3	1380	740	270	32	180	440	120	180	716	502

Table 3-3 Number of Plans/Sheets

EC No.	P & P Sheets	X-Section	Summarize	Mass Diagram	PQ	Drainage	Const. Sequence	Storm Control	Traffic	Details
56	72	133	4	2	1	16	3	12	22	10
57	76	125	20	0	14	1	30	0	20	58
6 Ph 1	14	68	3	1	2	7	3	5	24	15
6 Ph 2	11	43	2	1	2	5	3	5	16	13
6 Ph 3	17	74	3	1	2	6	3	8	23	23

4. Data Analysis

This chapter discusses the analysis of data based on the collected PE cost data. A component-based PE cost prediction approach which integrates the work effort (engineering hours) and number of sheets for major plan development tasks is selected for easier and more reliable estimation of PE cost. A data cleanup and preparation is performed to remove outliers and missing data points to increase the accuracy of the prediction models. The research team also conducted a principal component factor analysis to measure the relationship between the potential factors and their strengths.

4.1 Component-Based PE Cost System

This report presents a data-driven and component-based PE cost estimation system for roadway projects by utilizing a knowledge discovery in database (data mining) techniques. These components are based on the major plan development task outputs that result from the preliminary engineering phase conducted by either the in-house design team or the consulting firm which often constitutes the principal, project manager, senior engineer, project engineer, design technician, and CAD technician. These outputs primarily incorporate a) P&P Sheets, b) Cross-section Sheets, c) Summarizing Sheets, d) Mass Diagram Sheets, e) Pay Item Quantities (PQ) Sheets, f) Drainage Sheets, g) Storm Control Sheets, h) Construction Sequence Sheets, i)

Traffic Sheets, j) Detail Sheet, and k) Typical Section, Title & Alignment Sheets. Table 4-1 explains the outputs associated with the plan development sub-tasks.

The component-based estimation system consists of three functions or entities: a) engineering hours required per number of sheets, b) number of sheets, and c) cost per engineering hours for the selected major plan development task outputs. Once these entities are estimated, multiplying them and summarizing the whole plan develeopment task outputs would result in a PE cost for a certain roadway project (Eqn. 4.1). These classifications will allow the engineer to estimate the values for the respective entities. It could help engineers determine a reliable number of sheets, work effort (engineering hours) required per number of sheets, cost per engineering hour, and total cost for the selected plan development output.

Table 4-1 Plan Development Outputs

Plan Development Sheets	Typical Sub-Tasks
Plan & Profile	Prepare survey files, generate horizontal & vertical alignment, design super-elevation & driveways
Cross Section	Develop preliminary & draft cross-sections, end areas and volumes, generate & drain cross-sections
Summarize	Calculate quantities & summary sheets
Mass Diagram	Generate mass diagram
Pay Item Quantities	Assemble pay-items and notes
Drainage	Analyze existing drainage, determine size of cross-drain structure, design & finalize drainage structure
Storm Control	Generate site specific erosion control & storm water pollution prevention plan
Construction Sequence	Develop preliminary and final sequence of construction sheets
Traffic	Develop final construction traffic control plans, signing & striping plans, signal plans, & lighting plans
Details	Design & generate joint layout, drainage structure details, miscellaneous and survey data sheets
Typical Section	Create title sheet, & generate location map, draft typical section

In addition, this system not only allows engineers to easily manipulate the requirements for a specific task, but also helps them configure contingencies, as to whether any of the entities are either under/over-estimated. This system also helps to identify if there exists a misallocation of resources (engineering hours assigned to the respective skilled manpower or number of sheets assigned to each task) at a specific level, especially when negotiating PE costs with consulting firms. Figure 4-1 illustrates the component-based PE cost estimation system.

$$PECost = \frac{Eng.Hour}{No.of\ Sheets} * \frac{PE\ Cost}{Eng.Hour} * No.ofSheets \ldots\ldots\ldots.. Eqn. 4.1$$

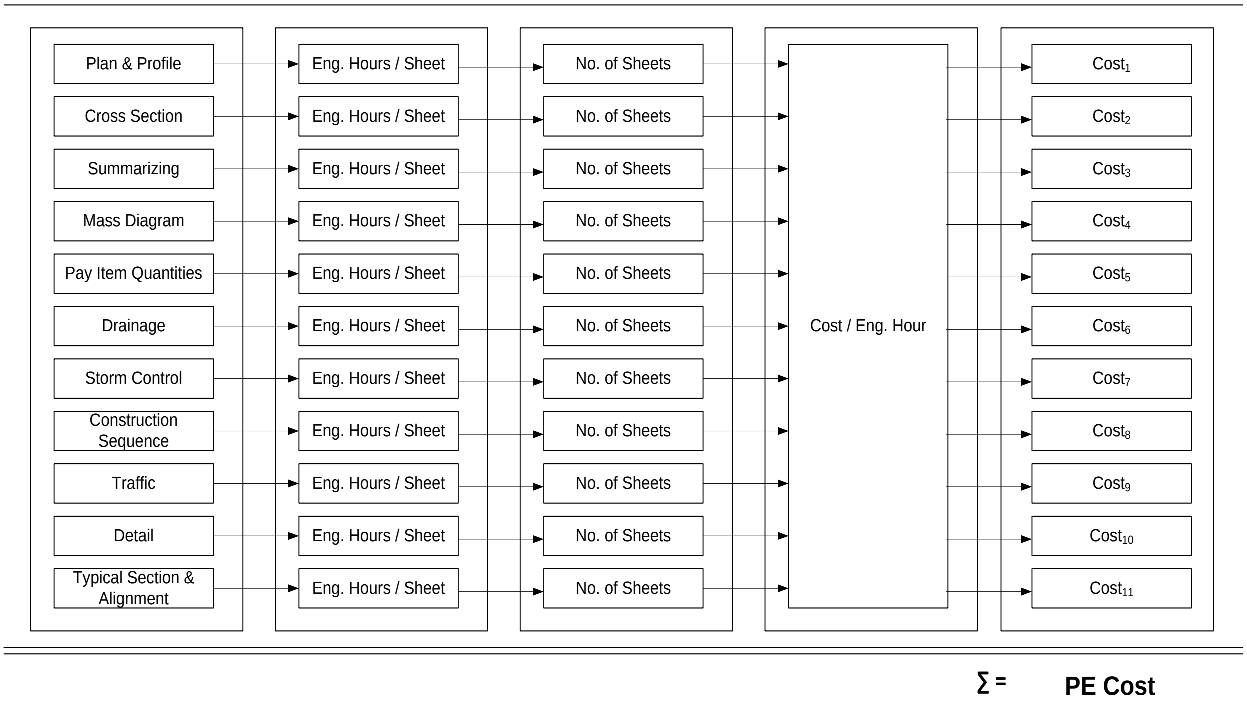


Figure 4-1 Component-Based PE Cost Estimation System

4.2 Data Preparation

The preparation of major plan development tasks such as the plan and profile, cross-section, drainage, construction sequence, summarizing, pay item quantities (PQ), and traffic works have sub-tasks used to estimate the engineering hours and number of sheets required for preliminary engineering cost estimation. For instance, a plan and profile development incorporates preparation of geometric data sheets, plan sheets, and profile sheets; while the drainage plan development includes the hydraulic drainage map, drainage area map, and storm sewer profiles. In order to approximate the contribution of these sub-tasks, a weighted mean is used to determine engineering hours that are required per sheet for the major plan development phases (Eqn. 4.2).

$$\overline{\chi} = \frac{\sum_{i=1}^{n} \omega_1 \chi_1}{\sum_{i=1}^{n} \omega_1} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots.. Eqn. 4.2$$

Where x_i represents the engineering hours per sheet, while w_i represents the number of sheets required. For example, developing plan and profile plans required 80, 60, and 40 engineering hours per sheet and 1, 4, and 2 numbers of sheets for preparing the geometric data sheets, plan sheets, and profile sheets respectively. Then, the weighted mean of engineering hours per sheet spent in developing the plan & profile plan is calculated as:

$$\overline{\chi} = \frac{80 * 1 + 60 * 4 + 40 * 2}{1 + 4 + 2}$$

For this report, the total aggregate cost or negotiated cost (with consulting firms) of the roadway plan development fee is utilized as the final PE cost. The PE cost incorporates the direct salary cost, payroll additive, an input percentage of direct salary cost (vacation, sick leaves, retirement, and FICA), direct non-payroll costs (materials and supplies, reproduction, data processing, travel expenses and equipment rental), indirect costs, an input percentage of all direct cots (administration, rent, utilities, and telephone) and 10% profit.

4.3 Data Classification

Based on meetings and interviews with ODOT engineers and a review of prior studies, 26 factors were identified as potential project-level factors affecting PE costs: a) project type, b) project length, c) design fees/quality, d) time constraints, e) region/location, f) organization structure (consulting firm), g) highway type, h) length of project, i) contract year, j) AADT, k) number of lanes, l) cross-section, m) pavement type, n) funding source, o) method of construction, p) shoulder type, q) lane width, r) soil type, s) type of permit, t) terrain type, u) area type, v) storm sewer, w) detour requirement, x) route type, y) NEPA document, and z) sidewalks.

Table 4-2 Classification of Project-level Factors

No	Categories	Factors
i.	Project Scope	Project type, project length, number of lanes, method of construction, contract type
ii.	Geographic Attributes	Area type, highway type, terrain type, location (division)
iii.	Design Attributes	AADT, pavement type, soil type, cross-section, shoulder type, lane width, storm sewer, sidewalk
iv.	Environmental Attributes	type of permit, NEPA document, detour requirement

No	Categories	Factors
v.	External Factors	type of consulting firm, time constraints, design fee (quality), Contract let year, funding source

Based on the definition, functionality, and availability of data, the research team classified project-level factors into five major categories: i) project scope, ii) geographic attributes, iii) design attributes, iv) environmental attributes, and v) external factors as shown in Table 4-2. Project scope is a function that characterizes a project and the work that needs to be accomplished by the highway agency. Geographical attributes refer to the location and place of the project with regard to the topographic and nature of the area. Design attributes imply that major engineering decisions are taken into consideration for designing a sustainable and economical highway project. Environmental attributes refer to the environmental aspects and permits required for the project. Additional factors, such as quality of consulting firm and time constraints, are categorized as external factors to consider their impact on PE costs. This chapter solely focuses on project-level factors even though there is an overlap of factors such as design fees and time constraints (external factors) which are mentioned as organizational-level factors in the literature review chapter. Table 4-2 shows the classification of project-level factors into 5 categories.

4.4 Data Analysis

The effects of various factors on PE cost are first analyzed using descriptive statistics. This method is an efficient technique to summarize, understand, and correlate project characteristics or potential factors with PE cost estimates. In addition, a factor analysis of principal components was conducted to explore the relationship and identify the correlation among these attributes.

4.4.1 Data Cleanup

During this process, a data partition is utilized to separate the data into a training-data set and a validation-data set. The training-data set is used as a preliminary model fitting to find the optimum model weights, while the validation-data set is used to assess the accuracy of the developed models. The first step in data partitioning is to specify the sampling method as simple random sampling, stratified random sampling, or cluster sampling. The second step is to assign the proportion of the sampled observation to each output data set. A stratified random sampling is used for categorical target variables, and simple random partitioning is utilized for interval variables to have the same probability of being assigned to one of the partitioned data sets. It should be noted that data partitioning can reduce the computation time of modeling runs; however, it may be inefficient for small data sets as a reduced sample size might affect the fit of the model. For this reason, 80% of the data (192) are allocated for training the data set to develop PE cost prediction models, while 20% (48) of the data is used to validate the models.

Table 4-3 Statistical Data for Categorical Variables

Variable Name (Factor)	Role	Number of Levels	Missing	% Missing	Mode 1	Mode 1 %	Mode 2	Mode 2 %
CONSULTING_FIRM	INPUT	50	0	0%	POE	9.07%	COBB	7.65%
ROUTE_TYPE	INPUT	4	3	1%	SH	39.66%	US	39.66%
LET_YEAR	INPUT	12	5	1%	2006	22.38%	2009	19.83%
DIVISION	INPUT	9	9	3%	4	20.96%	8	-
PROJECT_TYPE_1	INPUT	12	54	15%	BRIDGE & APPROACH ES	43.63%	-	15. 3%
FUND_TYPE	INPUT	9	55	16%	BRFY	22.1%	-	15.58%
NO_OF_LANES	INPUT	5	253	**72%**	-	71.67%	2	22.38%
LANE_WIDTH	INPUT	5	272	**77%**	-	77.05%	12’	20.4%
SHOULDER_WIDTH	INPUT	9	275	**78%**	-	77.9%	8’	7.93%
ALIGNMENT	INPUT	5	279	**79%**	-	79.04%	EXISTING	13.6%
SECTION	INPUT	11	282	**80%**	-	79.89%	4D	7.08%
SHOULDER_TYPE	INPUT	9	288	**82%**	-	81.59%	ASPHALT	5.1%
AREA_TYPE	INPUT	5	291	**82%**	-	82.44%	RURAL	13.6%
HIGHWAY_TYPE	INPUT	6	292	**83%**	-	82.72%	COLLECT	7.93%
PAVEMENT_TYPE	INPUT	6	292	**83%**	-	82.72%	ASPHALT	11.05%
TYPICAL_SECTION	INPUT	5	292	**83%**	-	82.72%	OPEN SECTION	15.86%
SIDEWALKS	INPUT	3	294	**83%**	-	83.29%	NO	16.15%
PERMIT_TYPE	INPUT	8	299	**85%**	-	84.7%	COE	5.38%
DETOUR	INPUT	8	301	**85%**	-	85.27%	NO	7.37%
TERRAIN_TYPE	INPUT	3	303	**86%**	-	85.84%	FLAT	7.08%
HIGHWAY_CLASSIF ICATION	INPUT	3	304	**86%**	-	86.12%	NON-NHS	11.61%

Table 4-4 Statistical Data for Interval Variables

Variable	Role	Mean	Standard Deviation	% Missing	Min	Med	Max	Skewness	Kurtosis

ADT	INPUT	3376.16	7955.97	**83.57%**	150	1750	60700	6.80	49.39
PROJ. LENGTH	INPUT	2.30	2.36	4.82%	0.07	1.39	13.5	1.55	2.37
PE_COST	TARGET	231160	255180.4	0.00%	25027	150500	227500	3.59	18.91

A 50% cutoff value is also assigned for all variables, which means that if more than half of the data points are missing, the variable is rejected from the analysis. However, a method for inputting data of median value for interval variables and mode value for categorical variables is utilized for the replacement of missing values with less than 50% of the missing data set which may have resulted from data collection errors, system failure, or incomplete response (Table 4-3). Based on the analysis, only seven variables or factors (consulting firm, length of project, contract let year, project type, route type, fund type, and location /division) have less than 50 % missing values. In addition, the remaining portions of the data are dominated by one type of condition or category (for instance highway classification has 86% missing data points and 11.61% contains a roadway project classified under a NON-NHS system.

There are several methods used for data imputations such as mean, median, tree surrogate, and midrange for interval variables and count (mode), distribution, and tree surrogate for categorical variables. This method is utilized to account for the missing important information while modeling or the observation that can be rejected which is contained in the non-missing variable. However it is noted that the inputting method has a central tendency to greatly affect the distribution of a variable; therefore, careful consideration is taken into account to minimize this effect.

In addition, some of the project attributes (variables) were transformed into various simple functions to attain a normal distribution and increase the fit of a model. Transforming variables enables the researcher to create new variables, transform the variables into categorical (class) variables, stabilize variances, remove non-linearity, and increase the normality of variables (Berry & Linoff, 1997). Transformations range from simple transformations to binning transformations to best power transformations for interval variables. Group-rare level transformations and dummy-indicator transformations are used for categorical variables. For this report, the research team decided to use simple transformation such as logarithmic, square root, and exponential functions based on the skewness and kurtosis values for interval variables.

Table 4-5 Imputed Categorical variables

Variable Name	Role	Number of Levels	Missing	Mode 1	Mode 1	Mode 2	Mode 2
CONSULT FIRM	INPUT	47	0	POE	9.93%	GARVER	7.45%
IMP DIVISION	INPUT	8	0	4	23.05%	8	18.44%
IMP FUND TYPE	INPUT	8	0	BRFY	38.65%	SSP	14.54%
IMP LET YEAR	INPUT	11	0	2006	23.76%	2009	20.92%
IMP PROJECT TYPE	INPUT	11	0	BRIDGES & APPROACHE S	59.22%	GRADE, DRAIN & SURFACE	11.70%
IMP ROUTE TYPE	INPUT	3	0	US	41.84%	SH	37.94%

Table 4-6 Imputed & Transformed Interval variables

Variable	ROLE	Mean	Deviation	Missing	Min	Med	Max	Skewness	Kurtosis
IMP_LOG_ PROJECT LENGTH	INPUT	0.97	0.59	0	0.07	0.92	2.48	0.48	-0.73
LOG_PE_ COST	TARG ET	11.98	0.84	0	10.13	11.92	14.64	0.42	-0.13

Based on the imputation, all missing data values were replaced with the most frequent occurrence (mode) for categorical variables and the middle value of the data sample (median) for interval variables. Logarithmic transformation of project length and PE cost has resulted in a lower skewness and kurtosis values (Table 4-5 and Table 4-6).

4.4.2 Descriptive Statistics

On average, 70% of roadway project PE costs fall in the range between $50,000 and $350,000, while more than 20% of the projects' costs range between $350,000 and $650,000; furthermore, less than 6% of the projects have a PE cost greater than $650,000. The mean PE cost of the collected data was $293,934 with a standard deviation of $252,844. This wide variation of PE cost might be a result of projects ranging from supplemental work of an existing contract to new design and complex rehabilitation highway projects. In addition more than 60% of the roadway project lengths fall between 1.4 miles and 7 miles. This might be due to limited funds available, method of construction as to not create an inconvenience to the public, a trend towards distributing/involving various consulting and construction firms, or the involvement of more rehabilitation projects. Figure 4-2 and 4-3 show the distribution plot (frequency) of project length across roadway projects and PE costs as a logarithmic function for the collected data respectively.

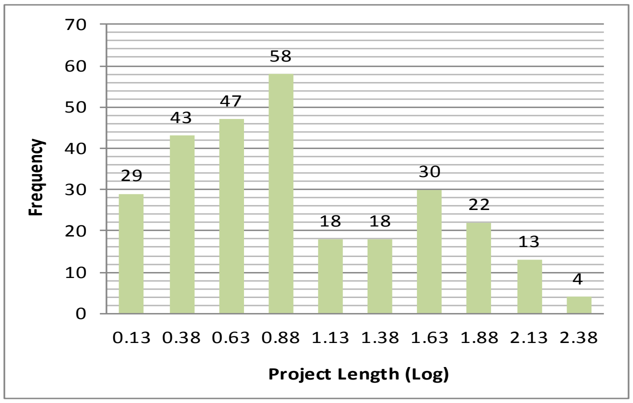


Figure 4-2 Distribution of Project Length

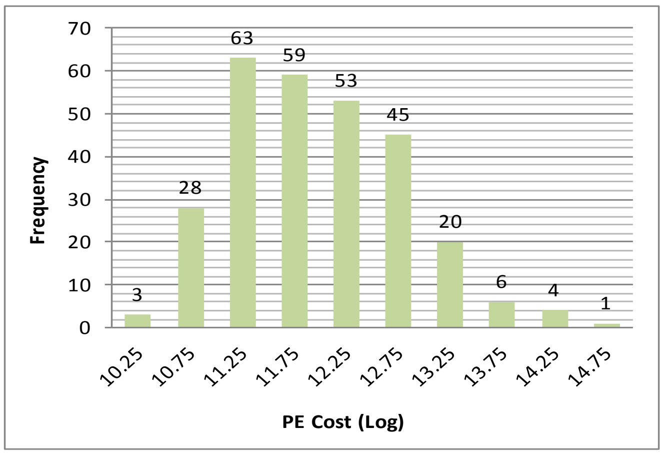


Figure 4-3 Distribution of PE Cost

ODOT classifies highway projects into three different categories as Tier I, II & III which are comprised of 17 divisions based on the scope and complexity of a project. However, the Roadway Division categorizes projects further based on the type of work performed by the division. Based on the frequency of data collected, almost 60% of projects fall under the construction of bridge and approaches. Projects in this category include replacing bridge structures, detours, roadway approaches, and traffic control. The grade, drain, surface, and bridge work incorporate 30% of the collected data (Figure 4-4). In addition, the research team considered contract let year as potential factor affecting PE cost to account for the value of the dollar or inflation. Based on the statistics, a higher number of projects were awarded to consulting firms in the years 2006 (67 projects) and 2009 (59 projects) with more than 70% of projects being contracted in the last five fiscal years. This might be due to a higher budget allocated for transportation agencies in an effort to enhance the economy (Figure 4-5).

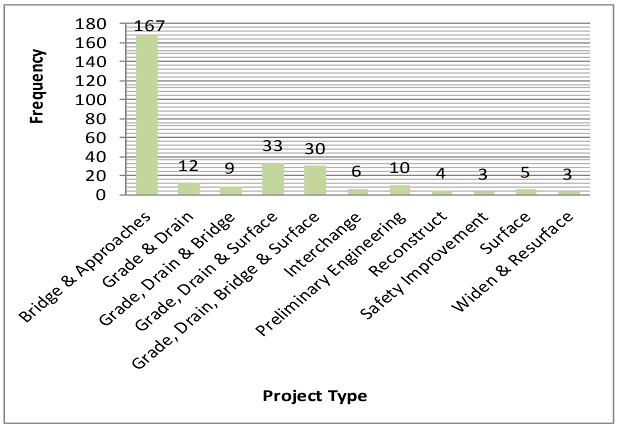


Figure 4-4 Distribution of Project Type

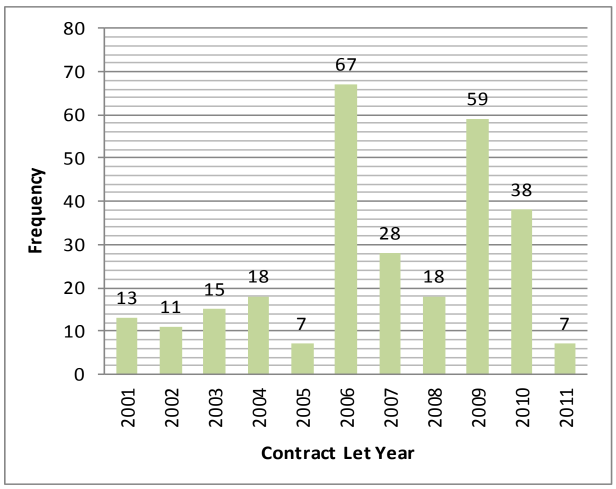


Figure 4-5 Distribution of Contract Let Year

ODOT also classifies the state into 8 divisions based on the geographical location. In order to develop a representative PE cost estimation model, projects performed in all divisions are incorporated in this study. Each division has more than 30 projects with a maximum of 65 (23%) projects in Division 4. The exceptions are Division 1 and Division 6 which have only 14 (5%) projects each. A higher number of US (118) and state highway (107) projects were contracted as compared to interstate highways (57). Figure 4-6 and 4-7 show representation of distribution plot of the data collected based on location and type of route respectively. In addition, consulting firms which have a good reputation in design work include Benahm, Cobb, EST, Garver, Macarthur, Poe, Triad, PEC, SRB, and Tetratech (each with more than 15 projects). It should be noted that all of these statistics are based only on the random selection of the training data (282 data points).

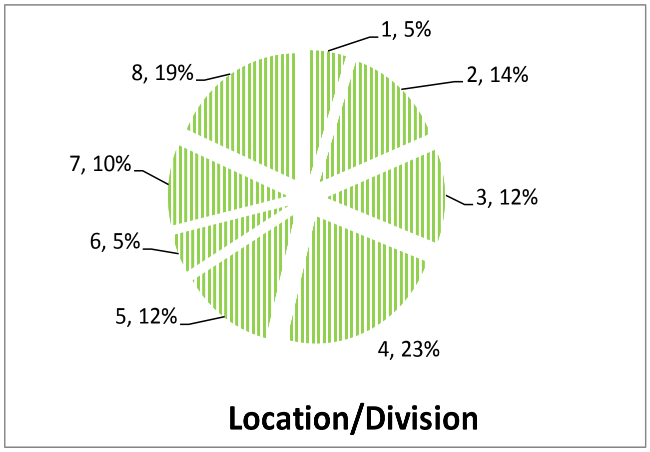


Figure 4-6 Distribution of Location

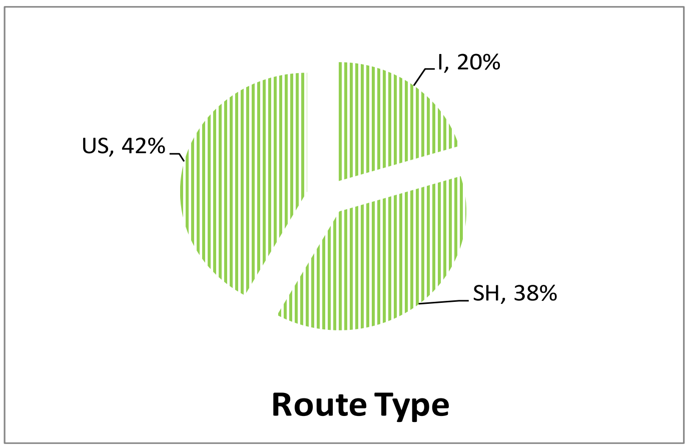


Figure 4-7 Distribution of Route Type

One method to examine the variable distribution and statistics of a data set is the variable-worth plot. The variable-worth plot ranks input factors (independent variables) according to their calculated worth. Based on the plot, the project factors affect PE cost in the order of project length (0.22401), project type (0.17124), fund type (0.13928), route type (0.12495), consulting firm (0.07643), contract let year (0.50124), and location /division (0.04123). Although this order holds true to most of the PE components, there might be a slight change of order depending on the type of plan development task. The variable worth of factors for the plan development task outputs are further discussed in Chapter 5: Model Development. Figure 4-5 shows the variable worth plot of PE cost.

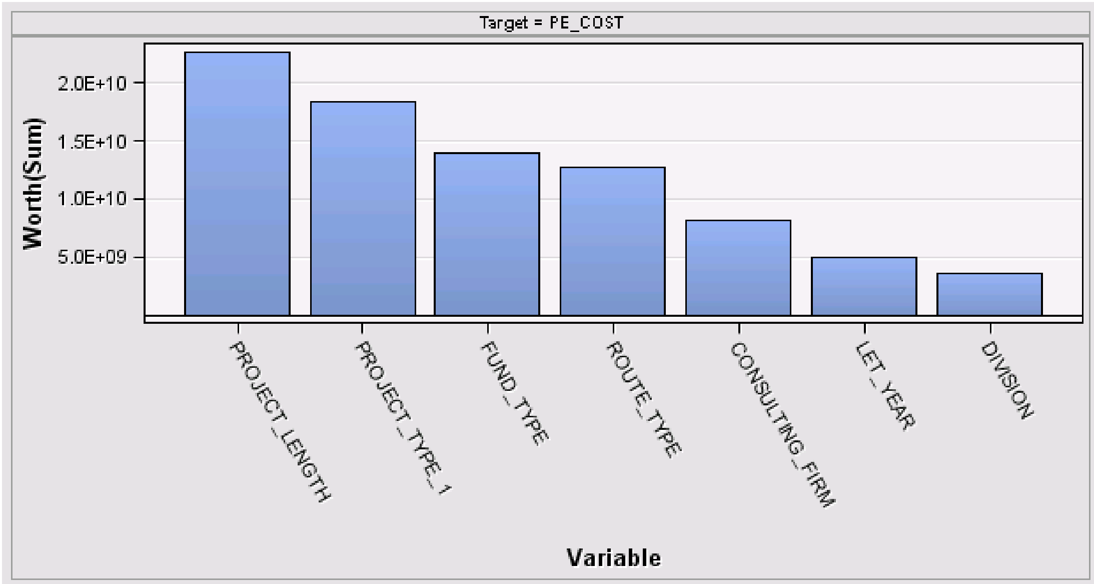
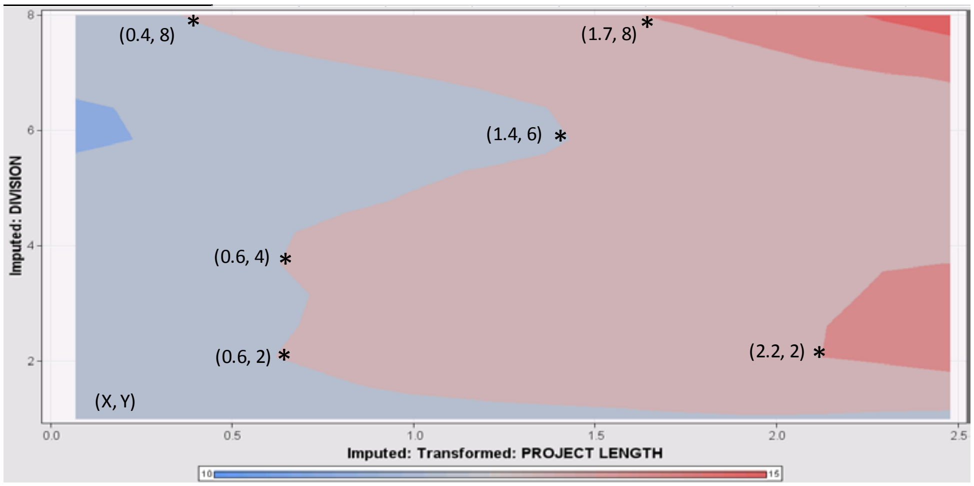


Figure 4-8 Variable Worth Plot of PE Cost

A bi-variate analysis of location and project length was also conducted to visually observe the effects of these factors on PE cost. Based on the result, higher PE costs are experienced in Divisions 2, 3, 4, and 8 as compared to Divisions 1, 5, 6, and 7. This is shown in a diagrammatic illustration of a contour analysis in Figure 4-6. For Divisions 2 and 3, as project length increases, PE costs also rise between the coordinates (0.6, 2) and (2.2, 0.6) and get higher as the project length passes 2.2 miles (2.2 and 2). For Division 4, PE costs get higher past the (0.6, 4) coordinate, while Division 8 has a higher PE cost between 0.4 miles and 1.7 miles and experiences the highest PE cost beyond 1.7 miles of a project length. On average, Divisions 2, 3, 4, and 8 have significantly higher PE costs with project lengths ranging from 0.55 miles to 2 miles and experience the highest PE cost between 2 miles and 2.5 miles. This may be because these regions are in close proximity to highly urbanized areas which makes the projects complex. A more detailed consideration of environmental issues, average daily traffic (ADT), design of roadway and route selections, planning right of way, and lane closure tactics should be taken into account when designing projects in these divisions. In addition, with the increase in the project length, the project might encounter an increased number of bridge work, different geographical and topographical features, and variation of availability of materials and skilled labor.



Where, X = length and Y = division

Figure 4-9 Bi-Variate Analysis of PE Cost

4.4.3 Factor Analysis

Factor analysis is a technique used to examine to what extent project-level factors explain PE cost and determine the factors that have the most influence on each data classification made in the previous section. Factor analysis is the identification of underlying factors that might explain the dimensions associated with large data variability (Hair et al. 2010). In this approach, the covariance and correlation matrix for all factors is first calculated to investigate the possible number of uncorrelated factors. Then, the number of uncorrelated factors is determined through the minimum eigenvalues criteria. The main purpose of factor analysis in this study is to reduce the large number of inter-correlated variables (26 project-level factors) to a smaller number of uncorrelated factors using Kaiser's eigenvalues criteria.

First, Kaiser's criterion requires taking the principal components for all factors and ranks the Eigen-values from largest to smallest. Based on the criteria, Eigen-values greater than 1.0 were retained or extracted for this study. Then, the variable loadings on each factor are calculated. A factor analysis results in a loadings ranging from -1 to 1 with loadings close to -1 or 1 indicating strong influence of the variable. For this study, loadings with ±0.3 are considered significant for interpreting the factors. Finally, an appropriate type of factor rotation is performed to create a better result or un-correlation among the factors. A varimax orthogonal rotation method was employed to maximize the correlations between these variables.

Table 4-7 Principal Component Factor Analysis of Correlation Matrix

Variable	PC1	PC2	PC3	PC4	PC5	PC6	PC7
Let year	0.003	-0.363	-0.675	-0.423	-0.476	-0.079	-0.025
Division	0.033	-0.714	-0.208	0.237	0.6	0.173	-0.003
Project Length	-0.183	0.188	-0.451	0.818	-0.212	-0.117	-0.01
AADT	0.31	0.409	-0.42	-0.187	0.536	-0.48	-0.087
No of Lanes	0.463	0.301	-0.265	0.004	0.012	0.787	0.067
Lane Width	0.57	-0.183	0.184	0.185	-0.222	-0.14	-0.711
Shoulder Width	0.575	-0.179	0.13	0.163	-0.179	-0.283	0.694
Eigen-value	2.3481*	1.1376*	1.076*	0.9224	0.8444	0.5119	0.1595
Proportion	0.335	0.163	0.154	0.132	0.121	0.073	0.023
Cumulative	0.335	0.498	0.652**	0.783	0.904	0.977	1

Table 4-7 shows the principal factor analysis of the correlation matrix (loadings along with the Eigen-values). According to Kaiser's criteria, three components (PC 1, PC 2, and PC 3) shown in the green box have Eigen-values greater than 1.0 (*) with 65.2 cumulative percentage (**), representing the data explained by these components. However, the results show that some factors load highly into other components that makes it difficult to interpret the subset containing the factors. For instance, lane width has a loading value of 0.575 in principal component 1 (PC 1) and -.711 in principal component 7 (PC 7). Therefore, a corrective measure such as equimax, quartimax, varimax, or orthomax rotation should be applied to decrease this overlapping effect on the loadings. For this report, the varimax rotation was selected to maximize the variance of the squared loadings. Based on an application of varimax orthogonal rotation, an increase in the correlation on each subset or component was observed. The loadings of the variables using the varimax rotation on each factor are shown in Table 4-8.

Table 4-8 Factor Analysis of Varimax Orthogonal Rotation with Total Variance

Variable	Factors			
	X_1	X_2	X_3	Communality
Project Type	**0.829**	0.163	-0.086	0.721
Project Length	**0.724**	-0.175	0.11	0.567
Fund type	0.384	**0.642**	-0.077	0.565
Let Year	-0.031	**0.741**	-0.252	0.614
Route type	0.195	**0.647**	0.228	0.509
Division	0.07	-0.218	**-0.829**	0.739
Consulting Firm	-0.027	0.216	**-0.508**	0.305

Variable	Factors			Communality
	X_1	X_2	X_3	
Eigen-value	1.8617	1.0893	1.0688	
Variance	1.6673	1.2676	1.085	4.0198
%	23.8%	18.1%	15.5%	57.4%
Cumulative %	23.80%	41.90%	57.40%	

Project type (0.829) and the length of a project (0.724) are greatly affected by, or load highly onto, X_1 while fund type (0.642), let year (0.741), and route type (0.647) are affected by X_2 and division (-0.829) and consulting firm (-0.508) are represented by X_3. Although the variables consulting firm and route type loadings are scattered between X_2 and X_3, there are strong relationships between project type and project length (X_1), fund type and let year (X_2), and division (X_3) which explain the scope of project, external factors, and geographical attributes respectively. All the communality values are more than 50%, except for consulting firm, which indicates that all variables are well represented by project scope, geographical attributes, and external factors. Based on a matrix plot, it was identified that a strong relationship exists between PE cost and the length of a project with a correlation matrix of 0.87. Therefore, project scope, geographical attributes, and external factors should be well defined or considered for estimating a reliable preliminary engineering cost.

For this report, it should be noted that a total of 22 nominal (categorical variables) and 3 interval variables were primarily considered as potential factors affecting PE costs. However, the research team reduced the number of variables from 26 to 7 due to missing data points and non-correlation of factors. This report will consider the individual 7 factors (consulting firm, length of project, contract let year, project type, route type, fund type, and location/division) for further analysis and development of PE cost prediction models.

5. Model Development

A reliable prediction of PE cost is essential for allocating budgets, minimizing errors and/or delays, and providing an efficient and effective management of projects for highway agencies. This chapter discusses the utilization of knowledge discovery in database (data mining) techniques to find meaningful patterns that can be used in developing PE cost prediction models. Three models, decision tree models, regression models, and neural network, are developed and compared to obtain the optimum PE cost prediction model. A validation of the optimum model is performed using recently completed engineering contract data.

5.1 Data Mining

Data mining can be defined as non-trivial extraction of implicit, previously unknown, interesting, and potentially useful information from data (Chen, 2001). It performs tasks such as classification, estimating, prediction, affinity grouping, clustering, and description. Data mining primarily involves six basic processes: problem statement and goal definition, identifying the sources (collection and understanding), preparing the data, build and train a model, validate the model, and implementation. The first three processes, or steps, are discussed in Chapters 1, 3, and 4. This chapter focuses on the remaining three steps: building and training the model, validating the model, and implementation.

Data mining can be classified as directed and undirected knowledge discovery. Undirected knowledge discovery recognizes relationships in data, while directed knowledge discovery explains those relationships once found (Berry and Linoff, 1997). In this report, a combination of directed knowledge discovery (decision tree and regression) and undirected knowledge discovery (neural network) is used to build, train, and compare the optimum preliminary engineering cost models. As mentioned in the Chapter 4, three types of components are used to predict the PE costs of roadway projects: a) engineering hours (man hours), b) aggregate cost/engineering hours, and c) PE cost for the various plan development task outputs. Models are developed for the 11 plan development task outputs using three components and are based on the project-level factors, decision tree models, multiple regression models, and neural network.. Then, a comparison of models is performed to obtain the best models followed by validation of the models. Figure 5-1 illustrates these model development phases.

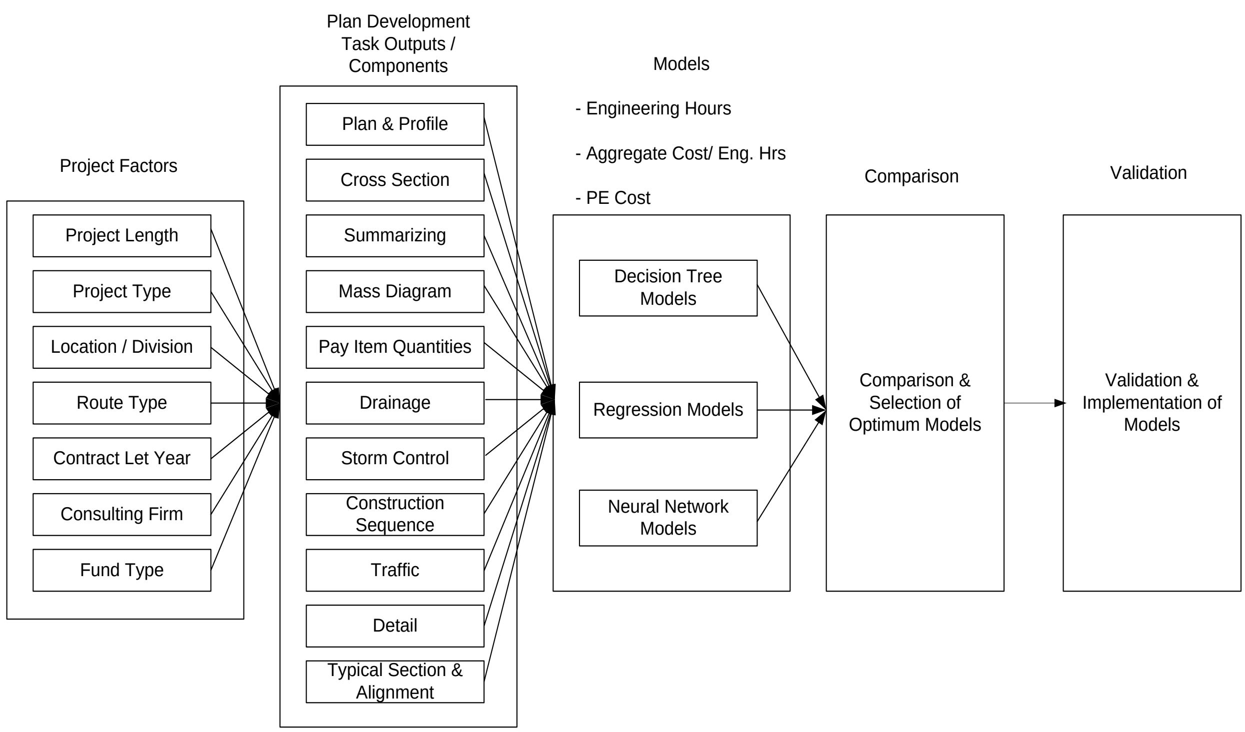


Figure 5-1 Model Development Phase

5.1.1 Decision Tree Model

A decision tree model is selected as one approach for directed knowledge discovery to develop PE cost prediction models. A decision tree model is a tree like structure that predicts target variables through a set of prediction rules (Berry and Linoff, 1997). Decision trees are drawn with a root node at the top by taking all the data and splitting it into branches or decision nodes. This process continues until it reaches the bottom node, or leaf node, based on the values of independent variables. During the splitting process, for each split or decision node or leaf, the number of observations is recorded, and the observation which has higher nodes is distributed to the lower nodes. A diagrammatic illustration of decision tree model is shown in Figure 5-2.

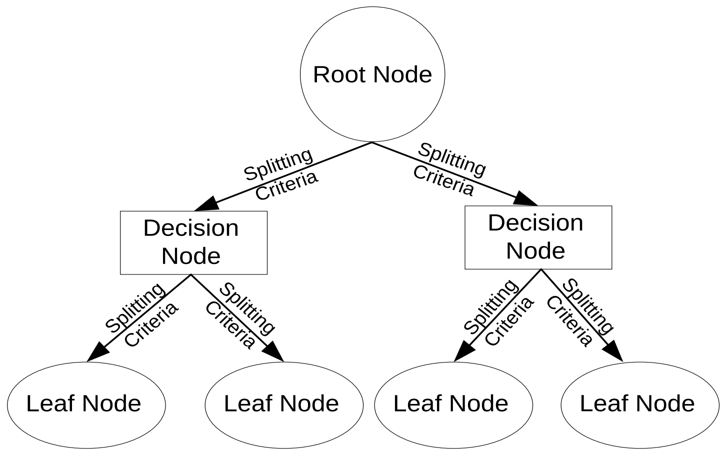


Figure 5-2 Decision Tree Model

In this study, a Pearson χ^2 statistic is used as a splitting criterion for the independent variables. The criterion uses the average value of the PE cost within each class for categorical variables and uses any potential value for interval variables as the splitting point. It should be noted that the χ^2 value is converted to a probability value for making a comparison with the χ^2 distribution. This probability value can be close to zero and is reported by its log-worth value. In addition, more complex models can be developed as a result of trees growing bigger which in turn results in over fitting of the model. Due to this reason, the validation set is used to select a sub-tree of the tree grown using the training set. This process is called "pruning". Pruning helps prevent over fitting and develops a simplified model. However, some splits have a sufficiently high worth (Chi-Square) value on the training data to enter the initial tree, and fail to improve the accuracy or error rate of the tree when applied to the validation data. Therefore, a set of fit statistics, average square error (ASE), and the root average squared error (RASE) is used to check the accuracy of the model. These fit statistics are also used for the remaining models and comparison as well.

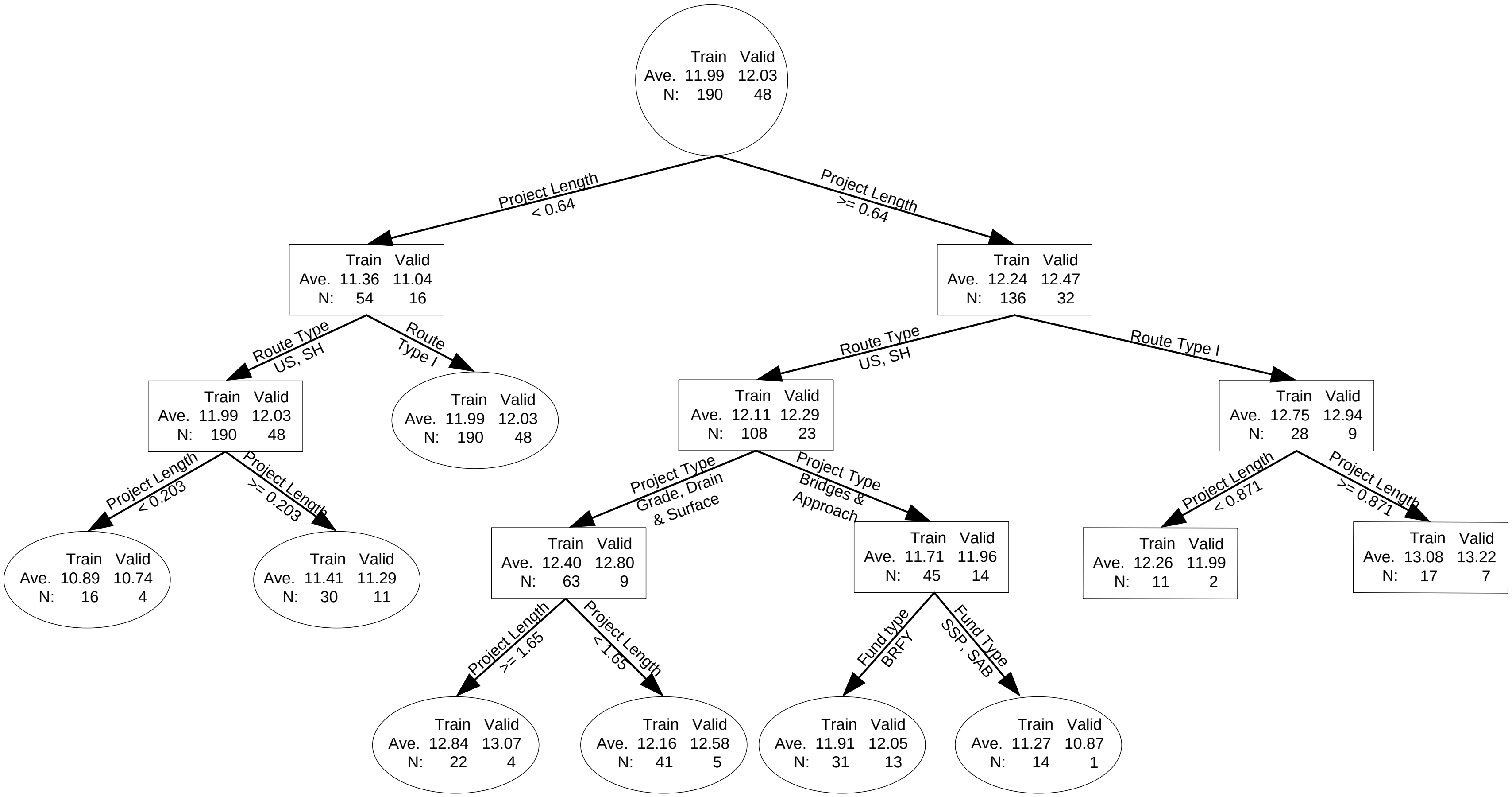


Figure 5-3 Decision Tree PE Cost Model

A decision tree model for estimating total PE cost is shown in Figure 5-3. Based on the model, the project length is the root node for splitting and consists of the average values for the PE cost and the number of cases used for both the training and validation process. For instance, if the project length is less than 0.64 miles, the average log-worth values are 11.36 and 11.04 with number of cases being 54 and 16 for training and validation, respectively. Route type is the next decision variable splitting either as interstate (leaf node) or US and state highway. The variable importance, in decreasing order, shows project length (1.00), route type (0.41), project type (0.33), and fund type (0.29). Based on the fit statistics, the ASE for the training-data set and validation-data set are 0.374683 and 0.351761, respectively. The RASE are 0.612114 and 0.593095. This close error rate shows that the developed model can perform well with new data set. Based on the analysis results, the decision tree models in predicting the PE cost components for major plan development task outputs in terms of the required engineering hours are attached as Appendix A.

Table 5-1 Decision Making Variables for Plan Development Task Outputs

Plan Development Task Outputs	Decision Making Variables/Attributes
Plan &Profile (P&P) Sheets	Project length, route type, let year
Cross-section Sheets	Project length, project type
Summarizing Sheets	Project length, fund type, consulting firm, let year, project type, route type
Mass Diagram Sheets	Project length, fund type
Pay Item Quantities (PQ) Sheets	-
Drainage Sheets	Project length, route type, consulting firm
Storm Control	-
Construction Sequence Sheets	Fund type, project length, let year, route type
Traffic Sheets	Project type, route type
Detail Sheet	Let year, route type, project length, project type
Typical Section & Alignment Sheet	Consulting firm, route type, project length, let year

Although the variable importance might slightly vary across the models, based on the decision tree models developed for the major development task outputs, project length seems to have a dominant effect on the majority of the PE cost components serving as a root node and decision node in the models. Route type also has a fairly significant effect on decision making as compared to project type, fund type, consulting firm, and contract let year. Project type affects traffic sheets, cross-section sheet, summarizing sheets, and detail sheets. The contact let year has an effect on developing plan and profile, summarizing, construction sequence, and typical section sheets work effort. The type of consulting firm is important to typical section, drainage sheets, and summarizing sheets; while fund type is important to construction sequence, mass diagram, and summarizing sheets. Although a representative model could not be developed for pay-item quantities and storm control plans, a summary of the decision making variables for the remaining models is shown in Table 5-1.

5.1.2 Regression Model

Multiple regression models are the second type of directed knowledge discovery used for this study. Regression models are the most widely accepted and frequently used models for the purpose of prediction and estimation due to the ease of understanding, training, and application of the model. This report utilizes a stepwise regression analysis to avoid reducndancy either due to low correlation or muliticolinearity by removing the least significant variables and adding the most significant variables step by step until no other effect in the model meets the significance level. For this report, significance levels of $\alpha = 0.15$ and $\alpha = 0.05$ are set to enter and leave the model, respectively. Akaike's Information Criterion (AIC) is used as a selection criteria for the models (models with the the smallest AIC value are chosen). AIC can be calculated using Eqn. 5.1,

$$AIC = n * \ln\left(\frac{SSE}{n}\right) + 2p \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots.\ldots\ldots\ldots Eqn. 5.1$$

Where, n is the number or cases considered, SSE is the error sum of squares, and p is the number of model parameters.

Based on the stepwise regression models, project length, route type, and fund type appear to be the dominating factors affecting the plan development task outputs. Project type, contract let year, and consulting firms are fairly important variables affecting the PE cost components. These results show almost a similar result when compared to the decision tree models. Likewise, a relationship between the project-level factors,pay-item quantities, and storm contol outputs were not identified. In addition, this confirms that project length is one of the most significant variables in the design and cost estimation of preliminary engineering of roadway projects. Table 5-2 shows the variable importance of the regression models.

Table 5-2 Variable Importance of Regression Models

Plan Development Task Outputs	Variable Importance
Plan &Profile (P&P) Sheets	Project length, route type, project type
Cross-section Sheets	Project length, project type, route type
Summarizing Sheets	Fund type, consulting firm, project type, project length, let year, route type
Mass Diagram Sheets	Fund Type, project length
Pay Item Quantities (PQ) Sheets	-
Drainage Sheets	Project length, route type, consulting firm
Storm Control	-
Construction Sequence Sheets	Fund type, project length, let year, route type
Traffic Sheets	Project type, route type
Detail Sheet	Let year, route type, project length, project type
Typical Section & Alignment Sheet	Consulting firm, route type, project length, let year

First, the study started with the assumption of developing linear models (with the hypothesis that the independent variable and dependent variables have a linear relationship). In most of the PE cost components, there appears to be a linear relationship between the project-level factors and the plan development task outputs. However, with a 95% confidence interval, the study resulted in a better model using the non-linear relationship assumption. Regresssion models developed as a function of logarithmic and polynomial functions better fit the actual value (training data). Once these models are developed, the perfomance is assessed using the coefficient of determination, R^2, to account for the sum of erros of the data. R^2 can be computed using Eqn. 5.2.

$$R^2 = \frac{\sum_{i=1}^{n}(y_i' - \overline{y})^2}{\sum(y_i - \overline{y})^2} \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots..\dots\dots.. Eqn. 5.2$$

Where, y_i is the actual value of the target variable, y_i' is the predicted value, and $\overline{y}$ is the mean value of the overall data. A model has a better fit whenthe R^2 value closer to 1.

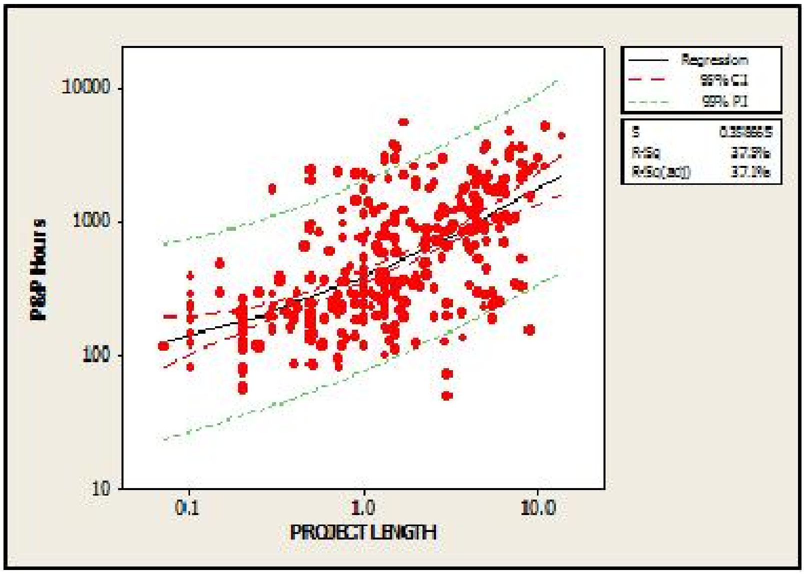


Figure 5-4 Fitted Line Plot for Plan & Profile Hours

For instance, the amount of work effort required to develop the plan and profile sheets for a project is expressed in terms of its project length as a logarithmic function with the coefficient of determination, R^2, of 37.5% (Figure 5-4). The regression model is set as Eqn. 5.3. The overall plan development task outputs resulted in R^2 values ranging from 21.2% to 40.2%. Although in the developed models R^2 is not the perfect fit, it indicates that the variability of the dataset is explained by the model as compared to the linear regression models which resulted in a coefficient of determination, R^2, less than 19.3%. Table 5-3 shows the PE cost prediction models for the various plan development task outputs based on the project length of a highway project.

$$Eng.Hours = 2.594 + 0.5496 \log_{10} PL + 0.1081 \log_{10} PL^2 \ldots\ldots\ldots\ldots Eqn. 5.3$$

Where, PL is the project length and Eng.Hours is the amount of work effort (engineering hours required to get a plan and profile sheet).

In the study, the number of records for engineering hours and number of sheets required did not match(with 353 total data points against 165 sheets) which made it difficult to compute the engineering hours per number of sheets used to develop the plan development task output models. Therefore, the researchers revised the component based PE cost prediction equation as Eqn. 5.4. Multiplying the two entities (engineering hours and aggregate cost per engeering hours) and summarizing for the whole plan develeopment tasks would result in a PE cost. In addition, a general model is developed for estimating PE cost to allow the engineer to estimate the values for the respective entities using the summation of the individual models and comparing it with the general PE cost model.

$$PE\ Cost = Eng.hours * \frac{Agg.Cost}{Eng.hours} \dots\dots\dots\dots\dots\dots\dots..\dots\dots Eqn. 5.4$$

Table 5-3 PE Cost Prediction Models Based on Project Length

Plan Development Task Outputs	Eng. Hours	R^2	Aggregate Cost /Eng Hour	PE Cost
Plan &Profile (P&P*	$2.594+0.5496\log_{10}(PL)+0.1081\log_{10}(PL)^2$	37.5%	1.900+ 0..00963 $\log_{10}$ (PL) -0.0276 $\log_{10}(PL)^2$ R^2 = 38.4%	5.142+ 0.4759 $\log_{10}(PL)$ R^2 = 40.2%
Cross-section*	$2.503+0.4748\log_{10}(PL)$	35.0%		
Summarizing*	$2.13+0.4042\log_{10}(PL)+0.06707\log_{10}(PL)^2$	28.9%		
Mass Diagram Sheets*	$1.402+0.2584\log_{10}(PL)$	22.6%		
Pay Item Quantities (PQ)*	$1.754+0.4136\log_{10}(PL)-0.08057\log_{10}(PL)^2$	27.3%		
Drainage Sheet*	$2.130+0.45476\log_{10}(PL)$	26.6%		
Storm Control*	$1.702+0.3861\log_{10}(PL)$	27.3%		
Construction Sequence*	$1.847+0.4234\log_{10}(PL)-0.07050\log_{10}(PL)^2$	21.2%		
Traffic Sheet*	$2.191+0.43301\log_{10}(PL)+0.09865\log_{10}(PL)^2$	23.9%		
Detail Sheet*	$2.145+0.4125\log_{10}(PL)$	25.2%		
Typical Section & Alignment*	1.796+0.0.4202$\log_{10}$(PL)	24.9%		

Where, PL = Project Length (Mile), * implies all the functions are $\log_{10}$ functions

A variety of non-linear regression models are developed to account for the various factors or attributes encountered in estimating PE cost. For instance, a project falls into a category of either interstate highway, state highway, or US highway with regard to its route type. As mentioned before, these route types have a significant effect on estimating PE cost. This is because projects near an interstate might have higher daily traffic, are in closer proximity to urbanized areas, or require special features in planning the right of way and lane closure tactics as compared to state highways. All of these reasons make the project design and PE cost higher. These effects of route type are shown in a logarithmic regression PE cost models across a project length (Figure 5-5). The models predict that there is a steep increase of PE cost followed by a gradual slope with the increase in project length. Although the coefficient of determination, R^2, for interstate highway projects is low with a value of 0.1214, PE costs tend to be higher for interstates (I) as compared to US highways (US) which have an R^2 value of 0.3105 and state highways (SH), 0.3086.

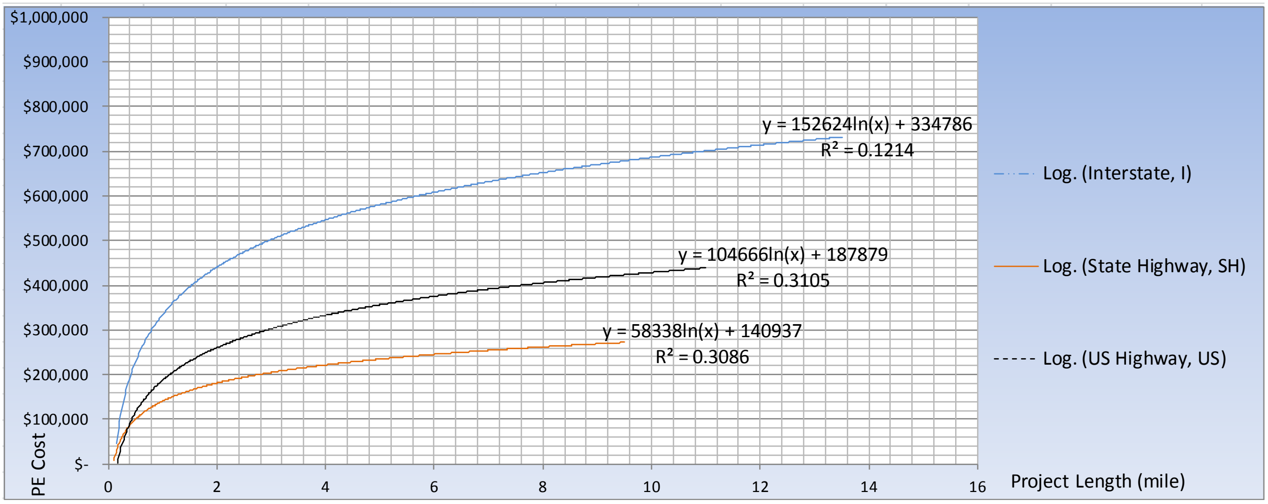


Figure 5-5 PE Cost Regression Models for Route Type

Similarly, PE costs tend to have a steep slope followed by a gradual increase for the various project types. The project types were broadly classified as either widen/reconstruct/interchange, or grade, drain, bridge and surface, or bridges and approaches. Although there is an overlap of PE costs between the project types for the first two miles of a project (with widen/reconstruct/interchange projects having higher PE costs), there is an increase of PE cost in the order of bridges and approaches, widen/reconstruct/interchange, and grade, drain, bridge and surface (Figure 5-6). This might be due to the low coefficient of determination, R^2, of widen/ reconstruct/interchange (0.0777) which might not be a representative logarithmic regression model as compared to the bridges and approaches model which resulted in an R^2 value of 0.187 and grade, drain, bridge and surface models of 0.3176. Grade, drain, bridge and surface projects involve all types of roadway work as compared to widen/reconstruct/interchange projects and bridge and approaches projects that might involve, but are not limited to, adding lanes or shoulders, constructing reinforced concrete boxes and drainage structures, or replacing existing bridge structures.

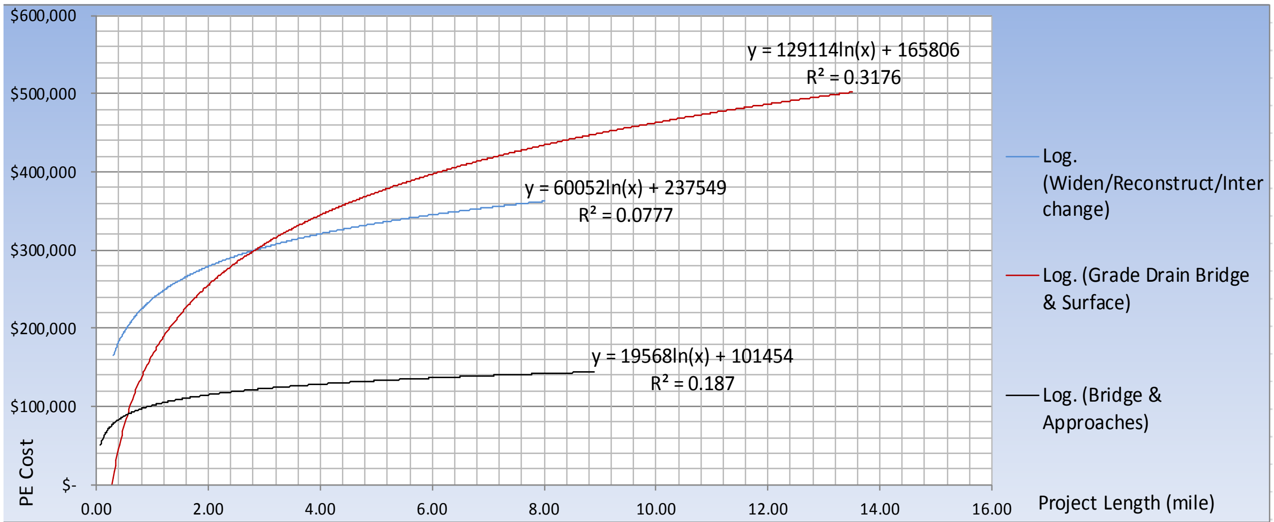


Figure 5-6 PE Cost Regression Models for Project Types

The effect of location on PE cost is associated with the daily traffic, geographical nature, environmental attributes in the area, designing the right-of-way, and the close proximity of the project for material delivery and supply. Based on ODOT's geographical classification of the state, Divisions 8, 3, 4 and 1 (decreasing order) have higher PE costs as compared to Divisions 2, 5, 7, and 6 (decreasing order). This may be attributed to the fact that some of these areas are densely populated and urbanized areas with higher traffic that require special considerations in designing and planning the right of way, lane closure tactics, and acquiring environmental documents. Figure 5-7 shows PE cost models for the various divisions. The developed models vary across the locations with Divisions 1, 3, 5, and 8 fitting polynomial regression models, while Divisions 2 and 6 fitting exponential regression models, and Divisions 4 and 7 fitting power regression models. The R^2 values for all location models vary from 0.2561 to 0.6602. An ODOT engineer can use these developed models to have a rough estimate of PE costs when various project attributes are encountered and compare the final PE costs with the component-based PE cost and negotiating purposes with consulting firms.

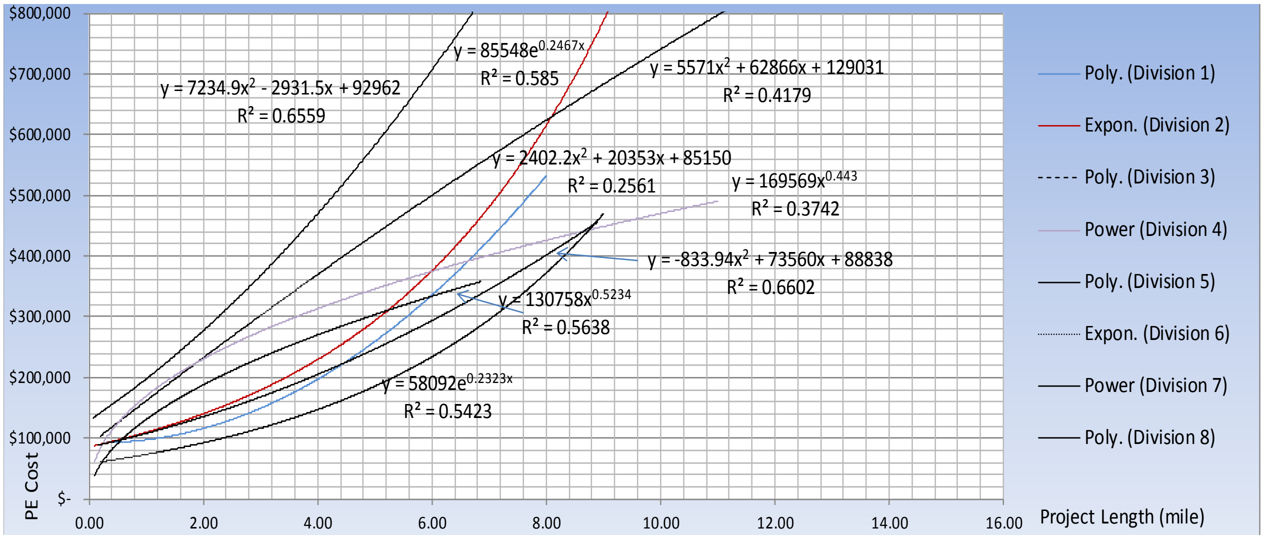


Figure 5-7 PE Cost Regression Models for Project Location

A score rankings matrix of the developed models is used to show the mean predicted and mean target values at various percentiles for the training-data set and the validation-data set. The score ranking matrix plot overlays a statistics model for standard, baseline, and best models in a lattice that is defined by the training- and validation-data sets (Berry and Linoff, 1997). Based on the plot, the training-data set of the predicted model almost fits the targeted models for most of the various plan development task outputs. However, there are discrepancies or errors with regard to the validation-data set. Therefore, the accuracy of the model is tested based on the ASE and RASE as used for the decision tree models. A sample score ranking matrix for plan and profile is shown in Figure 5-8. The ASE for the training-data set is 0.343877 while the validation is 0.529579. In addition, the RASE for the training- and validation-data sets are 0.58641 and 0.727722 respectively which implies that the error rate differences are not significant enough to predict new data. Regression models based on project length for predicting the engineering hours required to prepare the plan development outputs are included in Appendix B that resulted in a coefficient of determination ranging from 23.9% to 40.2%.

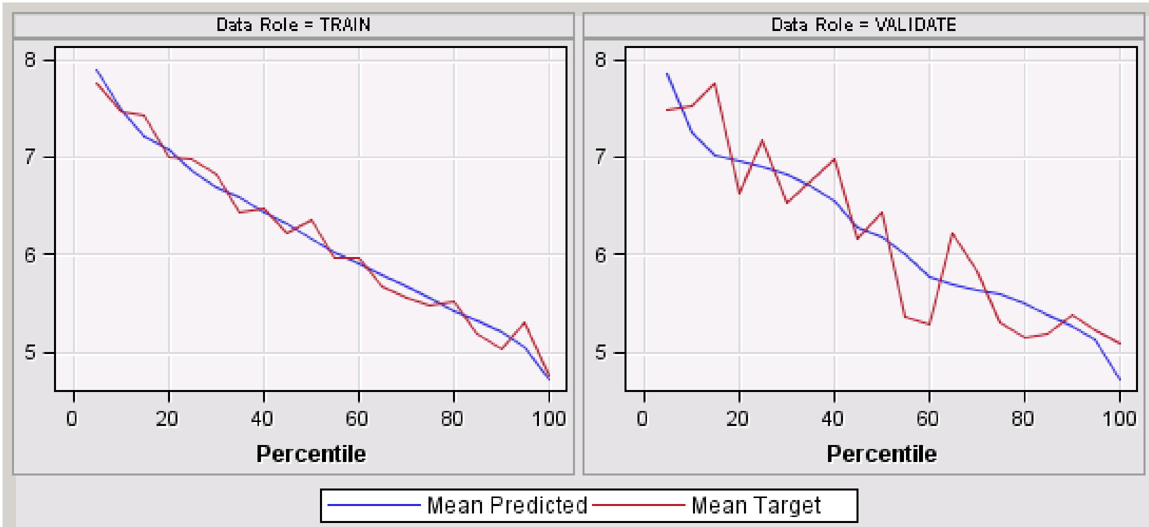


Figure 5-8 Mean Predicted Score Ranking Matrix

5.1.3 Neural Network Model

The third model that is utilized in this report is undirected knowledge discovery, or neural network models. The neural networks model is a very powerful, general purpose tool readily applied to prediction, classification, and clustering (Berry and Linoff, 1997). A neural network is a learning system which has the ability to generalize and learn from data by modeling the neural connections in human brains. Although the neural network is a powerful tool with higher accuracy and is capable of modeling nonlinear relationships among variables, its drawback is interpreting the variables as the result of training a neural network, and there are internal weights distributed throughout the network. Berry and Linoff (1997) state that "neural networks are best approached as 'black boxes' with mysterious internal workings".

A neural network model is comprised of three basic units: input layer, hidden layer, and output layer. Each unit or layer has many inputs that it combines into a single output value through nodes or neurons. For each of these units, a weight is assigned for its connection based on the training process. Figure 5-9 illustrates a representative neural network model in graphical format. The input layer represents the project-level factors (location, project length, etc.) that are discussed in the previous section, while the output layer represents the 11 PE cost components for the various plan development task outputs such as plan and profile sheets and drainage sheets. The selection of the optimum number of hidden layers depends on the type of model selection. In this report, Multi Layers Perceptron (MLP) is used to enable the neural network model to solve non-linear separable problems that cannot be attained using single layers. A supervised learning process of Back Propagation algorithm is utilized to train the MLP model.

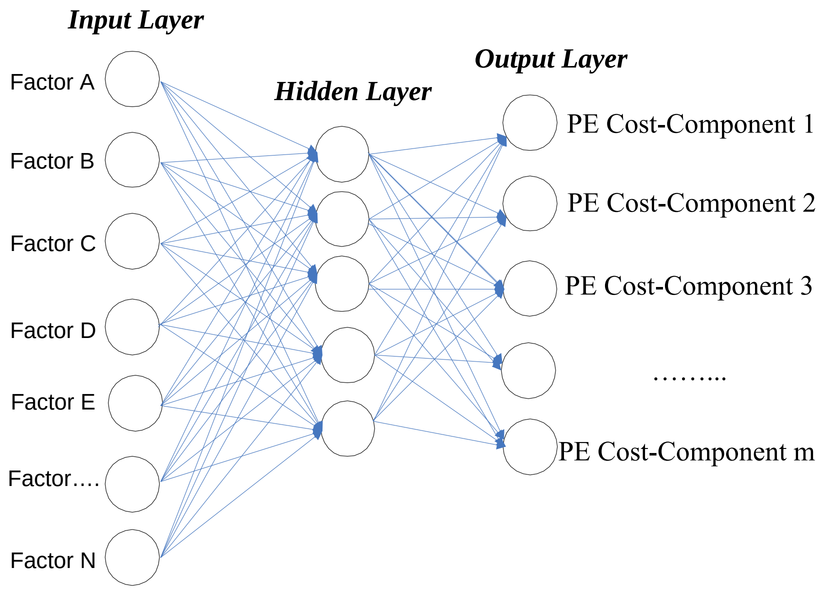


Figure 5-9 Representative Neural Network Model

Table 5-4 Important Variables for Neural Network Models

Plan Development Task Outputs	Decision Making Variables/Attributes
Plan &Profile (P&P) Sheets	Project length, route type, let year
Cross-section Sheets	Project length, project type
Summarizing Sheets	Project length, fund type, consulting firm, let year, project type, route type
Mass Diagram Sheets	Project length, fund type
Pay Item Quantities (PQ) Sheets	-
Drainage Sheets	Project length, route type, consulting firm
Storm Control	-
Construction Sequence Sheets	Fund type, project length, let year, route type
Traffic Sheets	Project type, route type
Detail Sheet	Let year, route type, project length, project type
Typical Section & Alignment Sheet	Consulting firm, route type, project length, let year

Based on the neural network models, the project length affects most of the plan development tasks; additionally, the route type also has a fairly significant importance to most tasks. Project type affects traffic sheets, cross-section sheet, summarizing sheets, and detail sheets. The contract let year has an effect on developing plan and profile, summarizing, construction sequence, and typical section sheets work effort. The type of consulting firm is important to typical section, drainage sheets, and summarizing sheets; while fund type is important to construction sequence, mass diagram, and summarizing sheets. A summary of important variables for the models is shown in Table 5-4.

5.2 Model Comparison

After all models are developed, these models are compared to obtain the most optimum models to predict preliminary engineering costs. As mentioned earlier in the report, the two fit statistics that were selected for this study are ASE and the RASE. The ASE is the measure of the variance or square difference between the actual and the predicted value divided by the number of records. RASE is simply the square root of ASE. The ASE can be calculated using Eqn. 5.5. The smaller the values of these errors imply that the accuracy of the model is better. The ASE and RASE is computed for the training and validation of all three models. Table 5-5 shows the summary of a comparison of the models.

$$ASE = \frac{1}{N}\sum_{i=1}^{n}(y_i - \overline{y}_i)^2 \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots Eqn. 5.5$$

Where, N is the number of observations, y_i is the actual target value of the i^{th} observation, and $\overline{y}_i$ is the predicted target value of the i^{th} observation.

The number of observations used for data partitioning purpose followed the 80:20 rule, with 282 data points for training and 71 data points for validation of the three models. Based on the comparisons, the neural network model performs better than the regression and decision tree models for 6 major plan development task outputs: summarizing sheets, storm control, construction sequence, detail sheets, pay item quantities, and PE cost (Table 5-5). With regard to plan and profile, cross-section sheets, and mass diagram sheets, the regression model outperforms decision tree and neural network models. The decision tree models outperform the neural network and regression models in drainage, traffic sheets, and typical section.

Table 5-5 Comparison of Models

Plan Development Tasks	Criteria	Decision Tree		Regression		Neural Network	
		Training	validation	Training	validation	Training	validation
Plan &Profile (P&P) Sheets	ASE	0.560625	0.626760	0.343877	0.529579	0.486864	0.5997712
	RASE	0.748749	0.791682	0.586410	0.7277219	0.697756	0.7744489
Cross-section Sheets	ASE	0.699933	1.548522	0.413793	1.452416	0.599597	1.5657284
	RASE	0.836620	1.244396	0.643267	1.2051623	0.774336	1.2512907
Summarizing Sheets	ASE	0.777591	1.644340	0.529002	2.4930812	0.916289	1.223656
	RASE	0.881811	1.282318	0.727325	1.5789494	0.957230	1.1061901
Mass Diagram Sheets	ASE	1.728887	2.200923	1.153416	1.743970	1.739437	1.8765761
	RASE	1.314872	1.483550	1.073972	1.3205945	1.318877	1.3698818
Drainage Sheets	ASE	0.728257	1.31240	0.593828	1.3480354	0.704935	1.4111320
	RASE	0.853380	1.145603	0.770602	1.1610492	0.839604	1.1879108
Storm Control	ASE	1.439551	1.178777	0.812367	1.8370611	1.203052	0.868334
	RASE	1.199813	1.085715	0.901314	1.3553822	1.096837	0.931844
Construction Sequence Sheets	ASE	1.034504	1.423953	0.679685	1.4453825	1.047435	1.305520
	RASE	1.017106	1.193295	0.824430	1.2022406	1.023443	1.142593
Traffic Sheets	ASE	2.435371	2.46068	2.129079	3.3593683	2.435494	2.672884
	RASE	1.560568	1.568657	1.459137	1.8328579	1.560607	1.634895
Detail Sheet	ASE	1.488488	1.644000	0.825476	2.3130791	1.185173	1.462168
	RASE	1.220036	1.282000	0.908557	1.5208810	1.088657	1.209201
Typical Section & Alignment Sheet	ASE	0.526249	0.56316	0.416617	0.6547077	0.562506	0.596513
	RASE	0.725430	0.750445	0.645459	0.8091401	0.750004	0.772343
Pay Item Quantities (PQ) Sheets	ASE	12753.00	8113.891	5417.290	9919.2295	6762.200	7573.977
	RASE	112.9290	90.07715	73.60230	99.595329	82.23260	87.02860
PE Cost	ASE	0.374683	0.351761	0.224495	0.2984472	0.210282	0.286551
	RASE	0.612114	0.593094	0.473809	0.5463032	0.458565	0.535305

This evaluation is based solely on ASE of the validation data as to test whether the developed model would perform better on a new data set. Based on the analysis and model comparison, the neural network models outperform regression models and decision tree models. Although decision tree and regression models outperform neural networks for some plan development tasks, the variation of errors is not significant. However, a comparison made based on the training-data set reveals that the regression model performs much better for all plan development task outputs. This might be due to the fact that a small number of validation data were used as compared to the training-data set, and the model might perform better with 50:50 training to validation ratio instead of 80:20. In addition, as discussed in the previous section, since the neural network is difficult to interpret the weights of the variables (hidden layers), the regression models might be the most suitable model to develop an estimation tool for predicting preliminary engineering cost.

6. PE Cost Estimation Tool

This chapter discusses the development of a component-based estimation tool for a reliable estimation of roadway PE costs which will help when facilitating the negotiation process with consulting firms. The tool, or system, is developed using Microsoft Excel Spreadsheet by utilizing the regression models developed in the previous chapter. The system, Roadway PE Cost Estimator, consists primarily of three sections or tabs:

- component-based PE cost estimator based on engineering hours,
- PE cost estimator based on project-level factors, and
- Number of sheets estimator.

Based on the input provided (length of the project) by the user, the model predicts PE cost along with the engineering hours, the number of sheets, and cost per engineering hour required for the various plan development task outputs. The system also predicts PE costs based on three project-level factors: project type, location, and route type.
PE cost estimator based on project-level factors

6.1 Component-Based PE Cost Estimator Tab

The component-based PE cost estimator tab consists of two parts. The first part consists of user inputs, where the user enters the project chracterstics (engineering contract number, project type, plan type, and date of estimation) and project length. The second part consists of the outputs of the engineering hours and cost per hour estimation for the various plan development tasks. Multiplying each entity's (plan development task outputs) engineering hours with the aggregate cost per engeering hours results in a component cost for each entity. Then, summing up the component costs for the whole plan develeopment task output results in a total PE cost. Figure 6-1 illustates a screenshot of the component-based PE cost estimator tab.

ROADWAY PE COST ESTIMATOR

E.C. NO.	533	PROJECT TYPE	RECONSTRUCT ROADWAY
DATE	5/4/2012	PLAN TYPE	CONSTRUCTION PLANS
		PROJECT LENGTH	2.2 Miles

No.	Plan Development Task Outputs	Engineering Hours	Aggregate Cost/Engineering Hour	PE Cost
1	Plan &Profile (P&P) Sheets	624	80	$ 49,920
2	Cross-section Sheets	464		$ 37,120
3	Summarizing Sheets	189		$ 15,120
4	Mass Diagram Sheets	31		$ 2,480
5	Pay Item Quantities (PQ) Sheets	77		$ 6,160
6	Drainage Sheets	194		$ 15,520
7	Storm Control	69		$ 5,520
8	Construction Sequence Sheets	97		$ 7,760
9	Traffic Sheets	224		$ 17,920
10	Detail Sheet	194		$ 15,520
11	Typical Section & Alignment Sheet	88		$ 7,040
			PE COST ∑ =	$ 180,080

Figure 6-1 Component-Based PE Cost Estimator Tab

6.2 Project-Level-Factor-Based PE Cost Estimator Tab

The second tab (project-level-factor-based PE cost estimator tab) allows the user to make three selections that describe the project from a drop down list. These three selections, or options, are based on project-level factors: project type (widen/reconstruct/ interchange, grade/drain/ bridge /surface, bridges & approaches), route type (interstate (I), US highway (US), State highway (SH)), and location (Divisions 1, 2, 3, 4, 5, 6, 7, 8). Once the user selects the appropriate project characteristic, the model predicts the PE cost for each selection. Since the project length is entered previously, the project-level-factor tab is updated automatically. This would allow the user to have an approximate estimate of the project based on the project-level factors and compare the results with the component-based PE cost estimator tab. Figure 6-2 shows the screenshot of the tab.

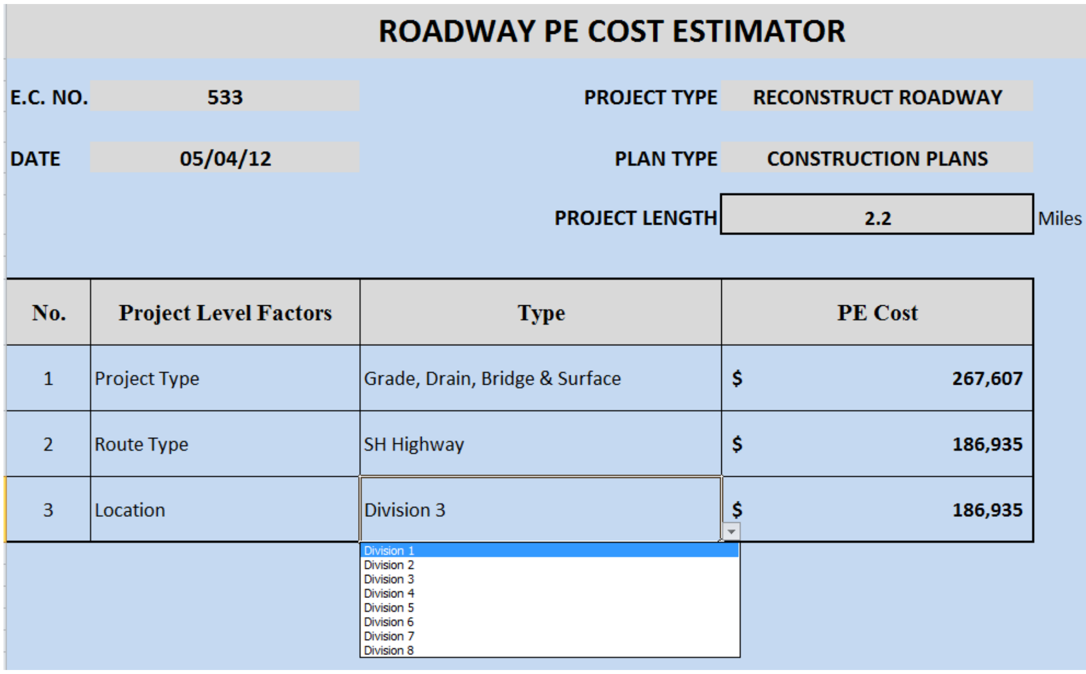


Figure 6-2 Project-Level-Factors-Based PE cost Estimator Tab

6.3 Number of Sheets Estimator Tab

As mentioned in the previous chapter, the number of records of engineering hours and number of sheets required did not match (lower number of records compared to number of sheets). However, regression models were developed to estimate the number of sheets for the plan development task outputs based on the 165 data points and project length. Based on the models, the coeffecient of determination, R^2, for all plan develeopment task outputs resulted in a range between 0.15 and 0.436, except for detail sheets and typical section and alignment sheets (Table 6-1). The number of sheets estimator tab allows the user to input roadway length for a certain engineering contract, and the tab calculates the number of sheets for each component and sums up the total number of sheets required for the project. Figure 6-3 shows a screen shot of the number of sheets estimator tab.

Table 6-1 Coefficient of Determination for Number of Sheets

Plan Development Task Outputs	R^2
Plan &Profile (P&P) Sheets	0.436
Cross-section Sheets	0.2764
Summarizing Sheets	0.2196
Mass Diagram Sheets	0.3063
Pay Item Quantities (PQ) Sheets	0.1566
Drainage Sheets	0.2276
Storm Control	0.2365
Construction Sequence Sheets	0.1928
Traffic Sheets	0.2454
Detail Sheet	0.0183
Typical Section & Alignment Sheet	0.0708

ROADWAY PE COST ESTIMATOR

E.C. NO. 533 PROJECT TYPE RECONSTRUCT ROADWAY

DATE 5/4/12 PLAN TYPE CONSTRUCTION PLANS

PROJECT LENGTH = 2.2 Miles

No.	Plan Development Task Outputs	Number of Sheets
1	Plan &Profile (P&P) Sheets	17
2	Cross-section Sheets	72
3	Summarizing Sheets	3
4	Mass Diagram Sheets	2
5	Pay Item Quantities (PQ) Sheets	2
6	Drainage Sheets	6
7	Storm Control	6
8	Construction Sequence Sheets	3
9	Traffic Sheets	15
10	Detail Sheet	-
11	Typical Section & Alignment Sheet	-
	Total No Of Sheets ∑ =	126

Figure 6-3 Number of Sheets Estimator Tab

6.4 Model Validation

A validation of the roadway PE cost estimator model was performed using random ODOT engineering contracts from the validation-data set which consists of engineers' estimated engineering hours and number of sheets. An engineering project with a contract let year of 2008 contracted in Division 7, state highway route, grade, drain, and surface project with 2.2 miles of project length was tested using the roadway PE cost estimator. Table 6-2 shows the comparisons of ODOT's estimate and component-based roadway PE cost estimator.

Table 6-2 Comparisons of PE Cost

Plan Develop. Tasks	Component- Based Prediction Model				ODOT			
	Eng. Hours /Sheet	No of Sheets	Cost/ Eng. Hour	PE Cost	Eng. Hours /Sheet	No of Sheets	Cost/ Eng. Hour	PE Cost
Plan & Profile	624	17	$ 80.00	$ 49,920	480	10	$ 82.82	$ 5,963
X-Section	464	72		$ 37,120	488	61		$ 39,754
Summarize	189	3		$ 15,120	340	4		$ 40,416
Mass Diagram	31	2		$ 2,480	24	1		$ 28,159
Pay Item Quant.	77	2		$ 6,160	90	1		$ 1,988
Drainage	194	6		$ 15,520	204	4		$ 7,454
Storm Control	69	6		$ 5,520	88	5		$ 16,895
Constr. Seq.	97	3		$ 7,760	28	1		$ 7,288
Traffic	224	15		$ 17,920	370	15		$ 2,319
Detail	194	-		$ 15,520	158	4		$ 30,643
Typical Section & Alignment	88	-		$ 7,040	72	23		$ 13,086
			PE Cost =	**$ 180,080**			**PE Cost =**	**$ 193,964**

Based on the model comparison, the roadway PE cost estimator resulted in a conservative preliminary engineering cost ($ 180,080) as compared to ODOT's estimate ($193,964). The model also predicted a relatively lower amount of work effort required (2,251 of engineering hours) as compared to ODOT's work effort of 2,342 engineering hours. However, higher estimation of engineering hours were estimated for plan & profile, mass diagram, construction sequence, and detail sheets. For comparison purposes of the PE cost components, engineering hours with greater or lower than 40 hours (1 week of work for 1 engineer) are considered significant to avoid over- or under-estimation of PE cost components. Based on this criterion, an over-estimation of engineering hours is experienced in plan & profile (difference of 144 engineering hours) and construction sequence (69 hours), while under-estimation has resulted in summarizing (151 hours) and traffic sheets (146 hours).

A comparison of total PE cost based on the project-level factors resulted in a close estimation with regard to route type ($ 186,935) and location ($ 197,557); however, it over-estimated with regard to the project type ($ 267,607). This over estimation might be due to the fact that the model considers project type as grade, drain, bridge, and surface, but the project considered might not have bridge work. Although the number of sheets estimator resulted in a slightly higher number as compared to ODOT's sheets, the margin of error is low (with most of the plan development task outputs higher by 1 or 2 sheets). This might be due to the low number of data points associated with developing the regression models. It should be noted that the number of sheets for detail and typical section sheets are rejected due to a low value of coefficient of determination, R^2. Overall, the model predicts a relatively closer preliminary engineering cost in which ODOT engineers can benefit for the purpose of estimation and negotiation with consulting firms.

7. Conclusion

7.1 Summary

A reliable estimate of preliminary engineering (PE) cost is critical for highway agencies for allocating budgets, completing projects on time, and reducing public inconvenience. However, many highway agencies do not have a prediction system or mechanism to estimate their PE costs and usually utilize engineers' judgment in estimating PE costs. This study presented data-driven and component-based roadway PE cost prediction models by utilizing a knowledge discovery in database (data mining) techniques. The study used a combination of knowledge discovery in database methods (regression models, decision tree models, and neural network models) and compared these three models to obtain the optimum roadway PE cost model as a function of a) engineering hours required per number of sheets, b) number of sheets and c) cost per engineering hours for the various plan development task outputs.

Based on the study, a comprehensive review of literature was first conducted to summarize the current practice of estimating preliminary engineering costs. A group of 353 roadway preliminary engineering contract data were acquired through interviews and meetings with engineers and ten years of engineering contract documents from the Oklahoma Department of Transportation (ODOT) engineering division. First, 26 critical factors affecting PE costs were identified and classified into five major categories of project-level factors: project scope, geographical attributes, design attributes, environmental attributes, and external factors. Then the study utilized factor analysis of a covariance and correlation matrix to investigate the significance and identify the correlation among these factors. Once the data analysis and classification were finalized, the study developed three models, (regression models, decision tree models, and neural network models) and compared them to obtain the optimum roadway PE cost model for the plan development task outputs. Finally, the study developed an Excel-based program called Roadway PE Cost Estimator and validated the program.

7.2 Findings

On average, 70% of roadway project PE costs fall in the range between $50,000 and $350,000, while more than 20% of the projects' costs range between $350,000 and $650,000. Furthermore, less than 6% of the projects have a PE cost greater than $650,000. This wide variation of PE cost might be a result of projects ranging from supplemental work of an existing contract to new design and complex rehabilitation highway projects. In addition more than 60% of the roadway project lengths fall between 1.4 miles and 7 miles. Of the ten years of data, a higher number of projects were contracted to consulting firms in the years 2006 (67projects) and 2009 (59 projects), with more than 70% of projects being contracted in the last five years. This might be due to the higher amount of budget allocated for transportation agencies for building the US economy.

A principal component factor analysis based on Kaiser's criterion retained components with eigenvalues greater than 1 and considered factor loading of ±0.3 significant for interpreting factors. The analysis resulted in three components that explain 57.4% of the total variance in the data. Project type and project length fall under project scope, while division and consulting firm belong to the geographical attributes. Additionally, fund type, contract let year, and route type fall under external factors category. However, due to the low correlation of factors and missing data points through the data cleanup and preparation stages, out of the 26 factors only 7 factors (consulting firm, length of project, contract let year, project type, route type, fund type, and location/division) were retained and used to develop roadway PE cost prediction models.

Although the variable importance might slightly vary across the developed models, based on the decision tree models developed for the major development task outputs, project length seems to have a dominant effect on the majority of the PE cost components serving as a root node and decision node in the models. Route type also has a fairly significant effect on decision making as compared to project type, fund type, consulting firm, and contract let year. Project type affects traffic sheets, cross-section sheet, summarizing sheets, and detail sheets. The contact let year has an effect on developing plan and profile, summarizing, construction sequence, and typical section sheets work effort. The type of consulting firm is important to typical section, drainage sheets, and summarizing sheets; while fund type is important to construction sequence, mass diagram, and summarizing sheets.

Regression models developed as a function of logarithmic and polynomial functions better fit the actual value or the training data. The regression models predict that there is a steep increase of PE cost followed by a gradual slope with the increase in project length with regard to route type and project type. Although there is an overlap of PE costs between the project types for the first one mile of a project (with widen/ reconstruct/ interchange projects having higher PE costs), there is an increase of PE cost in the order of bridges and approaches, widen/reconstruct/ interchange, and grade, drain, bridge, and surface. According to ODOT's geographical classification, Divisions 8, 3, 4 and 1 have higher PE costs as compared to Divisions 2, 5, 7, and 6. In addition, PE costs tend to be higher for projects that encounter interstate highways as compared to US highways and state highways.

Based on the neural network models, project length affects most of the plan development tasks, while route type has a fairly significant importance to most tasks. Project type affects traffic sheets, cross-section sheet, summarizing sheets, and detail sheets. The contact let year has an effect on developing plan and profile, summarizing, construction sequence, and typical section sheets work effort. The type of consulting firm is important to typical section, drainage sheets, and summarizing sheets; while fund type is important to construction sequence, mass diagram, and summarizing sheets.

Overall, project length is identified as the most significant factor affecting PE cost components using all three models. In addition, project type, fund type, and route type greatly affect PE plan development task outputs. Based on the decision tree and neural network models, representative models could not be developed for pay-item quantities and storm control plan development task outputs. A comparison of the developed models resulted in neural network models to outperform regression models and decision tree models based on the validation average squared error (ASE). Although the decision tree models and regression models outperform neural networks for some plan development task outputs, the variation of errors is not significant.

However, a comparison made based on the training-data set reveals that regression models perform much better for all plan development task outputs. This might be due to the fact that a small number of validation data were used as compared to the training-data set and the model might perform better with a 50:50 training to validation ratio instead of an 80:20 ratio. In addition, neural network models make it more difficult to interpret the variables as the results of training of the neural are distributed through internal weights across the network. Therefore, regression models have been selected to develop an Excel-based program called Roadway PE Cost Estimator. The program allows ODOT engineers to predict a relatively reliable estimate of PE cost as a function (entities) of engineering hours per number of sheets, number of sheets, and cost per engineering hours with coefficient of determination, R^2 , ranging between 21.2% and 41%.

7.3 Conclusion

An effective planning and estimating of PE cost and a well-defined scope in the preconstruction phase have a major impact throughout the project life cycle as it is difficult and costly to make changes as the project progresses through bidding and construction phases. An accurate preliminary engineering cost estimate and a well-defined scope are crucial for the success of any highway project. This study presents data-driven and component-based PE cost prediction models by utilizing critical factors retrieved from historical roadway project data. The data-driven and component-based estimation system, Roadway PE Cost Estimator consists of three functions or entities:

a) engineering hours required per number of sheets,
b) number of sheets, and
c) cost per engineering hours for the selected major plan development task outputs.

Once these entities are estimated, multiplying them, and summarizing them for the whole plan develeopment task outputs would result in a PE cost for a certain roadway project. The developed system allows the engineer to estimate the values for the respective entities. It could help engineers determine a reliable number of sheets, work effort (engineering hours) required per number of sheets, cost per engineering hour, and total cost for the selected plan development task output with respect to project length, location, route type, and project type. In addition, this system not only allows engineers to easily manipulate the requirements for a specific task, but also helps them configure contingencies, as to whether any of the entities are either under- or over-estimated. It also helps to identify if a misallocation of resources (engineering hours assigned to the respective skilled manpower or number of sheets assigned to each task) exists at a specific level especially when negotiating PE costs with consulting firms.

The results of this project are expected to significantly influence how efficiently and economically highway projects are planned, executed, and managed in the early stages of a project. Although this study is for the Roadway Division only, the results show that potential factors affecting PE cost and data-driven and component-based PE cost prediction models are influential in determining an efficient and reliable PE cost estimate. The developed model will not only allow ODOT to be equipped with a streamlined procedure for estimating PE costs, but will also facilitate consistent practices and a structured format of PE cost estimating.

7.4 Future Study

This study focused on developing a preliminary engineering cost prediction model for the Roadway Division. However, the Engineering Division consists of the Bridge Division, Survey Division, Right-Of-Way Division, Environmental Division, Utilities Division, Traffic Division, and Project Management Division. In addition, these divisions work with the capital programs, local government, and the Planning and Research Division. These ODOT divisions have the authority to perform the design work either by an in-house design team or outsource it to consulting firms depending on the type and size of the project. For instance, the Environmental Division outsources only 20% of their design work to consulting firms as compared to the Roadway and Bridge Divisions which outsource more than 50% of their design work to consulting firms. This would create a difficulty in managing, coordinating, and tracking the design work between the various firms and the highway agency.

Furthermore, these divisions utilize a variety of methods to estimate PE costs ranging from paper documents using engineers' judgment and experience to computer aided programs such as Microsoft Excel spreadsheets. However, there is not a system which integrates the methods/systems for easier data and information flow between these divisions. It should be noted that the data collection and preparation process for this study was a very time consuming process as the PE cost data is stored manually (paper format). For each project, the data was acquired by going through the engineering contract documents. Therefore, an innovative data and information integration framework should be developed to account for a) a smooth flow of data and information, b) collection and storage of digitalized and standardized data and information, c) easier communication and retrieval of information, and d) support decision making process between the various engineering divisions.

References

1. Anderson, Stuart D. and Blaschke, Byron C. (2004), "NCHRP Synthesis of Highway Practice 331: Statewide Highway Letting Program Management; A Synthesis of Highway Practice' Transportation Research Board, Washington D.C.

2. Anderson, Stuart D., Molenaar, Keith Robert, and Schexnayder, Cliff J. (2007*)* "Guidance for Cost Estimation and Management for Highway Projects During Planning, Programming, and Preconstruction" The National Cooperative Highway Research Program *(*NCHRP) Report 574, Transportation Research Board, Washington D.C.

3. ASCE (1972), "Manuals and Reports on Engineering Practice—No. 45 Consulting Engineering A guide for the Engagement of Engineering Services", New York.

4. ASCE (1981), "Manuals and Reports on Engineering Practice—No. 45 Consulting Engineering A guide for the Engagement of Engineering Services", New York.

5. ASCE (2003), "Engineering practice No. 45, How to Work Effectively with Consulting Engineers", New York

6. Berry, M. J.A, and Linoff G. (1997), "Data Mining Techniques: For marketing, Sales, and Customer Support", John Wiley & Sons, Inc. New York.

7. Carr, P. and Beyor, P. (2005). "Design Fees, the State of the Profession, and a Time for Corrective Action." Journal of Management Engineering, 21(3), 110–117.

8. Chang, Andrew Shing-Tao (2002), "Reasons for Cost and Schedule Increases for Engineering Design Projects". Journal of Management in Engineering, ASCE, January, 29-36.

9. Chou Jui-Sheng, Yang, I-Tung, and Chong, Wai Kiong (2009) "Probabilistic Simulation for Developing Likelihood Distribution of Engineering Project Cost" Elsevier B.V., Vol 18, 2009, pp. 570-577

10. Chen, Z. (2001). "Data mining and Uncertain Reasoning: an Integrated Approach", Wiley-Interscience, 1st Edition, New York.

11. Dysert, Larry R. (1997), "Scope Development Problems in Estimating", AACE International Transactions, Morgantown, WV

12. Gransberg, D.D., C. Lopez del Puerto and Humphrey, D. (2007) "Relating Cost Growth from the Initial Estimate Versus Design Fee for Transportation Projects," Journal of Construction Engineering and Management, ASCE, Vol. 133 (6), June 2007, pp. 404-408

13. Knight, K., and Fayek, A. R. (2002). "Use of Fuzzy Logic for Predicting Design Cost Overruns on Building Projects." Journal of Construction Engineering and Management, 128(6), 503-512.

14. Lafer, Gordon (2009) "The Hidden Cost of Contracting-out", Report for Analysis of Bridge Engineering Costs for the Oregon Department of Transportation.

15. Molenaar, K.R. (2005). "Programmatic Cost Risk Analysis for Highway Mega-Projects," ASCE Journal of Construction Engineering and Management, 131(3), 343-353.

16. Nassar, K. M., Hegab, M. Y., and Jack, N. W. (2005) "Design Cost Analysis of Transportation Projects" The proceedings of Construction Research Congress, San Diego, April 5-7, 2005, 949-956

17. Oberlender, Garold D. and Trost, Steven M. (2001) "Predicting Accuracy of Early Cost Estimates Based on Estimate Quality" Journal of Construction Engineering and Management, Vol 127, No. 3, May/June, 2001, pp. 173-182

18. Persad, K. R., O'Connor, J. T., and Varghese, K. (1995) "Forecasting Engineering Manpower requirements for Highway Preconstruction Activities" Journal of Management Engineering, ASCE, 11 (3), 41-47.

19. Saito Mitsuru, Sinha, Kumares C. and Anderson, Vigil L., "Statistical Models for the Estimation of Bridge Replacement Costs", Transportation Research Part A. General, Volume 25, Issue 6 pp. 339-350

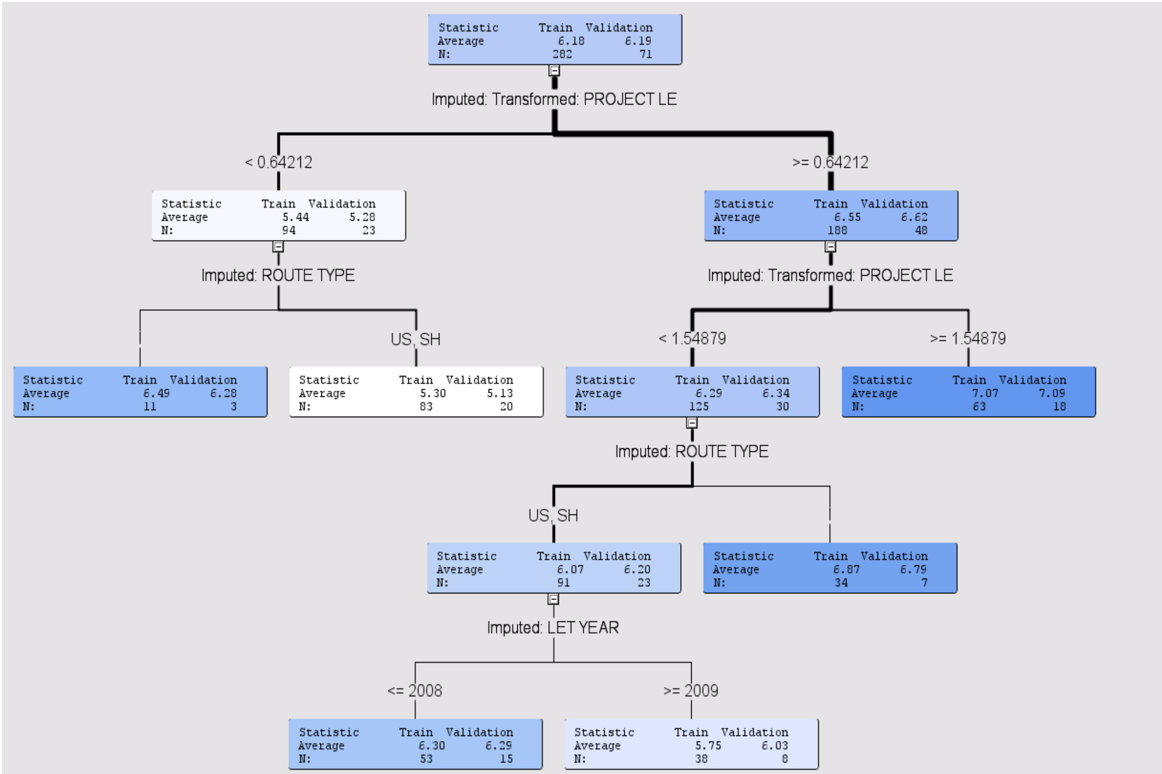
20. Schexnayder, C.J., Weber, S. L., and Fiori, C. (2003), "Project Cost Estimating: a synthesis of highway practice" National Cooperative research Program, Transportation Research Board

21. Trost, Steven M. and Oberlender, Garold D. (2003) "Predicting Accuracy of Early Cost Estimates using Factor Analysis and Multivariate Regression" Journal of Construction Engineering and Management, Vol 129, No. 2, April 1, 2003, pp. 198-204

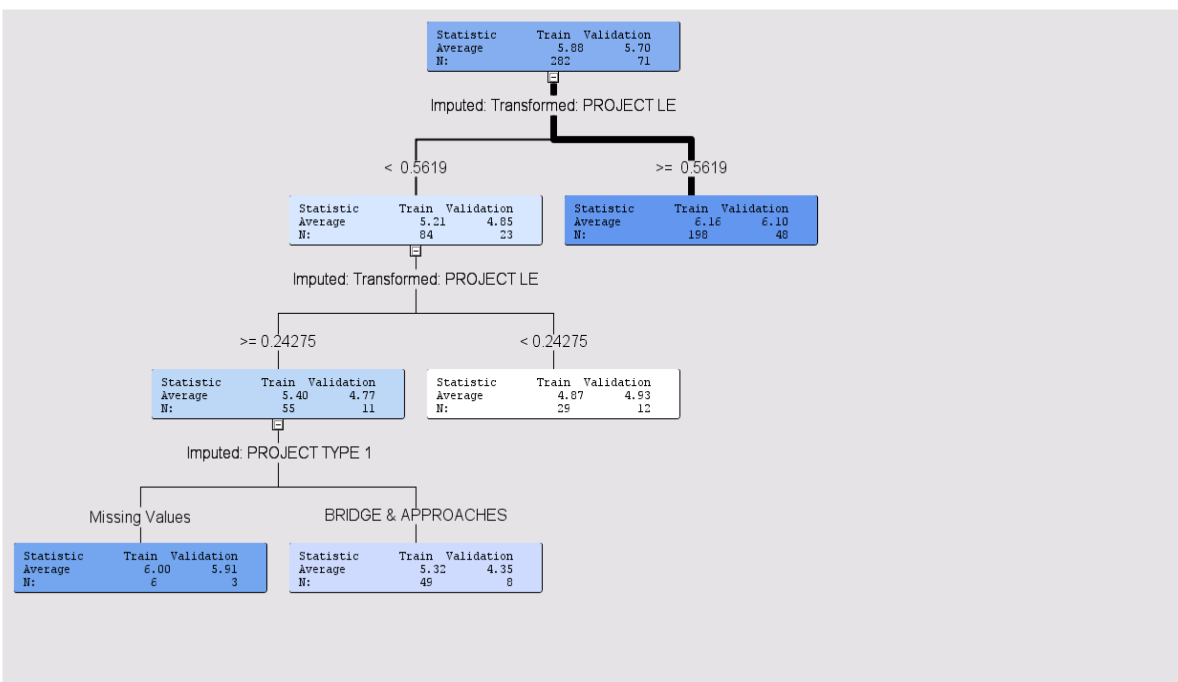
22. Wilmot, C.G., Deis, D.R., Schneider, H., and Coates, C.H. Jr. (1999) "In-House versus Consultant Design Cost in State Department of Transportation" Transportation Research Record 1654:153-160.

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

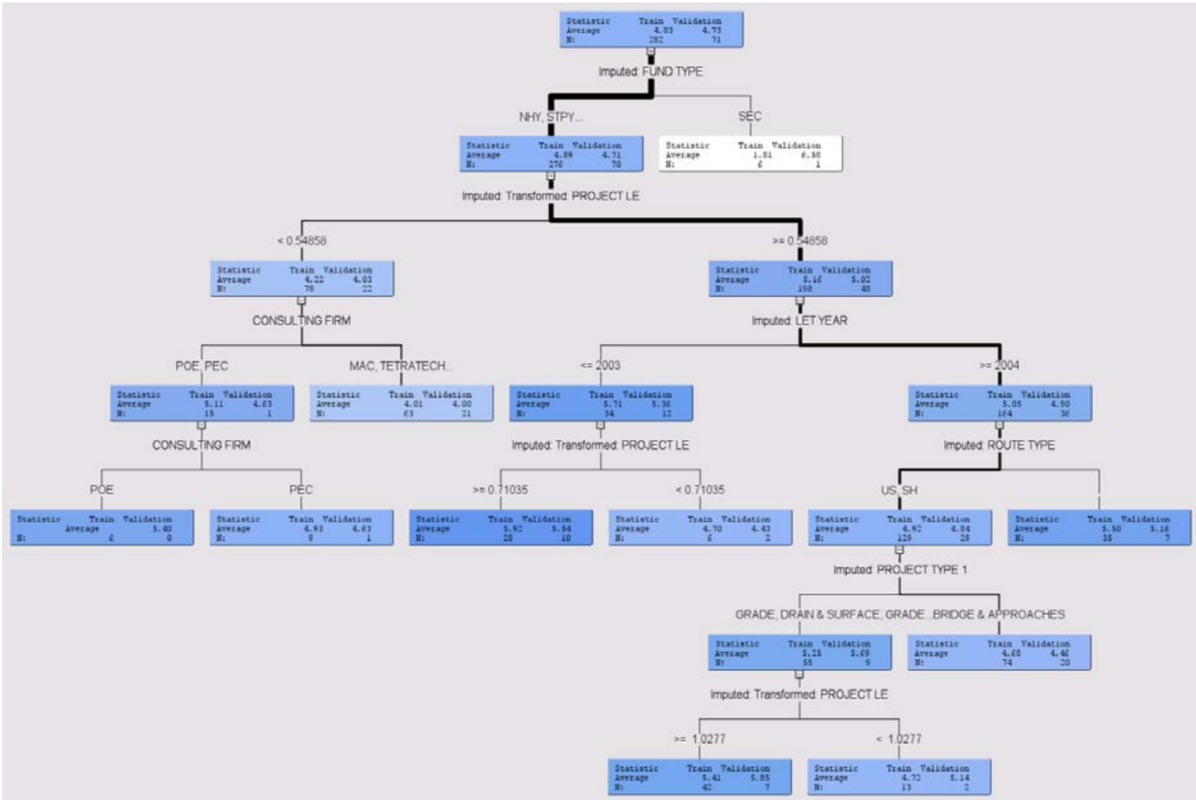
Appendix A – Decision Tree Models



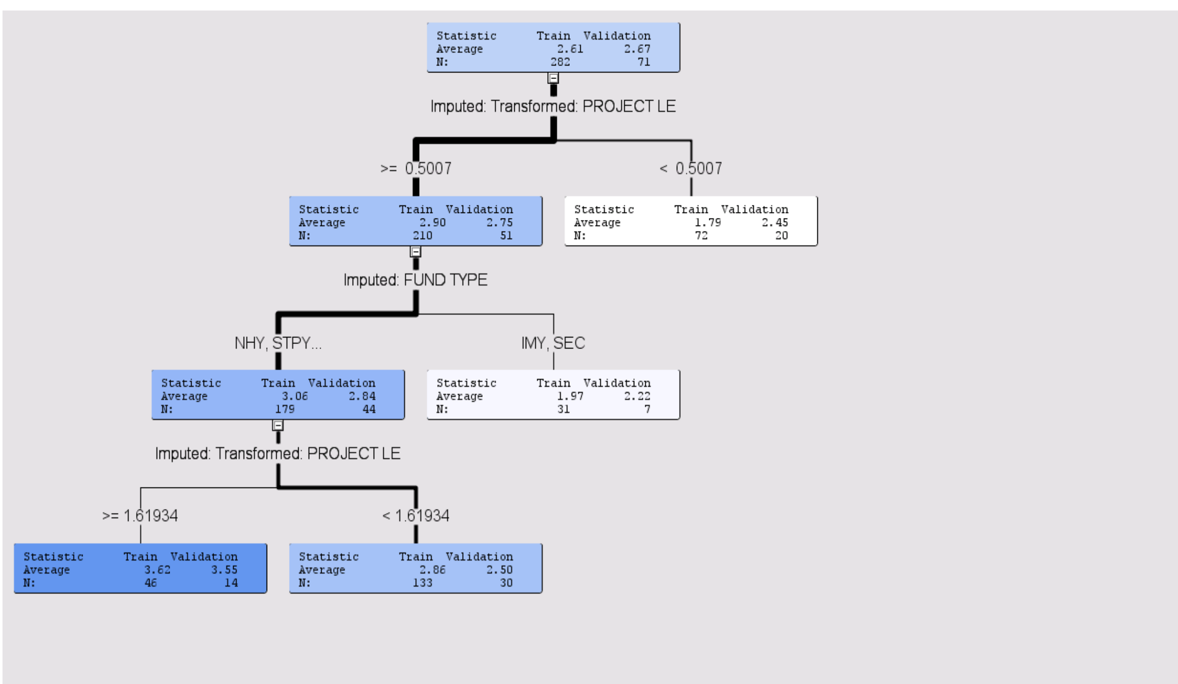
Decsion Tree Model for Plan & Profile Hours



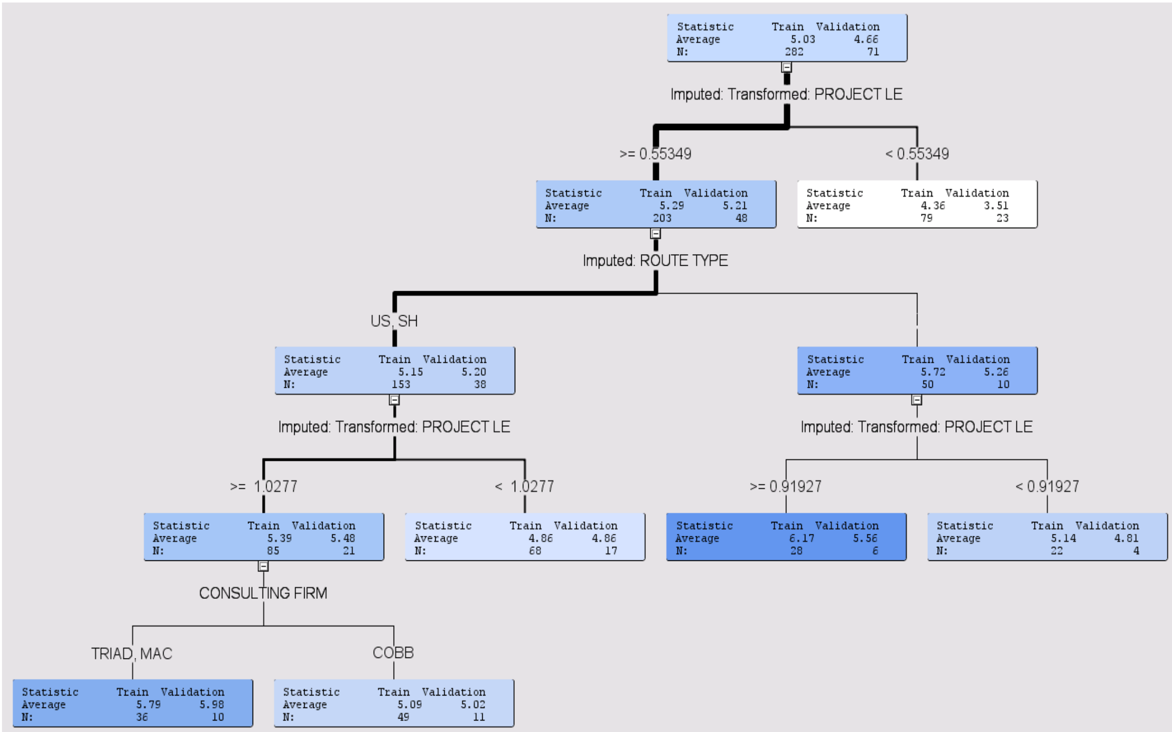
Decision Tree Model for Cross-Section Hours



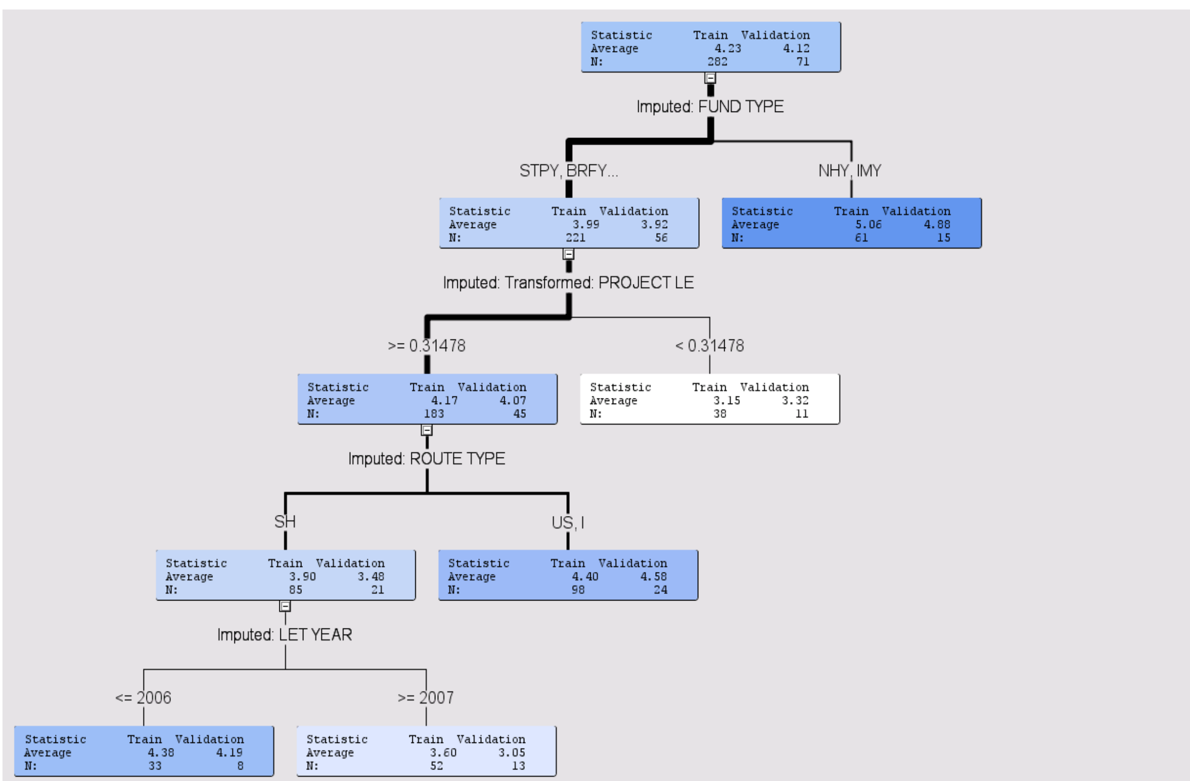
Decision Tree Model for Summarize Hours



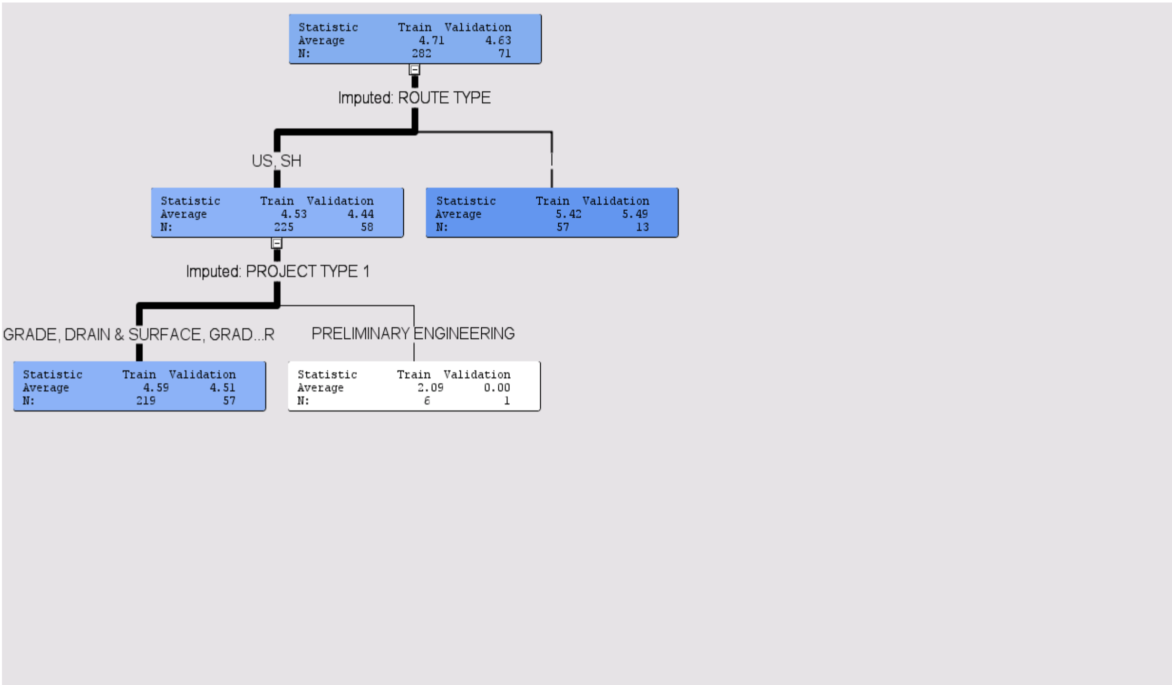
Decision Tree Model for Mass Diagram Hours



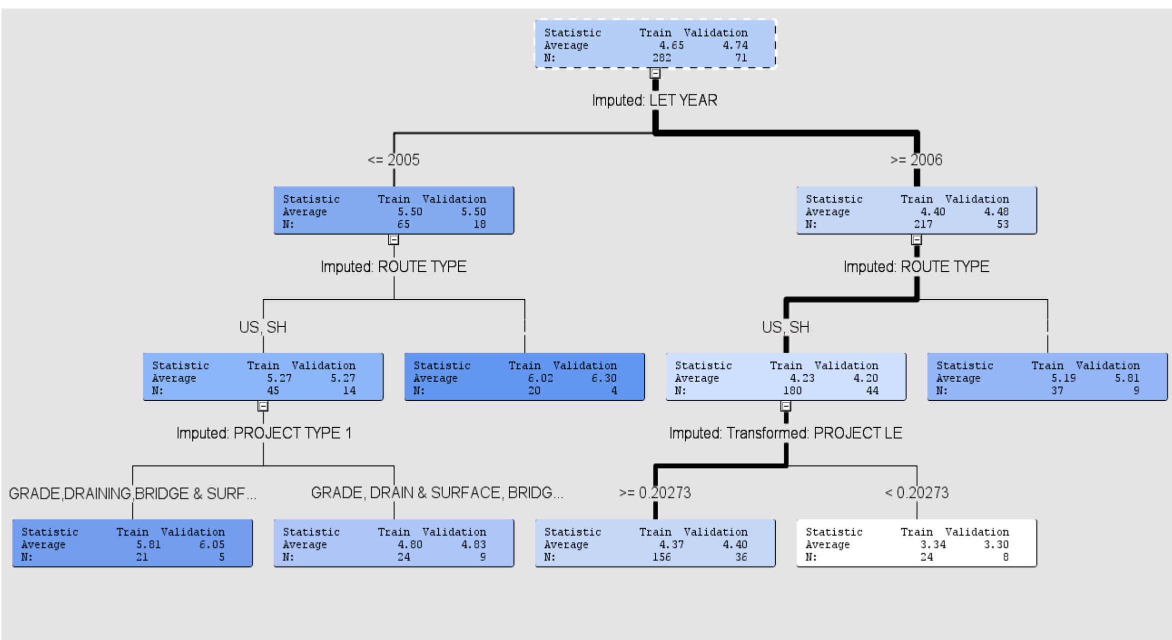
Decision Tree Model for Drainage Hours



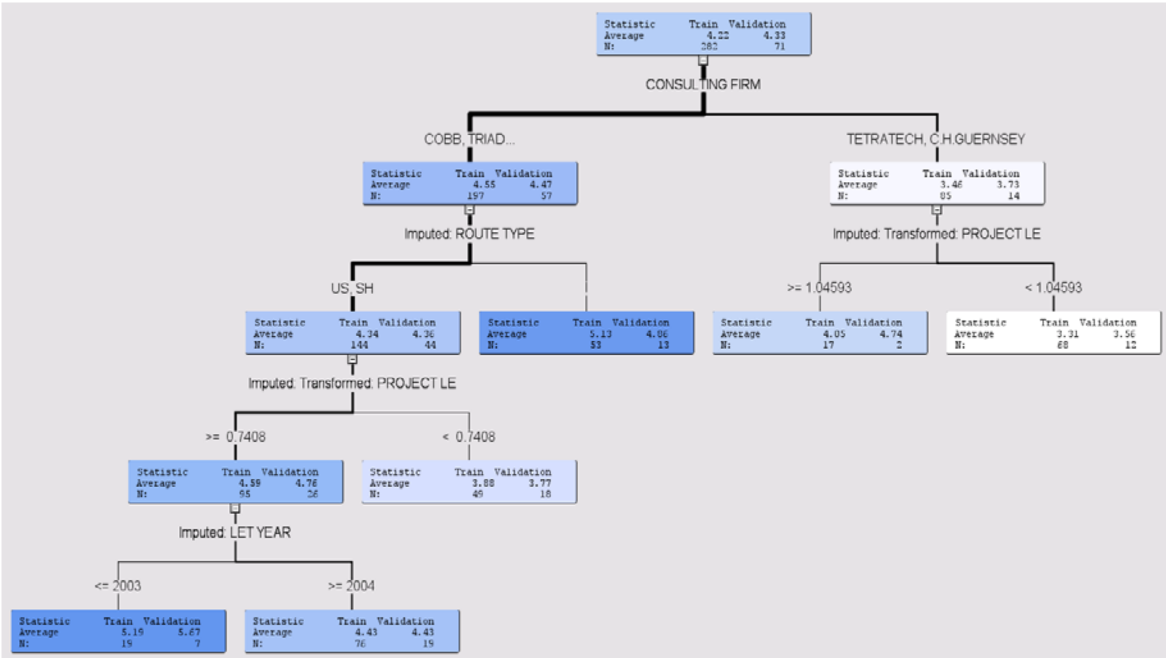
Decision Tree Model for Construction Sequence Hours



Decision Tree Model for Traffic Hours

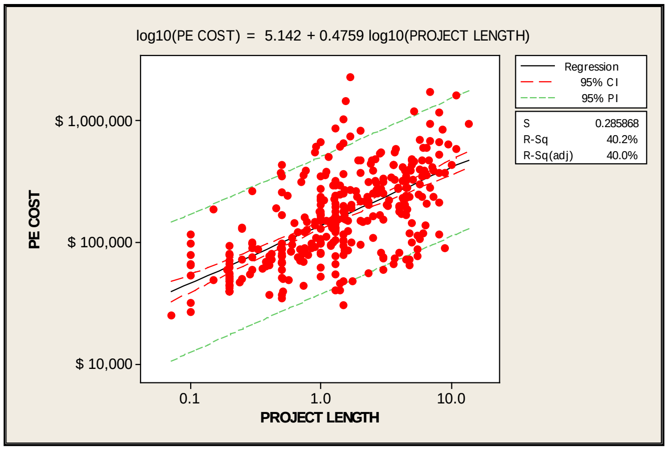


Decision Tree Model for Detail Sheet Hours

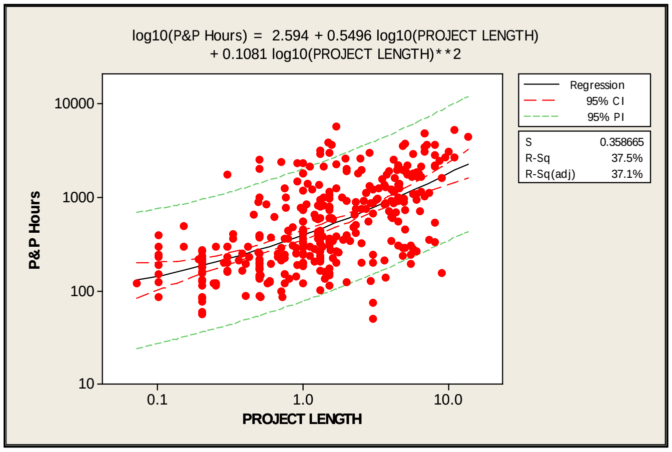


Decision Tree Model for Tilte Sheet Hours

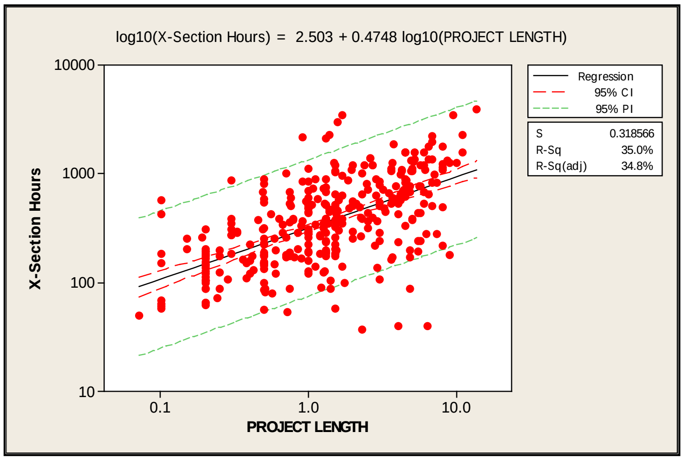
Appendix B - Regression Models



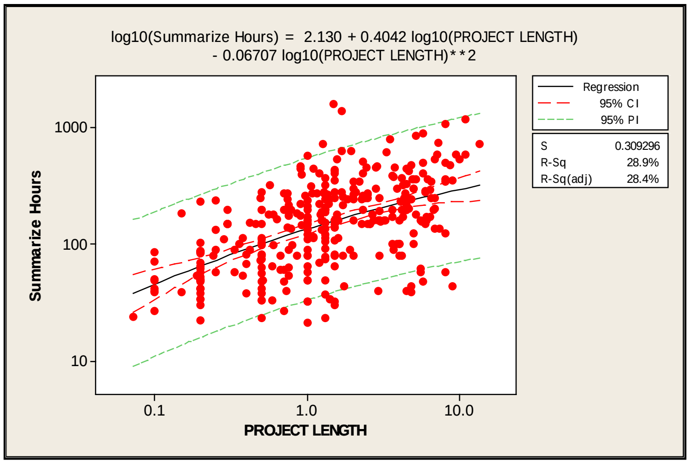
Fitted Line Plot of PE Cost



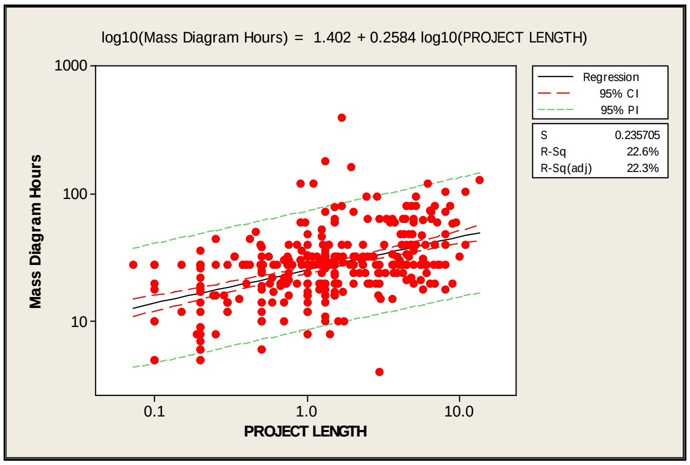
Fitted Line Plot of Plan & Profile Hours



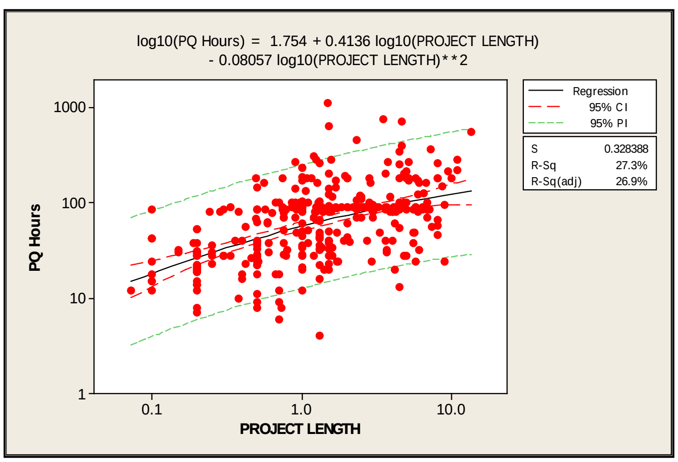
Fitted Line Plot of Cross-Section Hours



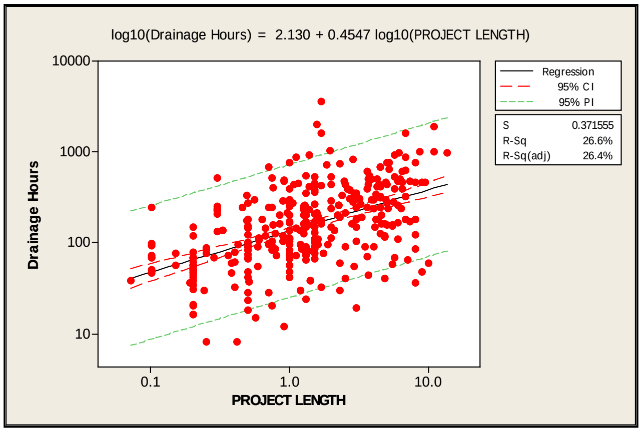
Fitted Line Plot of Summarize Hours



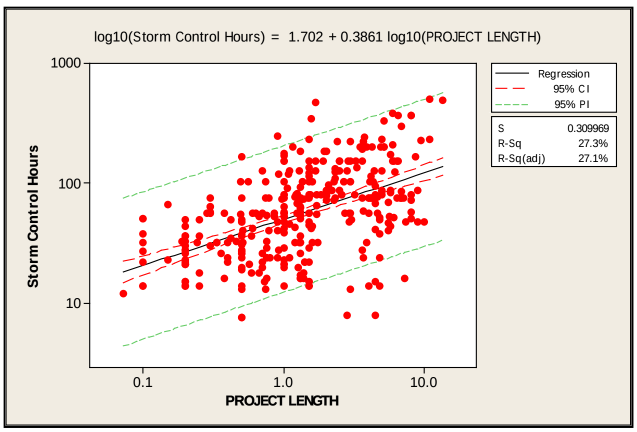
Fitted Line Plot of Mass Diagram Hours



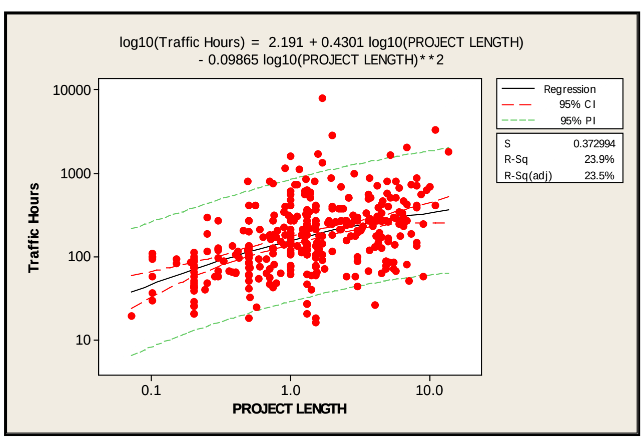
Fitted Line Plot of Pay-Item Quantities Hours



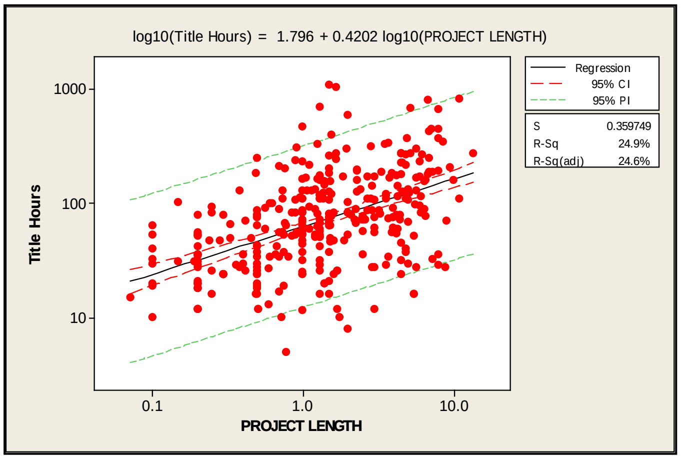
Fitted Line Plot of Drainage Hours



Fitted Line Plot of Construction Sequence Hours



Fitted Line Plot of Traffic Hours



Fitted Line Plot of Title Sheet Hours