Final Report

A Preliminary Cost Estimating Model for Transportation Projects

Sponsoring Agencies:

South Carolina Department of Transportation

Federal Highway Administration

Principal Investigators:

Kalyan R. Piratla, Ph.D.

Tuyen. Le, Ph.D.

Clemson University

January, 2024

Technical Report Documentation Page

1. Report No. FHWA-SC-24-01	2. Government Accession No.	3. Recipient's Catalog No.
4. Title and Subtitle A Preliminary Cost Estimating Model for Transportation Projects		5. Report Date *January, 2024* 6. Performing Organization Code
7. Author(s) *Kalyan R. Piratla, Tuyen Le, Md Shah Jamal, and Quan Do*		8. Performing Organization Report No.
9. Performing Organization Name and Address *Clemson University,* *109 Lowry Hall, Clemson SC 29634*		10. Work Unit No. (TRAIS) 11. Contract or Grant No. SPR No. 757
12. Sponsoring Agency Name and Address South Carolina Department of Transportation PO Box 191 Columbia, SC 29202-0191		13. Type of Report and Period Covered *Draft Final Report* 14. Sponsoring Agency Code
15. Supplementary Notes		

16. Abstract

As with any state highway agency (SHA), the construction needs exceed the budget limitations, which mandates that SCDOT prioritizes transportation project needs based on benefit/cost considerations. It is important to use every penny as wisely as possible. In such context, an early-stage cost estimate is imperative for evaluating the feasibility of large construction projects. There are currently no standard procedures implemented for developing planning-phase cost estimates across SCDOT offices. This research study developed a preliminary cost estimating tool (PCET) based on bid data gathered for over 325 past or ongoing transportation projects in South Carolina. The PCET tool enables planning personnel to rapidly develop both deterministic and probabilistic cost estimates. Three project categories were prioritized in this study, namely, widening, bridge replacement, and intersection improvement projects. Various project characteristics that transportation agencies would have data for in the planning phase are used as inputs to the PCET tool. The Microsoft Excel-based PCET tool is developed based on a combination of linear regression and machine learning algorithms. The PCET tool also includes an option to produce an early design stage cost estimate for the same three types of projects identified previously.

17. Key Word *Preliminary cost estimating, Microsoft excel-based cost estimating tool, Transportation projects*	18. Distribution Statement No restrictions.

19. Security Classif. (of this report) Unclassified.	20. Security Classif. (of this page) Unclassified.	21. No. of Pages *Total number of pages*	22. Price

Disclaimer

The contents of this report reflect the views of the author who is responsible for the facts and the accuracy of the data presented. The contents do not necessarily reflect the official views of the South Carolina Department of Transportation or Federal Highway Administration. This report does not constitute a standard, specification, or regulation.

The State of South Carolina and the United States Government do not endorse products or manufacturers. Trade or manufacturer's names appear herein solely because they are considered essential to the object of this report.

Acknowledgments

The project team acknowledges the support and guidance of the SCDOT's project steering committee led by Mr. Brad Latham. Other committee members include Mr. Craig Winn, Mr. Brian Dix, Mr. Tyke Redfearn, Mr. Hiram Sipes, and Ms. Shaquaisha Woods. We also acknowledge the input of representatives from various state departments to our survey conducted as part of this study. We thank Dr. Da Li and Ms. Sara Sadralashrafi for their contributions to this study. Finally, we also acknowledge the support of SCDOT's research staff Mr. Terry Swygert and Ms. Jade Watford.

Executive Summary

Construction needs far exceed the budget limitations of South Carolina Department of Transportation (SCDOT) like many other state highway agencies (SHAs). As a result, SCDOT is required to prioritize construction projects based on benefit to cost ratio. In this regard, early-stage cost estimates are significant for project feasibility. Planning phase is typically when these early cost estimates are developed to evaluate project feasibility. The challenge, however, is that no design detail is available at this stage, and the cost estimates would have to be based on broad project type, size, and location features. Sometimes, these estimates may need to be developed rapidly which is another challenge. While SCDOT currently develops and uses planning-stage cost estimates, they are not consistently done across the state and the procedure has not been recently evaluated. This research study developed a user-friendly preliminary cost estimating tool (PCET) for rapidly generating planning and early-design cost estimates for transportation projects.

The PCET tool is developed using linear regression and machine learning models to generate both deterministic and probabilistic cost estimates based on very few project features. Widening, bridge replacement, and intersection improvement projects are prioritized in this study. These prediction models are trained using bid data collected from over 380 past and current transportation projects managed by SCDOT. Project size is a key input, and it is characterized through length, number of (added) lanes, (added) shoulder width, and average side slope. Another key input is the year of letting which is significant because of the need to incorporate cost inflation. Using a linear regression model, the PCET tool can predict cost estimates with accuracy ranging from about 40% to 96% across the three project types. Further validation using new project data would increase confidence in the PCET tool and its utility for SCDOT.

Table of Contents

List of Figures

List of Tables

1. Introduction

The South Carolina Department of Transportation (SCDOT) is responsible for the systematic planning, construction, maintenance, and operation of the fourth largest state highway system in the U.S. (SCFOR, 2014). SCDOT invests hundreds of millions of dollars annually in maintenance, rehabilitation, and new construction of its statewide transportation infrastructure. As with any state highway agency (SHA), the construction needs exceed the budget limitations, which mandates that SCDOT prioritizes transportation project needs based on benefit/cost considerations. It is important to use every penny as wisely as possible. In such context, an early-stage cost estimate is imperative for evaluating the feasibility of large construction projects. Typically, earliest cost estimates are prepared during the planning phase of the project development when minimal scope and design details are available. It is a great challenge to prepare a cost estimate without many design details and it would naturally be somewhat inaccurate. The cost estimate is bound to change as the design gets completed and the construction timeline becomes more tangible. SHAs typically build a significant amount of contingency in these early cost estimates as a percentage of base estimate cost to account for the vast number of unknowns.

Nevertheless, these early estimates are important and often inform the project budgets that need to be maintained throughout the project development phase. If the project is considerably underestimated, there would be significant cost overruns in the later phases of the project development thereby delaying or limiting the investment into other prioritized projects. It is also possible that the project may lose support for further advancement at the stage of design completion due to lower benefit to cost ratio. It is however not uncommon for SHAs to underestimate project costs in the planning phase to keep the project alive, and this phenomenon is often referred as "optimism bias" (Jennings, 2012). In fact, it is statistically proven that pre-design cost estimates are deliberately low, and that this has led to 9 out of 10 projects ultimately having cost overruns (Gardener et al., 2017). Supporting this claim, another study reported that the final construction costs were 46% higher than the estimated costs at the time of programming based on analysis of data from Montana DOT (Alavi and Tavares, 2009). On the other hand, if the project is considerably overestimated, it would prevent the precious federal-aid money from being allocated to other timely needed high-priority projects (FHWA, 2015). It is therefore imperative to use a sophisticated estimating approach to systematically develop conceptual or preliminary estimates during the planning phase by rationally assigning contingency costs to account for the unknowns (Anderson et al., 2007). Furthermore, the time taken for many projects to mature from the planning phase through the letting phase can be multiple years, and this warrants the consideration of the risks associated with differed market conditions and price inflation.

SCDOT does not currently have an established agency-wide procedure for rapidly developing preliminary cost estimates for the variety of construction projects they plan and manage. There are also no agency-wide policies to account for differed market conditions and price inflation while preparing cost estimates. Addressing these gaps, this project developed a statistical cost estimating tool, namely preliminary cost estimating tool (PCET), to be useful in the planning and early design phases of transportation projects. The PCET tool is developed based on the empirical bid databases maintained by SCDOT in conjunction with accounting for cost inflation modeled using a state-level highway construction cost index (HCCI). The developed cost estimating tool will predict the range of total project cost based on a broadly defined scope with limited project characteristics identified. A survey of SHAs is also undertaken as part of this research study to support the cost estimating model development effort. The cost estimating tool is created in such a way that it can be easily updated to include data from additional projects in the future. The user-friendly tool developed in this project will enable SCDOT in rapidly developing a range of preliminary cost estimates with associated probabilities based on few project inputs. This ability will enable SCDOT to quickly respond to feasibility questions on large projects that may be considered for funding. The vision is to host the developed user-friendly tool on the SCDOT's Preconstruction Support web page for the project managers to use.

1.1 Research Objectives

The overarching goal of this project is to provide technical guidance to SCDOT in rapidly performing preliminary cost estimates of transportation projects with minimal design completed through a user-friendly computer program. The accuracy of the cost estimate is expected to improve with more design and scope data considered, and this accuracy would be reflected in the computer program through sophisticated statistical measures for SCDOT to be fully aware of the risk involved in using the preliminary cost estimates. The following are the specific objectives of this project:

1. Synthesize current state-of-the-art and state-of-the-practice in developing preliminary cost estimates for transportation projects.
2. Identify the highway construction cost index (HCCI) in South Carolina and demonstrate its use in accommodating cost inflation in transportation project estimates.
3. Develop, demonstrate, and validate a risk-based cost estimating approach for transportation projects that can be used in planning and preliminary design phases of the project development.
4. Develop a user-friendly Microsoft Excel-based computer program that will assist SCDOT personnel in generating rapid preliminary cost estimate ranges along with confidence levels using minimal design details.

1.2 Study Methodology

To accomplish the research objectives identified previously, six distinct tasks were conducted as illustrated in Figure 1. In Task-1, an extensive literature review was conducted to identify current state-of-the-art strategies for cost estimating focused on both preliminary and early-design estimating. In Task-2, several state DOTs in the U.S. were surveyed on their practices with respect to the type of approaches and tools used for developing preliminary cost estimates. In Task-3, historical bid data was collected for several past and ongoing construction projects of three types (i.e., widening, bridge replacement, and intersection improvements) to develop a comprehensive database that was later used for statistical analysis in the next task. In Task-4, data collected in Task-3 was analyzed using state-of-the-art statistical techniques in conjunction with the findings of the survey conducted in Task-2 to develop insights into generating cost estimates with minimal design detail and maximum accuracy possible. In Task-5, a computational tool was developed which embedded the statistical models from Task-4. The final task (Task-6) entailed preparation and submission of the final report that describes the study objectives, methodology and results, and highlights specific recommendations to SCDOT. Figure 1 highlights the value produced through all these different tasks in this study.

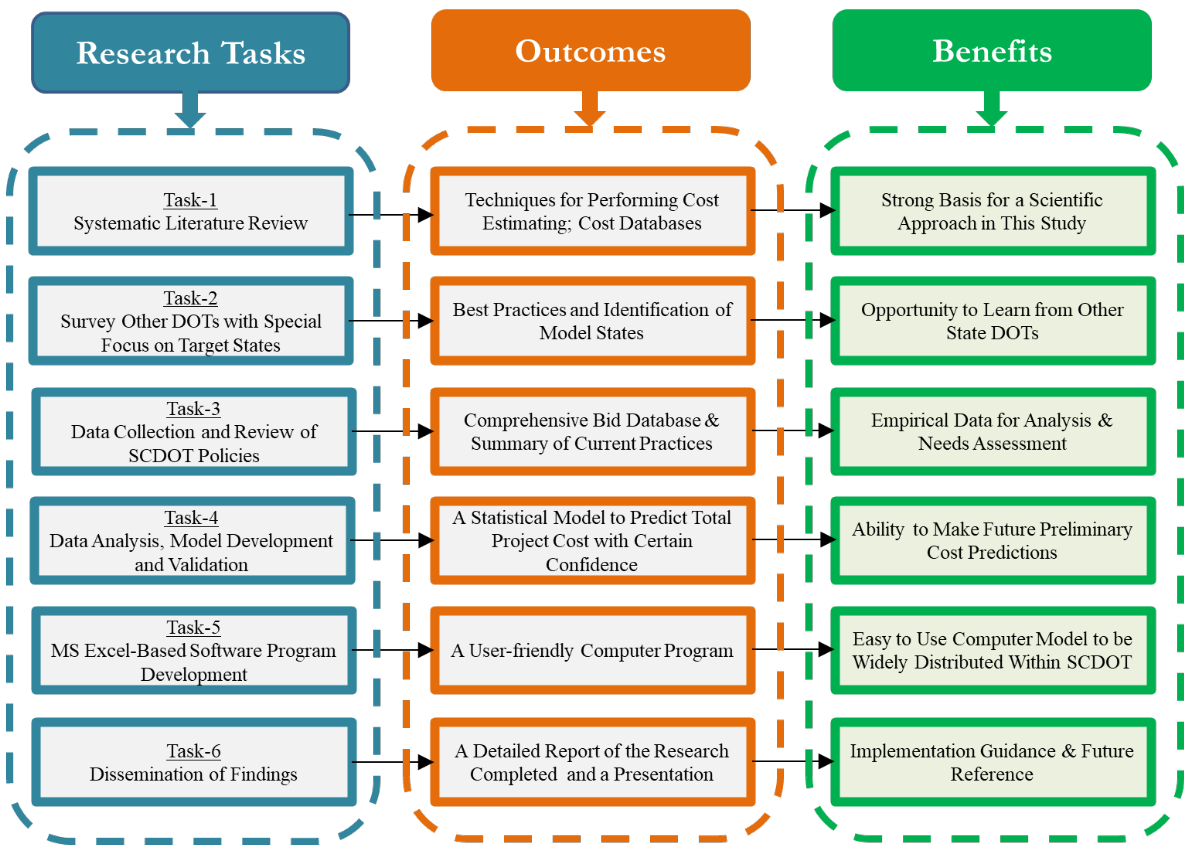


Figure 1. *Value Offered Through the Proposed Study*

1.3 Significance of this Research Study

This study explored multiple cost estimating approaches including a systematic risk-based approach for SCDOT to use. Although many studies were previously completed on this topic, those cost estimating models would not be readily suitable for SCDOT because the data used from different agencies may not have captured the unique challenges and policies implemented by SCDOT. Furthermore, the prevailing state of practice across state highway agencies are somewhat irrational in terms of assigning contingency costs in estimates and building false hopes with intentionally underestimated costs in the planning and scoping stages of project development. A more rational approach would be to assign contingency costs commensurate with the risk involved. Furthermore, adjustment of historic prices and accounting for inflation are loosely practiced across state highway agencies, and inconsistently implemented at SCDOT. This study presented a more systematic approach to account for inflation. Intellectually, the cost estimating approaches explored are sophisticated with potential for advancement of knowledge that would be beneficial to state highway agencies beyond SCDOT. The findings and the developed cost estimating tool will help SCDOT in rapidly producing preliminary cost estimates that are probabilistic and more accurate through a customized user-friendly computer tool.

2. Literature Review

A comprehensive review of literature was completed on the topic of preliminary cost estimating for transportation projects. Table 1 presents the list of most relevant studies we reviewed before the project work began. In addition, we have also reviewed cost estimating manuals of a few state DOTs including Connecticut, Montana, New Jersey, Washington, and Nevada. Additionally, we have reviewed a few graduate student theses and many conference papers.

Table 1. *List of Most Relevant Studies Reviewed*

Citation	Literature Type	Theme/Focus
Alavi and Tavares (2009)	Project Report (Montana DOT)	Recommendations for Highway Project Cost Estimation Approaches
Anderson et al. (2009)	Project Report (Texas DOT)	Construction Unit Cost Development
Van Dyke et al. (2017)	Project Report (Kentucky)	Review of Initial Project Estimates
Gransberg et al. (2017)	Project Report (Montana DOT)	Top-Down Cost Estimating Using Artificial Neural Networks
Liu et al. (2011)	Project Report (NCDOT)	Estimating Preliminary Engineering Costs
Anderson et al. (2007)	NCHRP Report	Fundamental Estimating Guidance
Paulson et al. (2008)	Project Report (NCHRP-Funded)	Cost Estimate Management Process Improvements
Skolnik (2011)	Project Report (NCHRP-Funded)	Price Indexing in Transportation Contracts
Pirece et al. (2012)	Project Report (SCDOT)	Price Indexing and Cost Adjustments Clauses
Turochy et al. (2001)	Project Report (Virginia)	Planning Stage Cost Estimating
AASHTO (2013)	Guidebook	Cost Estimating Guidebook
Adel et al. (2016)	Journal Paper	Parametric Cost Estimating
Asmar et al. (2011)	Journal Paper	PERT-like Cost Estimating Model
Bell and Kaminsky (1987)	Journal Paper	Database-Driven Cost Estimating
Chou (2009)	Journal Paper	Linearized Cost Estimating Model
Chou and O'Connor (2007)	Journal Paper	Internet-Based Highway Cost Database
Chou et al. (2006)	Journal Paper	Quantity-Based Estimating Approach
Fragkakis et al. (2010)	Journal Paper	Estimating Using Regression and Bootstrapping
Gardener et al. (2016)	Journal Paper	Reducing Data Collection Efforts for Conceptual Estimates
Gardener et al. (2017)	Journal Paper	ANNs with Bootstrap Sampling
Harper et al. (2014)	Journal Paper	Performance Measures for Cost Estimating
Hollar et al. (2013)	Journal Paper	Estimating Preliminary Engineering Costs for Bridges
Karaca et al. (2020)	Journal Paper	Improved Accuracy of Preliminary Estimates
Liu et al. (2013)	Journal Paper	Estimating Preliminary Engineering Costs
Petroutsatou et al. (2012)	Journal Paper	Tunnel Project Cost Estimation Using Neural Networks
Shane et al. (2009)	Journal Paper	Cost Escalation Factors

2.1 Literature Background and Overview

A cost estimate is the probable cost of a construction project. It serves two main purposes: (1) determine the probable project cost to evaluate feasibility and allocate funds, and (2) control the budget as the project is developed further and gets to letting. Three different types of estimates are typically developed and used in the context of construction projects. The first type is conceptual or preliminary estimate which is prepared with minimal design detail during the planning phase to serve the purposes of feasibility evaluation and rough budget establishment, as mentioned in Table 2. These estimates are expected to be less accurate. Most agencies use simple Microsoft Excel spreadsheets for preparing planning level estimates. Parametric estimates based on $/SF or $/lane mile are commonly employed. Examples of computer programs and spreadsheets developed for planning level estimates include the VDOT's Planning Cost Estimate Spreadsheet, Comparative Bridge Costs of CALTRANS, and Concept Cost Estimate Form of UDOT (Anderson et al., 2009). The preliminary estimates are especially challenging to develop given the vast number of unknowns in the earlier project stages. This difficulty is demonstrated in the AASHTO's *Practical Guide to Cost Estimating* (2013) manual, which provides a classification of estimates used in different transportation project development phases and suggests acceptable accuracies for the different estimates. Table 2 presents a detailed classification of cost estimates. As can be observed from Table 2, the acceptable estimate accuracy in the planning phase is -40% to +100% at best from the initial cost estimate to the final construction cost. Byrnes (2002) reported that SHAs add a contingency ranging from 5-45% depending on project type and uncertainty; similar contingency factors were also reported by Turochy et al. (2001).

The second type is design estimate which is prepared as the design is developed to ensure the project remains within the initially established budget. Design estimates during the scoping phase (refer to Table 2) are used to set the baseline costs and program the project. The design process is typically iterative and cost estimates are important criteria during the design. The design estimates continue to be used in the preliminary design and final design phases of the project, as can be noted from Table 2. Design estimates are more accurate than preliminary estimates as more design detail is available and accounted for in it. Parametric estimating and historic bid-price based estimating are common approaches for design estimating. Some agencies, however, use cost-based estimating using historic production rates, material, labor and equipment cost data for the critical (about 20%) pay items while using historic bid-based approach for the remaining (about 80%) pay items. Many agencies use sophisticated computer programs such as Estimator or Cost Estimating System (CES) for performing estimates during the design phase. An engineer's estimate is prepared during the final plans, specifications, and estimate (PS&E) phase. The engineer's estimate is crucial for committing the funds, inviting, and evaluating contractor bids. The design is completed at this stage and the engineer's estimate is expected to account for each cost aspect of the project.

Table 2. *AASHTO's Recommended Cost Estimate Classification (AASHTO, 2013)*

Development Phase	Scope/Design Completion	Estimate Purpose	Popular Methodology	Estimate Accuracy
Planning	0-2%	Preliminary Estimate: Estimate Potential Funds Needed (20-year plan)	Parametric Estimating (Range)	-50% to +200%
	1-15%	Preliminary Estimate: Prioritize Needs for Long Range Plans (10-year plan)	Parametric or Historic Bid-Based (Range)	-40% to +100%
Scoping	10-30%	Design Estimating: Establish a Project Baseline Cost	Historic Bid-Based or Cost-Based (Range)	-30% to +50%
Preliminary Design	30-90%	Design Estimating: Manage Budgets Against Baseline	Historic Bid-Based or Cost-Based (Smaller Range or Point Estimate)	-10% to +25%
Final Design (PS&E)	90-100%	PS&E Estimating: Bid Evaluation and Funds Allocation	Cost-Based or Historic Bid-Based (Point Estimate)	-5% to +10%

The third type of estimate is the detailed estimate which is typically developed by contractors after thoroughly considering the constructability aspects of the designed project. Detailed estimates are used for bidding on projects, and it is not uncommon to see deviation between detailed estimates and engineer’s estimates. Detailed estimates are typically prepared based on the quantities of different work items, anticipated production rates, labor, material and equipment costs, and it is important for the estimator to be knowledgeable about the construction process to prepare an accurate detailed estimate. Some design estimates may also be prepared in the same manner as the detailed estimates considering production rates and constructability aspects.

As one would expect, the contingency costs which account for the unknowns diminish as the project scope and design details become available in the later phases of the project development. Figure 2 illustrates how the base estimate grows and contingency costs diminish as the design details become available through the later project development phases.

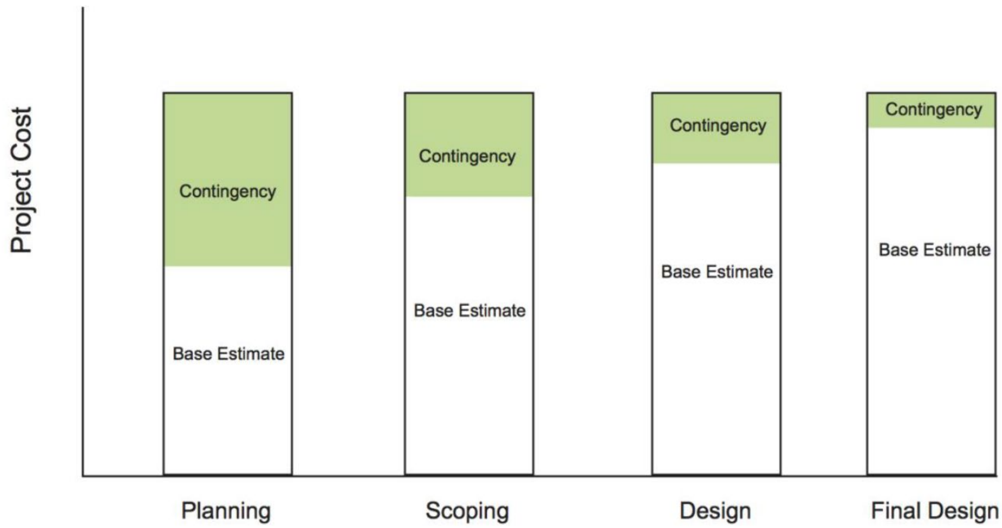


Figure 2. *Application of Contingency in Cost Estimates Over the Project Development Period (Adopted from Van Dyke et al., 2017)*

Two crucial requirements of developing cost estimates are historic bid prices and prevailing market conditions (Anderson et al., 2007). Historic data is often characterized in two ways: (a) unit price data – which is the historic unit price data for pay items most relevant to the prospective project; and (b) cost-based data – which includes production rates, crew sizes, material, labor, equipment and contractor markup costs. The cost-based approach is used for detailed estimating, as described previously, and it requires the conceptualization of the entire construction process. While the cost-based approach may be more accurate as it closely resembles the contractor's estimating approach, it is complicated, time-consuming, detail-oriented, and requires extensive construction knowledge. The historic unit price-based approach is most often used, especially in the earlier phases (scoping and preliminary design) of project development, as can be seen from Table 2. In this approach, simple unit price averages or weighted (by quantity) averages from past data are used as unit prices for future project for similar bid items. The historic unit price-based approach is proven to work best when multiple (three to be specific) lowest bids are considered for each past project as opposed to a single lowest bid (Schexnayder et al., 2003). Furthermore, parametric estimates which are simply based on itemized cost or entire project cost on per lane mile (or per sq. ft. where appropriate) of work basis are more commonly used in the planning phase estimates. The historic unit price data needs to be adjusted to suit project characteristics (i.e., complexity, region/location, size, etc.) and market conditions (i.e., bidding environment, economic situation, and inflation). Many agencies do not have formal guidelines for how to make these adjustments and it is often left to the individual estimator's engineering judgement (Anderson et al., 2009). Some SHAs developed own highway construction cost indices to track and account for inflation following Federal Highway Administration (FHWA)'s guidance.

Highway Construction Cost Indices (HCCI) are a critical measure of the purchasing power of road-building resources for highway agencies (Guerrero 2003; White and Erickson 2011). It represents actual contract bids made by contractors within a certain time-period (i.e., yearly) calculated as a function of unit bid prices and quantities of various bid items used in highway construction. Since HCCI is computed using contract bid prices rather than contract completion, it does not include escalated cost overrun due to unexpected seasonal events (e.g., flooding). Thus, it is considered a better tool to measure the overall construction market conditions. For this reason, state departments of transportation (DOTs) have widely used it to monitor the inflation in highway construction and reasonably forecast the preliminary expenditure need for a highway project (White and Erickson 2011; Guerrero 2003).

The concept of HCCI was first introduced by the FHWA in 1933. The index was originally named the Bid Price Index (BPI) which was later replaced with the term National HCCI (NHCCI) in 1991 (Whited and Alsamadani 2011). Subsequently, some DOTs have adopted FHWA's methodology to develop their state-level HCCIs (Wilmot and Cheng 2003). Currently, at least 21 DOTs compute their state-level HCCIs, but most of the current HCCI calculation methods adopted by DOTs are not sophisticated enough to assure that an HCCI can be used as a reliable indicator of the changing market conditions (Shrestha et al., 2016). One of the reasons is the use of a significantly insufficient sample size of bid items in HCCI calculation. Currently, State DOTs use as little as 14% to below 50% of the total construction bid prices to calculate their state-level HCCI (Shrestha et al., 2016). Moreover, current methodologies typically produce only one overall HCCI as a representative index to indicate the entire state's highway construction market condition. However, highway construction costs are heavily affected by the availability of local materials, equipment, and even specialty contractors. The project size and quantity of work would also significantly affect construction methods and productivity which are directly associated with project costs. The unique characteristics of highway construction and business environments in South Carolina require the development of a customized HCCI that correctly represents the market conditions and trends based on the state's local regions, project sizes, and project types. This study addressed this need by developing an advanced technique for determining HCCIs for different major types of highway construction projects and regions in South Carolina.

Finally, most state agencies use point preliminary estimates with a contingency assigned as a percentage of project cost. The point estimates result in a single cost value to be used for decision-making. While it is commonly known that the accuracy of these single value estimates in the planning and scoping phases is not great, it may lead to false sense of confidence among some project stakeholders as it does not indicate a confidence measure nor does it indicate the potential for cost growth (Garderner et al., 2017; AASTHO, 2013; Chelst and Canbolt, 2012). One concern is the lack of a rational approach for assigning the contingency costs dependent on the risk involved. A risk-based cost estimating approach would address this concern. Providing an estimate range is believed to capture the realistic possibility of variable cost as dependent on the unknowns (Gardener et al., 2017). Indeed, FHWA (2007) in their cost estimating guidance allows SHAs to express their preliminary estimates as a range with indicated confidence levels. In 2005, Molenaar developed a stochastic cost estimating approach for Washington DOT for projects costing over $100 million, and the SHA has been successfully using that approach since then. Molenaar opined that the risk-based stochastic method better conveyed uncertain nature of project costs at the planning level (Molenaar, 2005). Gardener et al. (2017) developed a bootstrap sampling based stochastic cost estimating approach based on historic bid price data where planning level estimates are presented as a range along with probability values.

2.2 Synthesis of Specific Studies

Numerous studies emphasize the significance of accurate cost estimation in transportation projects.

The research of Alavi & Tavares, (2009) addresses the prevalent issue of cost overruns in transportation infrastructure projects. The literature review highlights how cost overruns have a wide range of effects, such as modifications to project schedules, a narrowed scope, longer construction times, and a decrease in public trust. The research examines effective methods employed by other organizations and suggests improvements for the Montana Department of Transportation (MDT). These include creating a section dedicated to cost estimation, updating unit cost data regularly, creating a thorough manual, and implementing quality control and risk management programs. Future studies might investigate the implementation challenges and practical difficulties that may develop when implementing these strategies; however, a thorough examination of any potential constraints linked to the suggested approaches still needs to be completed in this report. Furthermore, investigating the recommended strategies' long-term efficacy would advance the current understanding of how they mitigate cost overruns in transportation projects (Alavi & Tavares, 2009).

The study by Anderson et al., (2009) investigates whether state highway agencies (SHAs) set project unit costs, including construction and maintenance. The study employs interviews with SHAs and a thorough online survey to determine standard practices. It demonstrates the need for established documented procedures for adjusting unit costs in response to project characteristics and market conditions. The short-term recommendations involve utilizing instruments like an estimator and a site manager database and considering cost-based estimating for specific items (S. Anderson et al., 2009). However, long-term recommendations include developing guidelines for adjusting unit prices and evaluating cost-based estimating for project phases. A drawback noted in the study is that the recommendations are based on observations and characteristics among SHAs; as such, it emphasizes the need for more comprehensive and systematic processes in unit cost generation across agencies. The report also recommends considering the work necessary to implement cost-based estimation into practice and looking into alternative information systems for adequate access to unit costs, which emphasizes the need for more research on the real-world challenges and system integration difficulties related to these recommendations (S. Anderson et al., 2009).

With a specific focus on the Kentucky Transportation Cabinet (KYTC), the article of Van Dyke et al., (2017) offers a thorough analysis of the methods and procedures employed by SHA for project cost estimation. SHAs usually employ a stepped approach to estimating, beginning with high-level programming, and ending with precise post-approval estimates. The article highlights the significance of straightforward, consistent procedures frequently made possible by proprietary or commercial software to ensure correctness and promote quick learning for new employees. Precise estimations require the incorporation of project-specific contingencies that consider environmental and geographical considerations. However, the paper acknowledges many difficulties, including the complicated nature of the estimation procedure, the requirement for multidisciplinary cooperation, and the dependence on historical data. Potential cost error, particularly in time-constrained scenarios, and inadequate access to and storage of previous data for more accurate projections are among the limitations. There may be discrepancies in how procedures are used, as evidenced by the different estimation techniques used by KYTC districts. The end of the paper emphasizes the significance of a systematic and transparent approach to estimating for successful project delivery (Van Dyke et al., 2017).

The article of Gransberg et al., (2017) thoroughly investigates how top-down estimating techniques—more significantly, using multiple regression models and artificial neural networks (ANN)—may improve the precision of building cost estimates. The study compares the predicted accuracy of the proposed methods with the agency's current practices. It offers a logical method for variable selection to address the challenges the MDT faces in early estimating and budgeting. Significant increases in the accuracy of predictions are found in the study, especially for new construction and bridge replacement projects, indicating advantages for more effective agency funding allocation. The supplied excel spreadsheet tool makes it easier to put the suggested approach into practice by providing easily accessible cost projections at the budgeting stage. Limitations may arise in the generalizability of the findings to diverse transportation agencies and project types, and the article suggests future research areas, such as integrating early project-level data and tailoring estimating systems for increased efficiency (Gransberg et al., 2017).

The research of Liu et al., (2011) addresses preliminary engineering (PE) cost and schedule estimation for transportation projects, emphasizing the North Carolina Department of Transportation (NCDOT). The study develops statistical models employing multiple linear regression, hierarchical linear models, Dirichlet process linear models, and multilevel Dirichlet process linear models by evaluating data from bridge and roadway projects. The research shows that project characteristics, such as location, project scope, and expected construction costs, impact how accurately PE cost ratios predict outcomes. The study contributes by offering an excel-based user interface for precise and effective PE cost estimations. Limitations include difficulties obtaining data, especially for projects involving roads, and difficulty interpreting outcomes from multilayer models (Liu et al., 2011).

According to the manual of Anderson et al., (2007), which overviews eight worldwide strategies and guidelines for managing and applying cost estimating. Although the article offers insightful information, it needs case studies or empirical data to support the suggested solutions' actual application in real-world situations. The focus on highway projects may limit generalizability to other construction environments, and the issues revealed call for additional investigation into mitigation techniques. The report should cover the economic and organizational consequences of implementing the suggested adjustments in greater detail (S. D. Anderson et al., 2007).

Comprehensive guidelines for improving cost estimating methods in SHA illustrated by Paulsen et al., (2008). The authors provide valuable guidance for the planning, scoping, and design stages, stressing the value of regular evaluations of estimates, cooperation with support offices, and comprehensive site assessments. Nevertheless, the article would benefit from case studies that show successful application and greater empirical assurance of the recommended guidelines. The organizational and economic consequences of the proposed changes that limit the article's depth also require controversy. The difficulties of incorporating these recommendations into SHA procedures should be investigated in future studies, considering institutional resistance and resource limitations (Paulsen et al., 2008).

A comprehensive study on "Price Indexing in Transportation Construction Contracts" by Skolnik, (2011) investigates how SHAs and contractors currently use and view Price Adjustment Clauses (PACs). According to the report, PACs are widely used, mainly for fuel and liquid asphalt, and are thought to have advantages like more bidders and stable markets. Nonetheless, the research recognizes certain obstacles, such as the burden of administrative work and resistance from contractors. Utilizing statistical models, the investigation evaluates the effect of PACs on bid prices, yielding findings that could be more compelling. Despite constraints on extrapolating outcomes and difficulties in measuring PAC efficacy, the research suggests utilizing PAC to improve estimation accuracy and mitigate industry hazards (Skolnik, 2011).

Unit price adjustment clauses (PACs) for construction materials were the subject of a thorough investigation by Pierce et al., (2012), who concentrated on the South Carolina Department of Transportation (SCDOT). In addition to investigating the viability of establishing PACs for ten more materials—including steel reinforcement—the research evaluates the procedural and financial ramifications of the current PACs for fuel and asphalt. The study acknowledges the widespread use of PACs nationwide and highlights their function in reducing financial risks related to variations in material costs during construction contracts. On the other hand, it recognizes the regional variations in PAC availability and draws attention to the many requirements encountered in clauses, such as trigger values and adjustment terminology. The limitations of the research include the need for a standardized approach to PACs across states and the complexities involved in developing clauses for diverse materials, particularly steel (Pierce et al., 2012).

The study of Turochy et al., (2001) investigate the cost estimating methodologies employed by SHA during the planning stage of highway project development. The report notes that SHAs should have greater national consensus and uniformity in adopting cost-estimating approaches. It attributes this diversity to factors such as topography, economy, and organizational structures. The study emphasizes the value of engineering judgment and experience in cost estimation, demonstrating a preference for skilled planners and engineers over sophisticated mathematical models. The results highlight the necessity for SHAs to invest substantial funding in planning-stage cost estimates to be subject to further investigation. The need for disclaimers to direct the appropriate use of planning cost estimate tables, front-loading resource allocation for cost estimates, and investigating oversight procedures are among the recommendations for additional action. Limitations of the study include the limited survey scope, and suggestions are made for expanding the survey, evaluating existing processes, and exploring the potential development of new cost estimation models based on project concepts (Turochy et al., 2001).

The AASHTO "Practical Guide to Cost Estimating (2013)" serves as a comprehensive and practical resource for SHA to develop realistic estimates of project costs, essential for successful program management (AASHTO, 2013). The AASHTO Technical Committee on Cost Estimating (TCCE) developed the handbook, filling the area's shortage by combining information gathered from NCHRP investigations. The guide provides structured approaches for estimators, project managers, and professionals involved in project development. It is divided into important estimate techniques (Conceptual, Bid-based, Cost-based, and Risk-based) and cost management activities (Inflationary considerations, letting strategies, Analysis of contractor bids, and Performance measures). Nonetheless, the focus needs to concentrate on any potential drawbacks or difficulties related to the recommended approaches. Recognizing the limitations and complexities of applying these strategies in various project situations is beneficial for future modifications. Furthermore, investigating innovative techniques and recent technology may improve the guide's relevance in a changing transportation project environment. Nevertheless, the guide provides valuable insights into cost estimating and management practices, laying a foundation for further advancements in the field (AASHTO, 2013).

The research conducted by Adel et al., (2016) addresses the critical need for applicable cost estimating models in the preliminary stages of highway projects. Their parametric model, allowing the use of a genetically optimized neural network, shows possibilities in conceptual cost estimation. A graphical user interface improves the model's applicability in practice, and identifying seven important components highlights the model's depth. However, the study's shortcomings include its reliance on historical data gathered from completed highway projects in Egypt between 2003 and 2013, which may limit the model's applicability to various project contexts and periods. Subjective model components like parameter selection and expert input may also introduce variability. However, the study substantially contributes to advancing cost estimation techniques in the early planning phases of highway projects (Adel et al., 2016).
The study by Asmar et al., (2011) presents a statistical methodology that is similar to program evaluation and review technique (PERT) and shows how well it can estimate construction costs conceptually. The study, however, fails to note the difficulties associated with highway incidentals and the limitations in data availability for contingency items. It is suggested that better data collecting, and a more thorough breakdown of incidentals should improve the accuracy of cost estimates (Asmar et al., 2011).

Bell & Kaminsky, (1987)'s microcomputer-based method for preliminary cost estimation in highway construction projects is insightful but has drawbacks. Previous bid data ignores changing industry dynamics since it makes assumptions about constants in the elements influencing costs. Furthermore, regional variances, project complexity, and technological improvements may limit the procedure's applicability. Uncertainty is introduced when material application rates are based on subjective estimations. Despite these drawbacks, the recommended approach offers a useful framework for preliminary cost estimation and valuable insights for planning and budgeting highway construction projects (Bell & Kaminsky, 1987).

With a focus on earthwork, pavement, and traffic control activities, Chou, (2009)'s research presents an expert system for early-stage cost estimation in Texas roadway construction projects based on the Generalized Linear Model (GLM). The study aims to enhance the accuracy of preliminary cost predictions by utilizing statistical models that account for project-specific characteristics. The proposed expert system offers a platform for ongoing quantity tracking throughout the project life cycle and considers historical unit prices. Although the methodology shows potential, there are some drawbacks as well. These include the possible reliance on previous bid data, the need for frequent updates as new project data becomes available, and the excellent use of the developed models to start quantity estimates in other states because of regional differences and project-specific factors (Chou, 2009).

By proposing a Web-based infrastructure cost estimating method, Chou & O'Connor, (2007) contribute to preliminary cost estimation in highway construction projects. Utilizing statistical models integrated into a Web-based relational database management system, the recommended approach seeks to improve accuracy, decrease variability, and simplify data storage. Although the study emphasizes users' effective compliance and satisfaction, there may be some drawbacks because it relies too much on historical district unit prices, which may not accurately reflect changing market conditions. Beyond the Texas Department of Transportation's Design and Construction Information System, the system's applicability to various project settings and areas requires investigation (Chou & O'Connor, 2007).

The quantity-based technique proposed by Chou et al., (2006) improves the preliminary cost estimation process for highway projects. The researchers emphasize the potential to differentiate quantity uncertainty from price uncertainty and acknowledge the crucial influence that preliminary estimates have on project viability. Although the automated estimating method shows possibilities with features including a comprehensive item-level initial assessment and monthly updates on unit bid pricing, there could be drawbacks because significant work items rely on previous data. Additional research is needed to determine whether the suggested approach can be applied to different project contexts and areas outside the Texas Department of Transportation. Furthermore, the system's efficiency depends on precise quantity prediction at the conceptual planning stage, which may be impacted by changing project dynamics (Chou et al., 2006).

The contributions of Fragkakis et al., (2010) to cost estimation approaches for bridge superstructures address the crucial requirement for accurate estimates at an early stage of the project. The recommended conceptual cost estimate method incorporates regression analysis and bootstrap resampling to forecast material quantities and related costs for the three main bridge deck construction methods. The paper effectively illustrates the models' satisfactory fit and their application to real-world data; however, there are some potential drawbacks, such as the method's generalizability to different bridge types and construction contexts, its reliance on assumptions underlying linear regression, and the requirement for precise input data during the preliminary study. The study also recognizes that estimations are inherently uncertain and that although the bootstrap technique reduces this uncertainty, differences in project-specific variables may cause uncertainties to continue (Fragkakis et al., 2010).

Gardner et al., (2017) contribute to advancing conceptual cost estimating for highway projects by introducing a stochastic approach that communicates uncertainty using bootstrap sampling. The study illustrates how crucial it is to provide estimates with confidence, particularly in the initial stages of a project when there is a need for more information. With the combination of artificial neural networks and bootstrap sampling, this proposed stochastic data-driven model offers a workable approach to generate empirical distributions, allowing a more accurate depiction of estimate ranges. The study emphasizes the advantages of this method. Still, it also raises some drawbacks, such as the neural network model's assumptions, the dependence on historical data, and the necessity of carefully evaluating each project's specific contingencies because they differ and need an appropriate basis for assessment. Furthermore, additional research is necessary to determine whether the approach is generalizable to different project contexts (Gardner et al., 2017).

The study of Gardner et al., (2016) presents a valuable perspective to conceptual cost estimating by challenging the common belief that more input variables necessarily enhance accuracy. Specializing in data-driven models with artificial neural networks and multiple-regression analysis, the work highlights the significance of adopting high-impact/low-effort input variables for conceptual estimates to be satisfactorily accurate. The research, in collaboration with the MDT, refutes popular belief. It presents a logical approach to input selection, demonstrating that adding variables indefinitely after 6–8 decreases returns on model performance. Although the study's limitations—such as its agency-specific focus—warn against extrapolating the findings to other agencies, even as it offers valuable insights for MDT. The use of perceptional data for input variable selection and evaluating how well these findings apply to various highway agencies represent future research directions (Gardner et al., 2016).

The underutilization of performance measures for highway cost estimating is a significant contribution Harper et al., (2014) offers to the discipline. The work synthesizes, classifies, and validates current actions, laying the groundwork for developing procedures and boosting estimate accuracy in state highway agencies. By identifying primary categories such as contingency amounts, estimating methods, competition effects, and bidding accuracy, the study offers valuable information to agencies seeking to create and disseminate new performance indicators. The research provides room for more study in these areas even though it provides a thorough list and notes that it focuses on something other than developing new standards or describing implementation procedures. The results underscore the significance of cost-estimating performance indicators given the decreasing federal financial assistance and the increasing number and cost of highway projects (Harper et al., 2014).

Hollar et al., (2013) provides insightful information when comprehending PE expenses in bridge projects. Their study tackles a frequently disregarded facet of PE costs and finds that cost estimates are significantly understated. The study builds statistical models based on project parameters from North Carolina DOT projects, emphasizing the necessity of employing a consistent percentage of construction costs for PE calculations. The models provide insightful information but have drawbacks, such as a 42.7% prediction error and difficulties with precision in data gathering, which call for standard operating procedures to increase accuracy in subsequent analyses. The research shows precise PE cost estimates are crucial for adequate infrastructure funding. Standardized data-gathering procedures and qualitative analysis are also necessary to overcome discrepancies in PE costs (Hollar et al., 2013).

Karaca et al., (2020) contribute to advancing early cost estimation practices in transportation infrastructure projects, focusing on top-down models. The study compares the accuracy of agency estimates with multiple regression and ANN estimates using a dataset of 996 MDT projects. The results challenge the common wisdom that bigger models always produce better results, suggesting that top-down models can improve prediction accuracy, particularly for complicated projects with smaller sample numbers. The study emphasizes the usefulness and effectiveness of top-down methods while highlighting the significance of finding a balance between bias and variation in model selection. One of the limitations is the 42.7% prediction error, which illustrates how difficult it is to accurately estimate the costs of construction projects because of shifting market pricing and project-specific variables. The study's recommendations for continued calibration and monitoring emphasize improving early estimation techniques (Karaca et al., 2020).

Liu et al., (2013) assist with PE cost estimation for road projects to offer SHAs effective budgeting techniques. Using data from 188 projects in North Carolina, the study challenges the conventional wisdom that Project-specific ratios, such as 13.3% for widening projects, 7.7% for rehabilitation, and 16.5% for new location/interchange projects, are revealed by calculating PE expenses, which are 10% of expected construction costs. Regression models are compared to historical means by the authors, who advise utilizing the latter for simplicity unless precise project-specific estimates are required. The limitations include the inability to predict PE duration and identify the causes of excessive PE cost ratios (Liu et al., 2013).

Petroutsatou et al., (2012) address the challenges of underground uncertainties and hazards in road tunnel construction cost estimation, particularly during the conception phase. The study utilizes neural networks particularly multilayer feed-forward and general regression neural networks to develop cost-estimating models using data from 33 twin tunnels of the Egnatia Motorway in Greece. The models demonstrate accuracy and robustness for early cost estimates, focusing on geology, geometry, and work quantities, and these characteristics render it an invaluable tool for evaluating alternative and cost-effective solutions in the early phases of tunnel projects. Nonetheless, significant differences in the geology of various tunnel projects and the requirement for ongoing model validation with multiple datasets to improve generalizability could be limited (Petroutsatou et al., 2012).

In exploring construction project cost escalation, Shane et al., (2009) undertake a comprehensive study, amalgamating insights from literature and interviews with over 20 state highway agencies to identify and categorize 18 primary factors influencing cost increases. The research emphasizes how transportation projects have been underestimated historically, especially in the public sector, where budget overruns affect infrastructure programs. Engineers, estimators, and project participants can improve cost-estimating accuracy and develop measures to mitigate the consequences of these escalation variables by utilizing the discovered factors, which are a valuable resource. The dynamic nature of construction sites, geographical differences, and the requirement for continual adaptability to changing project dynamics are a few examples of limitations (Shane et al., 2009).

Below is Table 3 which illustrates the summarizations of the synthesis literature review including research goal and outcomes.

Table 3*. Comprehensive Overview of Studies, Research Goals, and Outcomes*

SL	Research Goal	Research Outcomes	Citation
1	Evaluate MDT highway project cost estimating practices.	Proposed recommendations, including cost estimation, updated data, a manual, quality control, risk capture, inflation management, and training. Included an implementation timeline.	(Alavi & Tavares, 2009)
2	Improve state highway agencies' unit cost development practices.	Identified the absence of formal unit cost adjustment processes, suggesting short-term measures and long-term strategies for improvement.	(S. Anderson et al., 2009)
3	Explore state transportation agencies' project cost estimation approaches	Identified varied estimation methodologies, recommended consistent practices, and highlighted state-specific tools and challenges for accurate project cost estimates.	(Van Dyke et al., 2017)
4	Enhance MDT's early construction cost estimates using top-down estimating and artificial neural networks.	Improved prediction accuracy, identified key variables, proposed an Excel tool, and suggested future research areas.	(Gransberg et al., 2017)
5	Improve NCDOT highway project cost estimation	Developed accurate predictive models and a user-friendly interface for preliminary engineering costs.	(Liu et al., 2011)

6	Develop strategies, methods, and tools for effective cost estimation and management.	Eight global strategies, 30 recommended methods, and 90 tools to enhance cost estimation and management.	(S. D. Anderson et al., 2007)
7	Enhance state DOT cost estimation practices	Identified crucial tips and considerations for accurate estimates, covering planning, scoping, and design phases.	(Paulsen et al., 2008)
8	Evaluate PACs in transportation construction.	PACs are widely used (97% state DOTs), Contractors found PACs beneficial (90%), and PACs contributed to bid accuracy and market stability, with suggested improvements.	(Skolnik, 2011)
9	Assess PACs for construction materials, emphasizing SCDOT, focusing on financial and procedural aspects.	PACs are widely used (90% state DOTs), exhibit regional material preferences, with asphalt and fuel most prevalent. Mixed responses to steel PACs; recommends reinforcing steel PAC for SCDOT	(Pierce et al., 2012)
10	Evaluate state DOTs cost estimating methods for highway projects in the planning stage.	Diverse cost estimating methods among state DOTs, relying on engineering judgment. Limited use of sophisticated techniques observed. Recommendations include studying oversight processes and front-loading planning-stage cost estimates.	(Turochy et al., 2001)
11	Evaluate and enhance state DOTs' cost estimating practices for improved project management.	Provided practical techniques and strategies for estimators, project managers, and agency management, covering various cost estimation methods, inflation considerations, letting strategies, bid analysis, and performance measures.	(AASHTO, 2013)
12	Develop a parametric model for conceptual cost-estimation	Model development, validated by case study, and practical application.	Adel et al., (2016)
13	Develop a reliable methodology to analyze the historical bid data for cost estimation	Developed PERT type analysis, reliable estimating methodology with 20% accuracy.	(Asmar et al., 2011)
14	Developing a microcomputer-based cost-estimation	Unit price database, key factors identified, and systematic estimation	(Bell & Kaminsky, 1987)
15	Develop a generalized linear model-based expert system estimation	Identified factors, automated tracking, enhanced TxDOT cost estimation approaches	Chou, (2009)
16	Develop a web-based system for accurate preliminary cost estimation	Developed web-based system, mitigated cost estimates variability, and improved accuracy	Chou & O'Connor, (2007)
17	Develop a quantity-based system for preliminary cost estimates.	Developed automated estimating system with quantity models, detailed preliminary estimates	Chou et al., (2006)
18	Develop an early and reliable cost estimate method for bridge Construction	Cost estimate method using regression and bootstrap, validated, and mitigated uncertainty.	(Fragkakis et al., 2010)
19	Develop a stochastic conceptual cost estimating model for highway projects.	Data-driven model combining neural networks and bootstrap sampling, enabling the expression of estimate confidence and range.	(Gardner et al., 2017)
20	Evaluate data-driven models for cost estimating efficiency.	Optimized accuracy and mitigated data-collection efforts.	(Gardner et al., 2016)

21	Synthesize and validate performance measures for cost estimating.	Foundation for future measures and potential improvement in estimating accuracy.	(Harper et al., 2014)
22	Develop predictive models for PE cost estimation in bridge projects.	Identified underestimation of bridge PE costs, proposed predictive model, and recommended enhanced data collection procedures.	(Hollar et al., 2013)
23	Evaluate top-down estimating to enhance budgeting accuracy for public agencies.	Validated top-down model effectiveness, identified key variables, highlighted bias-variance trade-off for prediction accuracy.	(Karaca et al., 2020).
24	Evaluate strategies for PE cost estimation in roadway projects.	Identified historical mean PE cost ratios for project types, compared regression modeling to historical means, found correlation between PE cost ratio and PE duration.	(Liu et al., 2013)
25	Develop a neural network-based system for early cost estimation in road tunnel construction.	Identified key parameters affecting construction costs, collected and normalized real-world data, developed and compared neural network models (MLFN and GRNN), and demonstrated accuracy.	(Petroutsatou et al., 2012)
26	Identify and categorize cost escalation factors through literature review and agency interviews.	Categorized 18 cost escalation factors and verified through interviews with over 20 transportation agencies. Provided a basis for developing strategies and tools to enhance cost estimation.	(Shane et al., 2009)

Table 4 shows the DOT-centric toolset overview from previous studies. Most of the DOTs utilize excel based tool to determine the cost estimation. However, few of them used other tools, such as Estimate and Bid Analysis System (EBASE), Long Range Estimation (LRE), and Microsoft Visual C++ tools.

***Table 4**. DOT-Centric Toolset Overview from Previous Studies*

SL	Literature Type	Recommended Tools for DOT uses	Citation
1	Project Report (Montana DOT)	Excel tool	(Alavi & Tavares, 2009)
2	Project Report (Texas DOT)	TxDOT-Excel tool	(S. Anderson et al., 2009)
		WSDOT-EBASE & Excel	
		FDOT-LRE	
		UDOT-Excel tool	
		NYSDOT-Excel tool	
		MnDOT-Excel tool	
		Caltrans-Excel tool	
		VDOT-Excel tool	
3	Project Report (Kentucky)	Excel tool	(Van Dyke et al., 2017)
4	Project Report (Montana DOT)	Excel tool	(Gransberg et al., 2017)
5	Project Report (NCDOT)	**Interface Application:** Microsoft Visual C++	(Liu et al., 2011)

3. Survey of State DOTs

This chapter focuses on the preliminary cost estimation techniques used in transportation projects and provides the results of an extensive survey targeting transportation professionals at State Departments of Transportation (DOTs) in the United States. The purpose of the survey is to learn about the practices, methodologies, tools, and challenges associated with preliminary cost estimation as it is employed across various states. The survey's methodology and participant demographics are covered in full in the first section of the chapter. The survey findings are then explored in detail, indicating the variety of practices employed by DOTs in terms of estimation methods, data sources, and degrees of satisfaction with the current procedures. It also emphasizes the types of estimating tools utilized, the procedures applied to contingency cost estimation, and whether agencies adhere to federally authorized approaches. The survey further investigates how agencies handle unit costs, inflation, and cost indices, offering insights on regional influential factors. About 36% of State DOTs responded to the survey, and according to the survey, there was no federally prescribed preliminary cost estimating approach and the most common tool was an Excel-based tool. It is clear from the survey responses that there is wide variation in practices across the various State DOTs.

Based on the survey of 19 experts from 15 U.S. known and three unknown State DOTs, the present study contributes valuable insights by highlighting the necessity for standardization initiatives and addressing discrepancies to enhance the consistency and reliability of preliminary cost estimation procedures in transportation organizations.

3.1 Survey Methodology

To gain a comprehensive understanding of the preliminary cost estimating approaches employed by DOTs across the United States, this study conducted an extensive questionnaire survey. Appendix A includes the survey instrument used in this study for synthesizing practices across the various state DOTs. The primary goal of the questionnaire survey was to synthesize practices of other state DOTs in terms of approaches and tools used for developing cost estimates and how those estimates were used, considering the risks involved. The research methodology included two sections, which are participant selection, and survey sections.

Participant Selection. The survey was distributed electronically to 50 State DOTs' Value Engineering and Estimates Coordinators, Statewide Project Management Specialists, State Estimating Engineers, Research Implementation Managers, Project Managers, Independent Cost Estimating Coordinators, Engineering Supervisors, Engineers, Director of Preconstruction, Contracts and estimates Engineer, Civil Engineer IV, Chief Road Design Engineer, Bidding and Contract Services Engineer, Assistant State Materials Engineer, and Assistant Director of Planning across the United States. Participants were encouraged to provide detailed responses to maximize the richness of the data collected.

Survey Sections. The survey was distributed through secure online survey platforms, ensuring data integrity and confidentiality. It consisted of closed-ended and open-ended questions. The survey questionnaire comprised seven sections, each addressing specific dimensions of preliminary cost estimating approaches. These sections included participant's basic information, agency practices in preliminary cost estimation, contingency and risk management, types of cost estimates, inflation considerations, collaboration and knowledge sharing, and recommendations and suggestions.

3.2 Results and Discussion

Nineteen participants from fifteen different state DOTs and three unknown state DOTs in the United States of America (USA) provided comprehensive responses for this study. Sixty-three percent of the nineteen complete responders consented to a future follow-up interview. The location of the respondents is represented graphically in Figure 3. The data indicated that the 18 states that comprised 36% of the total states DOT represented by all survey respondents were Oklahoma, Wyoming, Missouri, North Carolina, Idaho, Utah, Delaware, Nevada, Kentucky, Georgia, Michigan, Minnesota, New York, Illinois, South Dakota, and three unknown states that were not included in the list.

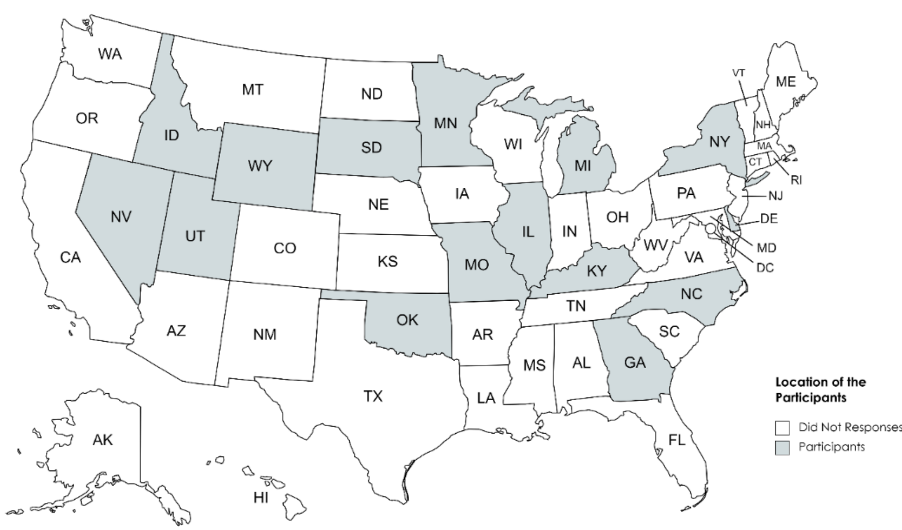


Figure 3. *Location of the Participants*

Seven SHAs (State Highway Agencies) had a systematic approach, seven did not have a standardized procedure, one did not respond, and four responded with detailed guidelines to cost estimation methods among transportation agencies (Figure 4). According to four comprehensive responses, the original STIP project estimates were based on SY unit costs and change as plan sets do. The scoping process and completed studies influenced detailed estimates. Notably, there's a distinction between long-range planning and Project Development groups in establishing project costs before advancing onto the STIP, emphasizing the complexity of the approaches.

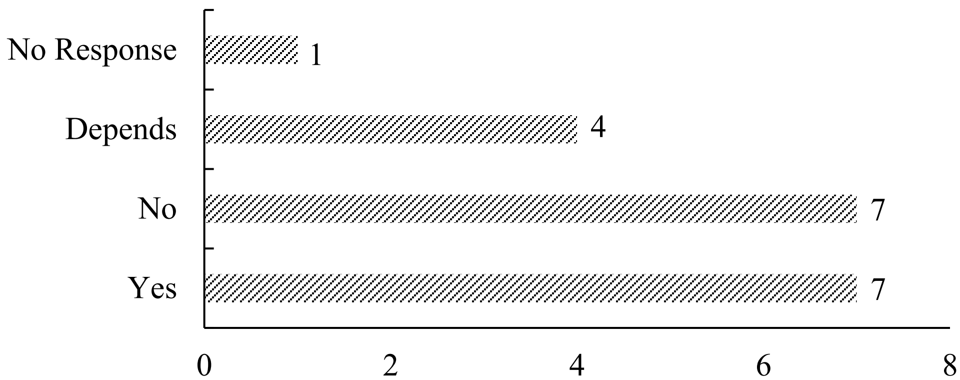


Figure 4. *Existence of Systematic Method for developing Preliminary Cost Estimation*

Responses to the survey regarding agency satisfaction with the initial cost-estimating method were different. Two respondents were neutral, two were highly satisfied, and six were moderately satisfied. Remarkably, one respondent expressed dissatisfaction, and eight did not react, adding uncertainties (Figure 5). This variety brought diverse viewpoints regarding the effectiveness of the agency's preliminary cost estimation procedure.

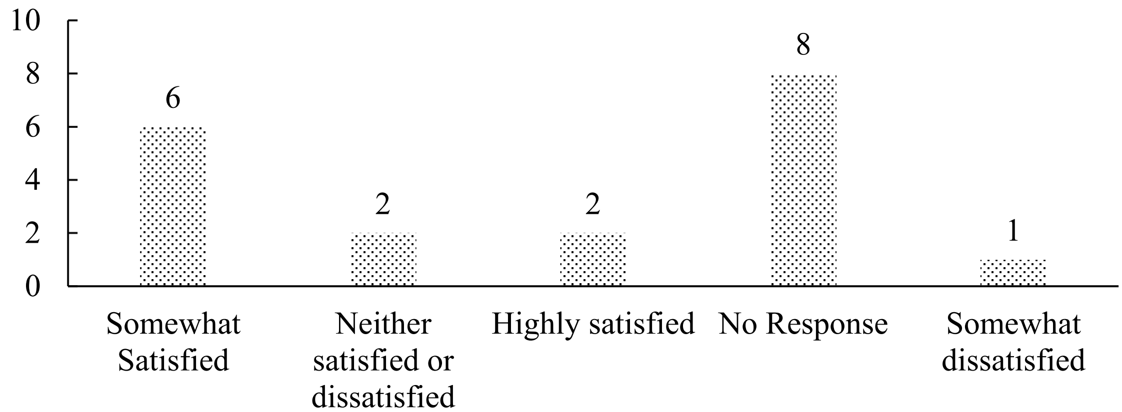


Figure 5. *Existence of Preliminary Cost Estimating Process's Satisfaction*

The majority of respondents (10) to the survey stated that their agency developed its preliminary cost-estimating approach in-state, either in-house or with a consultant or researcher (Figure 6). Eight responders, nevertheless, chose not to reply, indicating a lack of interest. One responder provided an extensive reply that was state-specific but might differ slightly depending on the

Area, Region, or TSC implementation because there was no set estimating tool and no requirement for particular Excel formats, which could result in variations in the forms used for implementation.

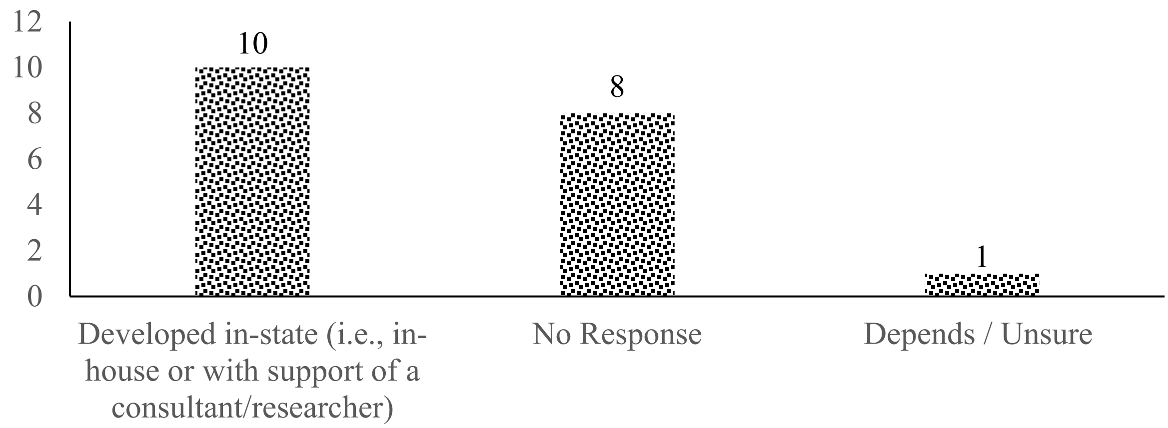


Figure 6. *Existence Preliminary Cost Estimating Approaches Developing Process*

According to the survey, the majority of the agencies (10 respondents) employed Excel tools for initial cost estimation (Figure 7). Only one responder mentioned using stand-alone software, and eight did not answer, indicating a lack of understanding or interest in the issue. The responses highlighted the general dependence in the surveyed setting on spreadsheet-based techniques for initial cost estimation.

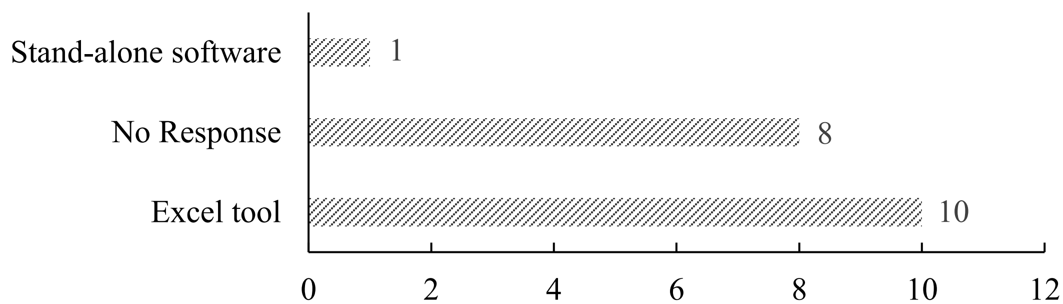


Figure 7. *Existence Preliminary Cost Estimating Approaches Formation*

Figure 8 shows that the agencies utilize a variety of preliminary cost-estimating techniques. One respondent employed bid-based and cost-based pricing, while six utilize a unit price approach. Two organizations set prices based on actual high-level quantities. Notably, twelve organizations chose not to reply, indicating a lack of participation or clarity in their approaches. These responses demonstrated how different early cost-estimating techniques are in the survey context.

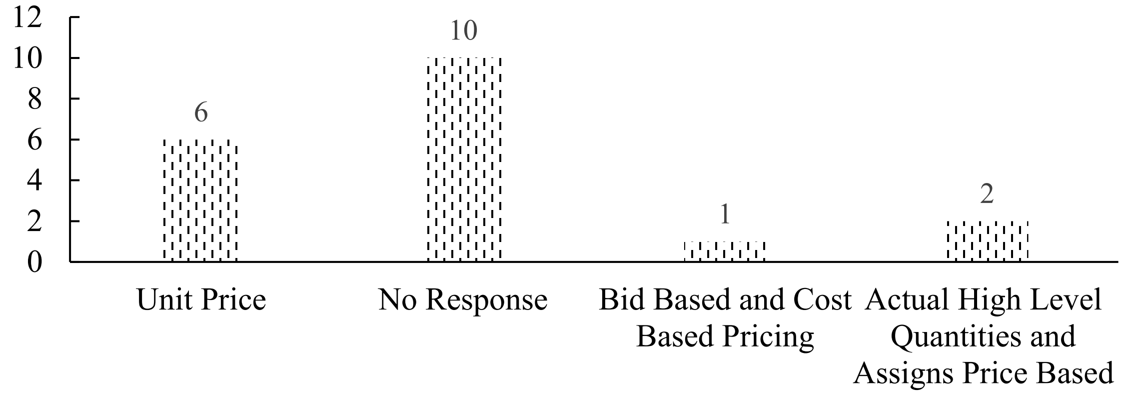


Figure 8. *Existence Preliminary Cost Estimating Approach*

The survey showed different approaches to estimating contingency costs in preliminary estimates. A proportionate relationship was demonstrated by the eight agencies that used a percentage of the base estimate (Figure 9). Three agencies use different techniques, highlighting the unpredictability of the process, and their grading plan estimated usually did not account for contingency. A percentage-based contingency calculation was included in the initial scoping estimates; however, as projects moved through the development phase, pay items were used to calculate updated construction costs. The nature of the project and its location were the main factors influencing the contingency percentage. Notably, eight agencies chose not to respond, suggesting a lack of understanding or involvement with this issue.

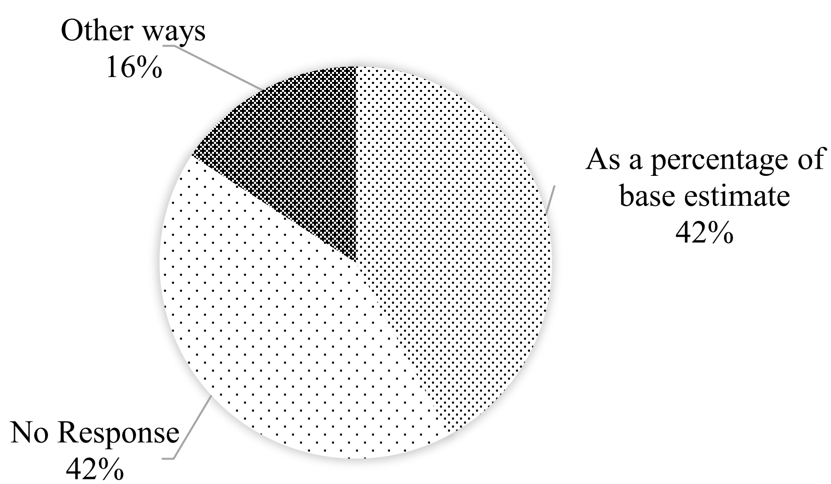


Figure 9. *Contingency Costs Estimated in the Preliminary Cost Estimates*

According to the survey, most agencies (11 out of the total) generated a deterministic cost estimate comprising a set value and a suitable contingency (Figure 10). Notably, eight agencies chose not to reply, suggesting a lack of understanding or involvement with this issue. These revealed a similarity in how the studied agencies approach preliminary cost estimates, emphasizing fixed value and contingency concerns.

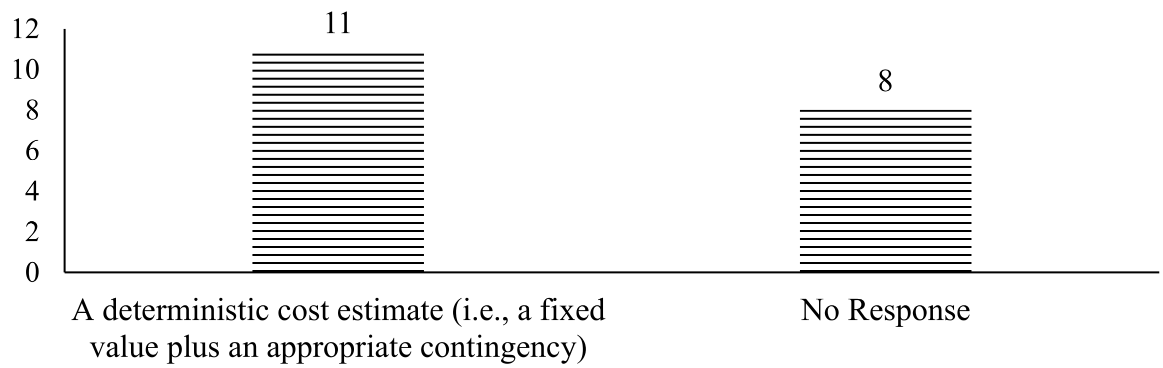


Figure 10. *Type of Existence Preliminary Cost Estimate Method*

The bid history from comparable projects, which captures market trends for the present circumstances, was the first step in the systematic procedure. The expected cost per unit computation was derived from historical regression curves for uncommon items and recent data for often-used products. Construction amounts were added to all bid prices after the bidding, which were then sorted based on the project's specifications. Then, unit prices were determined through linear regression. In order to offer current average unit costs internally and externally, quality assurance teams monitored bid history. Furthermore, bid prices from the three lowest bidders over the last 24 months could be seen in the quarterly updated Bid History Catalog.

Diverse agency techniques for creating unit costs for cost estimation are depicted in Figure 11. Eleven organizations emphasized systematic methods and follow a methodical process. On the other hand, six agencies didn't have a systematic procedure, which suggests possible variability. Using their bidtabs.net, one agency employed a unit price based on previous data. One agency's unit cost development procedures were unclear because they did not reply. These showed that different examined agencies had other methods for creating unit costs in a systematic way for cost estimation.

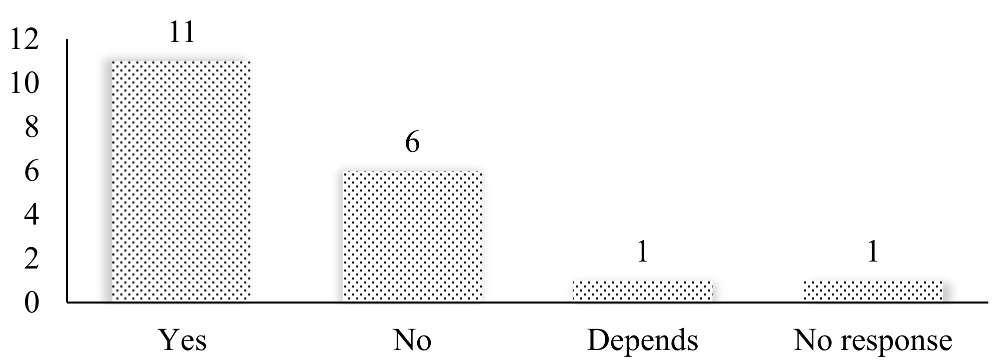


Figure 11. *Systematic Process for Developing Unit Costs for Cost Estimating*

According to the survey, agencies' maintained levels of historical unit prices increased significantly. Ten agencies concentrated at the state level, four at the district level, and four use alternative approaches, like combining their own data with the state level, working on a specific project with a particular location, and paying attention to a specific pay item (Figure 12). There was ambiguity since one agency did not reply. These demonstrated how different survey agencies manage historical unit costs in different ways.

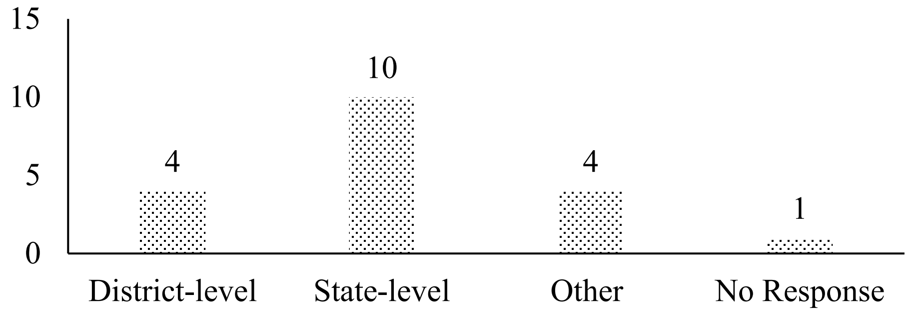


Figure 12. *Level of Historical Unit Cost*

According to the survey, agencies utilized different methods when estimating project costs to account for inflation in certain states or regions. Eleven agencies didn't have a structured strategy, which could lead to variability (Figure 13). On the other hand, five agencies used an organized approach for this. Two organizations pointed to a reliance on particular elements, such as the percent-based approach and the lack of a specific formula. Notably, confusion was introduced by one agency's lack of response. These demonstrated how different the investigated agencies' methods were when handling inflation issues in project cost estimation. The agency utilized tools like the Highway Construction Cost Index (HCCI) and a locally developed inflation calculator to account for inflation in the initial cost estimates. At the bottom of the forecast, inflation was included in and projected to the expected year of completion. A Composite Cost Index (CCI) was updated, and an on-staff economist ascertains the current inflation rate. In addition, they applied a percentage annually based on the project's delivery schedule, using suggestions from the statewide scoping manual during the yearly call for proposals.

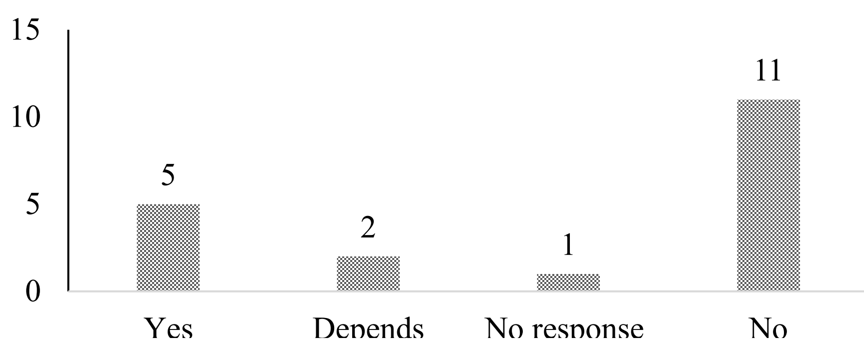


Figure 13. *Systematic Approach to Account for Inflation Specific to the Region/State*

Figure 14 showed limitations on adopting the Highway Construction Cost Index (HCCI) for inflation adjustment across agencies at the state or regional levels. Two agencies were working to develop the tool to track actual bid data, while three agencies currently use HCCI. Interestingly, thirteen agencies chose not to reply, indicating a lack of understanding or participation regarding using HCCI for inflation adjustments. These showed that different surveyed agencies had different policies regarding using HCCI to account for inflation when determining the cost of constructing new highways.

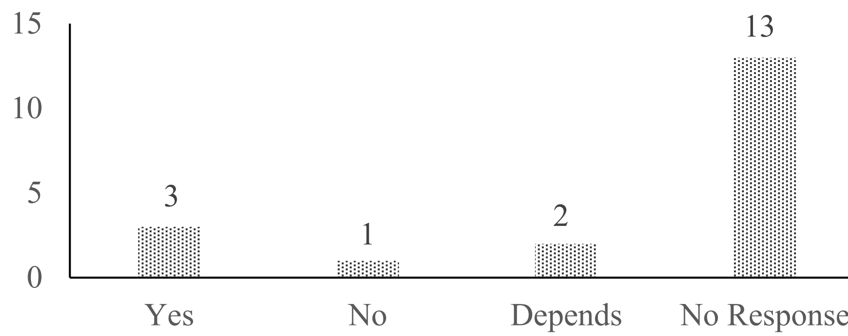


Figure 14. *State-wide or Region-wide Highway Construction Cost Index (HCCI)*

According to the survey, agencies had differing opinions about exchanging preliminary cost estimation tools with other state Departments of Transportation (DOTs). Figure 15 showed that nine agencies were eager to contribute, three were reluctant, five offer an "other" response that suggests further considerations, and two remained silent, creating confusion. They employed AASHTOW software, which differs depending on state configurations, and the study still needed to be completed. There's no official statewide cost estimating tool, but they could share regional samples and non-sensitive information with verification. Every district managed its initial approximation. These highlighted the various strategies the agencies assessed used to share tools and resources with other state DOTs cooperatively.

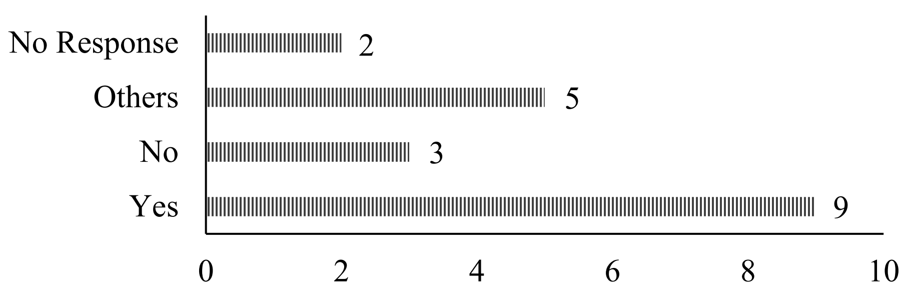


Figure 15. *Distribution of Preliminary Cost Estimation Tools*

3.3 Conclusions and Takeaways

The survey, which included responses from 19 experts in 15 U.S. known and three unknown State DOTs, clarifies a diverse environment in preliminary cost estimation procedures. Even while Excel tools and in-state development methodologies are widely used, non-responses in certain areas and the lack of standard methods indicate intrinsic diversity across the investigated organizations. Notably, different contingency estimating techniques demonstrate varied approaches, suggesting a new sophisticated approach to risk management would be useful. Comparably, agency-wide practices vary in how unit cost development and inflation factors are handled. The research highlights the need to tackle these discrepancies to improve the effectiveness and uniformity of preliminary cost estimation procedures in transportation organizations. Standardization initiatives could lead to a more consistent and dependable framework, especially regarding tools and methods. The results also point to possible areas for development, including improved data-sharing procedures, more cooperation, and more precise rules. These insights provide helpful guidance for improving preliminary cost estimation procedures and increasing the planning and development of infrastructure projects across various state agencies more efficiently and reliably.

The main recommendations of the survey are listed below:

- Stress the importance of defining project scope (90%) and managing risks and contingencies (10%) for accurate estimates.
- Use bid data from the last six months, prioritizing similar quantities, field districts, and project types for accurate analysis, and acknowledge reliance on historic prices, but emphasize the complexity of accounting for inflation.
- Use historical unit prices for similar work, adjusting based on current project specifics.
- Utilize recent project bids, consider cost indexes, inflation percentages, and contractor feedback for insights into market challenges, and seek perspectives from multiple subject matter experts.
- Recognize the simplicity in estimating construction costs versus challenges in estimating other expenses like Right of Way and Utilities, and be slightly conservative in estimates, rounding up for accuracy, especially for less frequently used items.
- Keep individuals engaged in bidding and estimation to anticipate changes and ensure an effective process and utilize historical databases for research projects.
- Regularly update estimates for accuracy, referencing consistent guidelines, and considering various funding sources. Include risk evaluations for factors like complex construction, and use "composite bid items" for preliminary estimates, combining multiple bid items for 'per mile' costs, updating prices based on recent relevant project quantities.

4. HCCI Development

4.1 Methodology

4.1.1 Background of HCCI

The Highway Construction Cost Index (HCCI) is a crucial indicator of the purchasing strength of highway agencies, as highlighted by Guerrero 2003. It is derived from the actual contract bids submitted by contractors over a specific timeframe, typically on an annual basis. The calculation involves unit bid prices and quantities of various items essential for highway construction. Unlike assessments based on contract completion, the HCCI excludes cost overruns resulting from unexpected seasonal events, such as flooding. Consequently, it is regarded as a more reliable tool for assessing overall construction market conditions. State departments of transportation (DOTs) have widely adopted the HCCI to monitor inflation in highway construction and to make reasonably accurate forecasts of the preliminary expenditure required for a highway project, as noted by Guerrero 2003 and White and Erickson 2011. Furthermore, certain DOTs utilize the HCCI as an inflationary gauge for preliminary and comprehensive cost evaluations, along with conducting lifecycle cost analyses (LCCA) for their highway projects. Additionally, HCCIs are recommended as a factor in determining gas tax rates to generate crucial revenue aimed at effectively maintaining the existing highway infrastructure system, according to (Shrestha et al. 2017).

4.1.2 Data Collection

The research team gathered historical bid data from the SCDOT for three types of projects: bridge replacements, intersection improvements, and widening. The dataset encompasses bid information from 380 projects, covering both completed and ongoing projects spanning the years 2013 to 2023, with a cumulative value exceeding $2.1 billion in construction projects. Table 1 illustrates that the team received a substantial number of projects for bridge replacements (130) and intersection improvements (204), whereas the count for widening projects is notably lower at 46. Notably, there is only one project in the dataset for intersection improvements in the year 2023. Examining the distribution of projects over the years, the dataset reveals a relatively limited number of widening projects annually (1, 2, 3, and 5 projects), with no bid data available for 2021 and 2023. This limited dataset could potentially impact the accuracy of the HCCI, a concern that is further explored in the calculation results section.

Table 5. *Summary of bid data from SCDOT*

Year	Number of projects by category			Total
	Bridge replacements	Intersection improvements	Widening	
2013	14	19	6	39
2014	10	29	13	52
2015	21	37	7	65
2016	10	26	7	43
2017	11	14	2	27
2018	18	21	1	40
2019	7	12	2	21
2020	16	28	5	49
2021	14	8	0	22
2022	9	9	3	21
2023	0	1	0	1
Total	130	204	46	380

4.1.3 Data Preprocessing

a. Outlier Removal

Before computing the Highway Construction Cost Index (HCCI), it is common practice to apply outlier removal to the bid item unit price data. This process is undertaken to reduce potential biases by identifying and excluding outliers from the dataset, as indicated by Jeong et al. 2021. In this project, the research team utilized two widely employed outlier determination methods, as outlined in (Jeong et al. 2021; Liu et al. 2021):

1. Outliers are those that deviate at least three standard deviations from the mean.
2. Outliers are considered as values greater than 1.5 times the Interquartile Range (IQR), calculated as the difference between the third quartile (Q3) and the first quartile (Q1), from Q3 or less than 1.5 times the IQR from Q1.

b. Special Items Removal

Items categorized as lump sum (e.g., mobilization) lack precision in representing the amount of work or materials required. Typically, these items are assigned a fixed quantity of 1, irrespective of the project's work volume or material quantity. Consequently, the bid prices for these items do not consistently align with their quantities. Additionally, some items had recorded quantities of 0 or text characters, likely stemming from input errors. To maintain accuracy and reliability in the analysis, these particular items were excluded from the calculation of HCCI.

4.1.4 HCCI Calculation

HCCIs are commonly expressed as equations involving bid prices and quantities over a specified time frame. In this study, the Laspeyres, Paasche, Fisher, and Chained indexing methods were employed, as these are widely recognized as the most commonly used formulas by State Highway

Agencies (SHAs) for calculating HCCIs (Shrestha et al. 2017). Given a dataset with *m* pojects, each has *n* work items, the indices are calculated using Equations (1) - (4):

$$Laspeyres\ index, L_{t,0}(p^0, p^t, q^0, q^t\ ) = \frac{\sum_{i=1}^{n} p_i^t q_i^0}{\sum_{i=1}^{n} p_i^0 q_i^0} \quad (1)$$

$$Paasche\ index, P_{t,0}(p^0, p^t, q^0, q^t\ ) = \frac{\sum_{i=1}^{n} p_i^t q_i^t}{\sum_{i=1}^{n} p_i^0 q_i^t} \quad (2)$$

$$Fisher\ index, F_{t,0}(p^0, p^t, q^0, q^t\ ) = \sqrt{L_{t,0} \times P_{t,0}} = \sqrt{\frac{\sum_{i=1}^{n} p_i^t q_i^0}{\sum_{i=1}^{n} p_i^0 q_i^0} \times \frac{\sum_{i=1}^{n} p_i^t q_i^t}{\sum_{i=1}^{n} p_i^0 q_i^t}} \quad (3)$$

$$Chained\ index, CI_{t,0} = \prod_{k=1}^{t} F_{k,k-1} \quad (4)$$

where *i* symbolizes a bid item, *p* denotes the average unit price, and *q* stands for quantity. The subscripts *0* and *t* designate the base year and the current year, respectively. The average unit price *p* of each item i for each year is determined as the weighted average of unit prices based on quantities, calculated by summing the products of unit prices and corresponding quantities and then dividing this sum by the total quantity.

$$Average\ unit\ price\ p_i = \frac{\sum_{j=1}^{m} p_{i,j} q_{i,j}}{\sum_{i=1}^{m} q_{i,j}} \quad (5)$$

Laspeyres index is the ratio of the total expenditure in the current period to the total expenditure in the base period, assuming that the same quantities. Paasche, on the other hand, utilizes the quantity vector for the current period and assumes it to be the same for the base period. The Fisher index is calculated as a geometric average of the Laspeyres and Paasche indexes. In this project, the Fisher index was computed on the annual basic between two consective periods *k* and *k-1*, which was later used for calculating the accumulated chained index for an longer period (see Eq. 4). The chained index, also known as the chained Fisher index, represents the inflation rate over a period of *t* years. The chained index formula is considered the ideal method for calculating a cost index which is used by FHWA for its NHCCI computation and is recommended for state DOTs' HCCI calculation (White and Erickson 2011).

In selecting bid items for the HCCI calculation above, the research team opted for the Dynamic Item Basket (DIB) method instead of the fixed item basket. Traditionally, the fixed item basket method involves choosing specific crucial bid items to form a consistent basket applied across all periods, simplifying the calculation process using basic spreadsheet tools. However, this approach, while convenient, poses a risk of inaccurately capturing changes in market conditions due to its reliance on a limited sample. The concept of the Dynamic Item Basket, introduced by Shrestha et al. 2017 (see Figure 16), aims to overcome the limitations associated with a fixed item basket. Rather than utilizing a fixed basket with a restricted set of bid items, the DIB method incorporates all bid items present in consecutive periods for computing the cost indexes. This inclusive approach allows for a significantly broader range of bid items, enabling HCCIs based on the DIB method to more accurately depict changes in costs within genuine market conditions.

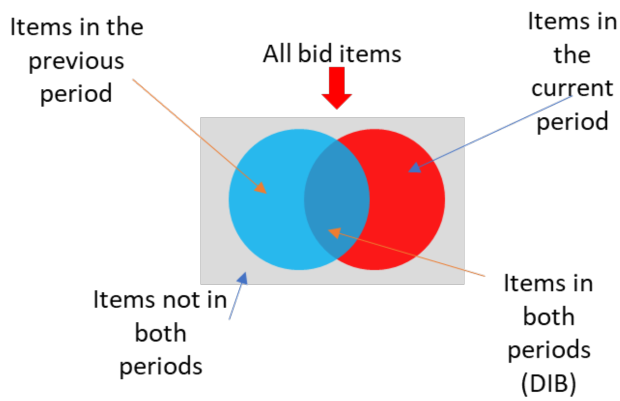


Figure 16. *Concept of DIB (Source: Shrestha et al. 2017)*

4.1.5 Sub-HCCIs

The calculation of HCCI can be performed across the complete historical project database for the entire state or for specific project groupings. Table 2 illustrates various HCCI types created in this research. These encompass a statewide HCCI derived from all accessible data, three sub-HCCIs based on contract characteristics, and six sub-HCCIs based on bid item characteristics. The statewide HCCI provides an overview of the overall market conditions in the state, while each sub-HCCI reflects the pricing trends within a particular group of projects.

To derive sub-HCCIs, our research team initially applied specific filtering criteria as detailed in Table 2, leading to the segmentation of the statewide database into sub-databases. Each of these sub-databases was then utilized independently to calculate the corresponding sub-HCCIs. For example, when computing sub-HCCIs based on varying contract sizes, we divided the statewide database into three distinct sub-databases representing small, medium, and large contracts. Subsequent to this segmentation, HCCI calculations were conducted separately for each sub-database (refer to the section "Contract size-based sub-HCCIs" below). Subsequent sections offer additional insights into the various sub-HCCIs.

Table 6. *Multidimensional HCCIs*

Category	**Data Filtering Criteria**	**Sub-HCCI**
Statewide	None	Statewide HCCI
Contract Characteristics Based HCCI	Project Work Type	Bridge replacements
		Intersection improvements
		Widening
	Contract Size	Small
		Medium
		Large
	Scope	Cluster 1
		Cluster 2
		Cluster 3
Bid Item Characteristics Based HCCI	Work Item division	Earthwork
		Bases and subbases
		Asphalt pavements
		Maintenance and control of traffic
		Structures
		Incidental construction

a. Contract Characteristics Based Sub-HCCIs

Contract characteristics-based sub-HCCIs include three types of sub-HCCIs, encompassing project type, scope, and contract size.

❖ Project work type-based sub-HCCIs

As indicated in Table 7, the research team gathered bid data for projects falling into three categories of work: bridge replacements, intersection improvements, and widening. The resulting database underwent filtration based on project work type criteria, leading to the creation of sub-databases for the computation of sub-HCCIs.

Table 7. *Project distribution based on work type*

Year	Number of projects			Total
	Bridge replacements	Intersection improvements	Widening	
2013	14	19	6	39
2014	10	29	13	52
2015	21	37	7	65
2016	10	26	7	43
2017	11	14	2	27
2018	18	21	1	40
2019	7	12	2	21
2020	16	28	5	49
2021	14	8	0	22
2022	9	9	3	21
2023	0	1	0	1
Total	130	204	46	380

❖ Contract size based sub-HCCIs

The monetary value of a contract can notably influence the prices of individual items. In this investigation, contracts were categorized into three groups based on their respective amounts, as outlined by Jeong et al. 2021: 1) small-sized contracts (below $700,000); 2) mid-sized contracts ($700,000-$8,500,000); and 3) large-sized contracts (exceeding $8,500,000).

Table 8 displays the distribution of projects based on contract size. The dataset comprises 380 contracts, totaling $2,105,997,073.58 in overall contract value. Small-sized contracts consist of 58 contracts (15.26%), contributing a total amount of $25,777,476.30 (1.22% of the overall contract amount). Mid-sized contracts, making up the majority with 266 contracts (70.00%), represent $701,193,956.55 (33.30% of the total). Large-sized contracts, numbering 56 (14.74%), possess the highest total value at $1,379,025,640.73 (65.48%). Refer to Table 9 for additional information on the sample size for each category in each year from 2013 to 2023.

Table 8. *Contract size-based-classification*

Contract size	Criteria	Contract count	Percentage (Number)	Total contract amount	Percentage (Dolar value)
Small-sized	<700,000	58	15.26%	$25,777,476.30	1.22%
Mid-sized	700,000≤amount≤8,500,000	266	70.00%	$701,193,956.55	33.30%
Large-sized	>8,500,000	56	14.74%	$1,379,025,640.73	65.48%
Grand total		380	100.00%	$2,105,997,073.58	100.00%

Table 9. *Project distribution based on contract size*

Year	Number of projects			Total
	Small-sized contract	Mid-sized contract	Large-sized contract	
2013	13	22	4	39
2014	12	32	8	52
2015	9	49	7	65
2016	5	30	8	43
2017	3	23	1	27
2018	6	31	3	40
2019	3	16	2	21
2020	6	35	8	49
2021	0	16	6	22
2022	1	12	8	21
2023	0	0	1	1
Total	58	266	56	380

❖ Scope based sub-HCCIs

The research team employed a Natural Language Processing (NLP)-based project vectorizing model, namely CW-TF-IDF (Do et al. 2023), and K-means clustering algorithm to cluster the projects into subcategories. This approach helps quantify the similarity between projects by capturing information about both the semantic similarity of pay item descriptions and the cost contribution. The clustering framework includes three main steps (as shown in Figure 17).

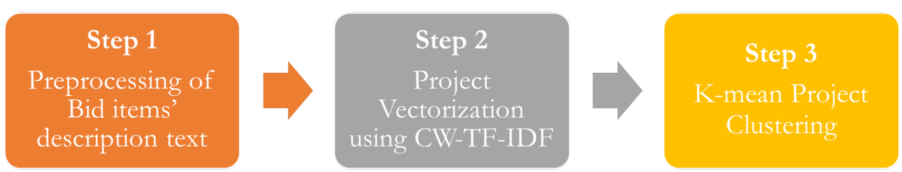


Figure 17. *Project clustering framework considering project scope*

<u>Step 1:</u> Several NLP techniques were employed to preprocess pay items' description text, encompassing tokenization (breaking text into smaller units), stop word removal (eliminating common words like "the," "a"), special character removal (such as numbers or symbols), lowercasing (converting text to lowercase), and lemmatization (reducing words to their base or canonical form) (see an example provided in Figure 18).

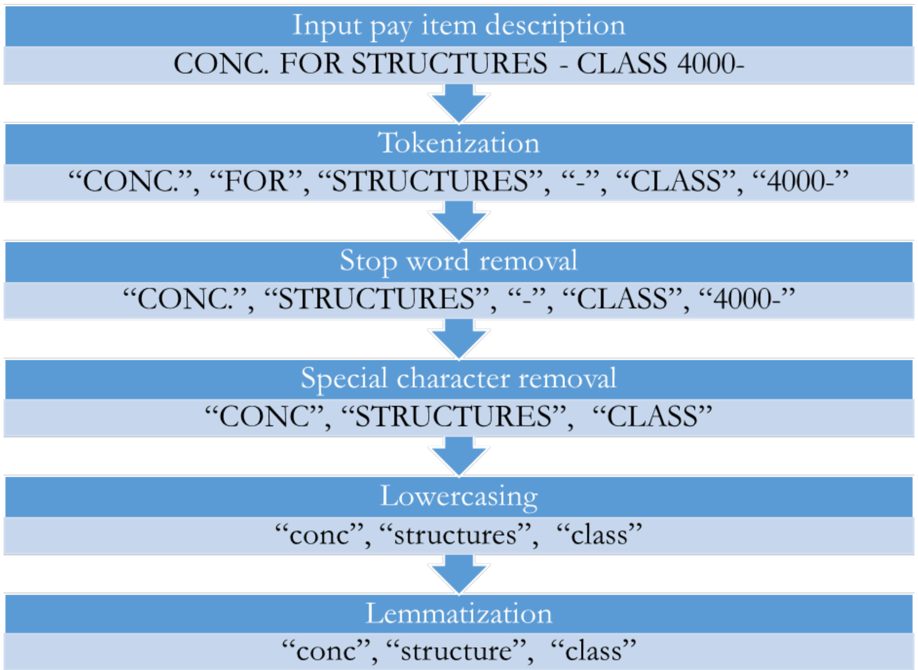


Figure 18. *Preprocessing pay item description text*

<u>Step 2:</u> Cost-Weighted TF-IDF (CW-TF-IDF) was employed to vectorize projects.

Table 10 shows the calculation of components and formation of project vectors.

Table 10. *CW-TF-IDF method for project vectorization*

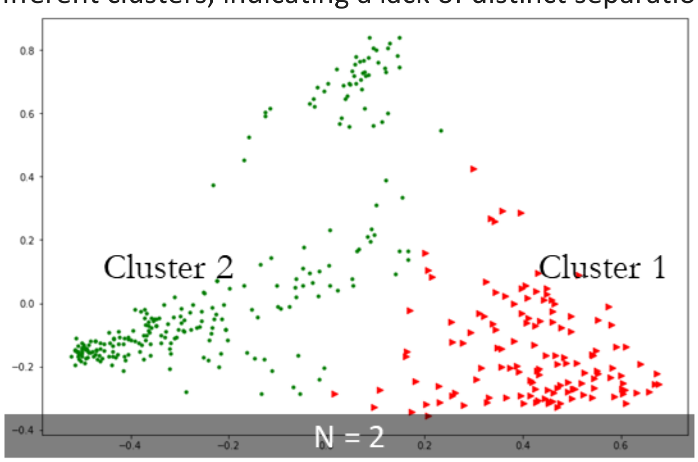
Item	Cost-Weighted TF-IDF (CW-TF-IDF)
Cost contribution of each pay item to the project amount	$cc_{i,j} = \begin{cases} \frac{c_{i,j}}{c_i} \times 100 & if\ \frac{c_{i,j}}{c_i} \times 100\ \geq \beta \\ 0 & otherwise \end{cases}$
Term Frequency (TF)	$ctf_i^t = \begin{cases} \sum_{j=1}^{np_i} n_{i,j}^t \times \frac{cc_{i,j}}{N_i} & if\ min\_df \leq \frac{D^t}{D} \times 100 \leq max\_df \\ 0 & otherwise \end{cases}$
Inverse Document Frequency (IDF)	$idf^t = \left(1 + log\frac{D}{D^t}\right)$

TF x IDF	$cw\text{-}tf\text{-}idf_i^t = \sum_{j=1}^{np_{i,j}} n_{i,j}^t \times \frac{cc_{i,j}}{N_i} \times \left(1 + log\frac{D}{D^t}\right)$
Project vector	$d_i = (cw\text{-}tf\text{-}idf_i^{t_1}, cw\text{-}tf\text{-}idf_i^{t_2}, \dots, cw\text{-}tf\text{-}idf_i^{t_m})$

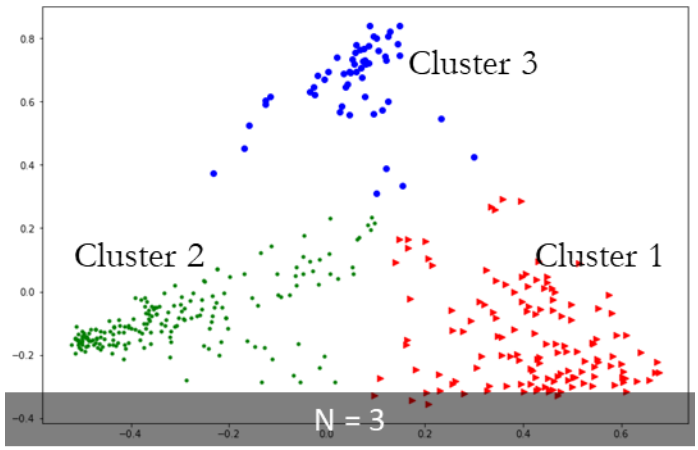
In the table above, $c_{i,j}$ = the cost of pay item *j* in the project *i*; c_i = the total cost of project *i*; $cc_{i,j}$ = cost contribution of pay item *j* in the project *i*; β = the minimum percentage threshold; $n_{i,j}^t$ = the number of occurrences of term *t* in pay item *j* of project *i*; np_i = the number of pay items in the project *i*; $cc_{i,j}$ = cost adjustment factor; N_i = the total number of terms in all pay item descriptions of project *i*; D^t = the number of projects that include term *t*; *D* = the total number of projects in the input dataset; *max_df* = maximum document frequency that term *t* occurs; and *min_df* = minimum document frequency that term *t* occurs; *m* = the dimension of project representation vectors, which is equal to the vocabulary size of the remaining terms in the entire historical project dataset.

Step 3: This study used K-means clustering algorithm to cluster projects into groups that contain similar projects. The project vectors obtained from step 2 were fed as input for this implementation.

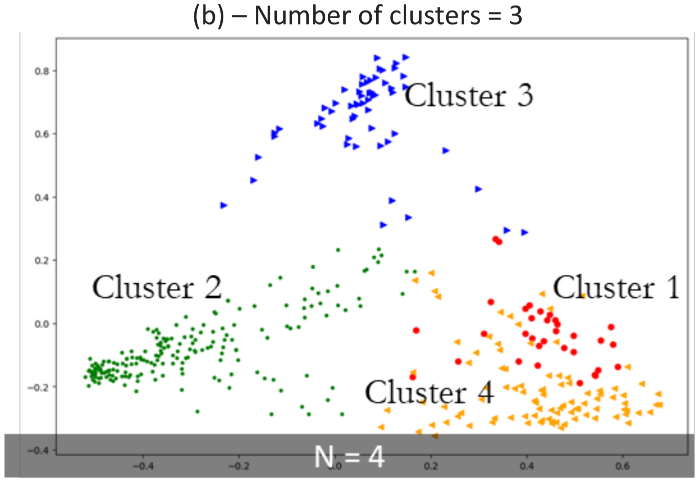
Multiple experiments were carried out by altering the number of clusters. The visual representations of the clustering strongly suggest that the ideal number of clusters is either 2 or 3, as depicted in Figure 19. This conclusion is drawn from the clearly defined boundaries observed in these cluster formations. In contrast, when employing 4 clusters, there is noticeable blending of data points from different clusters, indicating a lack of distinct separation.



(a) – Number of clusters = 2



(b) – Number of clusters = 3



(c) – Number of clusters = 4

Figure 19. *Scope-base project clustering results*

b. Bid Item Characteristics Based Sub-HCCIs

HCCI was also computed for various categories of work items, classified according to the divisions outlined in the 2007 SCDOT standard specification for highway construction. Consequently, six sub-HCCIs based on bid item characteristics were identified, encompassing earthwork, bases and subbases, asphalt pavements, maintenance and control of traffic, structures, and incidental construction.

4.1.6 HCCI Forecast

To address the absence of data for a specific period, the research team utilized linear regression and weighted time series analysis methods to create predictive models for HCCI based on historical data. These models can also serve as tools for forecasting future market conditions, assisting in the planning and budgeting of upcoming projects.

a. Linear Regression

In this study, we used linear regression to forecast the Chained index. Linear regression relies on the premise that the dataset can be adequately approximated by a straight line, often referred to as the best-fit line. This method assumes that forthcoming values will align within this linear trajectory. Mathematically, it is represented by a straightforward linear equation, as depicted in Eq. (6) below.

$$Chained\ HCCI_{t,0} = \alpha \times t + \epsilon \tag{6}$$

The equation presented predicts a Chained Index $HCCI_{t,0}$ for year t. The constants α and ϵ are determined through the regression analysis of historical data to establish the relationship between time and HCCI.

b. Weighted Time Series

The Weighted Time Series method assumes that upcoming trends will resemble recent history more than distant past data. Mathematically expressed as:

$$HCCI_{t,t-1} = \frac{\sum_{i=1}^{t-1} i \times HCCI_{i,i-1}}{\sum_{i=1}^{t-1} i} \tag{7}$$

Eq. (7) was used to forecast a Fisher Index $HCCI_{t,\,t-1}$ for a forthcoming year, subsequently utilized to calculate a Chained Fisher Index $HCCI_{t,0}$. This technique predicts one future year of HCCI. This forecasted value becomes part of the historical data, enabling continued forecasting of cost indexes for future years (Jeong et al. 2021).

c. Error Analysis

The efficacy of the forecasting models wre assessed through measures like Mean Absolute Percentage Error (MAPE) and Root Mean Square Error (RMSE). Both methods involve utilizing a segment of historical HCCI data to train the model, while the remaining data serve to evaluate the model's performance. This approach allows for an assessment of how well the model predicts unseen or future data based on its training. MAPE and RMSE were calculated using Eqs. (8) and (9) respectively, as detailed below.

$$MAPE = \frac{100}{n} \times \sum_{i=1}^{n} \left| \frac{HCCI_{i,actual} - HCCI_{i,forecasted}}{HCCI_{i,actual}} \right| \tag{8}$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n} (HCCI_{i,actual} - HCCI_{i,forecasted})^2}{n}} \tag{9}$$

Indeed, models exhibiting lower MAPE and RMSE values demonstrate a greater accuracy in forecasting HCCI values. A lower MAPE indicates a smaller average percentage difference between actual and forecasted values, while a lower RMSE signifies less variability or dispersion between the actual and predicted values, ultimately reflecting higher accuracy in the forecasting models.

4.2 Results

4.2.1. HCCIs

This section presents the key results of HCCI calculations for the statewide HCCIs and sub-HCCIs. Full tables of the results are presented in Appendix B: South Carolina Highway Construction Cost Index.

a. Statewide HCCIs

The year-over-year Fisher index values, representing the yearly inflation rate, are initially calculated for two consecutive years before computing the chained indexes, which depict the accumulated inflation rate over multiple years. Illustrated in Figure 20, the Fisher index exhibited fluctuations before the pandemic, including a singular deflation period in 2017-2018, with the highest recorded inflation rate at 1.15. During the pandemic, the index indicated deflation, marked at 0.83, attributed to the economic disruptions caused by the pandemic. Post-pandemic, the Fisher index displayed a recovery trend, with the inflation rate surging to 1.37 during 2021-2022.

Note that there is a notable difference in the number of bid items across periods. Specifically, the DIB data for the 2022-2023 period contains only 97 bid items, significantly lower compared to other periods. This discrepancy is primarily due to the presence of bid data for only one project in the year 2023, which could potentially impact the accuracy of the Fisher index for this duration. As depicted, the Fisher index for this period experienced a substantial drop when the market indicated an increasing trend after the pandemic.

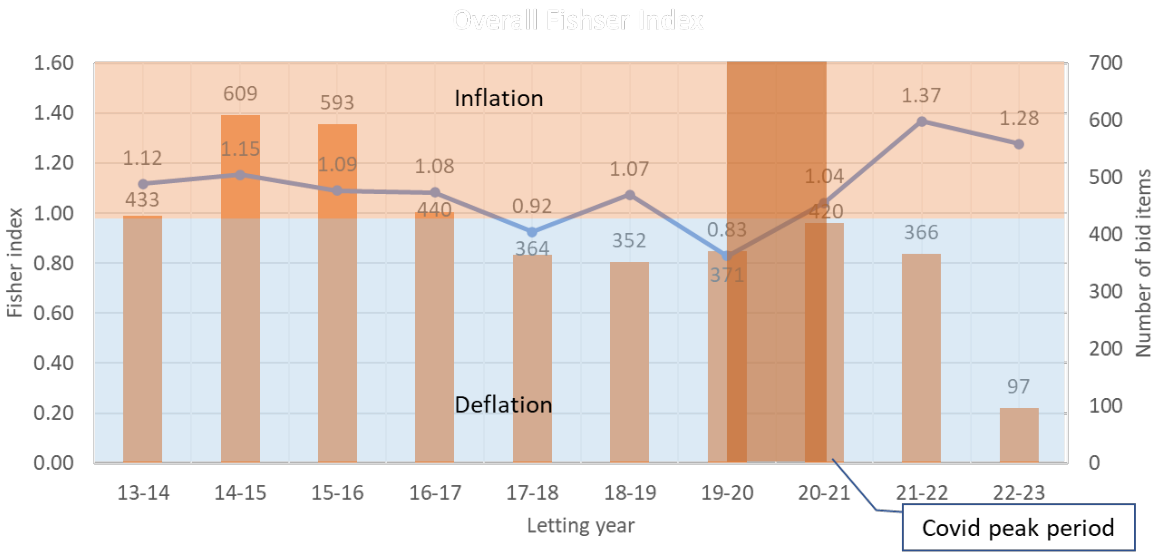


Figure 20. *Statewide Fisher index values for adjacent periods (yearly inflation rate)*

Figure 21 illustrates the cumulative inflation rates, as represented by the statewide chained index values, comparing each year to the base year of 2013. Before the pandemic, these values consistently showed an upward trend, signifying a general increase in costs related to highway construction projects. During the pandemic, a pronounced decline is evident, with the chained index dropping to 1.25 in 2020, followed by a modest recovery to 1.30 in 2021. This substantial decrease indicates a significant reduction in costs associated with highway construction projects during the pandemic. Post-pandemic, the chained index resumes its upward trajectory, suggesting a notable rise in construction costs, possibly reflecting changes in market conditions or economic factors following the crisis.

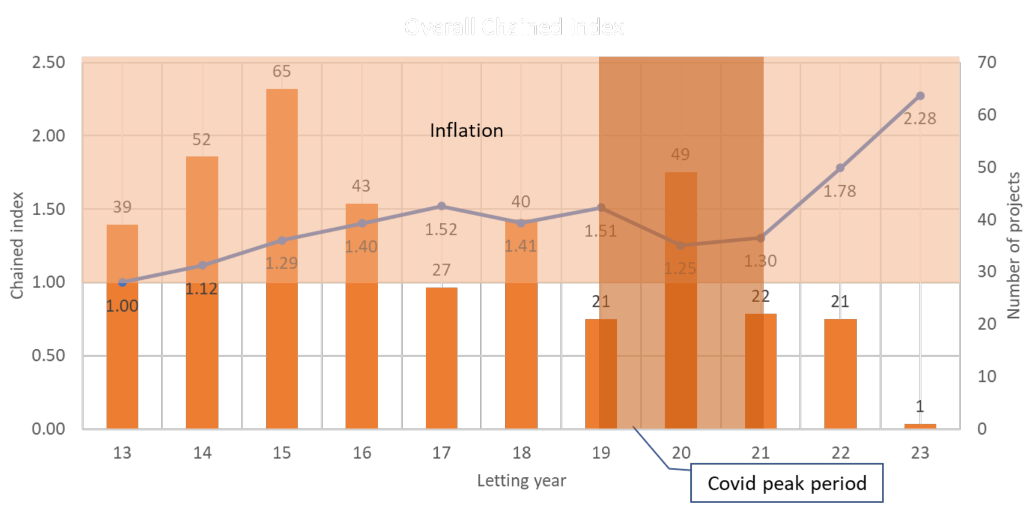


Figure 21. *Statewide chained index values (Accumulated inflated rate compared to the base year of 2013)*

b. Contract Characteristics-Based HCCIs

❖ Project type based sub-HCCIs

Figure 22 displays the chained HCCI index values for various project types across the period, along with the corresponding project counts. The chained index values for all project types consistently demonstrate an upward trajectory, indicating a continual rise in highway construction costs. Intersection improvements generally exhibited lower inflation rates compared to other project types. Widening projects consistently showed the highest values. As noted in the data collection section, the sample size obtained for widening is small, which could potentially affect the accuracy of HCCIs, resulting in anomalous values such as the chained index value of 2.26 derived from only one project in 2018.

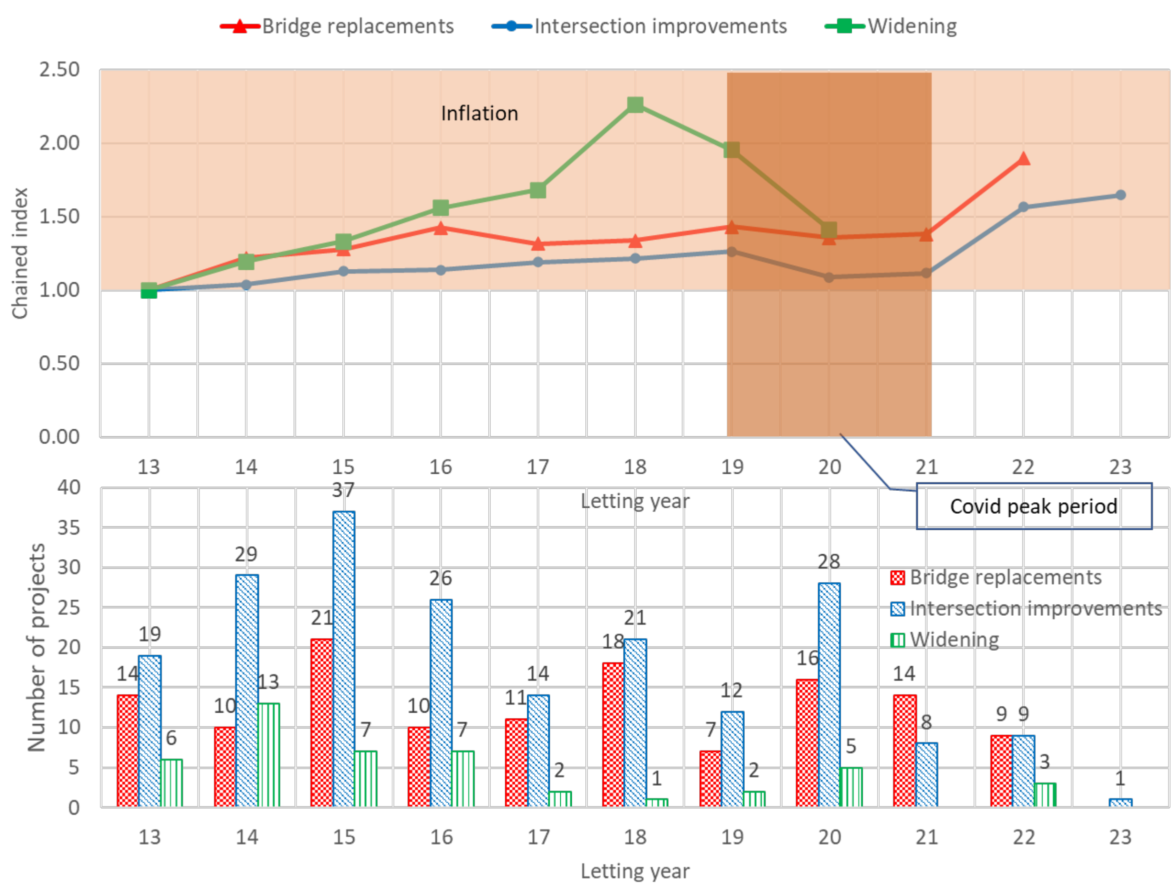


Figure 22. *Project type-based sub-HCCIs*

❖ Contract size based sub-HCCIs

As depicted in Figure 23, the chained index values for projects of different contract sizes consistently exhibit an upward trend, indicating a sustained increase in construction cost. Remarkably, large-sized contracts consistently showcase the highest values compared to the other contract sizes, signaling more pronounced inflation within this category. Interestingly, small-sized mid-sized contracts tended to display a similar trend at comparatively lower inflation rates.

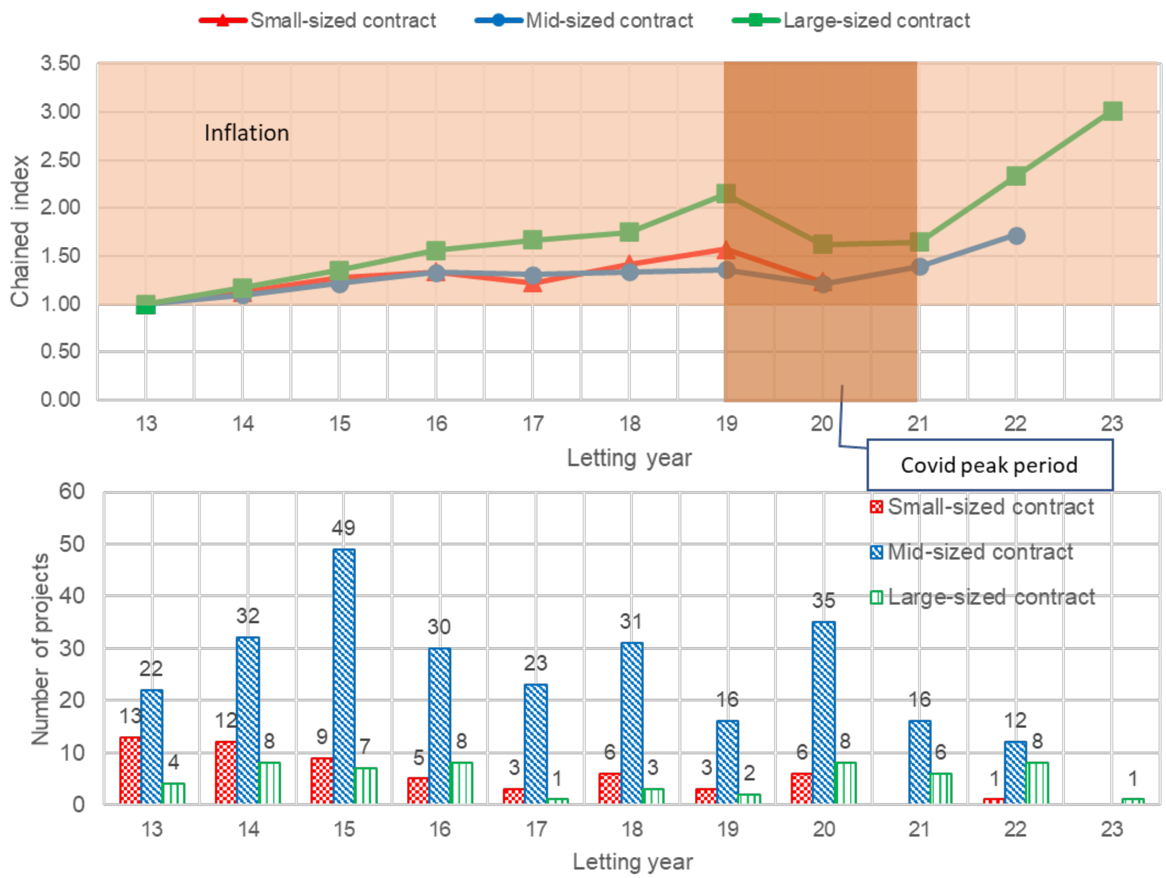


Figure 23. *Contract size based sub-HCCIs*

❖ Scope based sub-HCCIs

As mentioned in the method section, the clustering visualizations strongly indicate that it is appropriate to divide the project into 2-3 clusters (see Figure 19). Thus, the research team calculated and compared the HCCI results for the two options to determine the optimal one.

In Figure 24, the Chained HCCI index is presented with data clustered into two groups. As illustrated, the HCCI values for both clusters nearly doubled over the past decade. Before the pandemic, they exhibited a similar trend, with Cluster 2 experiencing a more substantial decline during the pandemic. However, this cluster managed to recover post-pandemic, reaching a Chained HCCI of 2 by the end of the year 2023.

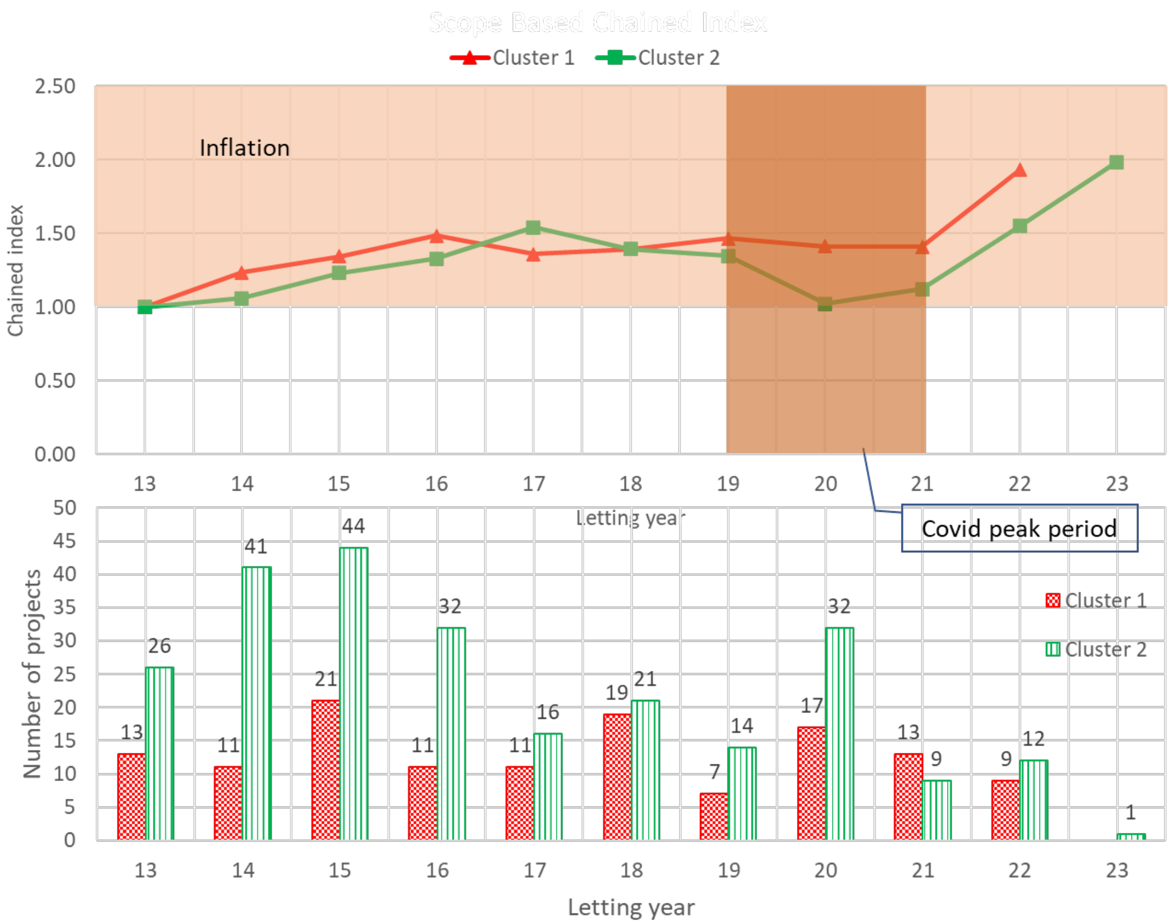


Figure 24. *Scope-based sub-HCCIs at the number of clusters of 2*

Figure 25 depicts the chained index values when the projects were categorized into three groups. There is a noticeable disparity in the values of Cluster 3 when compared to those of the other two clusters. Cluster 3 consistently exhibits the highest inflation rates over the past decade, while Clusters 1 and 2 undergo relatively similar changes during this period. This observation suggests that it might be more suitable to classify the projects into two categories.

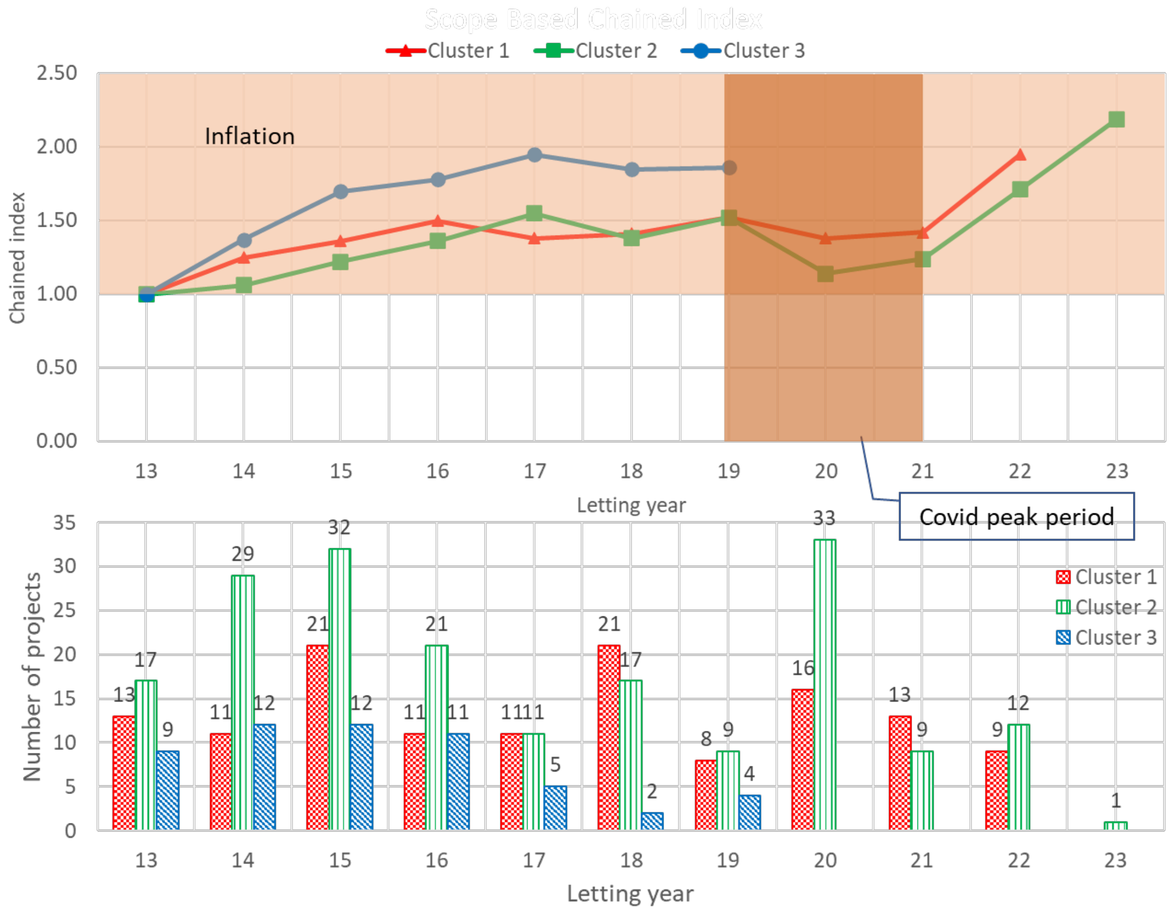
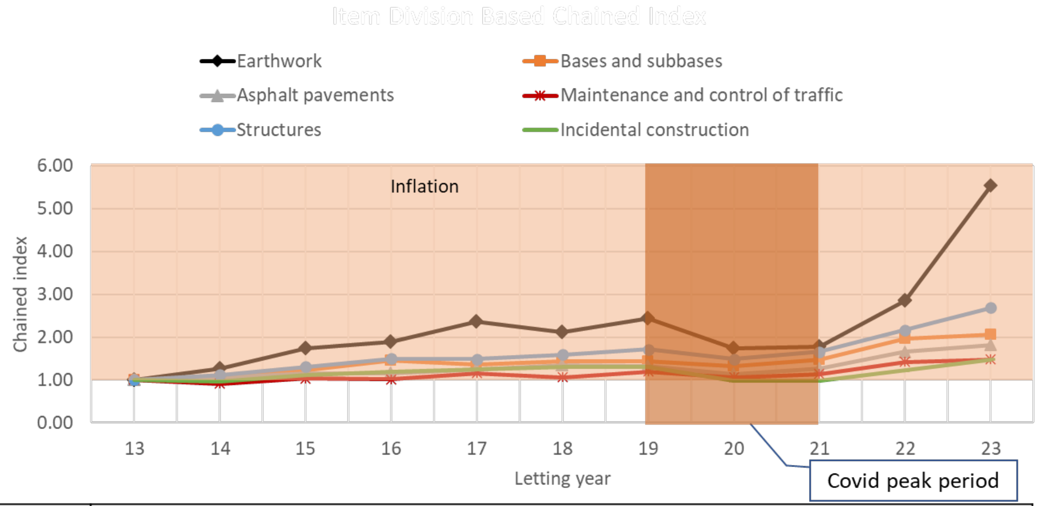


Figure 25. *Scope-based sub-HCCIs at the number of clusters of 3*

c. Bid Item Characteristics-Based HCCIs

This section provides the sub-HCCI values for various bid item divisions over the past 10 years. As shown in Figure 26, the inflation rate for every division consistently shows an upward trajectory. Notably, Earthwork repeatedly stands out with the highest chain index value every year, while the other divisions exhibit comparatively lower inflation rates.



Division	Number of project each letting year										
	13	14	15	16	17	18	19	20	21	22	23
Earthwork	37	50	65	41	27	38	21	47	22	21	1
Bases and subbases	37	45	64	41	27	36	19	49	22	21	1
Asphalt pavements	37	50	65	42	27	36	21	49	22	21	1
Maintenance and control of traffic	38	52	65	42	26	40	21	48	22	21	1
Structures	37	45	65	40	27	37	21	48	22	21	1
Incidental construction	39	51	65	41	27	40	21	47	22	21	1

Figure 26. *Work item division-based sub-HCCIs*

4.2.2. HCCI Forecasting

a. Forecast Error Analysis and HCCI Forecast Method Selection

The historical HCCI data in a certain timeframe was divided into a training dataset and a testing dataset. For instance, the HCCI data from 2013 to 2016 was employed to establish forecasting models. Subsequently, these models were utilized to compute HCCI values from 2017 to 2023. To assess accuracy, the forecasted HCCI values were compared to the actual HCCI values, allowing for the computation of errors.

Figure 27 and Figure 28 illustrate the influence of altering the historical data timeframe on forecasting accuracy, specifically in terms of MAPE and RMSE. These figures offer insights into the connection between the extent of historical information and the accuracy of predictive models. In both MAPE and RMSE measures, the error tends to diminish with an increase in the number of years of data utilized. Notably, the linear regression model outperforms the weighted time series method.

Based on the findings, the research team chose linear regression as the method for predicting HCCI to address missing data (refer to Table 11). The projected statewide HCCIs and sub-HCCIs for the next 15 years (2024-2038) were generated and are outlined in Appendix C: Forecasted HCCI Values.

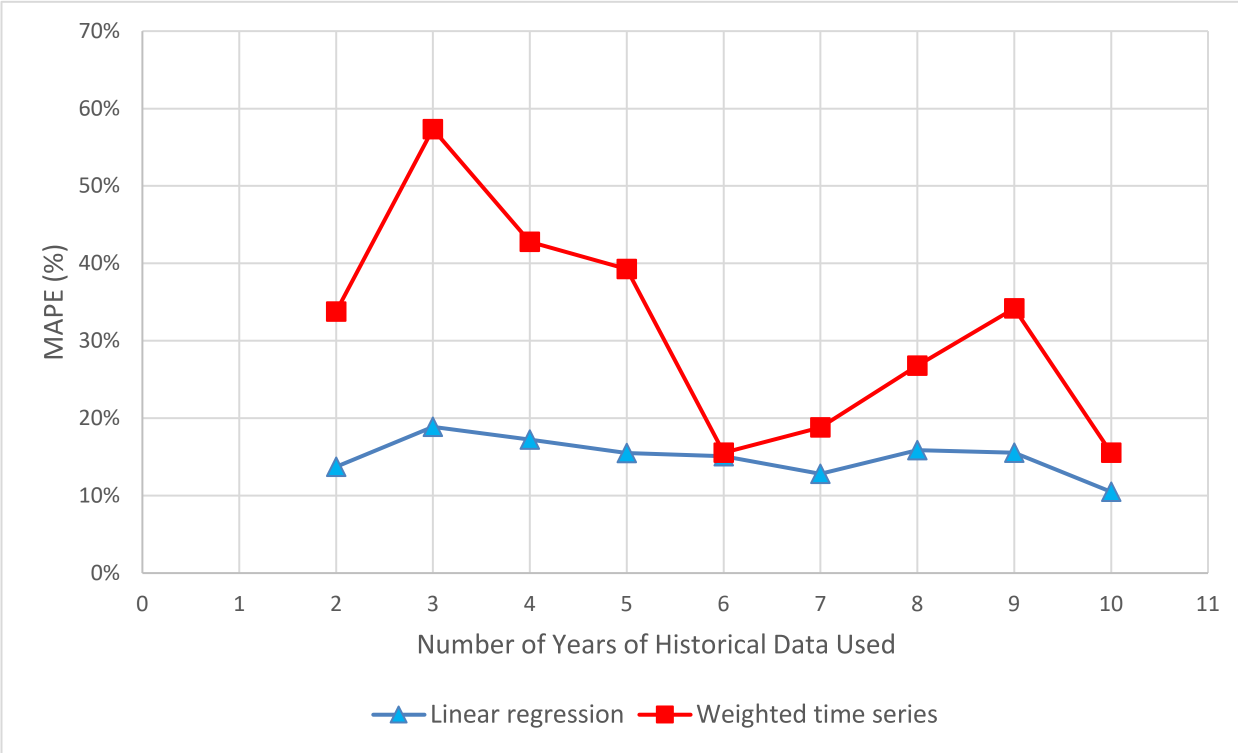


Figure 27. *Alteration in MAPE with the number of years of historical data used*

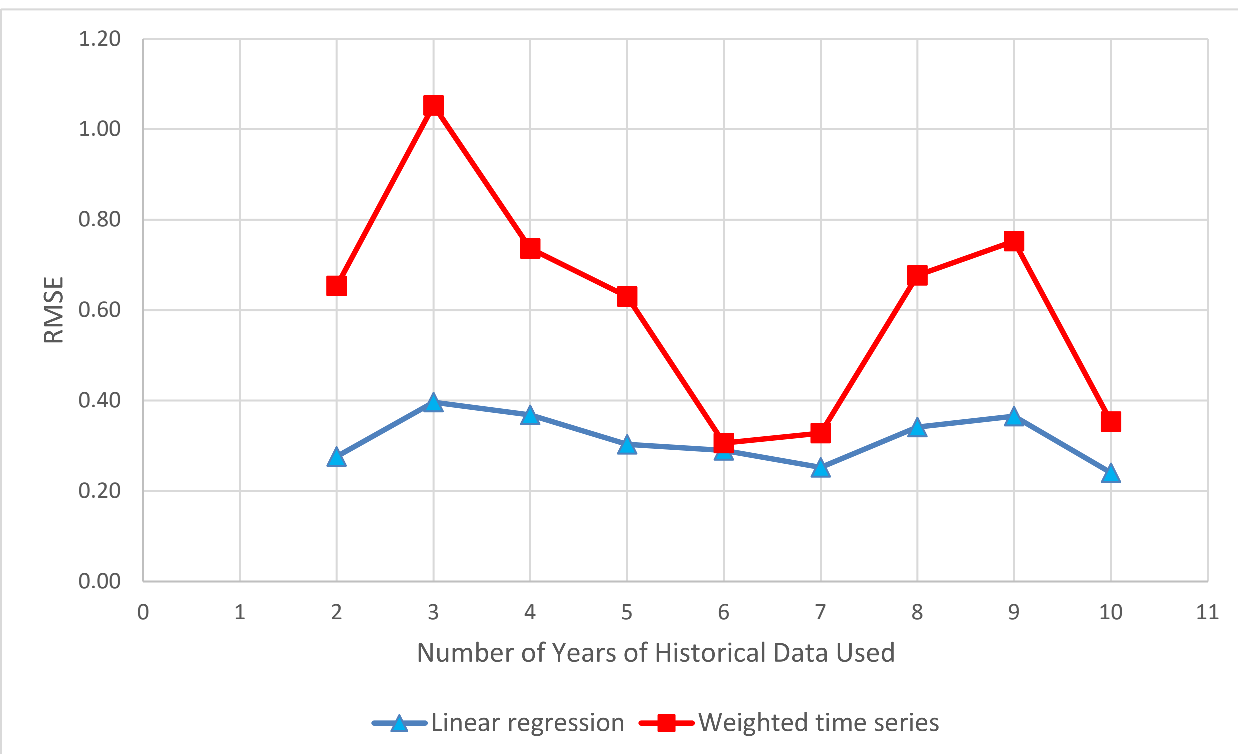


Figure 28. *Alteration in RMSE with the number of years of historical data used*

Table 11. HCCI values forecasted using linear regression

Contract Characteristics Based Sub-HCCIs										
		Project work type			Scope			Contract Size		
Letting year	Statewide HCCI	Bridge replacements HCCI	Intersection improvements HCCI	Widening HCCI	Cluster 1 HCCI	Cluster 2 HCCI	Cluster 3 HCCI	Small-sized contract HCCI	Mid-sized contract HCCI	Large-sized contract HCCI
2013	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
2014	1.12	1.22	1.04	1.20	1.25	1.06	1.37	1.13	1.10	1.17
2015	1.29	1.28	1.13	1.33	1.36	1.22	1.70	1.28	1.21	1.35
2016	1.40	1.43	1.14	1.56	1.50	1.36	1.78	1.34	1.33	1.56
2017	1.52	1.32	1.19	1.68	1.38	1.55	1.95	1.23	1.31	1.67
2018	1.41	1.34	1.22	2.26	1.41	1.38	1.85	1.42	1.33	1.75
2019	1.51	1.43	1.26	1.95	1.52	1.52	1.86	1.57	1.36	2.15
2020	1.25	1.36	1.09	1.41	1.38	1.14	2.27	1.23	1.21	1.62
2021	1.30	1.38	1.12	2.10	1.42	1.24	2.40	1.45	1.39	1.65
2022	1.78	1.90	1.57	2.21	1.95	1.71	2.54	1.50	1.72	2.33
2023	2.28	1.62	1.65	2.33	1.78	2.19	2.67	1.55	1.67	3.01

*Highlighted values indicate predictions from linear regression

4.3 Comparing South Carolina's HCCI to the National HCCI

4.3.1. Comparison Approaches

National Highway Construction Index (NHCCI) is a quarterly chained Finisher index designed to assess the national average fluctuations in highway construction costs over time. The FHWA relies on information from State web-posted winning bids for highway construction contracts. The most current NHCCI data can be obtained at https://www.fhwa.dot.gov/policy/otps/nhcci/. Given that the NHCCI (FHWA 2023) and the statewide HCCI of South Carolina (SCHCCI) in this project used different base years (2003 and 2013, respectively) for chained index calculation, it is crucial to adopt a systematic approach for their comparison. According to the guidance of the FHWA, States and stakeholders must transform the NHCCI into a format that allows for comparison with other indices. Direct comparisons might be misleading due to variations in methodologies and base years used by different indices. To accurately compare the NHCCI with other indices, it is crucial to convert both into percentage changes (year-over-year change), providing a more reliable platform for comparison (FHWA 2021). Besides, Liu et al. (2020) recommended two methods for HCCI comparison, including index trend visualization and statistical methods.

Additionally, FHWA calculated quarterly NHCCI that captures seasonal effects in the national market's conditions. This computation suffers from chain drift bias, especially in cases with significant seasonal fluctuations. For that reason, to mitigate this bias, experts recommend relying on the annual HCCI for a more stable and reliable representation of overall market trends and minimizing the impact of seasonal influences (Liu et al. 2020). When comparing to SCHCCI, our initial step involved calculating the annual NHCCI by averaging the quarterly NHCCI values for each fiscal year. This step allowed us to create an annualized representation for comparison purposes, ensuring a clearer assessment while minimizing the influence of short-term fluctuations before applying the selected comparison approaches.

a. Year-over-year Change Comparison

The year-over-year change in chained HCCI was used to compare NHCCI and SCHCCI. It measures the change in the Chained HCCI from a year to the previous year, reflecting year-over-year fluctuations. This approach serves as a means to portray yearly market fluctuations effectively, as calculated in the equation below.

$$year - over - year\ change,\ YC_{t,t-1} = \frac{CI_{t,0} - CI_{t-1,0}}{CI_{t-1,0}} \quad (10)$$

where *CI* is the chained HCCI; the subscripts *0, t-1,* and *t* designate the base year, the previous year, and the current year, respectively.

b. Chained Index Trend Comparison

The chained index was also used for the comparison purposes. In order to visualize trends for comparing HCCI, it is essential to standardize NHCCI and SCHCCI indices to a common timeframe (e.g., from 2013 to 2023). Specifically, we need to establish an NHCCI trend line with 2013 as the base year, aligning it with the SCHCCI. By computing the year-over-year changes in HCCI for consecutive periods, we can determine the chained index for this newly defined timeframe, as outlined in equation (7) below.

$$Chained\ index, CI_{t\prime,0} = \prod_{k=1}^{t\prime}(1 + YC_{k,k-1}) \quad (11)$$

where the subscripts *0 and t'* indicate the new base year (i.e., 2013) and the current year, respectively.

c. Statistical method

The research team adopted Kendall's Tau and Spearman's rank correlation coefficients to measure the relationship between NHCCI and SCHCCI. Both Kendall's Tau and Spearman's rank correlation coefficients are non-parametric measures used to assess the strength and direction of relationships between two variables. They both work well for monotonic relationships and do not require the variables to follow a specific distribution, making them more robust for non-linear relationships.

- Kendall's Tau correlation coefficient: this coefficient is derived from identifying the number of concordant and discordant pairs within observed data according to (Kendall 1948) and (Temizhan et al. 2022). The Kendall's Tau coefficient value between two variables, X and Y, is given by equation (8):

$$\tau = \frac{C - D}{\frac{n(n-1)}{2}} \tag{12}$$

  where *C* is the number of concordant pairs, *D* is the number of discordant pairs, and *n* is the number of observations.

  This coefficient varies between -1 and +1. The value of 1 indicates a strong positive association or agreement in rankings between variables. This suggests that an increase in one variable tends to be associated with an increase in the other variable, preserving their order. The value of -1 suggests a strong negative association or disagreement in rankings between variables. An increase in one variable corresponds to a decrease in the other variable, inversely preserving their order. The value of 0 implies a weak or negligible association between the rankings of variables. There is little to no consistent relationship in the ranks of the variables.

- Spearman's rank correlation coefficient: this coefficient quantifies the strength of a monotonic relationship between paired data (Fieller and Pearson 1961), computed using the following equation:

$$r = 1 - \frac{6\sum d_i^2}{n(n^2 - 1)} \tag{13}$$

  where *di* is *the* difference between the two ranks of each observation.

  Spearman's rank correlation coefficient also ranges between -1 and 1. The value of 1 indicates a perfect positive monotonic relationship between variables, while the value of -1 implies a perfect negative monotonic relationship between variables. The value of 0 suggests no monotonic relationship between the variables.

The strength of correlation can be assessed using the following guideline for the absolute value of the correlation coefficient (Obilor and Amadi 2018):

- $0 < |\text{correlation coefficient}| \leq 0.4$: weak
- $0.4 < |\text{correlation coefficient}| \leq 0.6$: moderate
- $0.6 < |\text{correlation coefficient}| \leq 1$: strong

To gain insight into the relationship between NHHCI and SCHCCI, the research team also performed hypothesis tests to examine the equality of NHCCI and SCHCCI. Due to the comparison setup, the HCCI data from independent groups (Nationwide vs South Carolina) were designed over the same period (2013-2023) and matched in pairs for each year. Consequently, we conducted both paired and unpaired tests. The unpaired test focused on assessing the overall difference between NHCCI and SCHCCI across all individual observations, while the paired test aimed to uncover the annual discrepancies between NHCCI and SCHCCI. Mann-Whitney U and Wilcoxon signed-rank tests were adopted to test the null hypothesis that there is no significant difference between NHCCI and SCHCCI. Mann-Whitney U (Mann and Whitney 1947) test compares two independent samples, whereas the Wilcoxon signed-rank test (Wilcoxon 1947) compares two paired samples. These are nonparametric alternatives to the unpaired and paired Student's t-tests, respectively. Unlike parametric tests, nonparametric tests do not rely on the assumption of normal distribution for the samples.

4.3.2. Results

Over the decade-long observations depicted in Figure 29, NHCCI and SCHCCI exhibit distinct patterns in their year-over-year changes, revealing noteworthy differences. NHCCI shows higher year-over-year change values for four consecutive periods between 2017 and 2021. Conversely, SCHCCI indicates higher values for the remaining six consecutive periods. Before the pandemic, SCHCCI demonstrates notable percent changes from 2013 to 2017 and during 2018-2019, hovering around 10%, while NHCCI reflects relatively smaller changes, averaging around 5% annually. During the pandemic in 2020, NHCCI records values around -2%, while SCHCCI registers notably larger decreases at approximately -10% to -20%. Following the pandemic, both NHCCI and SCHCCI show significantly increased year-over-year change values compared to pre-pandemic periods. The broader fluctuation range in SCHCCI may be attributed to its sensitivity to localized factors, often reacting strongly to local conditions and market downturns. In contrast, NHCCI, drawing nationwide data, presents a more balanced perspective across states.

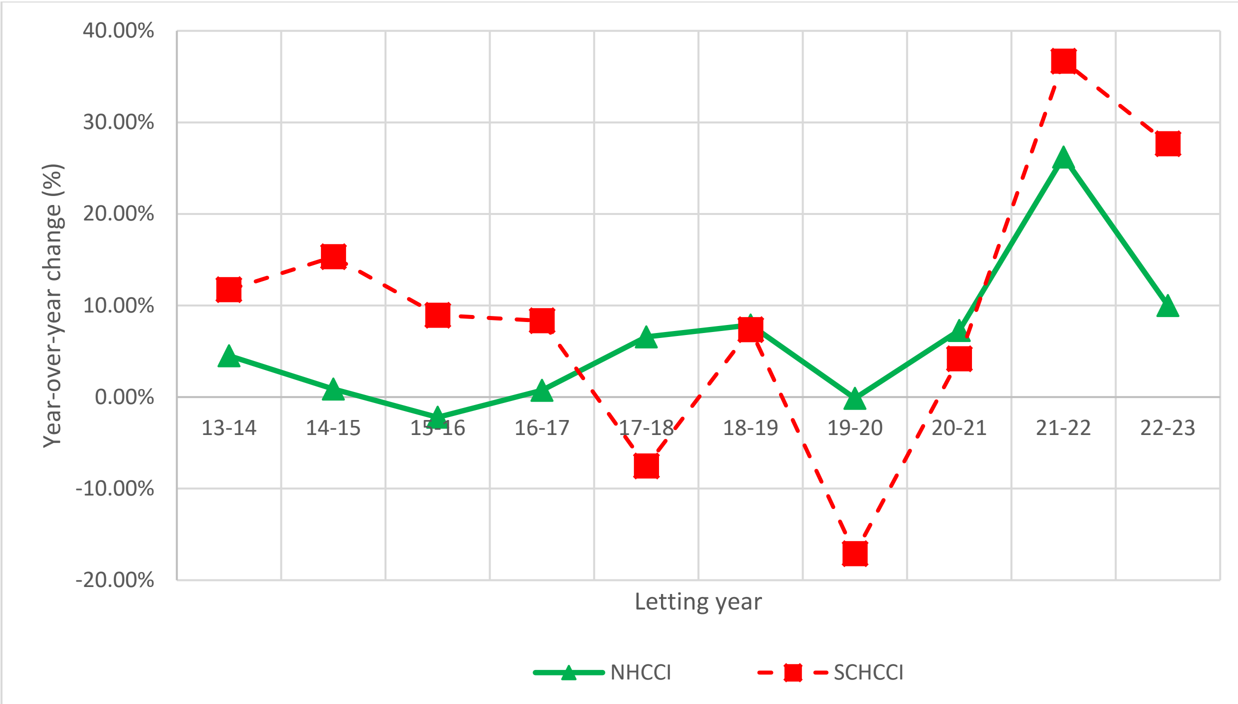


Figure 29. *Year-over-year change in chained NHCCI and SCHCCI (19-20: pandemic)*

Figure 30 illustrates the chained index values for NHCCI and SCHCCI from 2013 to 2023. SCHCCI follows a more erratic growth trajectory characterized by greater fluctuations, whereas NHCCI maintains a steadier, albeit slightly slower, upward trend. Notably, with the exception of the years 2014, 2020, and 2021, where NHCCI and SCHCCI demonstrate similar values, SCHCCI consistently exhibits notably higher values in the chained index.

At the onset of the pandemic, specifically from 2019 to 2020, SCHCCI undergoes a significant decline, reflecting the economic disruptions during that period. In contrast, NHCCI displays a relatively minor decrease during this phase. Post-pandemic, both NHCCI and SCHCCI depict rapid increases in their chained index values, indicating a swift rebound following the economic downturn induced by the pandemic. Overall, the visualization highlights a general alignment between the trends of SCHCCI and NHCCI, suggesting a parallel direction in their movement. However, SCHCCI demonstrates a notably wider range of fluctuations compared to NHCCI.

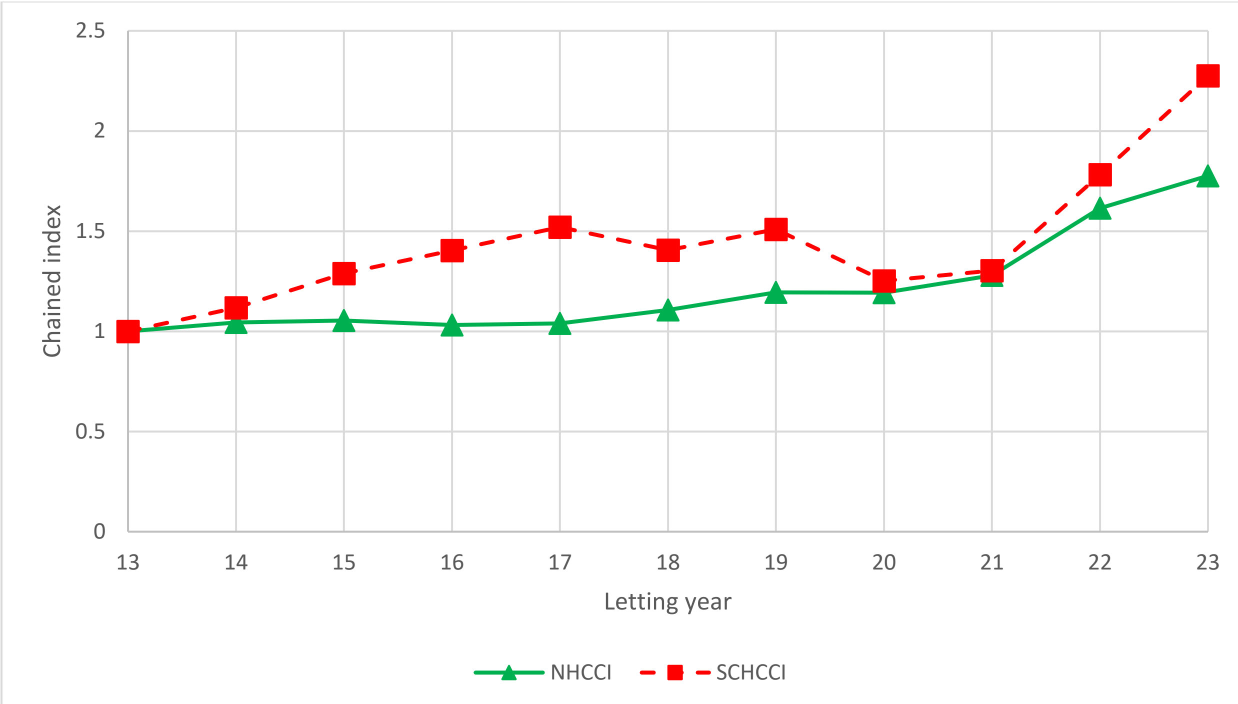


Figure 30. *Trend of NHCCI and SCHCCI (Accumulated inflated rate compared to the base year of 2013)*

Figure 31 presents the outcomes of the testing for Kendall's Tau and Spearman's rank correlation coefficients. The findings reveal a correlation between NHCCI and SCHCCI, with Kendall's Tau correlation being statistically significant at the 0.05 level, and Spearman's rank correlation demonstrating significance at the 0.1 level. Both coefficients, registering values of 0.491 for Kendall's Tau and 0.591 for Spearman's rank, fall within the range of 0.4 to 0.6, indicating a moderate relationship between NHCCI and SCHCCI.

Correlations

			NHCCI
Kendall's tau_b	SCHCCI	Correlation Coefficient	.491*
		Sig. (2-tailed)	.036
		N	11
Spearman's rho	SCHCCI	Correlation Coefficient	.591
		Sig. (2-tailed)	.056
		N	11

*. Correlation is significant at the 0.05 level (2-tailed).

Figure 31. *Results of correlation tests*

As depicted in Figure 32, an increase in NHCCI tends to correspond with a rise in SCHCCI. While the correlation is evident, variations or fluctuations in the pattern may occur due to other influences or factors affecting the relationship, such as local conditions of SCHCCI or bid data used for calculating SCHCCI, making the construction cost in SC inflate at a significantly higher rate than the national average.

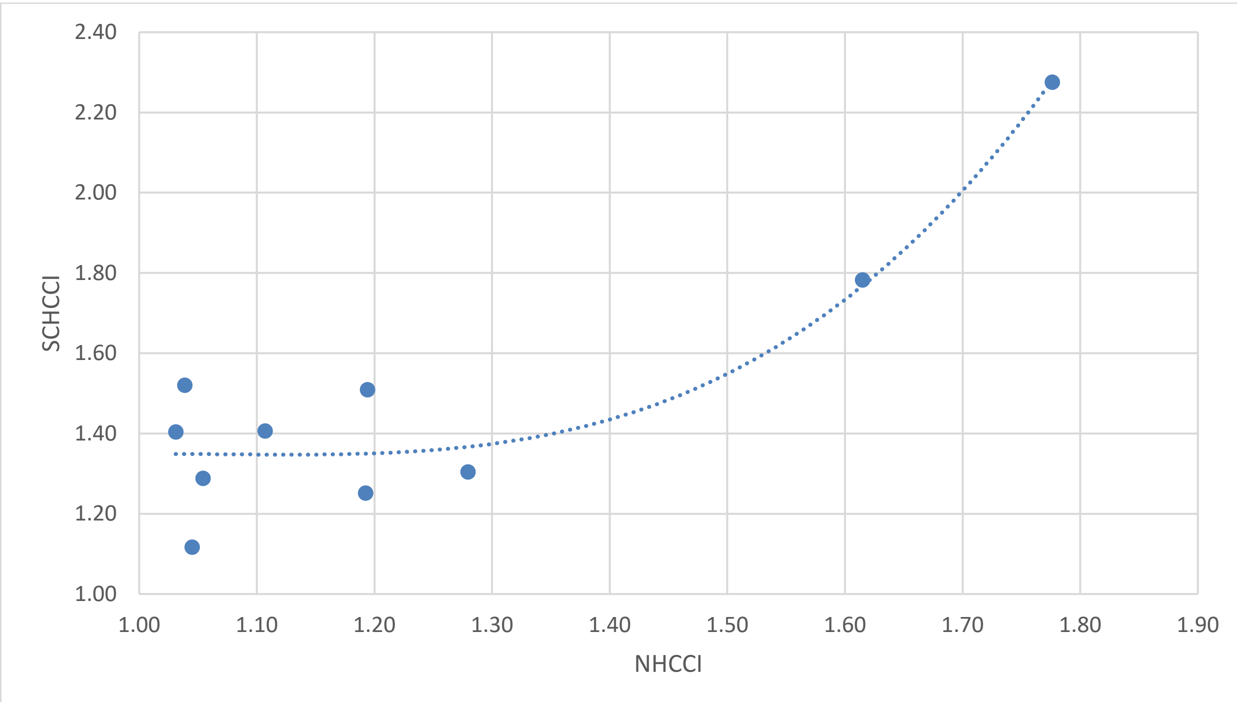


Figure 32. *Monotonic increasing relationship between NHCCI and SCHCCI*

Figure 33 displays the outcomes of the Mann-Whitney U and Wilcoxon signed-rank tests. Both tests produce p-values exceeding 0.05, specifically at 0.28 and 0.386, respectively. Consequently, we fail to reject the null hypothesis that there is no statistically significant difference between NHCCI and SCHCCI. In simpler terms, we lack adequate evidence to assert that the HCCI values statistically differ between Nationwide and South Carolina.

Hypothesis Test Summary

	Null Hypothesis	Test	Sig.[a,b]	Decision
1	There is no statistically significant difference between NHCCI and SCHCCI	Mann-Whitney U Test	.280[c]	Retain the null hypothesis.

a. The significance level is .050.

b. Asymptotic significance is displayed.

c. Exact significance is displayed for this test.

Hypothesis Test Summary

	Null Hypothesis	Test	Sig.[a,b]	Decision
1	There is no statistically significant difference between NHCCI and SCHCCI.	Wilcoxon Signed Rank Test	.386	Retain the null hypothesis.

a. The significance level is .050.

b. Asymptotic significance is displayed.

Figure 33. *Results of Mann-Whitney U and Wilcoxon signed-rank tests*

5. Cost Estimating Modeling

The focus of this chapter is to present the approach and results of the multiple modeling efforts undertaken for cost estimating during both the planning. While the target variable is the total project cost, several input variables were explored for being significant to the project cost. Cost estimating models were separately developed for widening, bridge replacement, and intersection improvement projects. The primary modeling approach focused on using the average of the three lowest bidders from the past projects in each of these project categories for predicting the project cost. Two types of models are proposed; one is a deterministic model where a single project cost is predicted, whereas the other is a probabilistic model where a range for the project cost is predicted with a distribution. To do so, linear regression and Neural Network prediction approaches are utilized. This chapter also presents a brief discussion on the preliminary cost estimating tool (PCET) that is developed as a project deliverable based on the models described in this chapter.

5.1 Planning-level cost estimate modeling

Cost estimates developed during the planning phase of a transportation project are derived based on very few project parameters for which data is available. These estimates are not meant to be highly accurate but are important for further project planning and budgeting purposes. Several input parameters are explored to be included in the planning-level cost estimate modeling in this research. These include: (1) SCDOT district #, (2) Number of project working days, (3) Year of letting, (4) SCDOT's HCCI, (5) Project sub-type, (6) Project length (miles), (7) Bridge length (ft), (8) Average shoulder width, (9) Terrian type, (10) Functional class, (11) Number of existing lanes, (12) Number of improved lanes, (13) Average side slope, (14) Pavement type, and (15) Urban/rural. Data may not be available for all these project parameters during the planning phase, but these parameters were nevertheless explored for their significance on project cost. Not all this data was readily available with SCDOT or other repositories. A significant effort was put into gathering as much data as possible for these parameters for all the past projects. The numbers of past projects of each category included in the model development effort are presented in Table 12. Total project cost is predicted based on the average of the three lowest bidders from past projects. The following sections present the results of the models explored in this study for each of the three project categories.

Table 12. *Modeling database features*

Project Category	# of Projects in the Database
Bridge Replacement	130
Intersection Improvement	204
Widening	46

5.1.1 Linear Regression Modeling: Input Parameter Screening

Appendix D presents all the individual linear regression models explored with numerous combinations of input parameters used to predict total cost of widening (Appendix D-1), bridge replacement (Appendix D-2), and intersection improvement (Appendix D-3) projects. Appendix D specifically highlights the input parameters considered, model performance (measured using R2, R2-adjusted and R2-predicted), and the analysis of variance that indicates the significance of the input parameters for each model presented. It should be noted that R2-predicted is a key measure of model's prediction accuracy and it is important that this measure be as high as possible to be able to rely on any model for future predictions. "Ave_3bid" parameter in all the models in Appendix D is the target parameter which is the average cost of the three lowest bidders. For each project type, variables that maximize the prediction model's accuracy are discussed in the following sub-sections. After defining these variables, the cost estimation tool including user-defined desired accuracy is developed.5.1.1.1 Widening projects

The scope of widening projects in the database varied considerably warranting to keep track of the project sub-types. It was not straightforward to categorize widening projects into different sub-types, but the proposal description, along with the bid items, were scrutinized to categorize them as comprising the scope presented in Table 13.

Table 13. *Scope variation in different widening projects*

Sub-categories	Count of Sub Type
Bridge Extension and Removal	21
Excavation and Pavement Treatment	1
Furnish and Install Wire	2
Intersection Improvement	1
Safety Section Improvement	6
Traffic Control and Clearing	5
Traffic Signal Improvement	5
Utility Relocation	5
Grand Total	**46**

As can be seen from Appendix D-1, a total of 10 models were explored using multiple combinations of input variables. The models presented in Appendix D-1 were iteratively modified to include those input parameters that are significant to maximize the prediction accuracy (R2-predicted). There were a few past projects that were identified as outliers which were affecting the model performance. These projects were deliberately removed from the project database to ensure higher model accuracy. It should be noted that there could have been unique circumstances that may have impacted the project costs to be somewhat extreme in these outlier projects and therefore they were deemed not fit to be used for making cost predictions for future projects. It is however important to not remove too many projects as outliers as the project category features may be lost, resulting in a model that is not suitable for the variety of projects in a particular project category.

Figure 34 presents the best performing model with a R2-predicted value of 83.76%. The significant input parameters highlighted in Figure 34 include SCDOT's HCCI for widening projects, road length, average side slope, average shoulder widened, and number of improved lanes. While HCCI is separately predicted using the model described in a previous chapter of this report, data for all other input parameters should be available during the planning stage of widening projects. A R2-predicted value of ~84% is deemed acceptable for a planning-level estimate which is not expected to be highly accurate.

Regression Analysis: Mean_Bid versus Number of Working days from SCD, Road Length (mile), Average Slope, Avg. Shoulder Widened (ft)

Regression Equation

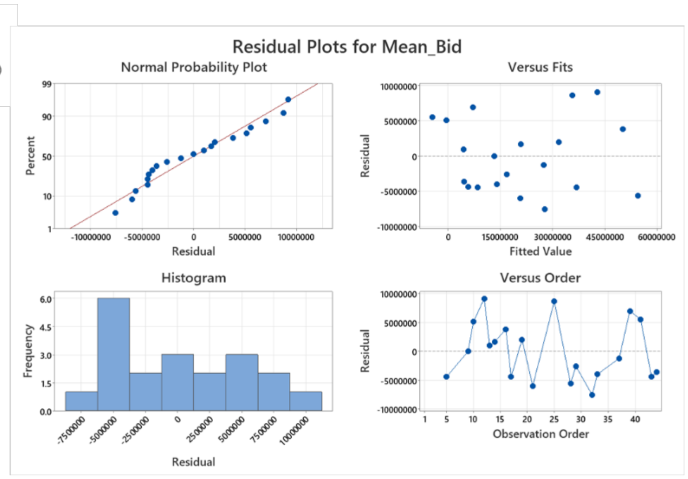
Mean_Bid = -15358341 + 51954 Number of Working days from SCD + 3357499 Road Length (mile) - 297876 Average Slope + 787461 Avg. Shoulder Widened (ft)

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
5821621	91.49%	89.22%	83.76%

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	4	5.46596E+15	1.36649E+15	40.32	0.000
Number of Working days from SCD	1	1.13948E+15	1.13948E+15	33.62	0.000
Road Length (mile)	1	7.53531E+14	7.53531E+14	22.23	0.000
Average Slope	1	5.47910E+12	5.47910E+12	0.16	0.693
Avg. Shoulder Widened (ft)	1	1.82066E+13	1.82066E+13	0.54	0.475
Error	15	5.08369E+14	3.38913E+13		
Total	19	5.97433E+15			



After removing outlier IDs 30 & 2

Figure 34. *Best performing deterministic linear regression model for widening mean bidder price estimation*

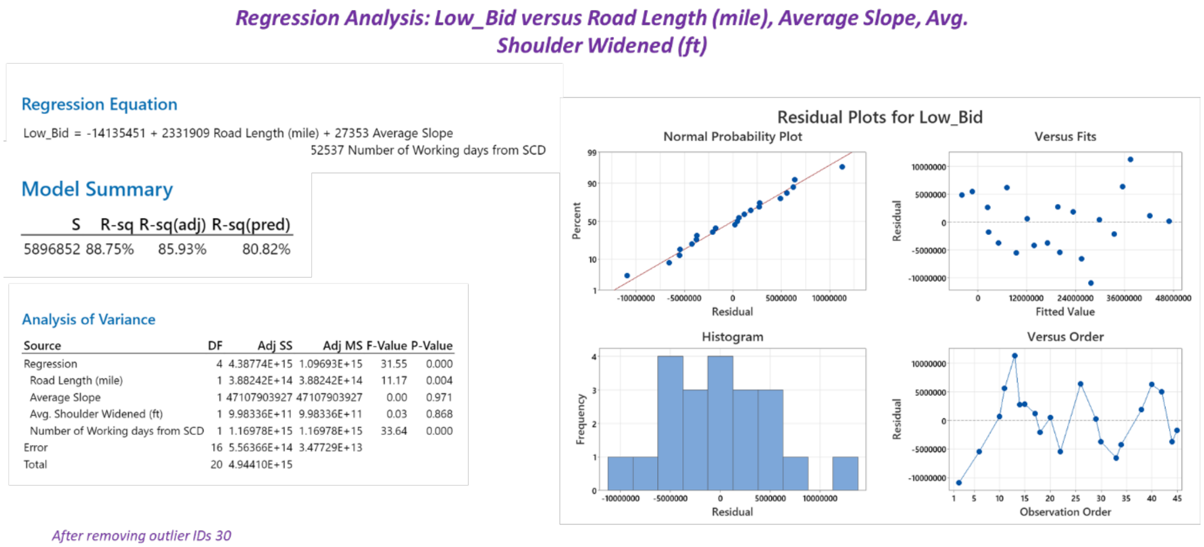


Figure 35. *Best performing deterministic linear regression model for widening lowest bidder price estimation*

5.1.1.2 Bridge replacement projects

Similar to the widening projects, bridge replacement projects are also categorized based on their scope using bid items and proposal description for the past projects. The sub-categories presented in Table 14 are considered as part of the cost estimate modeling.

Table 14. *Sub-categories of bridge replacement projects based on scope and description*

Row Labels	Count of Sub Type
Clearing & Grubbing, and Pavement Marking	1
Clearing & Grubbing, and Pavement Marking, and constructcurb and gutter	1
Pavement Marking, Traffic control and construct bike lane	1
Removal and Disposal of Existing Bridge and Construct Concrete Sidewalk	16
Removal and Disposal of Existing Bridge and construct curb and gutter	1
Removal and Disposal of Existing Bridge, clearing and construct curb and gutter	2
Removal and Disposal of Existing Bridge, Traffic Control	46
Removal and Disposal of structural obstacles and construct curb and gutter	1
Removal & Disposal of existing pavement and construct curb and gutter.	1
Traffic control, and clearing and grubbing	8
Traffic control, and clearing and grubbing, curb, and gutter	4
Traffic control, clearing and grubbing, and curb and gutter	3
Traffic control, clearing and grubbing, construct sidewalk	1
Traffic Control, Clearing, Pipe installing, and Concrete Sidewalk	1
Traffic Control, Clearing, Pipe installing, and Curb and gutter	2
Traffic control, installing pipe, and clearing and grubbing	1

Traffic control, installing pipe, and clearing and grubbing, construct sidewalk	6
Traffic control, installing pipe, and clearing and grubbing, curb, and gutter	33
Utility Staking and Clearing & Grubbing and Pavement Marking	1
Grand Total	**130**

As can be seen in Appendix D-2, a total of seven linear regression models were used to iteratively arrive at a reasonably performing model. Figure 36 presents an acceptably performing regression model after removing five outlier projects with a R2-predicted value 71.50%.

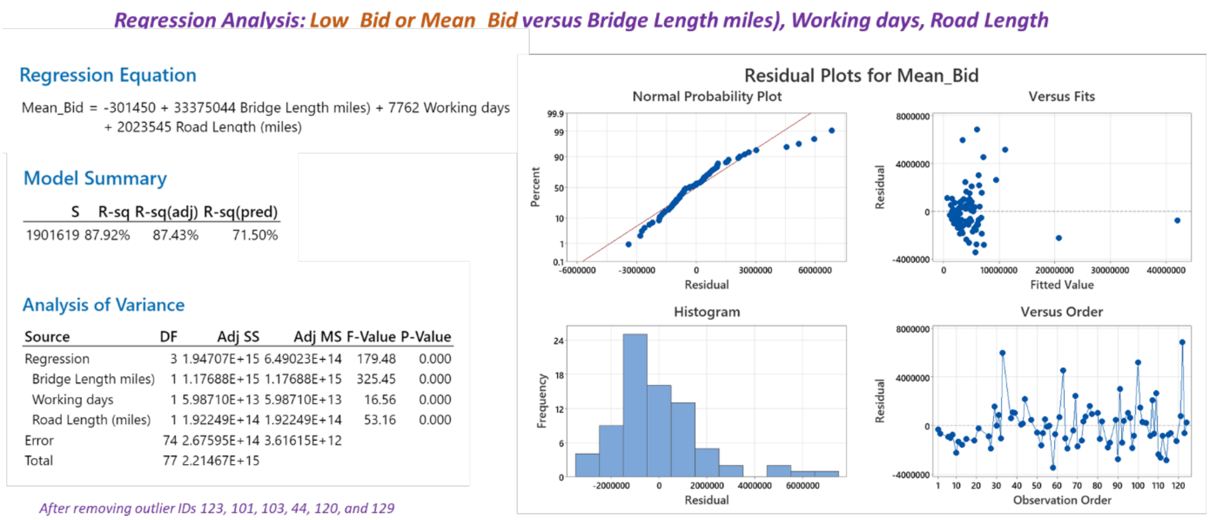


Figure 36. *Best performing deterministic linear regression model for bridge mean bid price estimate*

5.1.1.3 Intersection improvement projects

Past intersection improvement projects are initially categorized based on the scope from the bid item data and proposal description. Table 15 presents the different types of scope defined within all the intersection improvement projects data was made available for. As can be noticed from Table 15, the majority of the projects are in just four sub-categories, which are isolated and included in further analysis to make for a more meaningful interpretation of the past project data. Table 16 presents the refined database that is further analyzed considering just four sub-categories of intersection improvement projects.

__Table 15.__ Sub-categories of past intersection improvement projects based on scope

Project Sub-categories	# of Projects
Clearing, Traffic Control	1
Removal and disposal of existing pavement, Traffic Control	7
Removal and disposal of existing pavement, Traffic Control	1
Removal and disposal of existing pavement, Traffic Control, Clearing	69
Traffic Control, Clearing	4
Traffic Signal Installation	30
Traffic Signal Installation, Clearing	1
Traffic Signal Installation, Removal and disposal of existing pavement, Traffic Control	2
Traffic Signal Installation, Removal and disposal of existing pavement, Traffic Control, Clearing	77
Traffic Signal Installation, Traffic Control	1
Traffic Signal Installation, Traffic Control, Cleaning	11
Grand Total	**204**

__Table 16.__ Four dominant sub-categories of past intersection improvement projects

Project Sub-categories	# of Projects
Removal and disposal of existing pavement, Traffic Control, Clearing	69
Traffic Signal Installation	30
Traffic Signal Installation, Removal and disposal of existing pavement, Traffic Control, Clearing	77
Traffic Signal Installation, Traffic Control, Cleaning	11
Grand Total	187

Appendix D-3 presents all the individual regression models that were iteratively developed to predict total projects of intersection improvement projects considering various combinations of input parameters. A total of 15 models were developed with varying accuracies with the goal of improving the R2-predicted value of the model by including significant input parameters that influenced project cost. Figure 37 presents the results from the best performing model with the majority of the input parameters being significant. As can be seen from Figure 37, a R2-predicted value of about 61.14% was achieved, which is reasonable for a planning-phase cost estimate.

Regression Analysis: Mean_Bid versus Length, Working Days, HCCI

Regression Equation

Mean_Bid = -2610726 + 2128429 Length + 6436 Working Days + 1843649 HCCI

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
712441	64.37%	63.67%	61.14%

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	3	1.41211E+14	4.70704E+13	92.74	0.000
Length	1	6.24411E+13	6.24411E+13	123.02	0.000
Working Days	1	1.93618E+13	1.93618E+13	38.15	0.000
HCCI	1	5.87653E+12	5.87653E+12	11.58	0.001
Error	154	7.81662E+13	5.07573E+11		
Total	157	2.19377E+14			

After removing outlier IDs 112, 20, 77, 146, 137, 99, 153, 66

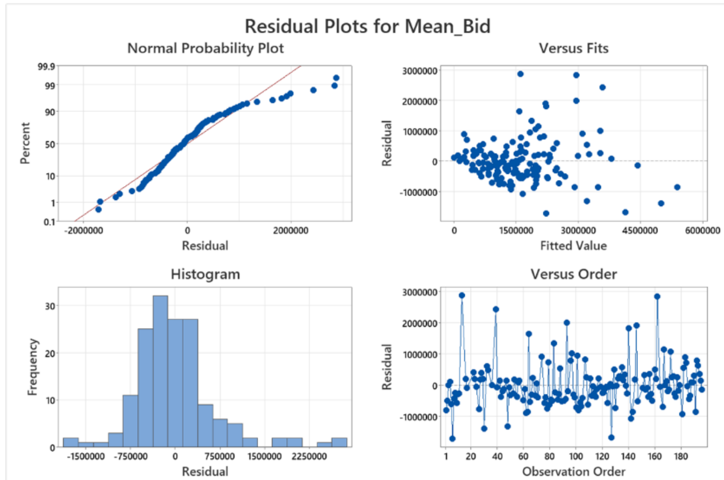


Figure 37. *Best performing deterministic linear regression model for intersection improvement project cost estimate prediction*

5.1.2 Linear Regression Modeling: Deterministic

In the last section, for each project type, variables that maximize the regression model's accuracy are defined. For intersection improvement projects, the variables are Number of Working Days, HCCI, and Road Length. For bridge projects, parameters include Number of Working Days, Road Length, and Bridge Length. As for widening projects, these variables consist of Number of Working Days, Average Shoulder Widened, and Road Length.

To construct a deterministic regression model, users specify the minimum acceptable accuracy threshold. The dataset is divided into two parts: 80% is used for training the model, and the remaining 20% is set aside for validation. Subsequently, a regression model is developed using the training data, and the testing accuracy of the model is assessed using the R-squared metric. If the model does not meet the minimum accuracy threshold defined by the user, it is discarded and replaced with a new model until the desired accuracy level is achieved.

For Intersection Improvement projects, the accuracy rate stands at 69%. Figure 38 presents the predicted value alongside their corresponding real value for validation dataset projects that have not been used for model training. According to this figure, the regression model for intersection projects exhibits the ability to capture the trend of bid values. This predictive model mostly overestimates bid values of less than $5 Million and underestimates the remaining bids.

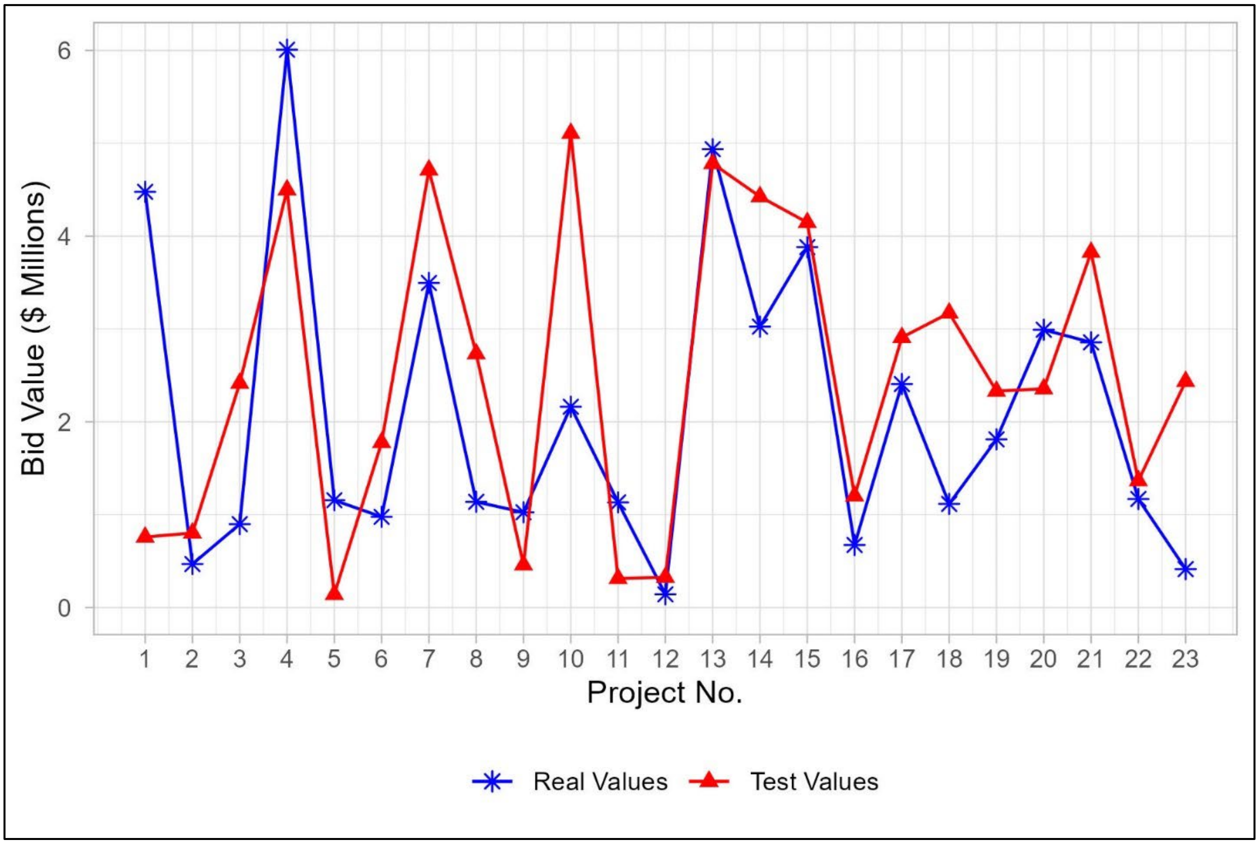


Figure 38. *Validation for deterministic regression cost estimation for intersection projects*

Bridge Replacement projects demonstrate an accuracy of 82%. As illustrated in Figure 39, the trend in bridge project bid values are captured, and the predicted values are exhibiting a good accuracy compared to the corresponding real bid values.

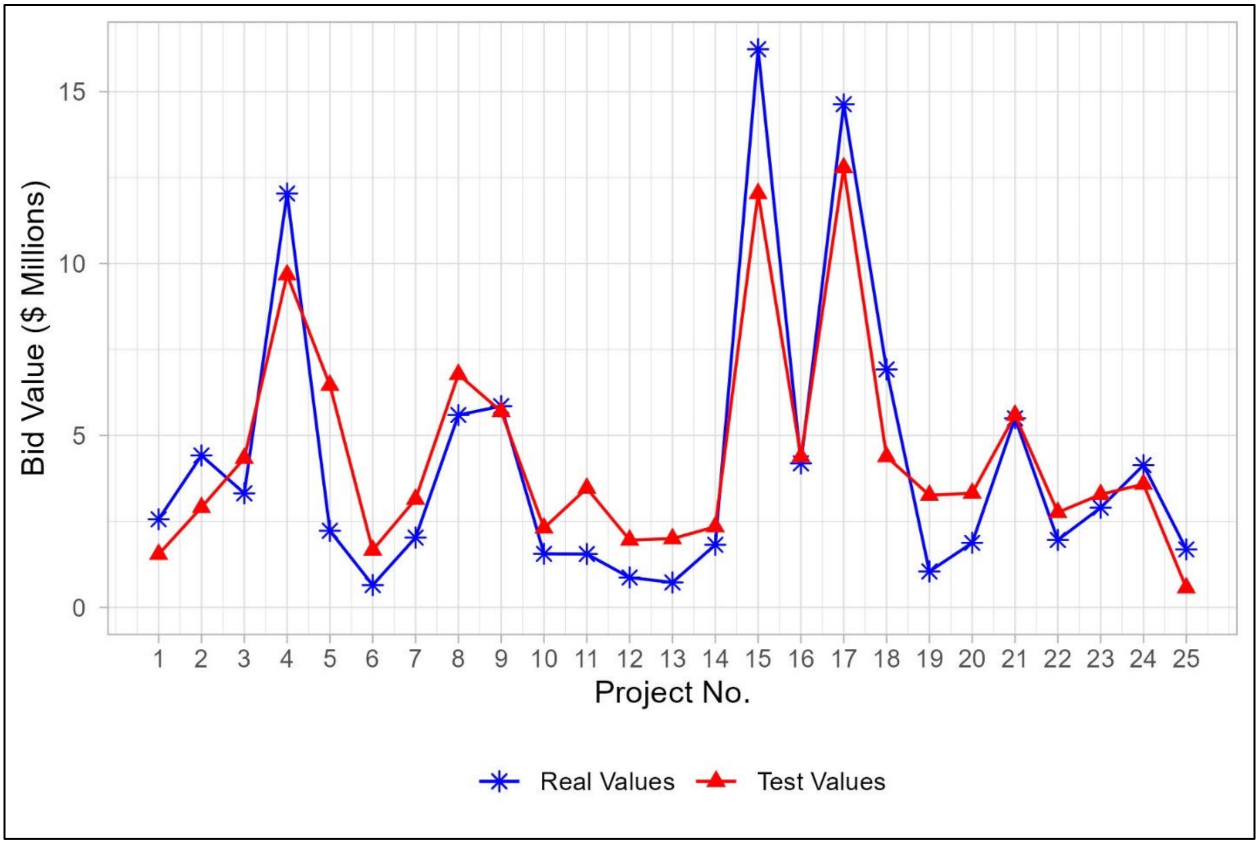


Figure 39. *Validation for deterministic regression cost estimation for bridge projects*

Moreover, Widening Projects achieve an accuracy of 78%. Despite capturing the trend, the difference between the predicted values and real bid values is discernible, as can be seen from Figure 40. This is due to the limited number of available data for training and also their highly diverse bid values.

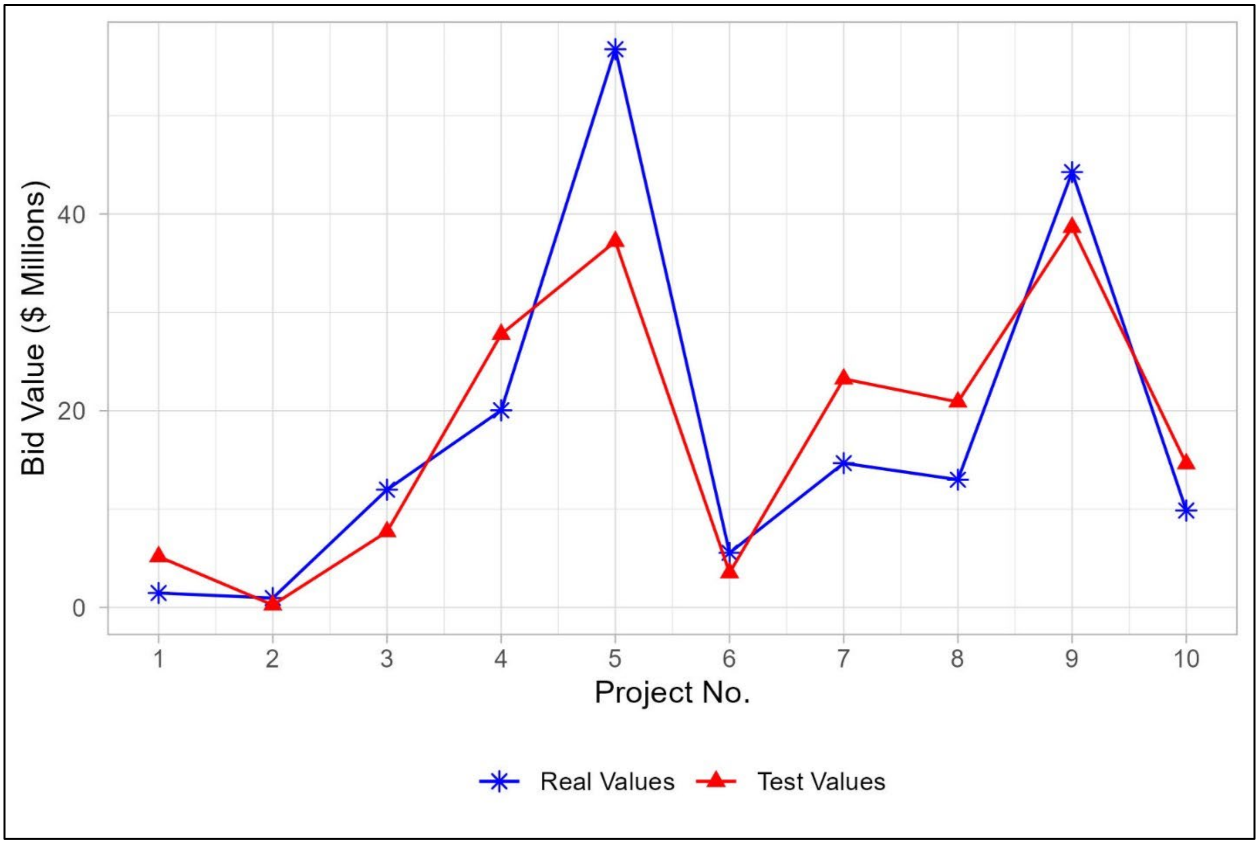


Figure 40. *Validation for deterministic regression cost estimation for widening projects*

5.1.3 Linear Regression Modeling: Probabilistic

The conventional approach that is still practiced by many SHAs to account for the unknowns is to add a fixed percentage of base estimate as contingency cost. The downside is that the required contingency should not be proportional to the project cost, but it should be truly reflective of the risk involved in the estimate. The conventional approach is criticized by scholars and alternative approaches have been explored (Baccarini, 2006; Gardener et al., 2017). Risk-based estimating combines traditional estimating for known work items with risk analysis techniques for uncertain work items. The Monte Carlo simulation approach is a popular technique for risk-based planning wherein the uncertain input variables (or work items/cost) are assumed to follow a certain probability distribution (e.g. normal). The output (i.e., project cost) prediction model is run multiple times (in thousands typically) with different random values each time for the uncertain input variable to generate a possible range of output values along with the probability of each result occurring. Examples of project uncertainties at the planning phase include insufficient right of way knowledge, utilities, environmental mitigation, traffic control challenges, inflation, and unforeseen events/changes. The downside of Monte Carlo simulation is that uncertain input variable is assumed to follow a certain probability distribution, which may not be the true case.

Another downside is that each uncertainty needs to be specifically identified and modeled in the simulation which may be challenging to do in the planning phase.

Alternatives to Monte Carlo simulation exist in the literature. For example, Garderner et al. (2017) successfully employed bootstrap sampling for risk-based estimating of a cost range in the planning phase. This approach seems highly promising and aptly suitable for this proposed study. A bootstrap dataset is a subset of the original dataset of historic projects identified for the analysis. A certain percentage of the original dataset of projects is identified for model building (through regression or ANNs) in each iteration to predict a range of project costs as outputs from multiple iterations combined. Subsequently, probability values can be easily assigned to indicate the probability (or likelihood) of the project cost to be less than a certain value or to be within a certain range (see Figure 41). This approach enables establishing the range of project costs to assign contingencies in a rational manner.

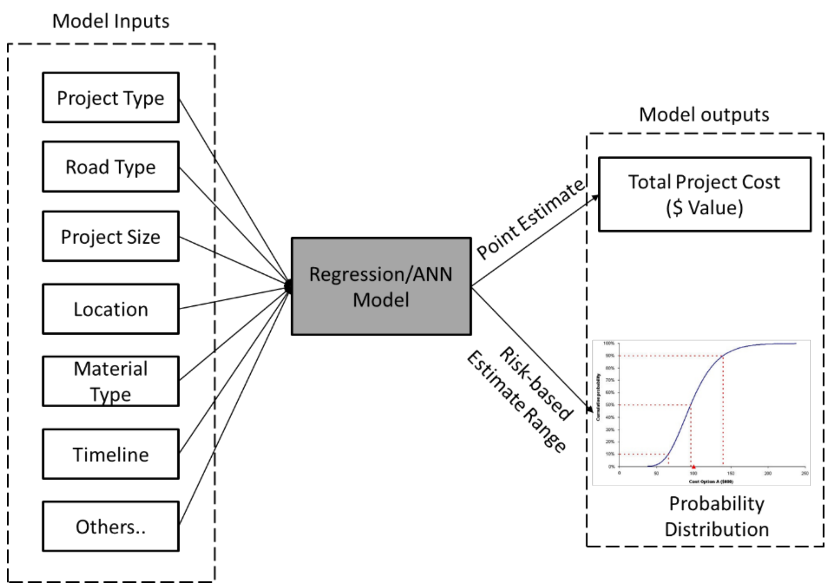


Figure 41. *Proposed Statistical Modeling Architecture*

To create a probabilistic linear regression model, users specify the desired number of models and a minimum acceptable accuracy threshold. The dataset is divided into two parts, validation and model building, by which 85% of the dataset is allocated for building models and the remaining 15% is set aside for the final validation assessment. The model-building dataset is used to generate multiple datasets using the bootstrap sampling approach. In all these generated datasets, 80% is allocated for training each model, and the remaining 20% is used for testing accuracy.

Then, linear regression models are developed using the training data, and the testing accuracy of each model is evaluated by the R-squared metric. Models failing to meet the user-defined accuracy threshold are discarded and are replaced with new ones until the desired accuracy level is achieved. After producing the desired number of models meeting the user-defined accuracy, the validation accuracy of each model is assessed using the validation dataset including 15% of the original dataset. Models in the top 50th percentile of validation accuracies are ultimately selected for the final probabilistic model.

To report output for predicting bid values from the selected models with top 50 percentile validation accuracy, the bid value will be reported. From these models, we gathered the average, minimum, and maximum predicted bid value, along with the average of top 20th percentile responses, for reporting in the planning-level cost estimating tool (PCET).

To generate models for each project, 20 models with a minimum testing accuracy of 70% are specified; therefore, all models have a minimum testing accuracy of above 70%. For Intersection Improvement projects, the validation accuracy spans from 40% to 56% among the constructed models. Figure 42 demonstrates the range of model outputs and the average of the top 20th percentile alongside their corresponding real value for validation dataset projects that were separated for validation purposes and were not used for training and testing. In this figure, each boxplot presents the range of predicted values of all models for a project, including minimum, first quartile (Q1), median, third quartile (Q3), and maximum values. It also excludes outlier predicted values by only considering the range from Q1-1.5(Q3-Q1) to Q3-1.5(Q3-Q1). The advantage of probabilistic models over deterministic ones is that they can provide a range for the bid value prediction. According to Figure 42, the probabilistic regression model has captured the trend of bid values for intersection projects, and in some predictions, the real value falls within the prediction ranges. However, the model overestimated most of the projects' bid values.

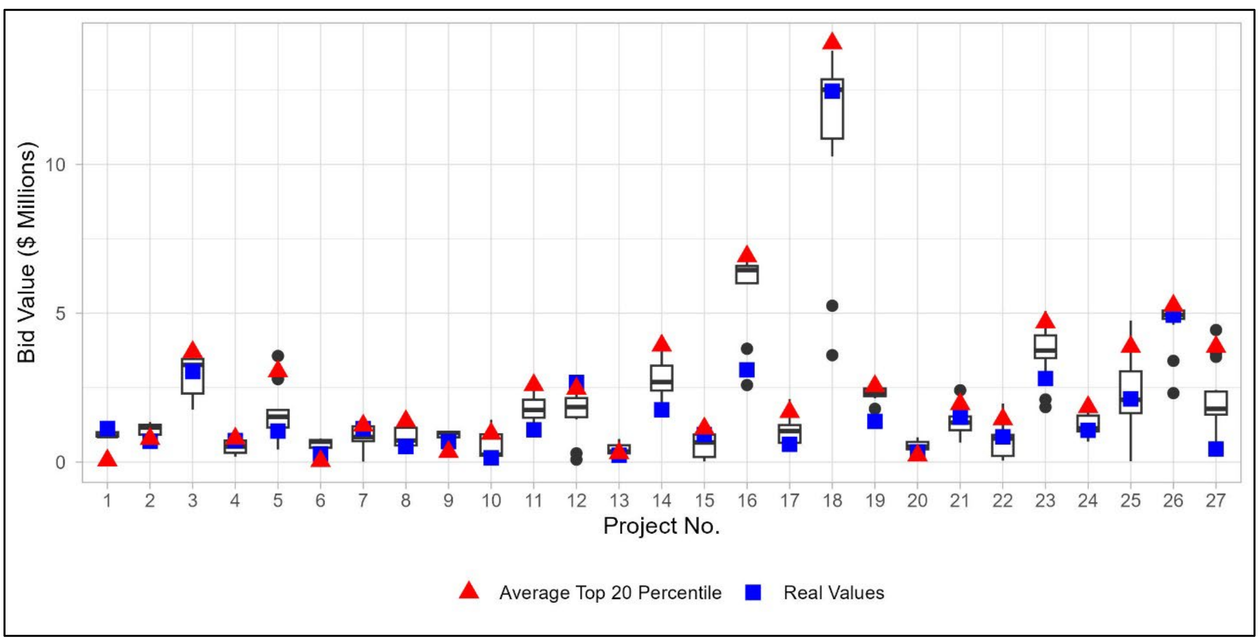


Figure 42. *Validation for probabilistic regression cost estimation for intersection projects*

Bridge Replacement projects display validation accuracies ranging from 64% to 96%. As indicated in Figure 43, the model effectively captured trends and predicted bid values with an acceptable accuracy. The simplicity of the regression model results in narrow ranges of predicted values for both intersection and bridge projects. Consequently, many real values do not fall within the predicted ranges for these projects.

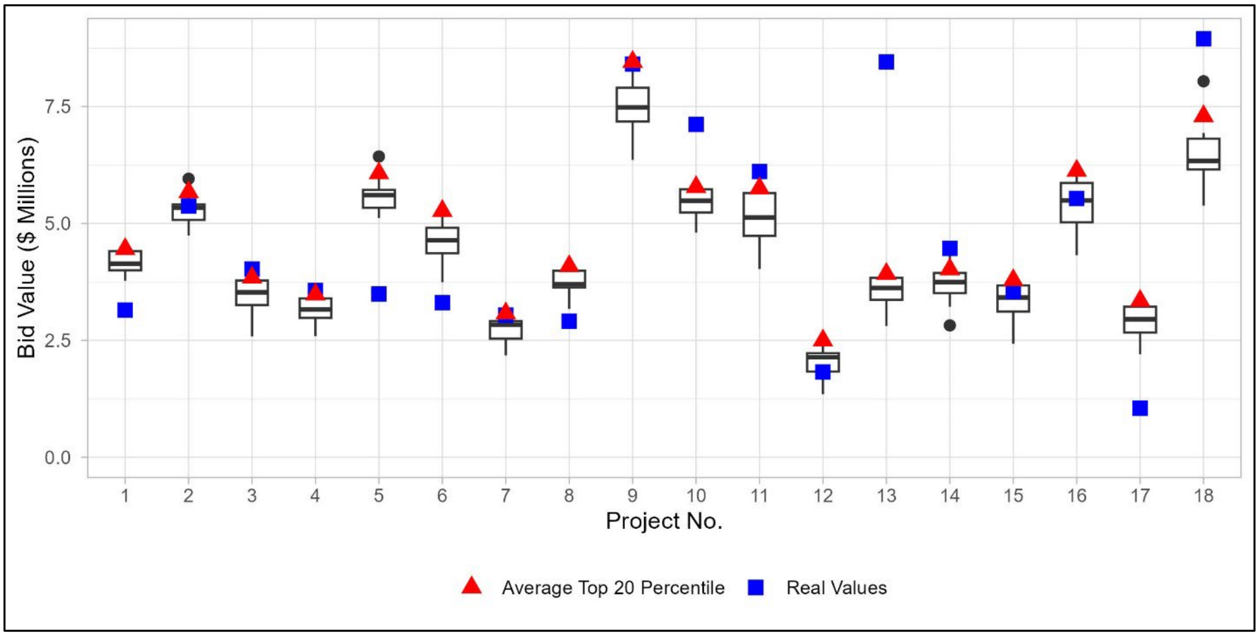


Figure 43. *Validation for probabilistic regression cost estimation for bridge projects*

Meanwhile, in the case of widening projects, validation accuracy spans from 53 % to 65%. Figure 44 presents prediction ranges and the average of the top 20th percentile alongside their corresponding real values for six projects. Here, the probabilistic regression model exhibits a wider range of predicted values, and because of that, most of the actual project bid values are encompassed by the predicted ranges.

Regarding the widening projects, as discussed in Section 5.1.2., the model can elevate its performance by adding new data from widening projects. The limited number of available data for bid values of widening projects, alongside the wide range of bid values in those data, does not allow a regression model to perform optimally.

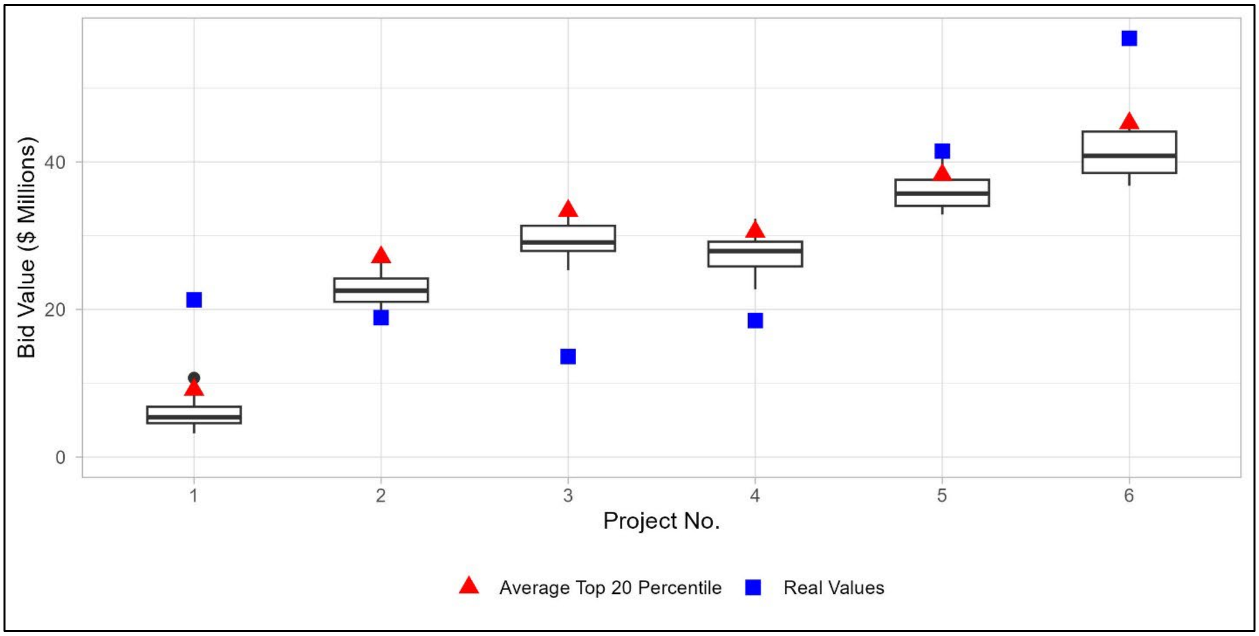


Figure 44. *Validation for probabilistic regression cost estimation for widening projects*

5.1.4 Neural Network Modeling: Deterministic

To construct Neural Network predictive models, the datasets for projects require modifications. The reported Excel datasets for bridge, intersection improvement, and widening projects have some missing data (N/A), removing of which results in losing a good part of the dataset and, consequently, having unreliable predictive models. To address this issue, all N/A cells are replaced with the average/mod value of their column. For instance, from the column "Number of Working days from SCDOT," blank cells are replaced with the average of this column. For the column "Number of Improved Lanes," the blank cells are replaced with the mod of this column. After addressing the missing data, the dataset is ready for developing predictive models.

To develop the deterministic model, 80% of the dataset is selected as the training data to develop the Neural Network model, and the other 20% is selected as testing data to evaluate the accuracy of the model. In this Neural Network model, a single hidden layer comprising four nodes is selected, along with a sigmoid activation function and linear output. Similar to the deterministic regression model, if the Neural Network model fails to meet the minimum accuracy threshold specified by the user, it is discarded and replaced with a new model. This process continues iteratively until the desired accuracy level is reached.

For intersection improvement projects, the accuracy rate achieved by the deterministic Neural Network model is 84%. Figure 45 presents the predicted (test) value vs. real bid values for 41 intersection improvement projects that were not used in model construction and training. According to this figure, the deterministic neural network model for intersection projects is able to predict the bid values with high accuracy, especially for projects with less than $5 Million bid value. This figure also shows the capability of the model to track the bid value rates, even when the real bid value is high and acts as an outlier.

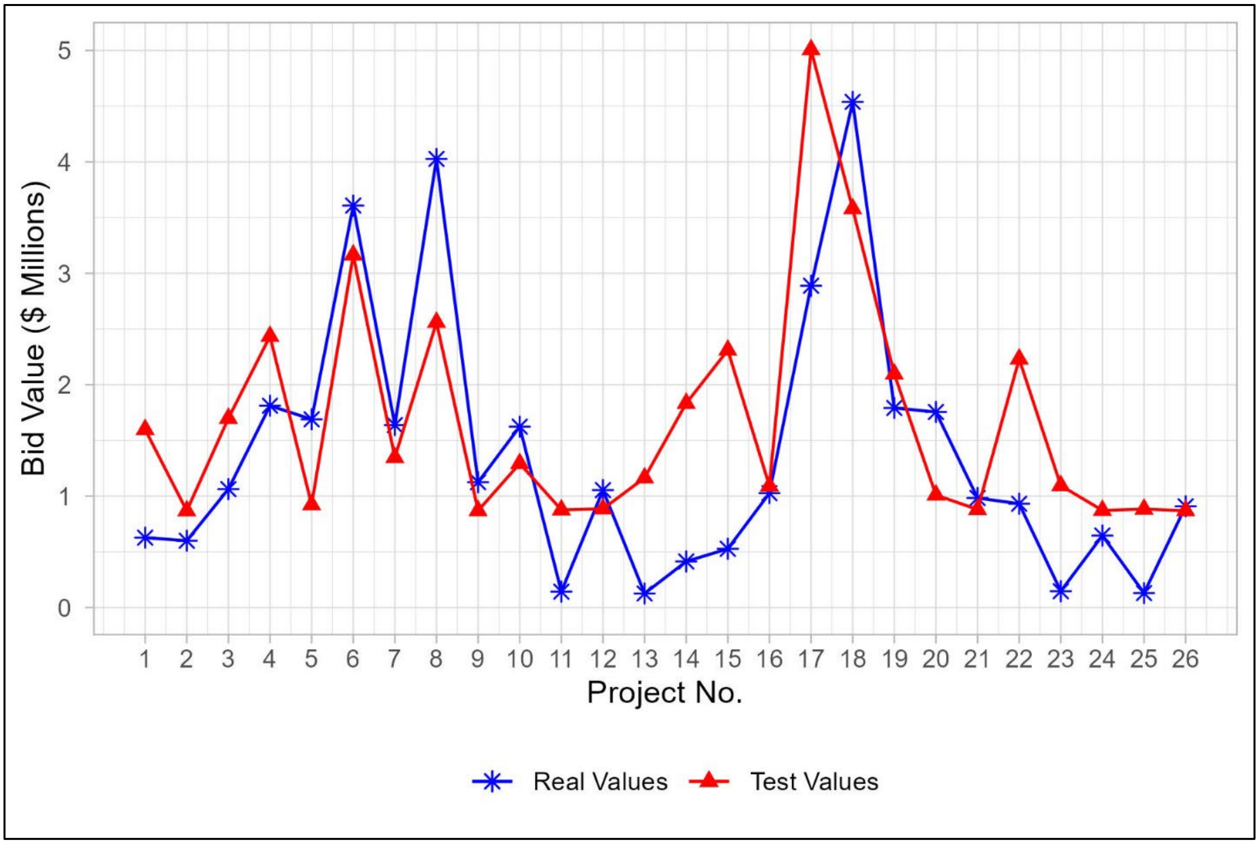


Figure 45. *Validation for deterministic neural network cost estimation for intersection projects*

Bridge replacement projects demonstrate a slightly higher accuracy of 90%. Figure 46 illustrates the model's capability for predicting bid values that were not used during model training. Out of these projects, most testing values are proximate to their corresponding real values, and the rest show acceptable predictions.

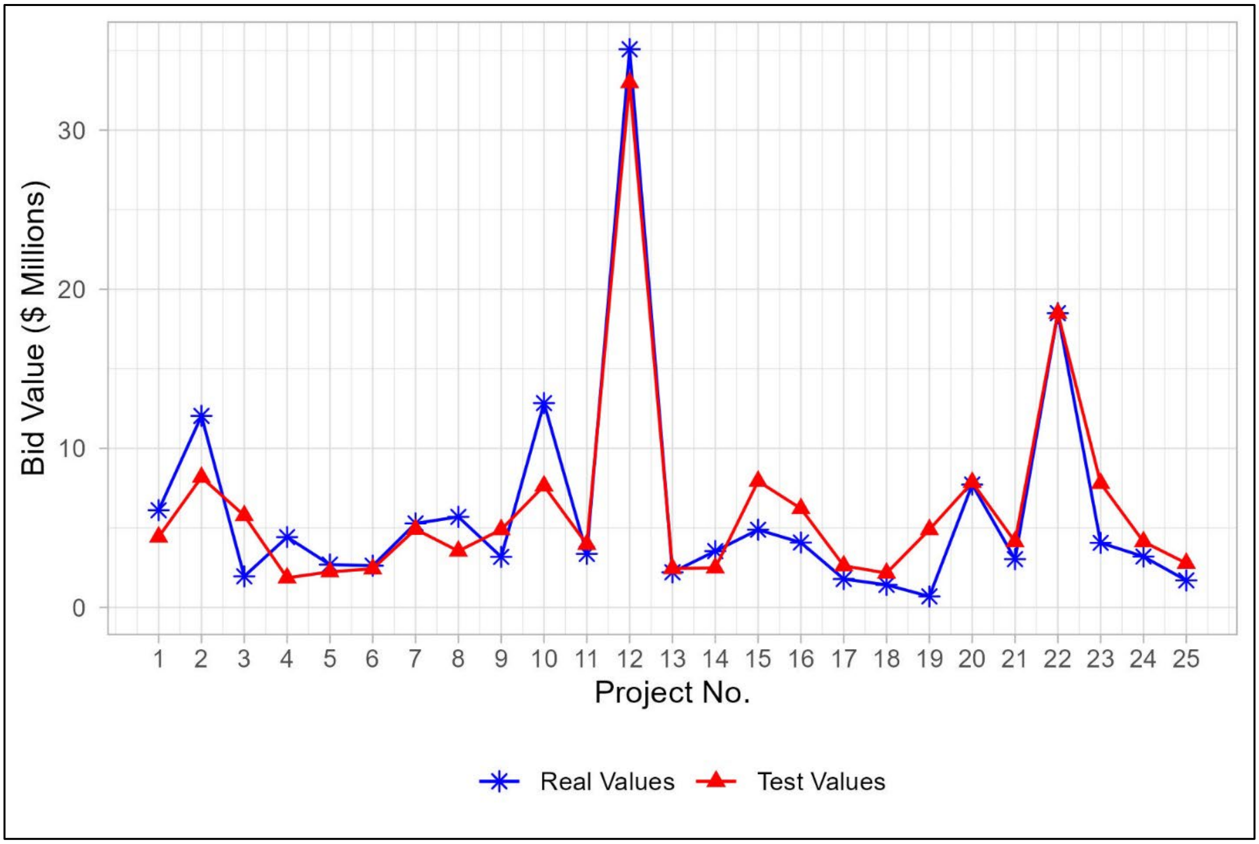


Figure 46. *Validation for deterministic neural network cost estimation for bridge projects*

The testing accuracy of the deterministic neural network model for widening projects is presented is Figure 47. This model achieved an accuracy of 86%, and according to the figure below, it perfectly tracks the bid values for widening projects, especially for projects with higher costs.

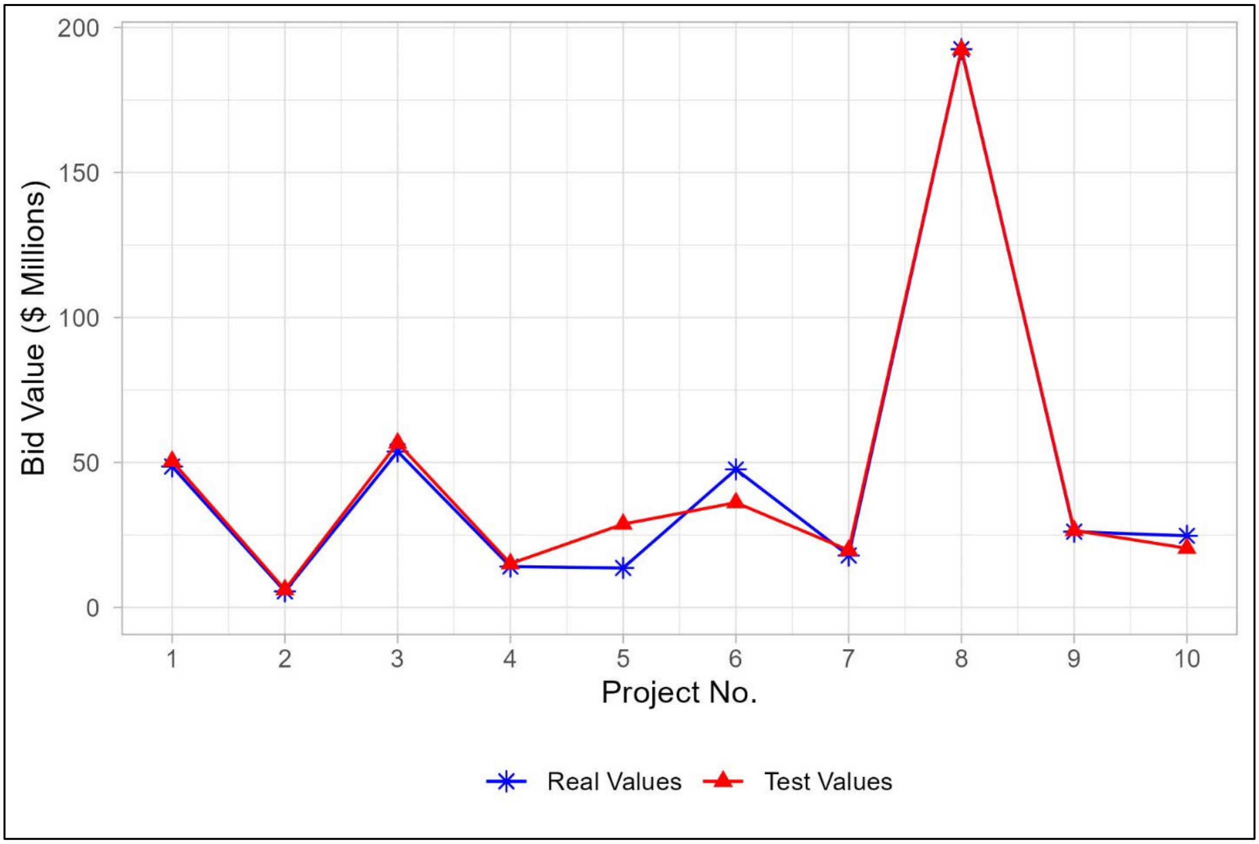


Figure 47. *Validation for deterministic neural network cost estimation for widening projects*

5.1.5 Neural Network Modeling: Probabilistic

Similar to linear regression, constructing a probabilistic model using a Neural Network involves specifying the number of models and a minimum accuracy threshold. The process includes allocating 15% of the data for final validation, generating multiple datasets from the remainder, and training models with 80% of each dataset. Models failing to meet the accuracy threshold are discarded and replaced until the desired level is reached. Models in the top 50th percentile of validation accuracy are chosen for the final model. Bid values are predicted from these models, with reporting including average, minimum, and maximum bid values, along with the average of the top 20th percentile responses. For each project, 20 models with a minimum testing accuracy of 70% are generated, ensuring all models meet the accuracy criterion.

For Intersection Improvement projects, the validation accuracy ranges from 40% to 55% across the constructed models. Figure 48 illustrates the model outputs along sides with the real value for 29 projects that are not used for training and testing and were separated initially for validation purposes. According to Figure 48, the probabilistic model for intersection improvement projects is able to capture the real bid value for most projects. In the majority of cases, the real value falls within the predicted range, with some predictions falling between Q1 and Q3. For projects where that predictions underestimate the real value, considering the average top 20 percentile provides a more accurate estimation.

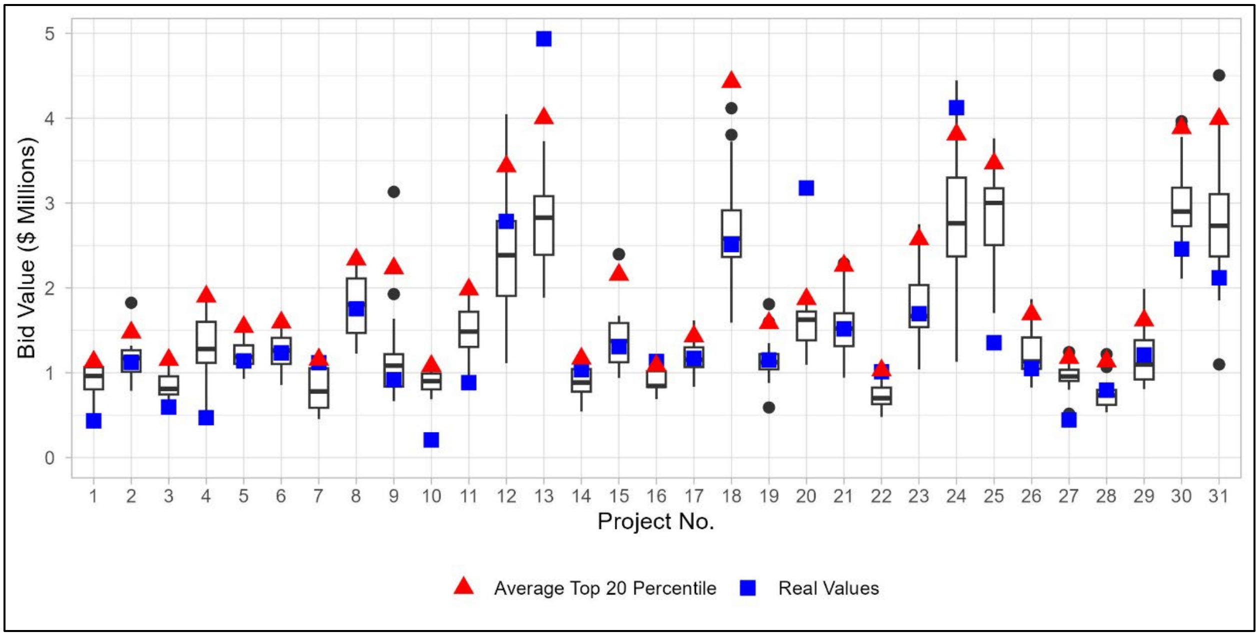


Figure 48. *Validation for probabilistic neural network cost estimation for intersection projects*

For Bridge Replacement projects, the validation accuracy ranges from 20% to 60%. According to Figure 49, the neural network probabilistic model overestimates the bid value for the majority of projects less than $5 Million. For the remaining projects with higher bid value, the average top 20 percentile is a better parameter for bid estimation.

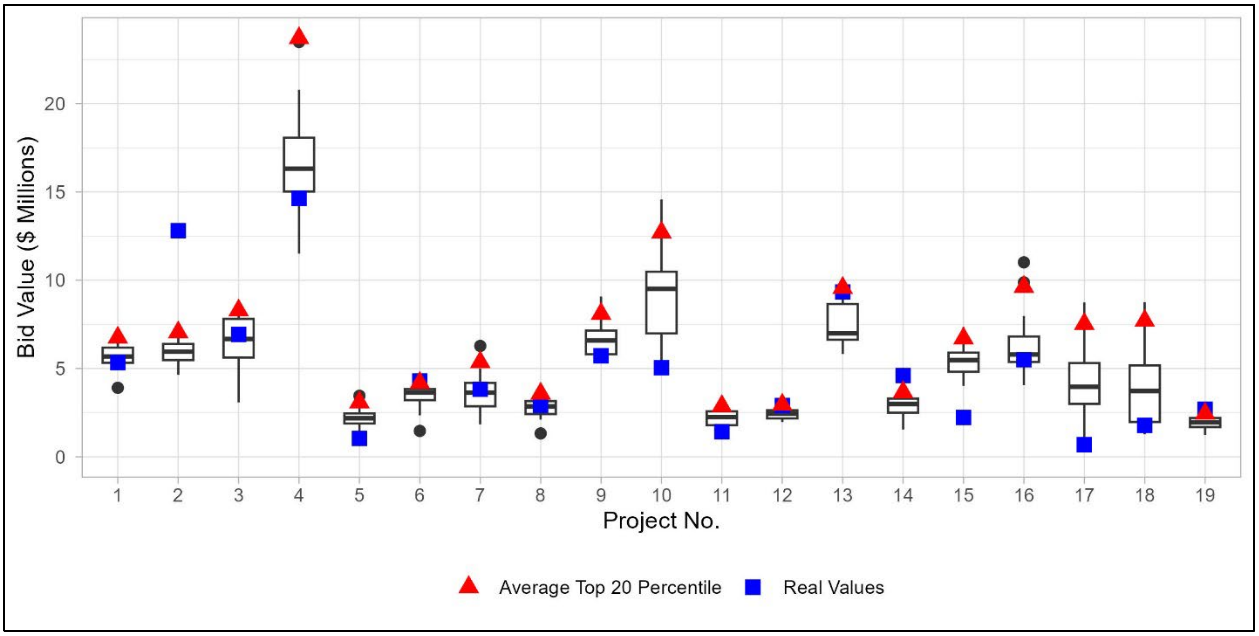


Figure 49. *Validation for probabilistic neural network cost estimation for bridge projects*

Meanwhile, for widening projects, the range of validation accuracy spans from 43% to 86%. Figure 50 illustrates the prediction ranges and the average of the top 20th percentile alongside their corresponding real value for six projects. In comparison to intersection and bridge projects, the neural network probabilistic model for widening encompasses a wider range of predicted values. Due to the limited number of widening projects available for training, the constructed models may produce bid values that vary noticeably from one another.

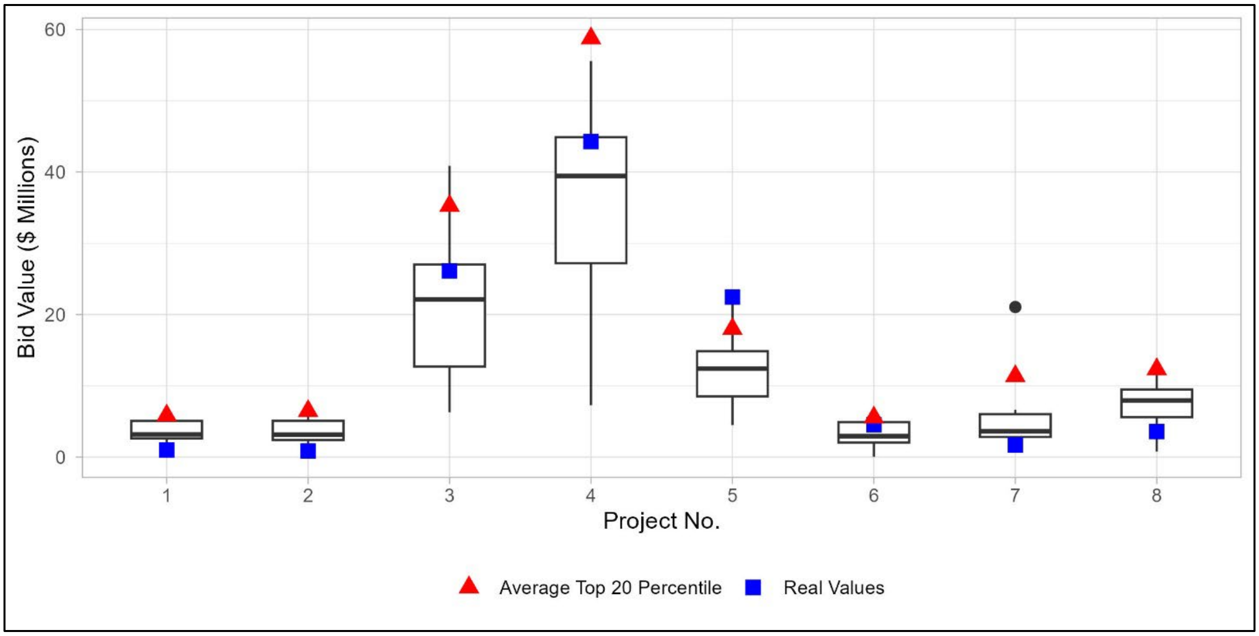


Figure 50. *Validation for probabilistic neural network cost estimation for widening projects*

5.2 Preliminary cost estimating tool (PCET) development

Figure 51 presents a snapshot image of the PCET tool that is developed as a deliverable in this study. The PCET tool is a user-friendly Microsoft Excel-based computational tool that allows users to: (1) Select cost estimating method: linear regression, neural network, or average of both, (2) Select project type: widening, bridge replacement, and intersection improvements, (3) Define project characteristics as inputs, and (4) Generate either a "Point Cost Estimate" or "Ranged Cost Estimate" along with model accuracies. The user can enter up to a total of 15 project parameters as input for running the PCET tool; however, not all the input parameters are used in the cost estimating models currently embedded in PCET. The input parameters and their possible values are presented in Table 17.

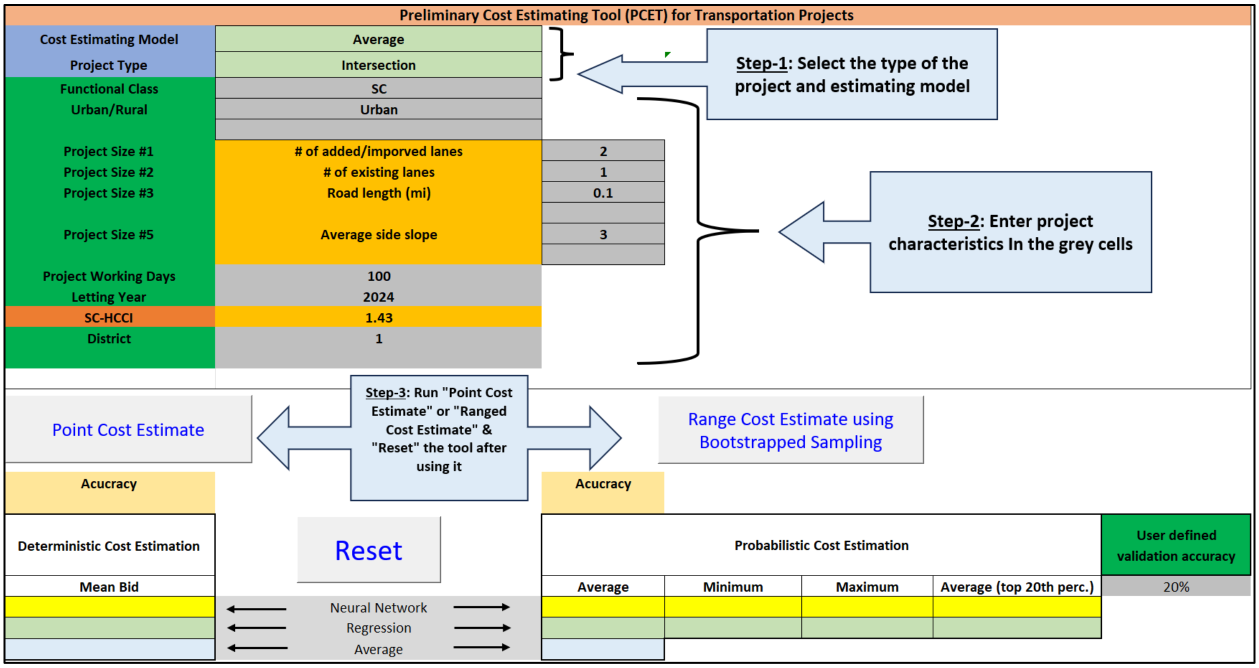


***Figure** 51. A snapshot outline of the PCET tool*

There are three instructional steps for using the PCET tool, as can be seen from Figure 51. Step-1 requires the user to select the cost estimate model of their choice. The options include linear regression, neural network, or the average of both models. Step-1 also requires the user to select the type of project they would like to develop an estimate for. The options include widening, bridge replacement, and intersection. Depending on the type of project selected in Step-1, relevant project characteristics will be shown in the tool. Step-2 requires the user to enter all the project characteristics in the gray colored cells either by typing in the values or choosing from the drop-down options. Users are encouraged to use their best judgement in case some characteristics are not known at the time of using the tool. It should however be noted that the accuracy of cost estimates is highly dependent on the accuracy of the project characteristic inputs. Moreover, in the case of probabilistic prediction, the user can enter a minimum validation accuracy to filter the probabilistic models used for prediction. Step-3 requires the user to either run the deterministic point cost estimate model or the probabilistic ranged cost estimate model by clicking on “Point Cost Estimation” or “Ranged Cost Estimate Using Bootstrapped Sampling” buttons, respectively. The user could also use these buttons one after another. Running these models will populate the cost estimate results in the bottom section of the tool. For the deterministic point cost estimation, a single estimate value is printed along with the associated accuracy measure. For the probabilistic ranged cost estimation, average, minimum and maximum estimates are printed along with mean model prediction accuracy. The user needs to reset the tool by clicking on the “Reset” button before closing the PCET tool. Resetting will erase all the results.

Table 17. *PCET input parameters*

Input Parameter	Possible values/Description
Project Type	• Widening • Bridge replacement • Intersection
Functional Class	• Secondary • SC • US • Interstate
Urban/Rural	• Urban • Rural
Estimate Type	• Planning • Early design
Project Size #1	• # of added lanes (0, 1, 2, 3, 4, 5, 6)
Project Size #2	• Road length (miles)
Project Size #3	• Added shoulder width (0ft, 2ft, 4ft, 6ft, 8ft, 10ft, 12ft)
Project Size #4	• Average side slope (0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10)
Project Size #5	• Bridge length (miles)
Base Course Type	• Brief description
Intermediate Course Type	• Brief description
Surface Course Type	• Brief description
Project working days	• Expected # of working days (days)
Letting Year	• Enter expected year of letting
Topography	• Flat, rolling, or mountainous

6. Conclusion, Recommendations, and Implementation

6.1. Conclusions

This research study primarily focused on developing planning level cost estimating models for three types of transportation projects namely, road widening, bridge replacement, and intersection improvement projects. Linear regression, artificial neural networks, and combination of both these approaches were explored to estimate total project cost using project size, location, and other features as inputs. Past bid data available in SCDOT's repositories was used for developing the cost estimating models in this study. Specifically, bid prices averaged over three lowest bidders for 46 widening projects, 130 bridge replacement projects, and 204 intersection improvement projects were collected and used for model building.

Challenges identified early on in this research study include lack of design detail and the need to base these planning level cost estimates on broad project features, which naturally makes these estimates somewhat less accurate. Additionally, the occasional need to rapidly produce planning-level cost estimates is also noted. Attempting to address these challenges, the project goal is to develop a user-friendly tool namely, preliminary cost estimating tool (PCET), for SCDOT to rapidly generate planning cost estimates for transportation projects. It is expected that the produced estimates are not highly accurate but will support budgeting and other planning-level project goals dependent on cost estimates.

The conclusions of this study are as follows:

1. South Carolina-specific highway construction cost index (SCHCCI) was developed based on past bid data, and it was found that the general trend matches that of NHCCI, but SCHCCI exhibited greater fluctuations in some time periods
2. Project size features such as road length, bridge length, expected number of working days were found to be significant influencers of the total project cost
3. SCHCCI was found to be a significant influencer of total project cost for intersection projects, but not so for widening and bridge replacement projects
4. Estimating the range of project costs instead of a point estimate is deemed more useful, and an approach for developing such estimates was developed and validated using linear regression and artificial neural networks
5. Linear regression provided accuracies (measured as R2 predicted) of 83.76%, 71.5%, and 61.14% for parameter selection, and 78%, 82%, and 69% for pointed estimates for widening, bridge replacement, and intersection projects, respectively

6. For pointed cost estimates, artificial neural networks resulted in average accuracies (measured as R2 of predicted vs. actual) of 86%, 90%, and 84% for widening, bridge replacement, and intersection projects, respectively
7. For ranged cost estimates, linear regression produced somewhat erratic results mainly because of the small sample size
8. For ranged cost estimates of bridge replacement projects, linear regression produced an average accuracy (measured as R2-predicted) of 64% to 96%
9. For ranged cost estimates of intersection projects, linear regression produced an average accuracy (measured as R2-predicted) of 40% to 56%
10. For ranged cost estimates of widening projects, linear regression produced an average accuracy (measured as R2-predicted) of 53 % to 65%
11. For ranged cost estimates of widening projects, neural networks produced accuracies (measured as R2 of predicted vs. actual costs) of 43% to 86%.
12. For ranged cost estimates of bridge replacement projects, neural networks produced accuracies (measured as R2 of predicted vs. actual costs) of 20% to 60%.
13. For ranged cost estimates of intersection projects, neural networks produced accuracies (measured as R2 of predicted vs. actual costs) of 40% to 55%.

Further validation using new project data would increase confidence in the PCET tool and its utility for SCDOT. The bigger takeaway from this study is that total project cost can be predicted, albeit less accurately in some cases, based on few project characteristics that are available in the planning stage of a transportation project. One parameter that was found highly influential and included in most of the developed models is the number of working days; this parameter need to be reasonably estimated for higher accurate cost estimates. The models for widening project type have performed poorly compared to other project types mainly because of the smaller project sample size made available to the research team; therefore, widening project estimates need to be cautiously developed and used based on the PCET tool produced in this study. Another disclaimer is that these planning level estimates are expected to be less accurate, and therefore should be treated cautiously. In the interest of project budgeting, some contingency (as a chosen %) may be added to the PCET estimates.

6.2. Recommendations & Implementation Guidance

Based on the findings of this research study, it is recommended that SCDOT adopt and use the project deliverable – the PCET tool – in a phased approach. In the first phase, the tool needs to be further validated using comparisons with cost estimates developed using SCDOT's conventional approach. It is recommended that the three project types – widening, bridge replacement, and intersection improvements – be included in this phase-1 validation that could span six months to a year. Phase-1 validation ideally will inform the pre-construction office of the practical merits and limitations of the PCET tool and assess their preference between linear regression, neural network, or the combination along with any necessary adjustments (e.g., add 20% contingency) that may be needed. In phase-2, the PCET tool along with the needed adjustments may be broadly used across the three project types. In addition, SCHCCI which was developed as part of this research study may also be used to adjust cost estimates outside of the PCET tool.

The PCET tool is developed in such a way that the embedded cost prediction models can be re-trained on a need basis as and when additional bid data for the three project types is available. Therefore, SCDOT is strongly recommended to re-train the PCET tool using the provided instructions at least once in every 2 years. Updating the models would make future cost estimates more accurate and informed from recent bid prices.

References

AASHTO. (2013). Practical guide to cost estimating.

Abdelaty, A., Jamal, M. S., & Mountain-Plains Consortium. (2022). Guidelines for Developing and Reviewing Baseline Schedules for Wyoming Transportation Projects. Mountain-Plains Consortium. https://rosap.ntl.bts.gov/view/dot/64092

Adel, K., Elyamany, A., Belal, A. M., & Kotb, A. S. (2016). Developing parametric model for conceptual cost estimate of highway projects. International Journal of Engineering Science, 6(7), 1728–1734.

Alavi, S. and Tavares, M.P., 2009. Highway project cost estimating and management.

American Association of State and Highway Transportation Officials (AASHTO). 2013. Practical Guide to Cost Estimating. Washington, D.C.

Anderson, S. D., Molenaar, K. R., & Schexnayder, C. J. (2007). Guidance for cost estimation and management for highway projects during planning, programming, and preconstruction (Vol. 574). Transportation Research Board.

Anderson, S., Damnjanovic, I., Nejat, A. and Ramesh, S., 2009. Synthesis on construction unit cost development: technical report (No. FHWA/TX-09/0-6023-1). Texas Transportation Institute.

Asmar, M. E., Hanna, A. S., & Whited, G. C. (2011). New approach to developing conceptual cost estimates for highway projects. Journal of Construction Engineering and Management, 137(11), 942–949. https://doi.org/10.1061/(ASCE)CO.1943-7862.0000355

Awuku, B., Asa, E., Baffoe-Twum, E., & Essegbey, A. (2022). Conceptual cost estimation of highway bid items–A systematic literature review. Engineering, Construction and Architectural Management, ahead-of-print. https://doi.org/10.1108/ECAM-03-2022-0266

Baccarini, D. (2006). “The maturing concept of estimating project cost contingency—A review.” 31st Australasian Univ. Building Educators Association Conf. (AUBEA 2006), Curtin Univ. of Technology, Bentley, WA, Australia.

Bell, L. C., & Kaminsky, A. (1987). Data base for preliminary cost estimating. Journal of Transportation Engineering, 113(4), 341–347.

Byrnes, J. E. (2002). “Best practices for highway project cost estimating.” M.S. thesis, Arizona State Univ., Mesa, AZ.

Chelst, K., and Canbolat, Y. B. (2012). Value-added decision making for managers, CRC Press, Boca Raton, FL, 1–545.

Chou, J.-S. (2009). Generalized linear model-based expert system for estimating the cost of transportation projects. Expert Systems with Applications, 36(3), 4253–4267.

Chou, J.S. and O'Connor, J.T., 2007. Internet-based preliminary highway construction cost estimating database. Automation in Construction, 17(1), pp.65-74.

Chou, J.S., Peng, M., Persad, K.R. and O'Connor, J.T., 2006. Quantity-based approach to preliminary cost estimates for highway projects. Transportation Research Record, 1946(1), pp.22-30.

Do, Q., Moriyani, M. A., Le, C., and Le, T. 2023. "Cost-Weighted TF-IDF: A Novel Approach for Measuring Highway Project Similarity Based on Pay Items' Cost Composition and Term Frequency." J. Constr. Eng. Manag., 149(8), 4023069.

Federal Highway Administration (FHWA). 2015. National Review of State Cost Estimation Practice. U.S. Department of Transportation. Washington, D.C.

FHWA (Federal Highway Administration). (2007). "Major project program cost estimating guidance." Available at: https://www.fhwa.dot.gov/ipd/project_delivery/tools_programs/cost_estimating/guidance.aspx

Flyvbjerg, B., Ansar, A., Budzier, A., Buhl, S., Cantarelli, C., Garbuio, M., Glenting, C., Holm, M. S., Lovallo, D., & Lunn, D. (2018). Five things you should know about cost overrun. Transportation Research Part A: Policy and Practice, 118, 174–190. https://doi.org/10.1016/j.tra.2018.07.013

Fragkakis, N., Lambropoulos, S. and Pantouvakis, J.P., 2010. A cost estimate method for bridge superstructures using regression analysis and bootstrap. Organization, technology & management in construction: an international journal, 2(2), pp.182-190.

Gardner, B. J., Gransberg, D. D., & Jeong, H. D. (2016). Reducing data-collection efforts for conceptual cost estimating at a highway agency. Journal of Construction Engineering and Management, 142(11), 04016057.

Gardner, B. J., Gransberg, D. D., & Rueda, J. A. (2017). Stochastic conceptual cost estimating of highway projects to communicate uncertainty using bootstrap sampling. ASCE-ASME Journal of Risk and Uncertainty in Engineering Systems, Part A: Civil Engineering, 3(3), 05016002.

Gransberg, D., Jeong, H. D., Karaca, I., & Gardner, B. (2017). Top-down construction cost estimating model using an artificial neural network.

Guerrero, P. F. 2003. Comparison of States' Highway Construction Costs. US General Accounting Office.

Harper, C. M., Molenaar, K. R., Anderson, S., & Schexnayder, C. (2014). Synthesis of performance measures for highway cost estimating. Journal of Management in Engineering, 30(3), 04014005.

Hollar, D. A., Rasdorf, W., Liu, M., Hummer, J. E., Arocho, I., & Hsiang, S. M. (2013). Preliminary engineering cost estimation model for bridge projects. Journal of Construction Engineering and Management, 139(9), 1259–1267.

Jennings, W., 2012. Why costs overrun: risk, optimism and uncertainty in budgeting for the London 2012 Olympic Games. Construction Management and Economics, 30(6), pp.455-462.

Jeong, D. H., Shrestha, K. J., Steadman, M., Kuzio, J., and others. 2021. Develop an Improved SDDOT Construction Cost Index.

Joseph Shrestha, K, H David Jeong, and Douglas D Gransberg. 2016. "Current Practices of Highway Construction Cost Index Calculation and Utilization." In Construction Research Congress 2016, 351–60.

Karaca, I., Gransberg, D. D., & Jeong, H. D. (2020). Improving the accuracy of early cost estimates on transportation infrastructure projects. Journal of Management in Engineering, 36(5), 04020063. https://doi.org/10.1061/(ASCE)ME.1943-5479.0000819

Le, C., Le, T. and Jeong, H.D., 2019a. Geographic Information Systems (GIS) Based Visual Analytics Framework for Highway Project Performance Evaluation. In Advances in Informatics and Computing in Civil and Construction Engineering (pp. 719-724). Springer, Cham.

Le, T., Le, C. and David Jeong, H., 2018a. Lifecycle Data Modeling to Support Transferring Project-Oriented Data to Asset-Oriented Systems in Transportation Projects. Journal of Management in Engineering, 34(4), p.04018024.

Le, T., Le, C., Jeong, H.D. and Jahren, C., 2018b. Visual Exploration of Large Transportation Asset Data using Ontology-Based Heat Tree. International Journal of Transportation, 6(1), pp.47-58.

Liu, H., Kwigizile, V., and Huang, W.-C. 2021. "Holistic Framework for Highway Construction Cost Index Development Based on Inconsistent Pay Items." J. Constr. Eng. Manag., 147(7), 4021052.

Liu, M. and Liu, M., 2011. Preliminary engineering cost trends for highway projects (No. FHWA/NC/2010-10). North Carolina Department of Transportation Research and Development Group.

Molenaar, K. R. (2005) "Programmatic cost risk analysis for highway megaprojects." J. Constr. Eng. Manage., 10.1061/(ASCE)0733-9364(2005) 131:3(343), 343–353.

Paulsen, C., Gallivan, F., Chavez, M., & Venner, M. (2008). NCHRP 8-36 Task 72: Guidelines for Cost Estimation Improvements at State DOTs.

Petroutsatou, K., Georgopoulos, E., Lambropoulos, S. and Pantouvakis, J.P., 2012. Early cost estimating of road tunnel construction using neural networks. Journal of construction engineering and management, 138(6), pp.679-687.

Pierce, C. E., Huynh, N. N., & Guimaraes, P. (2012). Cost indexing and unit price adjustments for construction materials. University of South Carolina. Dept. of Civil & Environmental Engineering.

SCDOT (2017) South Carolina Department of Transportation Structural Efficiencies Study

SCFOR, "Highway Facts," <http://www.sctransportation.com/highway-facts.html> (Accessed on August 11, 2016).

Schexnayder, C.J. and Scholar, P.E., 2003. PROJECT COST ESTIMATING.

Shane, J. S., Molenaar, K. R., Anderson, S., & Schexnayder, C. (2009). Construction project cost escalation factors. Journal of Management in Engineering, 25(4), 221–229.

Shrestha, K. J., Jeong, H. D., & Gransberg, D. D. (2017). Multidimensional highway construction cost indexes using dynamic item basket. Journal of Construction Engineering and Management, 143(8), 04017036. https://doi.org/10.1061/(ASCE)CO.1943-7862.0001307

Skolnik, J. (2011). Price indexing in transportation construction contracts. Jack Faucett Associates Bethesda, MD.

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

Van Dyke, C., Gibson, B., Jasper, J. and Kreis, D., 2017. Review of Initial Project Estimates for Design, Right of Way, Utilities, and Construction.

White, K., and Erickson, R. 2011. "New cost estimating tool." Public Roads, 75(1).

Whited, Gary, and Rayyan Alsamadani. 2011. "New Methodology for Developing Cost Indexes for Highway Construction."

Wilmot, Chester G, and Guangxiang Cheng. 2003. "Estimating Future Highway Construction Costs." Journal of Construction Engineering and Management 129 (3): 272–79.

Appendixes

Appendix-A: Survey Instrument Used to Synthesize State of Practice Across Various States

SPR 757: SCDOT's Survey on Preliminary Cost Estimating Approaches for Transportation Projects

The purpose of this survey is to solicit inputs on successful preliminary cost estimating approaches used across the State DOTs along with identifying best practices for developing regional highway construction cost indices (HCCIs). For each State DOT, this survey may be completed by personnel working in the pre-construction division or other relevant divisions.

This survey has up to 23 questions with the ability to add comments in addition to your answers for each question. This survey is estimated to take about 15 minutes to complete. Please contact Dr. Kalyan Piratla at kpiratl@clemson.edu for any questions or concerns regarding this survey.

1. Participant's Name: ____
2. Participant's Email: ____
3. Participant's Phone Number: ____
4. Participant's Agency (e.g. South Carolina DOT): ________________________________
5. Participant's Job Title: ______________________________
6. Does your agency currently implement a systematic method for developing preliminary cost estimates in the planning phase of transportation projects?
    a. Yes
    b. No
    c. Depends/Unsure (explain in comments): ____

If "Yes" or "Depends/Unsure" to Q.6, proceed to Q.7 and if "No," proceed to Q. 15

7. How satisfied are you with the preliminary cost estimating process at your agency?
    a. Highly satisfied
    b. Somewhat satisfied
    c. Neither satisfied or dissatisfied
    d. Somewhat dissatisfied
    e. Highly dissatisfied

8. Is the preliminary cost estimating approach used by your agency developed in-state or adopted from federal guidelines?
    a. Developed in-state (i.e., in-house or with support of a consultant/researcher)
    b. Adopted from federal guidelines.
    c. Depends/Unsure (explain in comments):
    ____

If "Adopted from federal guidelines" to Q.8, proceed to Q. 9, otherwise to Q. 10

9. Briefly identify the federally-prescribed preliminary cost estimating approach used by your agency: ____

10. Is the preliminary cost estimating approach used by your agency in the form of an excel tool or a stand-alone software?
    a. Excel tool
    b. Stand-alone software
    c. Other (explain in comments): _____________________________________

11. Briefly describe the preliminary cost estimating approach used by your agency (e.g., unit price, linear regression, machine learning-based).

____

12. How are contingency costs estimated in the preliminary cost estimates used by your agency?
    a. As a percentage of base estimate
    b. As a risk-based measure related to the specific project
    c. Other ways (explain in comments): _______________________

If "risk-based measure" is selected for Q. 12, proceed to Q. 13; otherwise to Q. 14

13. What kind of risk-based measure does your agency use for preliminary cost estimates? (e.g., Monte-carlo) _________________________________

14. What type of preliminary cost estimate is produced by your agency?
    a. A deterministic cost estimate (i.e., a fixed value plus an appropriate contingency)
    b. A probabilistic cost estimate (i.e., a distribution of values along with their probability)
    c. A cost range (i.e., a lower and upper value)
    d. Other (explain in comments): _________________________

15. Does your agency have a systematic process for developing unit costs for cost estimating purposes?
    a. Yes
    b. No
    c. Depends/Unsure (explain in comments): _________________________________

If "Yes" to Q. 15, proceed to Q. 16; if not, proceed to Q. 17

16. Please describe the systematic process: ____

17. At what level are the historical unit costs maintained by your agency?
    a. State-level
    b. District-level
    c. County-level
    d. Other (explain in comments): _________________________

18. Does your agency have a systematic approach to account for inflation specific to the region/state for various project types?
    a. Yes
    b. No
    c. Depends/Unsure (explain in comments): ________________________

If "Yes" or "Depends/Unsure" to Q. 18, proceed to Q. 19; if not, proceed to Q. 21

19. How does your agency account for inflation while developing and using preliminary cost estimates? ____

20. Does your agency use a state-wide or region-wide highway construction cost index (HCCI) to account for inflation?
    a. Yes
    b. No
    c. Depends (explain in comments): __________________________

21. Are you able to share the preliminary cost estimating tool(s) along with other relevant tools and manuals with other state DOTs?
    a. Yes
    b. No
    c. Depends/Unsure (explain in comments): ____

22. What recommendations/suggestions do you have for developing accurate preliminary cost estimates for transportation projects?

____

23. Would you entertain a brief one-on-one discussion over a phone call as a follow-up to your survey responses?

   a. Yes (If "Yes," provide the best phone number to reach you at: ______________)
   b. No

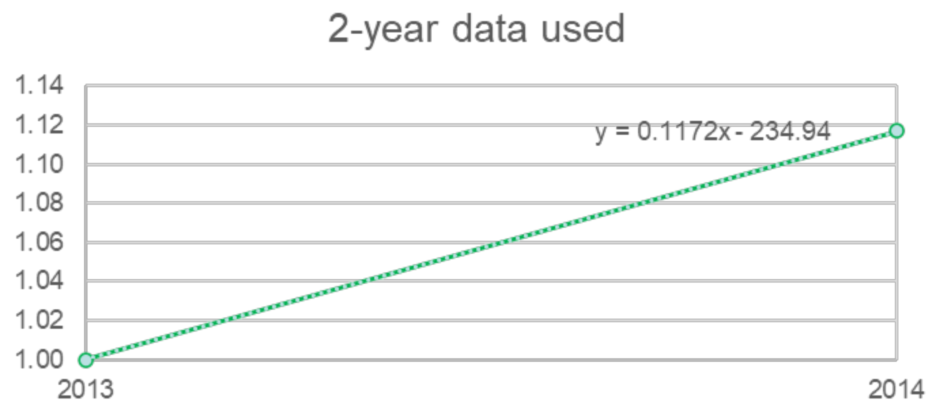
Appendix B: South Carolina Highway Construction Cost Index

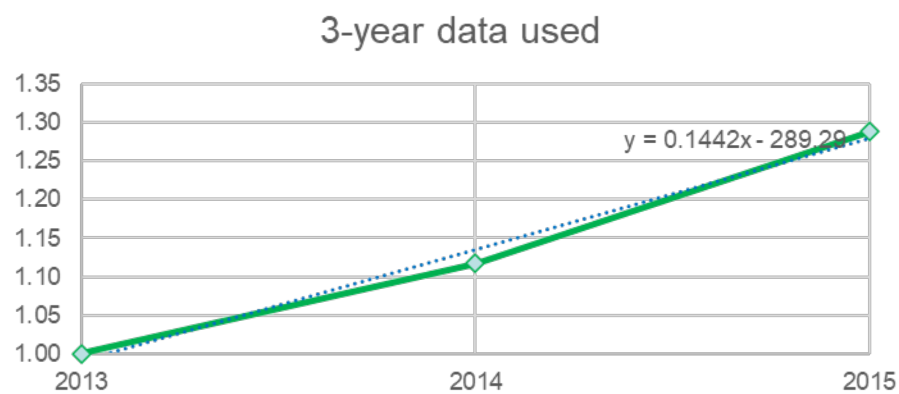
Values of statewide HCCIs and contract characteristics based HCCIs

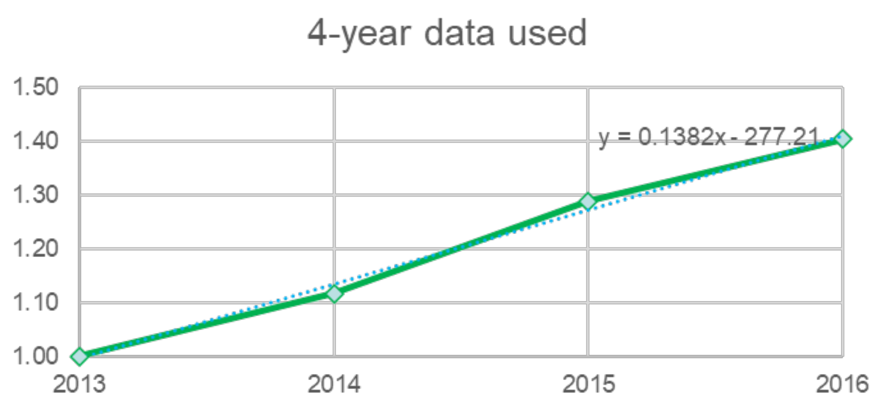
Letting year	Statewide HCCI	Project work type			Scope			Contract Size		
		Bridge replacements HCCI	Intersection improvements HCCI	Widening HCCI	Cluster 1 HCCI	Cluster 2 HCCI	Cluster 3 HCCI	Small-sized contract HCCI	Mid-sized contract HCCI	Large-sized contract HCCI
2013	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
2014	1.12	1.22	1.04	1.20	1.25	1.06	1.37	1.13	1.10	1.17
2015	1.29	1.28	1.13	1.33	1.36	1.22	1.70	1.28	1.21	1.35
2016	1.40	1.43	1.14	1.56	1.50	1.36	1.78	1.34	1.33	1.56
2017	1.52	1.32	1.19	1.68	1.38	1.55	1.95	1.23	1.31	1.67
2018	1.41	1.34	1.22	2.26	1.41	1.38	1.85	1.42	1.33	1.75
2019	1.51	1.43	1.26	1.95	1.52	1.52	1.86	1.57	1.36	2.15
2020	1.25	1.36	1.09	1.41	1.38	1.14	N/A	1.23	1.21	1.62
2021	1.30	1.38	1.12	N/A	1.42	1.24	N/A	N/A	1.39	1.65
2022	1.78	1.90	1.57	N/A	1.95	1.71	N/A	N/A	1.72	2.33
2023	2.28	N/A	1.65	N/A	N/A	2.19	N/A	N/A	N/A	3.01

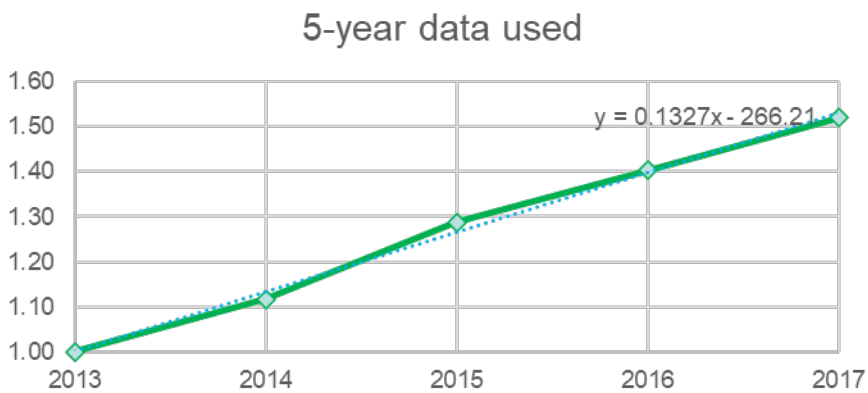
Values of statewide HCCIs and bid item characteristics based HCCIs

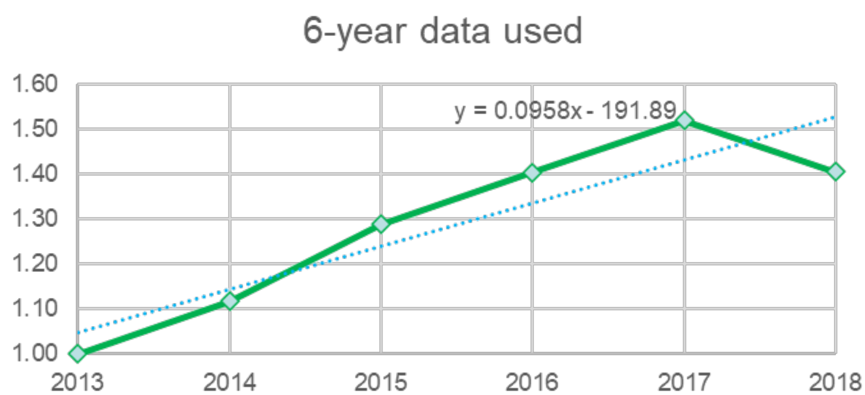
Letting year	Statewide HCCI	Work item division					
		Earthwork HCCI	Bases and subbases HCCI	Asphalt pavements HCCI	Maintenance and control of traffic HCCI	Structures HCCI	Incidental construction HCCI
2013	1.00	1.00	1.00	1.00	1.00	1.00	1.00
2014	1.12	1.26	1.04	1.06	0.91	1.11	0.97
2015	1.29	1.74	1.22	1.15	1.03	1.30	1.11
2016	1.40	1.90	1.45	1.16	1.01	1.49	1.19
2017	1.52	2.35	1.36	1.24	1.16	1.49	1.24
2018	1.41	2.12	1.44	1.34	1.06	1.59	1.30
2019	1.51	2.43	1.44	1.32	1.20	1.71	1.30
2020	1.25	1.74	1.31	1.13	1.06	1.48	0.98
2021	1.30	1.78	1.47	1.26	1.14	1.65	0.99
2022	1.78	2.86	1.97	1.65	1.42	2.16	1.23
2023	2.28	5.54	2.06	1.82	1.48	2.69	1.48

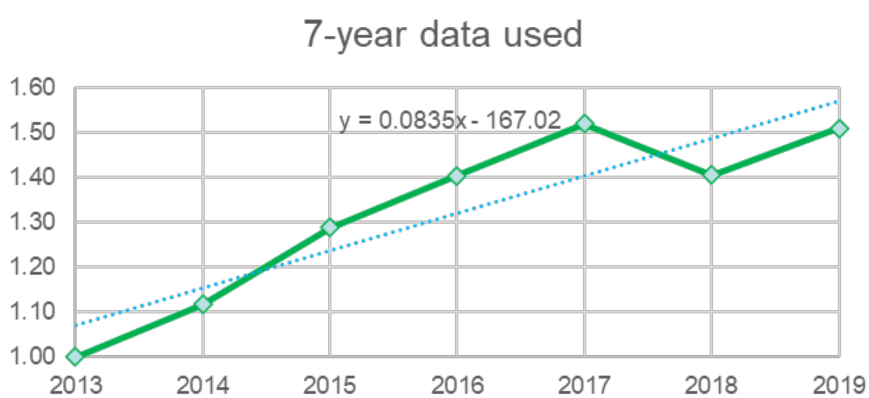

2-year data used
y = 0.1172x - 234.94
1.14
1.12
1.10
1.08
1.06
1.04
1.02
1.00
2013
2014

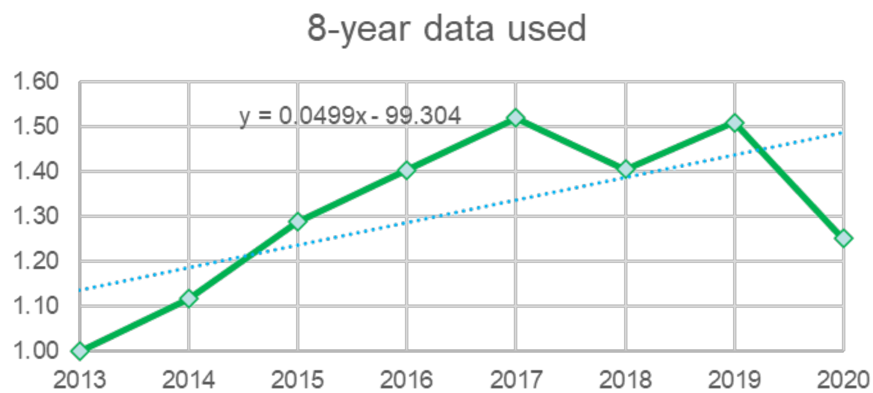

3-year data used
y = 0.1442x - 289.29
1.35
1.30
1.25
1.20
1.15
1.10
1.05
1.00
2013
2014
2015

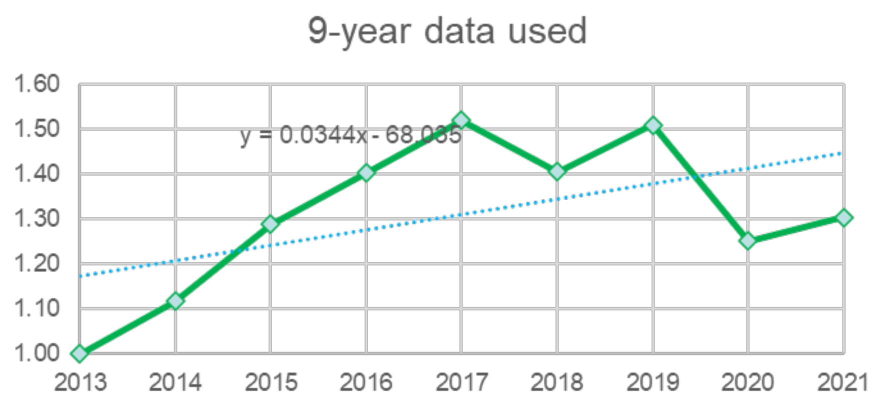

4-year data used
y = 0.1382x - 277.21
1.50
1.40
1.30
1.20
1.10
1.00
2013
2014
2015
2016

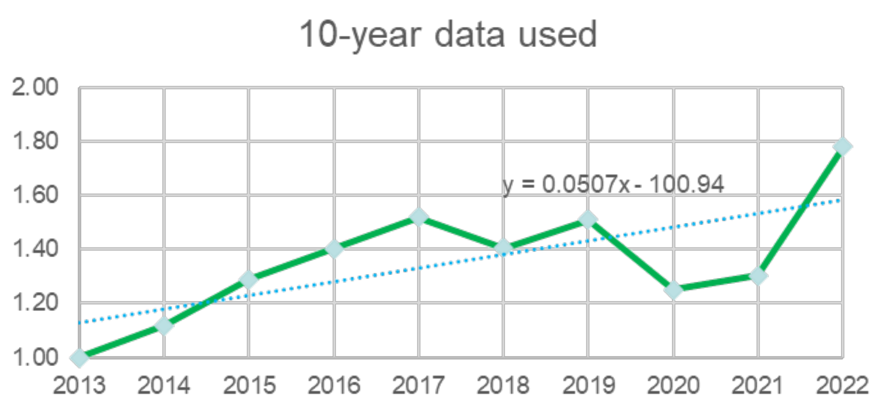

5-year data used
y = 0.1327x - 266.21
1.60
1.50
1.40
1.30
1.20
1.10
1.00
2013
2014
2015
2016
2017


6-year data used
y = 0.0958x - 191.89
1.60
1.50
1.40
1.30
1.20
1.10
1.00
2013
2014
2015
2016
2017
2018


7-year data used
y = 0.0835x - 167.02
1.60
1.50
1.40
1.30
1.20
1.10
1.00
2013
2014
2015
2016
2017
2018
2019


8-year data used
y = 0.0499x - 99.304
1.60
1.50
1.40
1.30
1.20
1.10
1.00
2013
2014
2015
2016
2017
2018
2019
2020


9-year data used
y = 0.0344x - 68.035
1.60
1.50
1.40
1.30
1.20
1.10
1.00
2013
2014
2015
2016
2017
2018
2019
2020
2021


10-year data used
y = 0.0507x - 100.94
2.00
1.80
1.60
1.40
1.20
1.00
2013
2014
2015
2016
2017
2018
2019
2020
2021
2022

Appendix C: Forecasted HCCI Values

Forecasted values of statewide HCCI and contract characteristics based HCCI

Letting year	Statewide HCCI	Project work type			Scope			Contract Size		
		Bridge replacements HCCI	Intersection improvements HCCI	Widening HCCI	Cluster 1 HCCI	Cluster 2 HCCI	Cluster 3 HCCI	Small-sized contract HCCI	Mid-sized contract HCCI	Large-sized contract HCCI
2024	1.86	1.72	1.43	2.34	1.73	1.85	2.74	1.67	1.67	2.68
2025	1.94	1.77	1.48	2.46	1.80	1.92	2.88	1.72	1.73	2.83
2026	2.02	1.83	1.53	2.58	1.86	1.99	3.03	1.76	1.78	2.98
2027	2.10	1.88	1.57	2.70	1.92	2.07	3.18	1.81	1.84	3.12
2028	2.18	1.94	1.62	2.82	1.98	2.14	3.32	1.85	1.89	3.27
2029	2.26	1.99	1.67	2.93	2.04	2.22	3.47	1.89	1.95	3.42
2030	2.34	2.05	1.72	3.05	2.10	2.29	3.61	1.94	2.01	3.57
2031	2.42	2.10	1.77	3.17	2.17	2.36	3.76	1.98	2.06	3.71
2032	2.50	2.16	1.81	3.29	2.23	2.44	3.91	2.03	2.12	3.86
2033	2.58	2.21	1.86	3.41	2.29	2.51	4.05	2.07	2.17	4.01
2034	2.66	2.27	1.91	3.53	2.35	2.59	4.20	2.12	2.23	4.15
2035	2.74	2.33	1.96	3.64	2.41	2.66	4.34	2.16	2.29	4.30
2036	2.82	2.38	2.01	3.76	2.47	2.73	4.49	2.21	2.34	4.45
2037	2.90	2.44	2.05	3.88	2.54	2.81	4.63	2.25	2.40	4.60
2038	2.98	2.49	2.10	4.00	2.60	2.88	4.78	2.30	2.46	4.74

Forecasted values of statewide HCCI and bid item characteristics-based HCCIs

Letting year	Statewide HCCI	Work item division					
		Earthwork HCCI	Bases and subbases HCCI	Asphalt pavements HCCI	Maintenance and control of traffic HCCI	Structures HCCI	Incidental construction HCCI
2024	1.86	3.86	2.05	1.59	1.38	2.46	1.33
2025	1.94	4.13	2.13	1.65	1.42	2.59	1.36
2026	2.02	4.39	2.22	1.71	1.47	2.71	1.38
2027	2.10	4.65	2.31	1.77	1.51	2.84	1.40
2028	2.18	4.92	2.39	1.83	1.56	2.97	1.43
2029	2.26	5.18	2.48	1.90	1.60	3.09	1.45
2030	2.34	5.44	2.57	1.96	1.65	3.22	1.48
2031	2.42	5.71	2.65	2.02	1.69	3.35	1.50
2032	2.50	5.97	2.74	2.08	1.74	3.47	1.53
2033	2.58	6.23	2.83	2.15	1.78	3.60	1.55
2034	2.66	6.50	2.91	2.21	1.83	3.72	1.58
2035	2.74	6.76	3.00	2.27	1.87	3.85	1.60
2036	2.82	7.02	3.09	2.33	1.92	3.98	1.63
2037	2.90	7.28	3.18	2.40	1.96	4.10	1.65
2038	2.98	7.55	3.26	2.46	2.01	4.23	1.67

Appendix-D: Cost Estimating Model Exploration

Appendix D-1: Cost Estimate Models for Widening Projects

Regression Analysis: Ave_3bid ($) versus Number of Working days, HCCI, Road Length (mile), Bridge Length (feet), Year_letting, Sub Type

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
20813072	81.28%	62.56%	*

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	17	3.19781E+16	1.88106E+15	4.34	0.002
Number of Working days from SCD	1	3.08024E+14	3.08024E+14	0.71	0.411
HCCI	1	2.45188E+15	2.45188E+15	5.66	0.029
Road Length (mile)	1	2.70514E+15	2.70514E+15	6.24	0.023
Bridge Length (feet)	1	2.37544E+14	2.37544E+14	0.55	0.469
Year_letting	7	6.73372E+15	9.61959E+14	2.22	0.085
Sub Type	6	2.00808E+15	3.34679E+14	0.77	0.602
Error	17	7.36413E+15	4.33184E+14		
Total	34	3.93422E+16			

Figure 52. *Widening Cost Estimate Model D-1.1*

Regression Analysis: Ave_3bid ($) versus Number of Working days from SCD, HCCI, Road Length (mile), Year_letting, Sub Type

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
20770532	72.86%	54.77%	*

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	18	3.12692E+16	1.73718E+15	4.03	0.001
Number of Working days from SCD	1	7.65562E+14	7.65562E+14	1.77	0.194
HCCI	1	9.95629E+14	9.95629E+14	2.31	0.140
Road Length (mile)	1	6.55515E+15	6.55515E+15	15.19	0.001
Year_letting	8	9.44462E+15	1.18058E+15	2.74	0.024
Sub Type	7	3.33001E+15	4.75716E+14	1.10	0.390
Error	27	1.16482E+16	4.31415E+14		
Total	45	4.29174E+16			

Figure 53. *Widening Cost Estimate Model D-1.2*

Regression Analysis: Ave_3bid ($) versus Number of Working days from SCD, HCCI, Road Length (mile), Sub Type

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
24548974	50.85%	36.81%	*

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	10	2.18246E+16	2.18246E+15	3.62	0.002
Number of Working days from SCD	1	1.72627E+15	1.72627E+15	2.86	0.099
HCCI	1	3.96217E+15	3.96217E+15	6.57	0.015
Road Length (mile)	1	3.72459E+15	3.72459E+15	6.18	0.018
Sub Type	7	3.92642E+15	5.60917E+14	0.93	0.495
Error	35	2.10928E+16	6.02652E+14		
Total	45	4.29174E+16			

Figure 54. *Widening Cost Estimate Model D-1.3*

Regression Analysis: Ave_3bid ($) versus Number of Working days from SCD, HCCI, Road Length (mile), Average Slope, Sub Type

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
21007690	84.88%	73.54%	*

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	9	2.97359E+16	3.30399E+15	7.49	0.001
Number of Working days from SCD	1	9.80439E+14	9.80439E+14	2.22	0.162
HCCI	1	3.66865E+15	3.66865E+15	8.31	0.014
Road Length (mile)	1	2.78982E+15	2.78982E+15	6.32	0.027
Sub Type	5	2.68558E+15	5.37117E+14	1.22	0.359
Average Slope	1	8.73004E+14	8.73004E+14	1.98	0.185
Error	12	5.29588E+15	4.41323E+14		
Total	21	3.50318E+16			

Figure 55. *Widening Cost Estimate Model D-1.4*

Regression Analysis: Ave_3bid ($) versus Number of Working days from SCD, HCCI, Road Length (mile), Average Slope

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
21667896	77.22%	71.86%	18.99%

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	4	2.70503E+16	6.76258E+15	14.40	0.000
Number of Working days from SCD	1	7.62102E+13	7.62102E+13	0.16	0.692
HCCI	1	3.65214E+15	3.65214E+15	7.78	0.013
Road Length (mile)	1	7.10924E+15	7.10924E+15	15.14	0.001
Average Slope	1	1.19297E+15	1.19297E+15	2.54	0.129
Error	17	7.98146E+15	4.69498E+14		
Total	21	3.50318E+16			

Figure 56. *Widening Cost Estimate Model D-1.5*

Regression Analysis: Ave_3bid ($) versus HCCI, Road Length (mile), Average Slope

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
21157703	77.00%	73.17%	29.08%

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	3	2.69741E+16	8.99137E+15	20.09	0.000
HCCI	1	5.15168E+15	5.15168E+15	11.51	0.003
Road Length (mile)	1	1.47226E+16	1.47226E+16	32.89	0.000
Average Slope	1	1.30422E+15	1.30422E+15	2.91	0.105
Error	18	8.05767E+15	4.47648E+14		
Total	21	3.50318E+16			

Figure 57. *Widening Cost Estimate Model D-1.6*

Regression Analysis: Ave_3bid ($) versus HCCI, Road Length (mile), Average Slope, Sub Type

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
21972563	82.08%	71.06%	*

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	8	2.87555E+16	3.59443E+15	7.45	0.001
HCCI	1	5.34316E+15	5.34316E+15	11.07	0.005
Road Length (mile)	1	7.02554E+15	7.02554E+15	14.55	0.002
Average Slope	1	1.21604E+15	1.21604E+15	2.52	0.137
Sub Type	5	1.78136E+15	3.56271E+14	0.74	0.608
Error	13	6.27632E+15	4.82794E+14		
Total	21	3.50318E+16			

Figure 58. *Widening Cost Estimate Model D-1.7*

Regression Analysis: Ave_3bid ($) versus HCCI, Road Length (mile)

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
26376784	30.29%	27.05%	0.00%

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	2	1.30008E+16	6.50040E+15	9.34	0.000
HCCI	1	2.61628E+15	2.61628E+15	3.76	0.059
Road Length (mile)	1	8.80222E+15	8.80222E+15	12.65	0.001
Error	43	2.99166E+16	6.95735E+14		
Total	45	4.29174E+16			

Figure 59. *Widening Cost Estimate Model D-1.8*

Regression Analysis: Ave_3bid ($) versus HCCI, Road Length (mile), Average Slope, Number of improved lanes, Avg. Shoulder Widened (ft)

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
16453911	87.63%	83.77%	32.39%

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	5	3.07001E+16	6.14002E+15	22.68	0.000
HCCI	1	2.18599E+15	2.18599E+15	8.07	0.012
Road Length (mile)	1	7.61772E+15	7.61772E+15	28.14	0.000
Average Slope	1	1.14300E+15	1.14300E+15	4.22	0.057
Number of improved lanes	1	5.39379E+14	5.39379E+14	1.99	0.177
Avg. Shoulder Widened (ft)	1	2.41377E+15	2.41377E+15	8.92	0.009
Error	16	4.33170E+15	2.70731E+14		
Total	21	3.50318E+16			

Figure 60. *Widening Cost Estimate Model D-1.9*

Regression Analysis: Ave_3bid ($) versus HCCI, Road Length (mile), Average Slope, Avg. Shoulder Widened (ft), Number of improved lanes

Regression Equation

Ave_3bid ($) = -63944581 + 46573727 HCCI + 6804308 Road Length (mile) - 4857211 Average Slope + 6013369 Avg. Shoulder Widened (ft) + 4591049 Number of improved lanes

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
11262397	94.86%	93.03%	64.32%

Upon removing 2 outlier projects

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	5	3.27975E+16	6.55951E+15	51.71	0.000
HCCI	1	1.66171E+15	1.66171E+15	13.10	0.003
Road Length (mile)	1	5.62368E+15	5.62368E+15	44.34	0.000
Average Slope	1	1.78030E+15	1.78030E+15	14.04	0.002
Avg. Shoulder Widened (ft)	1	2.48015E+15	2.48015E+15	19.55	0.001
Number of improved lanes	1	6.68131E+14	6.68131E+14	5.27	0.038
Error	14	1.77578E+15	1.26842E+14		
Total	19	3.45733E+16			

Figure 61. *Widening Cost Estimate Model D-1.10*

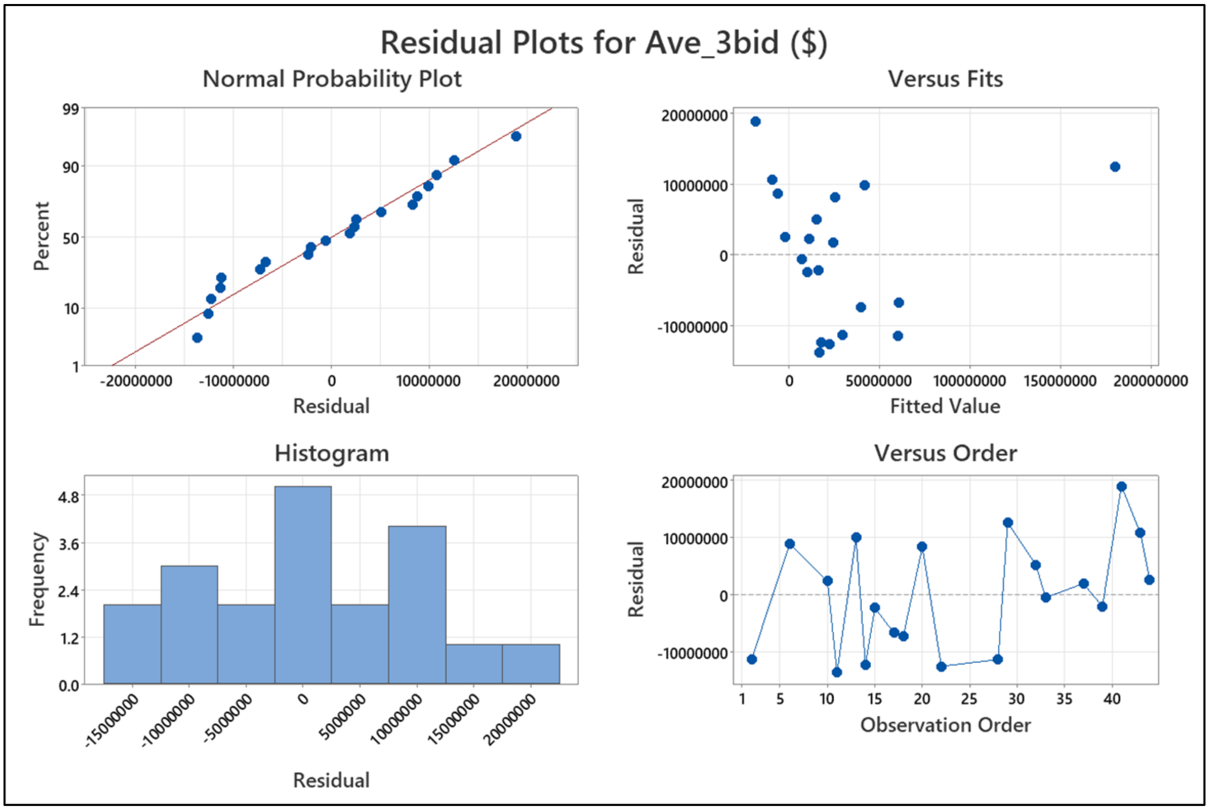


__Figure 62.__ Residual plots for the regression model D-1.10

Appendix D-2: Cost Estimate Models for Bridge Replacement Projects

Regression Analysis: Ave_3bid ($) versus Number of Working days from SCD, HCCI, Road Length (miles), Bridge Length miles), Number of existing lanes, Number of improved lanes, Average Slope, Average Shoulder Width (ft), Cluster, District, Urban/ rural, On Land/Over water, Flat/ Rolling/ Mountainous

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
1494593	97.53%	85.21%	*

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	20	3.53465E+14	1.76733E+13	7.91	0.029
Number of Working days from SCD	1	1.32765E+13	1.32765E+13	5.94	0.071
HCCI	1	6.05759E+11	6.05759E+11	0.27	0.630
Road Length (miles)	1	1.73323E+13	1.73323E+13	7.76	0.050
Bridge Length miles)	1	3.29752E+12	3.29752E+12	1.48	0.291
Number of existing lanes	1	1.39511E+12	1.39511E+12	0.62	0.474
Number of improved lanes	1	31614492085	31614492085	0.01	0.911
Average Slope	1	7.00154E+11	7.00154E+11	0.31	0.605
Average Shoulder Width (ft)	1	3.16617E+12	3.16617E+12	1.42	0.300
Cluster	10	1.16440E+13	1.16440E+12	0.52	0.816
Urban/ rural	1	4.44336E+12	4.44336E+12	1.99	0.231
On Land/Over water	1	9.45924E+12	9.45924E+12	4.23	0.109
Error	4	8.93523E+12	2.23381E+12		
Total	24	3.62400E+14			

Figure 63. *Bridge Replacement Cost Estimate Model D-2.1*

Regression Analysis: Ave_3bid ($) versus Number of Working days from SCD, HCCI, Road Length (miles), Bridge Length miles), Number of existing lanes, Number of improved lanes, Average Slope, Average Shoulder Width (ft)

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
1673696	87.63%	81.45%	0.00%

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	8	3.17580E+14	3.96975E+13	14.17	0.000
Number of Working days from SCD	1	3.31245E+13	3.31245E+13	11.82	0.003
HCCI	1	1.18907E+13	1.18907E+13	4.24	0.056
Road Length (miles)	1	6.58690E+13	6.58690E+13	23.51	0.000
Bridge Length miles)	1	340415230	340415230	0.00	0.991
Number of existing lanes	1	2.09868E+13	2.09868E+13	7.49	0.015
Number of improved lanes	1	4.61216E+12	4.61216E+12	1.65	0.218
Average Slope	1	1.59380E+12	1.59380E+12	0.57	0.462
Average Shoulder Width (ft)	1	1.03039E+13	1.03039E+13	3.68	0.073
Error	16	4.48201E+13	2.80126E+12		
Total	24	3.62400E+14			

Figure 64. *Bridge Replacement Cost Estimate Model D-2.2*

Regression Analysis: Ave_3bid ($) versus Number of Working days from SCD, HCCI, Road Length (miles), Bridge Length miles), Number of existing lanes, Average Shoulder Width (ft)

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
3618657	75.99%	73.59%	26.26%

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	6	2.48690E+15	4.14483E+14	31.65	0.000
Number of Working days from SCD	1	2.21982E+14	2.21982E+14	16.95	0.000
HCCI	1	3.79267E+12	3.79267E+12	0.29	0.592
Road Length (miles)	1	1.42165E+14	1.42165E+14	10.86	0.002
Bridge Length miles)	1	9.20480E+14	9.20480E+14	70.29	0.000
Number of existing lanes	1	1.17399E+13	1.17399E+13	0.90	0.348
Average Shoulder Width (ft)	1	1.08514E+11	1.08514E+11	0.01	0.928
Error	60	7.85681E+14	1.30947E+13		
Total	66	3.27258E+15			

Figure 65. *Bridge Replacement Cost Estimate Model D-2.3*

Regression Analysis: Ave_3bid ($) versus Number of Working days from SCD, HCCI, Road Length (miles), Bridge Length miles), Number of existing lanes

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
3589121	75.99%	74.02%	29.15%

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	5	2.48679E+15	4.97357E+14	38.61	0.000
Number of Working days from SCD	1	2.22473E+14	2.22473E+14	17.27	0.000
HCCI	1	4.20918E+12	4.20918E+12	0.33	0.570
Road Length (miles)	1	1.42058E+14	1.42058E+14	11.03	0.002
Bridge Length miles)	1	9.24693E+14	9.24693E+14	71.78	0.000
Number of existing lanes	1	1.16599E+13	1.16599E+13	0.91	0.345
Error	61	7.85789E+14	1.28818E+13		
Total	66	3.27258E+15			

Figure 66. *Bridge Replacement Cost Estimate Model D-2.4*

Regression Analysis: Ave_3bid ($) versus Number of Working days from SCD, Year_letting, Road Length (miles), Bridge Length miles), Number of existing lanes

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
3594406	75.92%	73.94%	31.54%

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	5	2.48447E+15	4.96894E+14	38.46	0.000
Number of Working days from SCD	1	2.28214E+14	2.28214E+14	17.66	0.000
Year_letting	1	1.89347E+12	1.89347E+12	0.15	0.703
Road Length (miles)	1	1.15440E+14	1.15440E+14	8.94	0.004
Bridge Length miles)	1	1.13560E+15	1.13560E+15	87.90	0.000
Number of existing lanes	1	1.85427E+13	1.85427E+13	1.44	0.236
Error	61	7.88105E+14	1.29198E+13		
Total	66	3.27258E+15			

Figure 67. *Bridge Replacement Cost Estimate Model D-2.5*

Regression Analysis: Ave_3bid ($) versus Number of Working days from SCD, Road Length (miles), Bridge Length miles)

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
3623850	72.24%	71.20%	32.00%

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	3	2.73419E+15	9.11398E+14	69.40	0.000
Number of Working days from SCD	1	4.48499E+14	4.48499E+14	34.15	0.000
Road Length (miles)	1	2.06065E+14	2.06065E+14	15.69	0.000
Bridge Length miles)	1	1.18414E+15	1.18414E+15	90.17	0.000
Error	80	1.05058E+15	1.31323E+13		
Total	83	3.78478E+15			

Figure 68. *Bridge Replacement Cost Estimate Model D-2.6*

Regression Analysis: Ave_3bid ($) versus Road Length (miles), Bridge Length miles), Number of Working days from SCD

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
2605952	81.18%	80.46%	56.76%

**<u>*Upon removing 2 outlier projects*</u>

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	3	2.28479E+15	7.61597E+14	112.15	0.000
Road Length (miles)	1	1.55595E+14	1.55595E+14	22.91	0.000
Bridge Length miles)	1	1.24801E+15	1.24801E+15	183.77	0.000
Number of Working days from SCD	1	2.18832E+14	2.18832E+14	32.22	0.000
Error	78	5.29697E+14	6.79098E+12		
Total	81	2.81449E+15			

Figure 69. *Bridge Replacement Cost Estimate Model D-2.7*

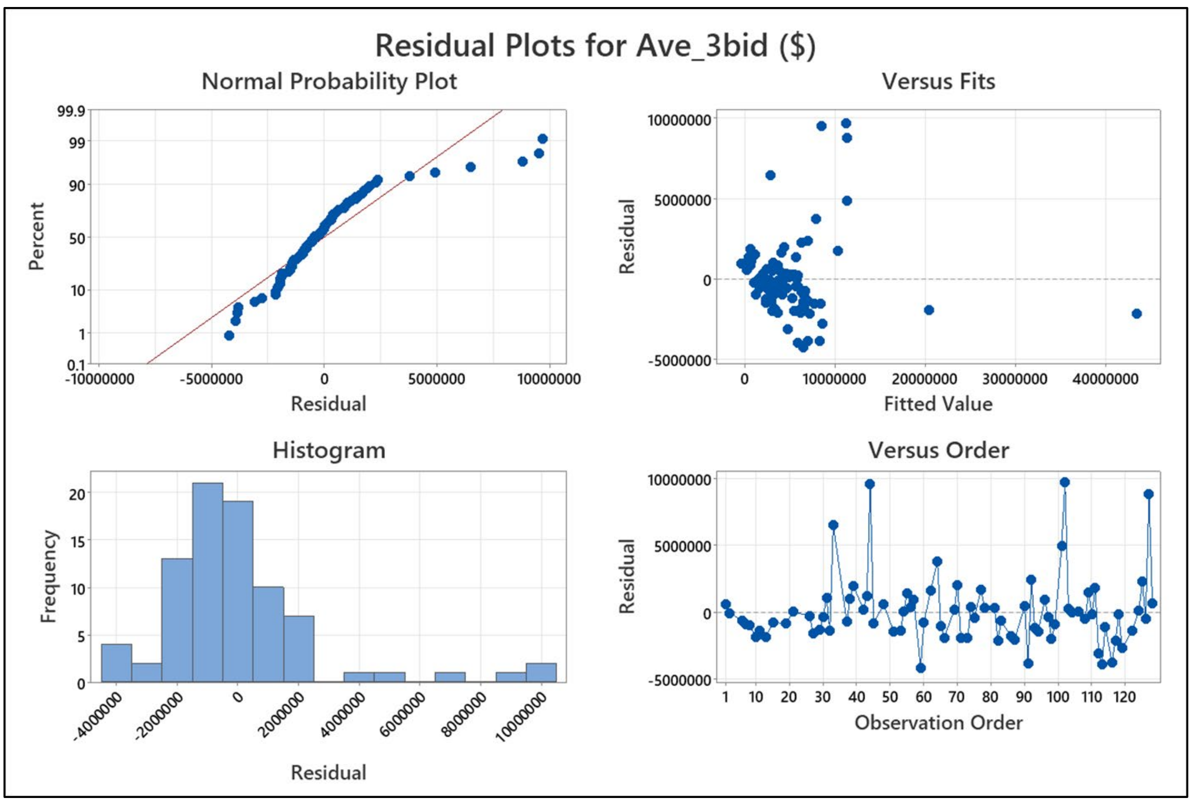


Figure 70. *Residual plots for model D-2.7*

Regression Analysis: Ave_3bid ($) versus Road Length (miles), Bridge Length miles), Number of Working days from SCD

Regression Equation

Ave_3bid ($) = -522977 + 2027315 Road Length (miles) + 33187664 Bridge Length miles)
+ 7910 Number of Working days from SCD

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
1725050	89.69%	89.28%	77.47%

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	3	1.94170E+15	6.47235E+14	217.50	0.000
Road Length (miles)	1	1.93437E+14	1.93437E+14	65.00	0.000
Bridge Length miles)	1	1.17476E+15	1.17476E+15	394.77	0.000
Number of Working days from SCD	1	6.94140E+13	6.94140E+13	23.33	0.000
Error	75	2.23185E+14	2.97580E+12		
Total	78	2.16489E+15			

<u>Upon removing 5 outlier projects</u>

Figure 71. *Bridge Replacement Cost Estimate Model D-2.8*

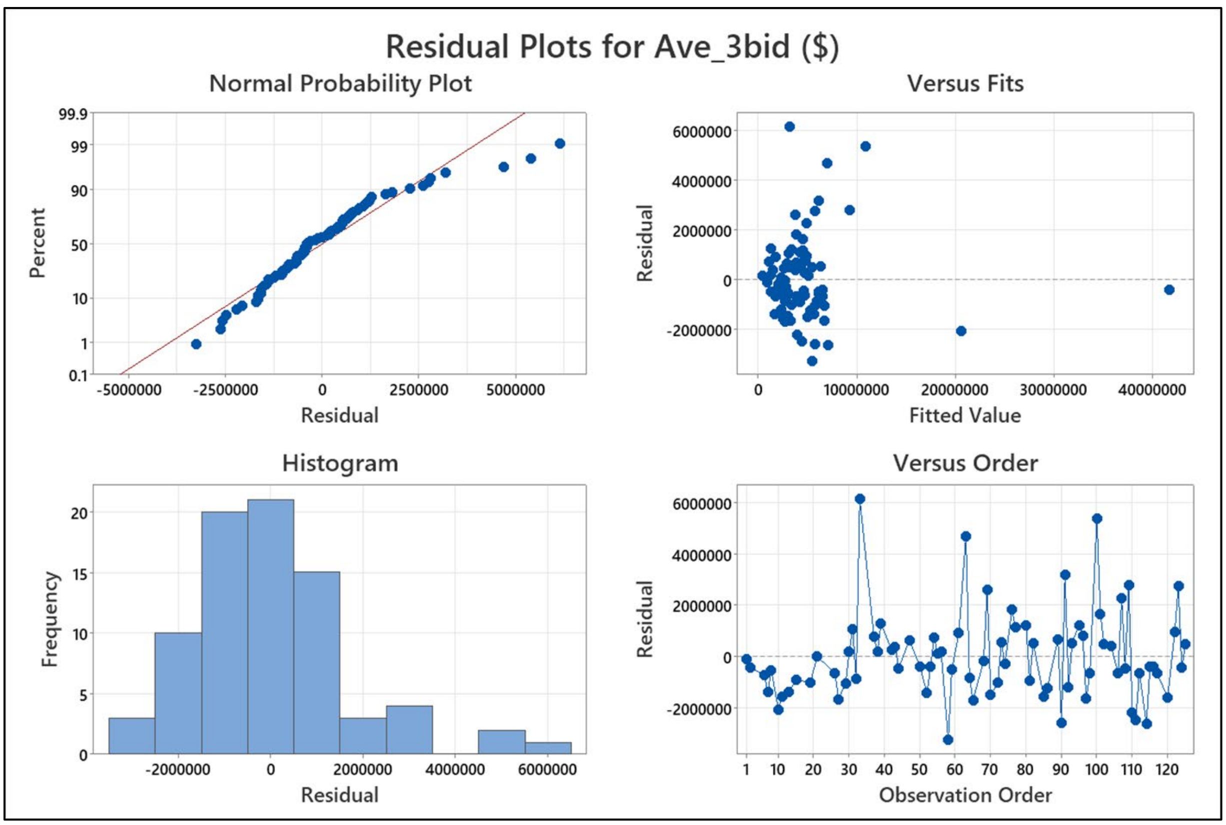


Figure 72. *Residual plots for model D-2.8*

Regression Analysis: Ave_3bid ($) versus Road Length (miles), Bridge Length miles), Number of Working days from SCD

Regression Equation

Ave_3bid ($) = -790138 + 1962631 Road Length (miles) + 33310037 Bridge Length miles) + 8613 Number of Working days from SCD

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
1577938	91.40%	91.05%	81.70%

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	3	1.95879E+15	6.52931E+14	262.23	0.000
Road Length (miles)	1	1.80378E+14	1.80378E+14	72.44	0.000
Bridge Length miles)	1	1.18295E+15	1.18295E+15	475.10	0.000
Number of Working days from SCD	1	8.11691E+13	8.11691E+13	32.60	0.000
Error	74	1.84252E+14	2.48989E+12		
Total	77	2.14304E+15			

*****Upon removing 6 outlier projects***

Figure 73. *Bridge Replacement Cost Estimate Model D-2.9*

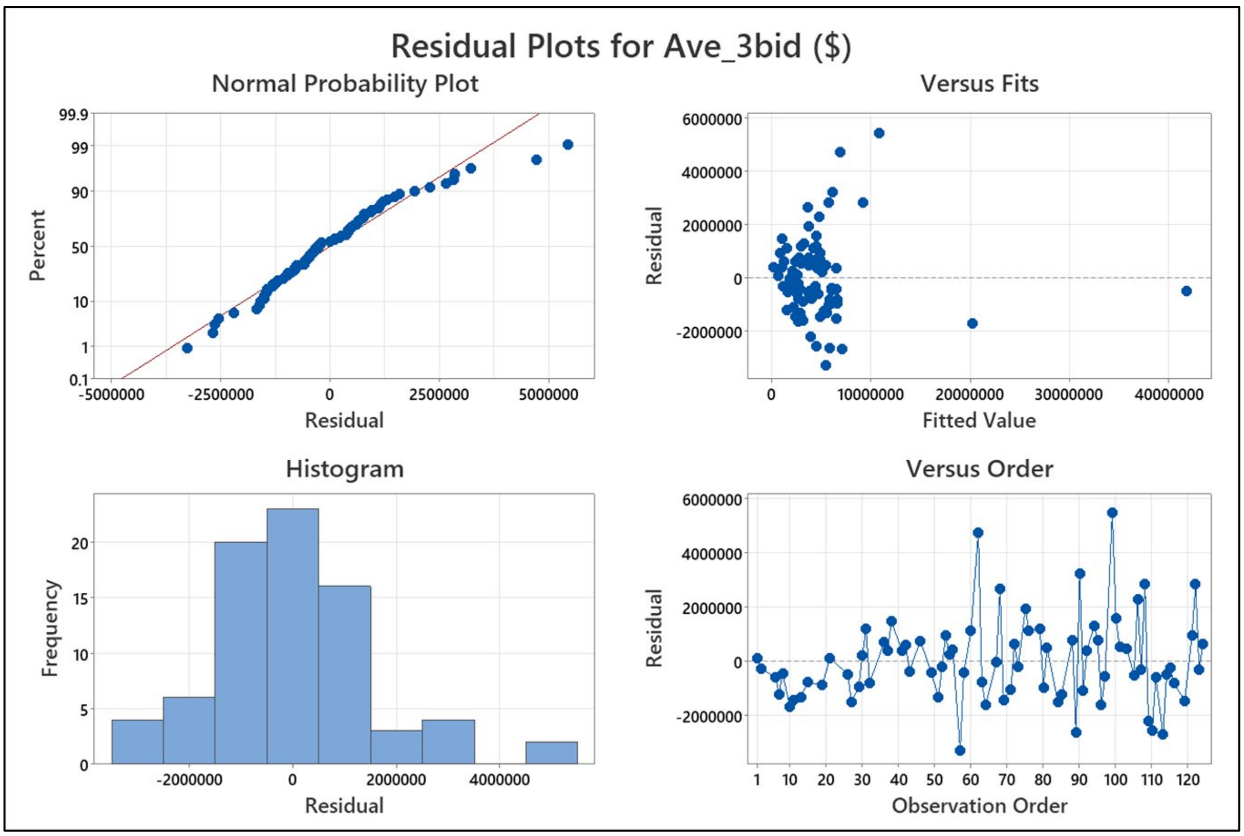


Figure 74. *Residual plots for model D-2.9*

Appendix D-3: Cost Estimate Models for Intersection Improvement Projects

Regression Analysis: Ave_3bid ($) versus Number of Working days from SCD, HCCI, Length, Number of Existing lanes, Number of improved lanes, Average Slope, Cluster, District, Sub Type, Functional Class, Urban/ rural

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
448363	76.31%	64.23%	19.85%

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	25	3.17350E+13	1.26940E+12	6.31	0.000
Number of Working days from SCD	1	1.45760E+12	1.45760E+12	7.25	0.010
HCCI	1	2.28826E+11	2.28826E+11	1.14	0.291
Length	1	4.00845E+12	4.00845E+12	19.94	0.000
Number of Existing lanes	1	3303156937	3303156937	0.02	0.899
Number of improved lanes	1	1.91538E+11	1.91538E+11	0.95	0.334
Average Slope	1	7.02707E+11	7.02707E+11	3.50	0.068
Cluster	5	3.10715E+12	6.21429E+11	3.09	0.017
District	6	3.40962E+11	56826948518	0.28	0.942
Sub Type	3	5.16336E+11	1.72112E+11	0.86	0.470
Functional Class	3	1.05715E+12	3.52383E+11	1.75	0.169
Urban/ rural	2	2.53143E+12	1.26571E+12	6.30	0.004
Error	49	9.85043E+12	2.01029E+11		
Total	74	4.15854E+13			

Figure 75. *Intersection Project Cost Estimate Model D-3.1*

Regression Analysis: Ave_3bid ($) versus Number of Working days from SCD, HCCI, Length, Average Slope, Cluster, Sub Type, Functional Class, Urban/ rural

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
488099	71.61%	64.81%	40.51%

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	17	4.26585E+13	2.50932E+12	10.53	0.000
Number of Working days from SCD	1	3.39914E+12	3.39914E+12	14.27	0.000
HCCI	1	8.11138E+11	8.11138E+11	3.40	0.069
Length	1	6.41578E+12	6.41578E+12	26.93	0.000
Average Slope	1	1.66911E+12	1.66911E+12	7.01	0.010
Cluster	5	3.47821E+12	6.95642E+11	2.92	0.019
Sub Type	3	1.02491E+12	3.41637E+11	1.43	0.240
Functional Class	3	2.90156E+12	9.67186E+11	4.06	0.010
Urban/ rural	2	3.65436E+12	1.82718E+12	7.67	0.001
Error	71	1.69151E+13	2.38241E+11		
Total	88	5.95736E+13			

Figure 76. *Intersection Project Cost Estimate Model D-3.2*

Regression Analysis: Ave_3bid ($) versus Number of Working days from SCD, HCCI, Length, Average Slope, Sub Type, Functional Class, Urban/ rural

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
518009	65.77%	60.36%	28.69%

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	12	3.91803E+13	3.26502E+12	12.17	0.000
Number of Working days from SCD	1	3.59520E+12	3.59520E+12	13.40	0.000
HCCI	1	1.92800E+12	1.92800E+12	7.19	0.009
Length	1	7.12259E+12	7.12259E+12	26.54	0.000
Average Slope	1	1.68595E+12	1.68595E+12	6.28	0.014
Sub Type	3	1.23701E+12	4.12336E+11	1.54	0.212
Functional Class	3	3.62270E+12	1.20757E+12	4.50	0.006
Urban/ rural	2	2.83172E+12	1.41586E+12	5.28	0.007
Error	76	2.03933E+13	2.68333E+11		
Total	88	5.95736E+13			

Figure 77. *Intersection Project Cost Estimate Model D-3.3*

Regression Analysis: Ave_3bid ($) versus Number of Working days from SCD, HCCI, Length, Average Slope, Functional Class, Urban/ rural

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
523261	63.69%	59.56%	24.94%

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	9	3.79433E+13	4.21592E+12	15.40	0.000
Number of Working days from SCD	1	4.09347E+12	4.09347E+12	14.95	0.000
HCCI	1	1.99377E+12	1.99377E+12	7.28	0.009
Length	1	7.37878E+12	7.37878E+12	26.95	0.000
Average Slope	1	1.78721E+12	1.78721E+12	6.53	0.013
Functional Class	3	3.63814E+12	1.21271E+12	4.43	0.006
Urban/ rural	2	2.42540E+12	1.21270E+12	4.43	0.015
Error	79	2.16303E+13	2.73802E+11		
Total	88	5.95736E+13			

Figure 78. *Intersection Project Cost Estimate Model D-3.4*

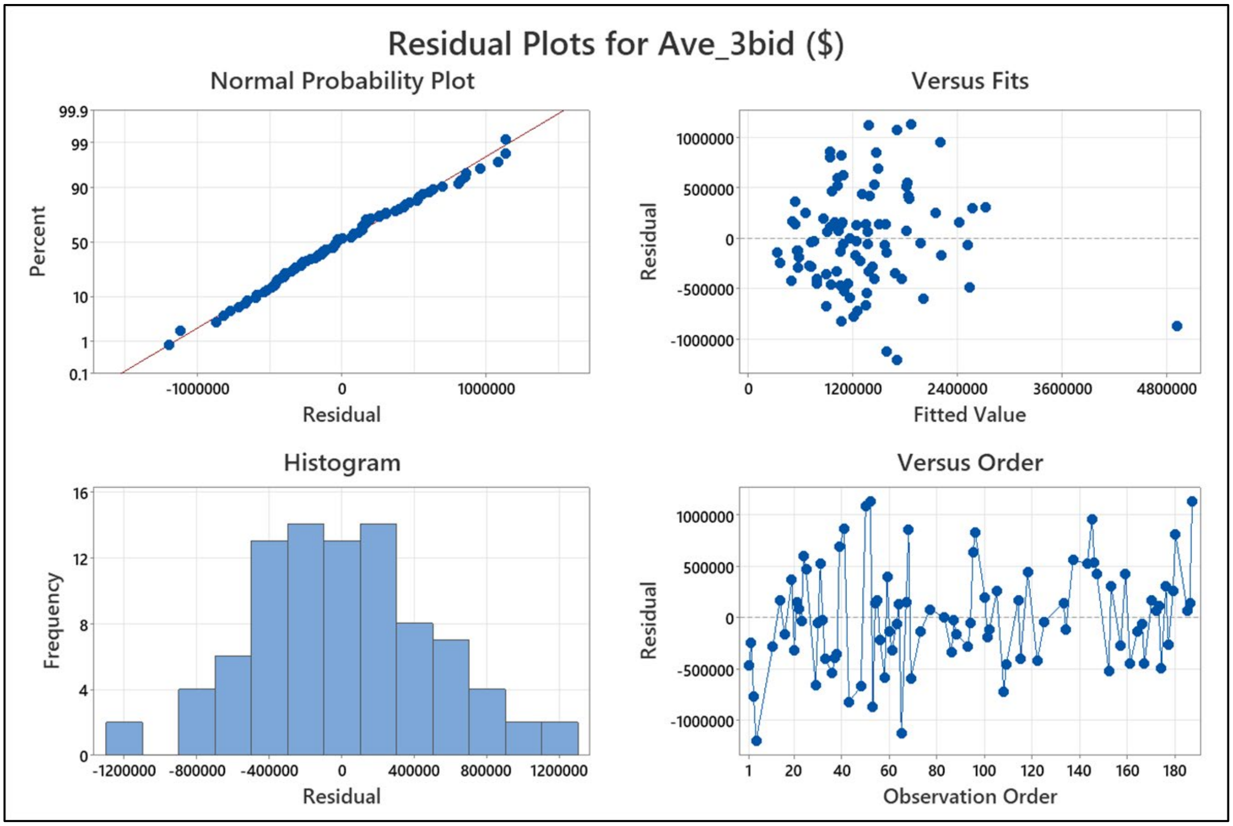


Figure 79. *Residual plots for model D-3.4*

Regression Analysis: Ave_3bid ($) versus Number of Working days from SCD, HCCI, Length, Average Slope, Functional Class, Urban/ rural

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
478024	65.79%	61.85%	56.04%

After removing one outlier project

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	9	3.42813E+13	3.80903E+12	16.67	0.000
Number of Working days from SCD	1	3.51771E+12	3.51771E+12	15.39	0.000
HCCI	1	1.25000E+12	1.25000E+12	5.47	0.022
Length	1	9.44075E+12	9.44075E+12	41.31	0.000
Average Slope	1	9.81835E+11	9.81835E+11	4.30	0.041
Functional Class	3	1.60268E+12	5.34228E+11	2.34	0.080
Urban/ rural	2	2.08599E+12	1.04299E+12	4.56	0.013
Error	78	1.78235E+13	2.28507E+11		
Total	87	5.21048E+13			

Figure 80. *Intersection Project Cost Estimate Model D-3.5*

Regression Analysis: Ave_3bid ($) versus Number of Working days from SCD, HCCI, Length, Average Slope, Functional Class, Urban/ rural

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
456185	68.88%	65.24%	60.17%

After removing 2 outlier projects

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	9	3.54612E+13	3.94013E+12	18.93	0.000
Number of Working days from SCD	1	4.28534E+12	4.28534E+12	20.59	0.000
HCCI	1	7.64208E+11	7.64208E+11	3.67	0.059
Length	1	9.49284E+12	9.49284E+12	45.62	0.000
Average Slope	1	9.43833E+11	9.43833E+11	4.54	0.036
Functional Class	3	1.85567E+12	6.18558E+11	2.97	0.037
Urban/ rural	2	2.44911E+12	1.22455E+12	5.88	0.004
Error	77	1.60240E+13	2.08105E+11		
Total	86	5.14852E+13			

Figure 81. *Intersection Project Cost Estimate Model D-3.6*

Regression Analysis: Ave_3bid ($) versus Number of Working days from SCD, HCCI, Length, Average Slope, Functional Class, Urban/ rural

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
435891	70.68%	67.21%	62.10%

After removing 3 outlier projects

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	9	3.48138E+13	3.86819E+12	20.36	0.000
Number of Working days from SCD	1	4.29491E+12	4.29491E+12	22.60	0.000
HCCI	1	5.33399E+11	5.33399E+11	2.81	0.098
Length	1	1.00636E+13	1.00636E+13	52.97	0.000
Average Slope	1	9.37112E+11	9.37112E+11	4.93	0.029
Functional Class	3	1.62024E+12	5.40081E+11	2.84	0.043
Urban/ rural	2	2.21510E+12	1.10755E+12	5.83	0.004
Error	76	1.44401E+13	1.90001E+11		
Total	85	4.92538E+13			

Figure 82. *Intersection Project Cost Estimate Model D-3.7*

Regression Analysis: Ave_3bid ($) versus Number of Working days from SCD, HCCI, Length, Average Slope, Functional Class, Urban/ rural

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
411600	72.56%	69.26%	64.35%

After removing outlier 4 projects

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	9	3.35932E+13	3.73258E+12	22.03	0.000
Number of Working days from SCD	1	4.63692E+12	4.63692E+12	27.37	0.000
HCCI	1	3.10947E+11	3.10947E+11	1.84	0.180
Length	1	1.02893E+13	1.02893E+13	60.73	0.000
Average Slope	1	6.89985E+11	6.89985E+11	4.07	0.047
Functional Class	3	1.48887E+12	4.96289E+11	2.93	0.039
Urban/ rural	2	1.94065E+12	9.70327E+11	5.73	0.005
Error	75	1.27061E+13	1.69415E+11		
Total	84	4.62993E+13			

Figure 83. *Intersection Project Cost Estimate Model D-3.8*

Regression Analysis: Ave_3bid ($) versus Number of Working days from SCD, Length, Average Slope, Functional Class, Urban/ rural

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
413856	71.88%	68.93%	63.75%

After removing 4 outlier projects

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	8	3.32822E+13	4.16028E+12	24.29	0.000
Number of Working days from SCD	1	4.69795E+12	4.69795E+12	27.43	0.000
Length	1	1.09431E+13	1.09431E+13	63.89	0.000
Average Slope	1	5.53814E+11	5.53814E+11	3.23	0.076
Functional Class	3	1.52707E+12	5.09023E+11	2.97	0.037
Urban/ rural	2	1.92983E+12	9.64914E+11	5.63	0.005
Error	76	1.30170E+13	1.71277E+11		
Total	84	4.62993E+13			

Figure 84. *Intersection Project Cost Estimate Model D-3.9*

Regression Analysis: Ave_3bid ($) versus Year Letting, Number of Working days from SCD, HCCI, Length, Average Slope, Sub Type, Functional Class, Urban/ rural

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
384520	77.33%	73.17%	66.13%

**After removing 4 outlier projects*

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	13	3.58015E+13	2.75396E+12	18.63	0.000
Number of Working days from SCD	1	2.79013E+12	2.79013E+12	18.87	0.000
Length	1	1.00552E+13	1.00552E+13	68.01	0.000
Year Letting	1	1.60545E+12	1.60545E+12	10.86	0.002
HCCI	1	1.24579E+11	1.24579E+11	0.84	0.362
Average Slope	1	1.70992E+11	1.70992E+11	1.16	0.286
Sub Type	3	1.04814E+12	3.49380E+11	2.36	0.078
Functional Class	3	1.50233E+12	5.00778E+11	3.39	0.023
Urban/ rural	2	1.34845E+12	6.74227E+11	4.56	0.014
Error	71	1.04978E+13	1.47856E+11		
Total	84	4.62993E+13			

Figure 85. *Intersection Project Cost Estimate Model D-3.10*

Regression Analysis: Ave_3bid ($) versus Year Letting, Number of Working days from SCD, Length, Average Slope, Sub Type, Functional Class, Urban/ rural

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
384100	77.06%	73.23%	66.48%

**After removing 4 outlier projects*

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	12	3.56769E+13	2.97308E+12	20.15	0.000
Number of Working days from SCD	1	3.05409E+12	3.05409E+12	20.70	0.000
Length	1	1.00209E+13	1.00209E+13	67.92	0.000
Year Letting	1	1.74810E+12	1.74810E+12	11.85	0.001
Average Slope	1	2.97073E+11	2.97073E+11	2.01	0.160
Sub Type	3	9.53940E+11	3.17980E+11	2.16	0.101
Functional Class	3	1.51769E+12	5.05897E+11	3.43	0.021
Urban/ rural	2	1.45816E+12	7.29082E+11	4.94	0.010
Error	72	1.06223E+13	1.47533E+11		
Total	84	4.62993E+13			

Figure 86. *Intersection Project Cost Estimate Model D-3.11*

Regression Analysis: Ave_3bid ($) versus Year Letting, Number of Working days from SCD, Length, Average Slope, Functional Class, Urban/ rural

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
392875	75.00%	72.00%	66.70%

After removing 4 outlier projects

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	9	3.47230E+13	3.85811E+12	25.00	0.000
Number of Working days from SCD	1	3.38169E+12	3.38169E+12	21.91	0.000
Length	1	1.10918E+13	1.10918E+13	71.86	0.000
Year Letting	1	1.44076E+12	1.44076E+12	9.33	0.003
Average Slope	1	4.84423E+11	4.84423E+11	3.14	0.081
Functional Class	3	1.18307E+12	3.94358E+11	2.55	0.062
Urban/ rural	2	1.39955E+12	6.99777E+11	4.53	0.014
Error	75	1.15763E+13	1.54350E+11		
Total	84	4.62993E+13			

Figure 87. *Intersection Project Cost Estimate Model D-3.12*

Regression Equation

Functional Class	Urban/ rural	
IS	Rural	Ave_3bid ($) = -113271594 + 56403 Letting Year + 1040857 Length + 4029 Number of Working days from SCD + 36119 Average Slope
IS	urban	Ave_3bid ($) = -113783212 + 56403 Letting Year + 1040857 Length + 4029 Number of Working days from SCD + 36119 Average Slope
IS	Urban	Ave_3bid ($) = -113341610 + 56403 Letting Year + 1040857 Length + 4029 Number of Working days from SCD + 36119 Average Slope
SC	Rural	Ave_3bid ($) = -114001721 + 56403 Letting Year + 1040857 Length + 4029 Number of Working days from SCD + 36119 Average Slope
SC	urban	Ave_3bid ($) = -114513339 + 56403 Letting Year + 1040857 Length + 4029 Number of Working days from SCD + 36119 Average Slope
SC	Urban	Ave_3bid ($) = -114071737 + 56403 Letting Year + 1040857 Length + 4029 Number of Working days from SCD + 36119 Average Slope
SEC	Rural	Ave_3bid ($) = -113799053 + 56403 Letting Year + 1040857 Length + 4029 Number of Working days from SCD + 36119 Average Slope
SEC	urban	Ave_3bid ($) = -114310672 + 56403 Letting Year + 1040857 Length + 4029 Number of Working days from SCD + 36119 Average Slope
SEC	Urban	Ave_3bid ($) = -113869069 + 56403 Letting Year + 1040857 Length + 4029 Number of Working days from SCD + 36119 Average Slope
US	Rural	Ave_3bid ($) = -113921924 + 56403 Letting Year + 1040857 Length + 4029 Number of Working days from SCD + 36119 Average Slope
US	urban	Ave_3bid ($) = -114433543 + 56403 Letting Year + 1040857 Length + 4029 Number of Working days from SCD + 36119 Average Slope
US	Urban	Ave_3bid ($) = -113991940 + 56403 Letting Year + 1040857 Length + 4029 Number of Working days from SCD + 36119 Average Slope

Figure 88. *Regression Equations for Model D-3.12*

Regression Analysis: Ave_3bid ($) versus Year Letting, Number of Working days from SCD, Length, Average Slope, Urban/ rural

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
404452	72.44%	70.32%	64.76%

After removing 4 outlier projects

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	6	3.35399E+13	5.58999E+12	34.17	0.000
Number of Working days from SCD	1	3.04293E+12	3.04293E+12	18.60	0.000
Length	1	1.44453E+13	1.44453E+13	88.31	0.000
Year Letting	1	1.78476E+12	1.78476E+12	10.91	0.001
Urban/ rural	2	1.00909E+12	5.04543E+11	3.08	0.051
Average Slope	1	3.91662E+11	3.91662E+11	2.39	0.126
Error	78	1.27594E+13	1.63582E+11		
Total	84	4.62993E+13			

Figure 89. *Intersection Project Cost Estimate Model D-3.13*

Regression Analysis: Ave_3bid ($) versus Year Letting, Number of Working days from SCD, Length, Average Slope, Urban/ rural

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
377783	76.03%	74.17%	70.61%

After removing 5 outlier projects

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	6	3.48652E+13	5.81087E+12	40.72	0.000
Number of Working days from SCD	1	2.34704E+12	2.34704E+12	16.45	0.000
Length	1	1.59602E+13	1.59602E+13	111.83	0.000
Year Letting	1	1.90537E+12	1.90537E+12	13.35	0.000
Urban/ rural	2	3.37041E+11	1.68520E+11	1.18	0.313
Average Slope	1	3.33359E+11	3.33359E+11	2.34	0.131
Error	77	1.09894E+13	1.42720E+11		
Total	83	4.58546E+13			

Figure 90. *Intersection Project Cost Estimate Model D-3.14*

Regression Analysis: Ave_3bid ($) versus Letting Year, Length, Number of Working days from SCD, Average Slope

Regression Equation

Ave_3bid ($) = -126452190 + 62632 Letting Year + 1288726 Length
+ 3272 Number of Working days from SCD + 25780 Average Slope

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
378646	75.30%	74.05%	71.78%

After removing 5 outlier projects

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	4	3.45282E+13	8.63204E+12	60.21	0.000
Letting Year	1	2.09921E+12	2.09921E+12	14.64	0.000
Length	1	1.57979E+13	1.57979E+13	110.19	0.000
Number of Working days from SCD	1	2.24379E+12	2.24379E+12	15.65	0.000
Average Slope	1	2.65487E+11	2.65487E+11	1.85	0.177
Error	79	1.13265E+13	1.43373E+11		
Total	83	4.58546E+13			

Figure 91. *Intersection Project Cost Estimate Model D-3.15*