PREDICTING COSTS FOR BRIDGE REPLACEMENT PROJECTS

by

Patrick Stuart Timothy Phillips

A thesis submitted to the faculty of
The University of North Carolina at Charlotte
in partial fulfillment of the requirements
for the degree of Master of Science in
Construction & Facilities Management

Charlotte

2017

Approved by:

____
Dr. Tara Cavalline

____
Dr. Matthew Whelan

____
Dr. Thomas Nicholas

ABSTRACT

PATRICK STUART TIMOTHY PHILLIPS. Predicting costs for bridge replacement projects. (Under the direction of DR. TARA CAVALLINE)

The North Carolina Department of Transportation (NCDOT) uses historical highway bridge records to make cost-effective decisions on which maintenance, repair, or replacement action is appropriate for a deficient bridge. The current method for estimating total bridge replacement cost does not provide reliable and consistent estimates, which impairs forecasting efforts. Updating the current prediction models to include additional factors that may influence cost would theoretically improve the fidelity of the models. Prior studies on bridge cost estimation models for NCDOT and INDOT (Indiana) served as a starting point for the modeling effort detailed in this study. A dataset of recent NCDOT bridge replacement projects was compiled to serve as a foundation for the updated models. Statistical software was used to perform multivariate regression analysis to identify statistically significant predictors and to build models to predict the geometric characteristics of new, replacement bridges (such as new bridge length, width, and span length), as well as right of way costs, engineering costs, construction costs, and total replacement costs. Two approaches were explored in order to predict cost: 1) predicting new bridge characteristics from old bridge characteristics, then predicting bridge replacement costs from predicted new bridge characteristics, and 2) predicting bridge replacement costs directly from old bridge characteristics. New models developed as part of this work were compared to the previously utilized models based on how well the model fit the data (R^2) and the confidence interval of the prediction. When both sets of models were used with current bridge replacement data, the

new models achieved better fits and yielded narrower confidence intervals than the previously utilized models. Comparing the residual error distributions for the different modeling approaches, the models developed to predict costs directly from the bridge characteristics of the structure being replaced were found to out-perform the models developed to predict cost using forecasted characteristics of the replacement structure. Predicting replacement project costs as a total cost (instead of summing the predicted right of way, engineering cost, and construction cost amounts) avoided introducing compounded error from aggregated component cost predictions. For future work, it is recommended that changes with respect to how bridge information is logged into databases would streamline the data conditioning process and increase the usable number of entries for creation of the models.

DEDICATION

I would like to thank all the friends, family, and coworkers that had a hand in supporting me throughout this project. I would like to thank my parents for encouraging me to return to school after finishing my bachelor’s degree and for providing moral (and some financial) support along the way. I would also like to give a special thanks to Dr. Tara Cavalline for giving me the opportunity to work on a project for the NCDOT as I worked on completing my thesis and my degree in Construction and Facilities Management. I would like to thank the members of my committee (Dr. Matthew Whelan and Dr. Thomas Nicholas) for all their assistance over the course of this project and for providing me with valuable feedback. I would also like to thank Corey Rice and Aidan Alar for their assistance during the project. Lastly, I would like to thank our project team’s various contacts at NCDOT for making this project possible.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1: INTRODUCTION

1.1 Introduction to Current NCDOT Bridge Management

The North Carolina Department of Transportation (NCDOT) is responsible for the maintenance of over 13,500 bridge structures across the state (NCDOT, 2017). In order to effectively manage all of these structures, the NCDOT stores inspection data and other pertinent information for each bridge in a series of databases. As with most state transportation agencies, this data is used in a risk-based approach to prioritize future bridge projects and to make cost-effective maintenance, repair, rehabilitation (MR&R), and replacement decisions.

Many decisions made by NCDOT with regards to bridge project selection and prioritization are driven by cost. To make fair comparisons, the BMS needs to associate a dollar value with each remediation alternative. When replacement of a bridge is a possible option (or identified as the necessary option), an accurate estimate of the replacement cost is needed. Within the Bridge Management System software utilized by NCDOT, estimates for bridge replacements are made at a conceptual level, meaning that the estimates only consider a few known project parameters since a detailed design has not yet been made. Although only a conceptual estimate, an accurate estimated bridge replacement cost allows for state transportation agencies to prioritize upcoming projects and to determine which projects can be reasonably covered within a budget.

NCDOT personnel indicate that current bridge replacement cost prediction methods and models used by NCDOT are reasonably accurate for medium-sized bridge projects. However, replacement cost estimates for smaller and larger bridge projects are typically inaccurate. Current cost prediction models employed within the BMS are quite simple and are based upon roadway system classification (primary, secondary, or interstate), with a simple unit cost (dollars per square foot) multiplied by the deck area of the existing bridge (Table 1.1). The inclusion of additional project factors within improved bridge replacement cost models could potentially improve the accuracy of the bridge replacement cost predictions. When utilized in bridge management for thousands of potential highway bridges projects, the needs forecasting analysis would be much improved at the network level. Currently, NCDOT desires a single dynamic model that considers additional project parameters, provides more accurate total bridge replacement project cost estimates, and can be readily updated when necessary.

Table 1.1: Current bridge replacement unit costs in NCDOT BMS

Roadway System Classification	Unit Cost ($/SF deck area)
Interstate	$704.00
Primary	$664.00
Secondary	$529.00

1.2 Research Significance

Using an inaccurate cost estimation model that does not consider many important factors will produce highly variable results, affecting the ability of a state highway agency to effectively evaluate MR&R alternatives for bridges, to identify when replacement is the desired option, to prioritize projects, and to forecast agency needs. Estimates generated with a wide confidence interval make it difficult for state transportation agencies to correctly anticipate funding needs when requesting state resources for bridge replacement. Significantly overestimated bridge replacement costs may delay the letting of additional bridge projects. Conversely, if a replacement cost is significantly underestimated, the agency is at risk of having to delay work on projects that have already been let or otherwise address this shortcoming. Use of accurate bridge replacement cost models, based on recent bridge characteristics and replacement cost data, will aid in both project prioritization and budget forecasting.

As mentioned previously, current bridge replacement cost prediction models in the BMS utilize only roadway system classification and deck area of the existing bridge as predictor variables. Changes in design loads and required capacity of bridges, waterway and floodplain requirements, and other factors often require replacement bridges to be longer and wider than the original bridge, causing the simplified replacement cost method programmed into the BMS to be inaccurate. Since bridge replacement costs are influenced by the design of the structure, the ability to make reliable predictions for the characteristics of the replacement structure could be useful in strengthening the accuracy of the final cost estimates. Another way to improve the accuracy of these models would be consideration of additional variables that can be statistically shown to be linked to bridge replacement cost. These could potentially include factors such as location, design type, bridge materials, average daily traffic (ADT), and type of route carried. Additional factors that may affect bridge replacement cost are already stored in the BMS and other auxiliary databases available to NCDOT's Structures Management Unit. Since much of this data is collected regularly, these factors would be relatively easy to integrate into the forecasting models, if deemed to be significantly related to bridge replacement costs.

Due to the changing nature of infrastructure design and construction cost prediction models should also be dynamic and easily updated. Changes in design loads, traffic demands, and highway regulations can render a static prediction model obsolete. These requirements also dictate bridge design, which ultimately has a driving influence over cost. Providing a clear methodology for developing prediction models based on a number of years of recent data would allow for models to be adjusted and refined as necessary. The result of updating bridge replacement cost models over time could have effects as minor as changed coefficients, or as extensive as adding or removing variables from the equation.

With a more accurate cost prediction model (or models), the NCDOT could make more informed decisions on funding and prioritizing their projects. On a single-project basis, a more accurate replacement cost estimate should lead to a lower likelihood of actual project cost exceeding the projected cost during the forecasting stage. From a network standpoint, improved bridge replacement cost models would help improve the overall condition of the bridges owned and maintained by NCDOT by allowing for more projects to be let each year with fewer delays caused by inadequate funding. Successful development and implementation of bridge replacement cost models could also provide guidance to other state transportation agencies interested in adopting improved cost estimating models for their asset management programs.

1.3 Objectives

Below is a list of the objectives addressed in this thesis:

1. Compile records of recently updated bridge characteristic and cost data from several NCDOT databases into a central dataset that is suitable for use in developing new bridge prediction models.
2. Use data from the central dataset to evaluate the accuracy and appropriateness of older prediction models for project costs and new bridge attributes developed by other researchers.

3. Consider using one of the two approaches for estimating bridge replacement project costs: (1) by predicting the bridge characteristics for the replacement structure from those of the structure being replaced and then predicting cost from those estimated values, or (2) predicting cost directly from the bridge characteristics of the structure being replaced.
4. Create a set of updated bridge replacement prediction models that improve the accuracy of estimates by inclusion of additional, statistically significant predictor variables. These models can be used to predict project costs or changes in bridge dimensions between the existing and replaced structures.
5. Select one "recommended" model for each dependent (predicted) variable
6. Identify the best approach for predicting cost, whether it be to predict directly from old bridge characteristics or to predict with forecasted new bridge characteristics (per Objective 2)

1.4 Organization of Thesis

This thesis consists of six chapters. The first chapter serves as an introduction to forecasting for bridge management and presents the need for accurate cost estimation models. The second chapter is a literature review of cost estimation methods for highway construction projects, with a focus on bridge replacement costs. The literature review also includes information on how costs that occur before, during, and after construction have an impact on the total project cost. This chapter also includes information on how other state agencies estimate highway bridge project costs. In the third chapter, the research methodology used for this project, including how the data was sourced, compiled, and conditioned, is presented. The mathematical models created for each dependent variable associated with new bridge characteristics and bridge replacement costs are presented in the fourth chapter. In this chapter, the recommended model for each dependent variable based on model complexity and standard error of the estimate is also identified. The fifth chapter describes the model validation process, in which two different modeling approaches are considered for each predicted variable. Final recommendations for each model and approach are included in this chapter. The sixth chapter provides a conclusion to the work performed by summarizing the steps taken to update the prediction models, the results, and future research needs that were identified during this study.

CHAPTER 2: LITERATURE REVIEW

2.1 Cost Estimation for Construction Projects

Cost estimation for construction projects has been described as a combination of art and science (Gould 2005). An estimator should be able to think creatively and use their experience and judgement to make assumptions for uncertain conditions. Estimators should also be methodical, organized, and able to manage complex calculations. The estimator's strengths come in part from experience with similar projects and from an ability to visualize how conditions may change in the future, whether it be within the term of the project or years down the road. Collier (1984) describes this kind of knowledge as "experiential" information. In the absence of detailed design information, the estimator injects his or her experiential judgements that are made based on any general project information that is available. In the early stages of a project there is little design information available, so the estimator must rely heavily on their own personal experiences and rules-of-thumb to determine the general cost for the project (Collier, 1984).

Reliable cost estimates are an asset for owners. Even the most basic preliminary estimates can give the owner an idea of whether the project is economically feasible. As the design is developed, more detailed estimates can help an owner find a reasonable tradeoff between scope and quality. For projects procured with a bidding stage, a final estimate based on the completed design gives owners an idea of the project's value to benchmark the contractor's bid estimates (Gould, 2005).

2.1.1 Types of Cost Estimates

When discussing cost estimates for construction, it is important to differentiate between the terms *cost* and *price.* For the owner, the price that they pay for a completed project is usually greater than the cost to construct the project. That is because the cost to the contractor includes more than just the materials and manpower needed to complete the project. The contractor also has to pay for mobilization, demobilization, idle time, small tools, insurance, and permitting. These "direct costs" have to be accounted for by the contractor and charged to the owner as "reimbursable" costs. Additionally, the contractor also charges the owner for indirect costs that are "non-reimbursable," since they cannot be attributed to specific items of work at the site. Common indirect costs for the contractor include operational (home office) costs, contingency, and the contractor's profit. Since these costs are less tangible than direct costs, the contractor estimates these as either a fixed percentage or as part of a lump-sum amount, depending on the type of contract with the owner. For the owner, the price that they pay for a project is the sum of the direct and indirect project costs charged to them by the contractor (Collier, 1984).

It is important to remember that not all cost estimates are equal. Two estimates made at different points on a project's lifespan are different because the quality of the project information improves along the pre-construction timeline. Even the least accurate type of estimate serves a purpose to the owner. The following sections introduce and describe the different types of estimates used by owners during each phase of the pre-construction and construction stages of a project.

2.1.1.1 By Project Phase

The type of cost estimate that can be generated for a project is dependent on the amount of information available to estimators. As the design for the project matures, more information becomes available to estimators, which allows for more detailed estimates. Figure 2.1 shows the type of estimate used for each phase of the project (Schexnayder et al., 2003).

Project Stage	Concept Development	Design	Advertisement	Bid/ Award	Construction
Time	→				
Estimate	Conceptual Estimate	Design Estimates	Prebid Estimates	Bid Analysis	Change Orders

Figure 2.1: Estimate development in relation to project development (Schexnayder et al, 2003)

In the conceptual stage of a project, not enough details are available to create a detailed cost estimate based upon material takeoffs or design documents. Conceptual estimates are typically developed based on the estimator's knowledge and experience, and are calculated based upon cost per square foot, previous projects, or order of magnitude (i.e. cost per room, cost per parking space) (Levy 2006). Typically, the only known attributes for a bridge replacement project at this stage of the project are forecasted structure dimensions based on location and anticipated traffic demand (Abed-al-Rahim & Johnston, 1995). To get a basis of understanding for how much the replacement project will cost, a mathematical model can be used to compute the estimated cost based on available known variables.

Conceptual estimates can be created from gross historical bidding data. Without a complete design, there are many unknown factors still present that may affect cost. Estimates generated after design and before bid are sometimes referred to as a state's or engineer's estimate and are detailed enough to finalize project funding prior to bid solicitation (Schexnayder et al., 2003). When estimators create these estimates several key assumptions are made. Some of these key assumptions may be that the project scope will not change, inflation has been accounted for, no unanticipated regulatory changes will occur, no strikes, no damaging weather, and that the project will not be mismanaged (Schexnayder et al., 2003).

During the design development phase, the owner's design team makes decisions on certain aspects of the design. For a bridge project, this may be the substructure design (piles versus post and sill), deck material, or number of spans. Each component of the design can be priced based on historic data and calculated as a percentage of the total project cost. The owner can make decisions on whether one of the components would cost too much and if there is a more economically feasible alternative for that part of the design. In some cases, the owner may elect to reduce the scope or size of the project to preserve quality (Gould, 2005).

Before advertising a project for bidding, the owner or the owner's construction manager will create a more detailed estimate for the project's cost. Since the design is almost complete by this point, the estimator can use more accurate unit prices for each component of the project. Not only does this allow the owner to determine the "fair" price for the project, but it also helps familiarize the owner with the contents of the contract documents and allows the owner to project day-to-day cash flow needs with a cash flow analysis (Gould, 2005).

Bidders for a construction contract prepare detailed pre-bid estimates based on the contract documents provided with the bid advertisement. The contractor's estimators create material takeoffs from design information found in the specifications and plans, such as cubic yards of concrete or linear feet of guardrail. Breaking down the project into smaller operations also allows estimators to estimate the manpower and equipment required for that operation. A well-organized and comprehensive list of operations with item codes reduces the likelihood of an estimator omitting part of the project in their estimate. Additional amounts are added to each subtotal to cover overhead costs and profit, which are included in the final price that the owner pays for a work item. Typically, the contractor's operational overhead costs and profit are calculated as a fixed percentage of the direct costs while job overhead items can be represented as a unit cost or lump-sum amount (Peurifoy, 1975; Collier, 1984).

As with the preliminary estimates prepared by the owner, the contractor or the contractor's estimator must also consider project-specific factors that affect the material and labor rates for a project. A site visit allows for the estimator to identify site problems, such as accessibility, location, and site clearing, that would lead to higher mobilization costs. Knowledge of local material prices, wages, and availability of skilled workers helps estimators make more informed decisions when they assume a unit price for an operation (Foster, 1972).

Change orders are a way for contractors to seek equitable adjustment for lost time or money during the construction phase. This typically happens whenever there are circumstances that delay the final completion date of the project. The cause of the delay will dictate whether the contractor is owed additional time or money from the owner. These causes can range from severe weather, worker illnesses, and labor shortages to inadequate drawings and delays in permitting (Levy, 2006).

For instances where the delay was out of the control of the contractor but caused by the owner or members of the design team, the contractor can recover costs associated with that delay. This includes direct cost items, such as equipment rentals, labor, materials, stocking, subcontractors, and transportation. Contractors can also be reimbursed for indirect costs incurred from the delay, which includes the additional operations costs for both their field office staff and any home office staff involved with the project. A third compensable cost category, known as impact costs, includes losses in productivity, shortages of skilled workers, and extended warranties that resulted from the delay in construction. When added up, the apparent and "hidden" costs of a compensable delay can have an extensive impact on the project's budget (Levy, 2006).

2.1.1.2 Top-Down versus Bottom-Up Estimating

Top-down estimates are made by looking at the project from a macro level. These estimates can be made when most of the design has not yet been developed, which makes top-down estimating ideal for creating conceptual estimates. While top-down estimates can be helpful for understanding the "big picture" of the project, the reliability of the estimation is more difficult to control. In the absence of specific design information, the estimator must make educated assumptions about the project based on any general project parameters that are available. The quality of these assumptions can depend on the experience of the estimator (Gransberg et al., 2013).

As the project's design becomes more developed, bottom-up estimating can be used to create more accurate predictions for both the total project cost and individual work items. Bottom-up estimating works in a similar fashion as top-down but on a much smaller scale. After the project has reached a point where the work items can be organized into a work breakdown structure (WBS), a top-down estimate is performed on each WBS item. The total predicted cost of the project can be found by adding up the individual estimates for all the items in the WBS (Gransberg et al., 2013).

Figure 2.2 shows how top-down and bottom-up estimates are performed for pre-construction services (PCS), which includes engineering and right-of-way acquisition costs (Gransberg et al., 2013). The process shown in Figure 2.2 could be applied to other individual aspects of a project. Both estimating methods produce an overall cost estimate, however the individual sources used to produce each estimate have different levels of detail. The three-point estimates in the bottom-up method are essentially smaller scale top-down estimates for specific tasks (Gransberg et al., 2013).

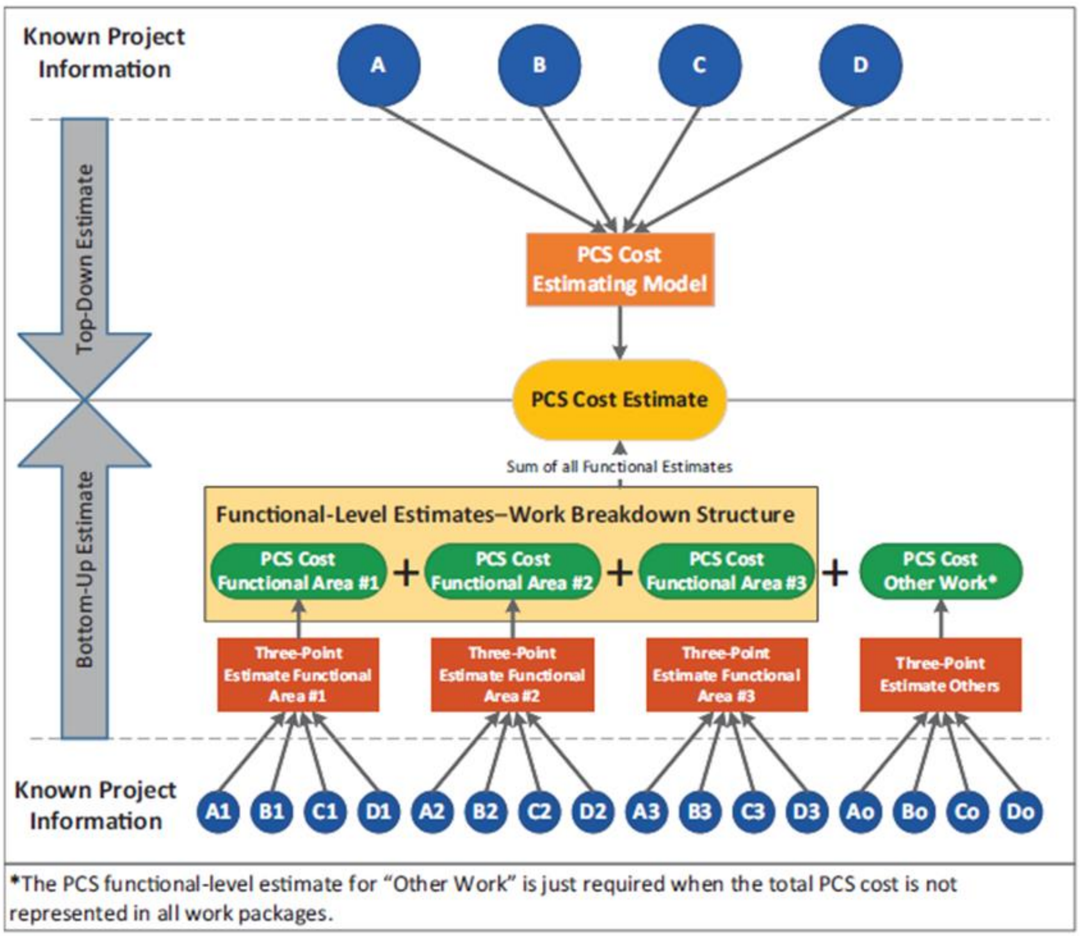

Figure 2.2: Top-down and bottom-up estimating methods for PCS costs (from Gransberg et al., 2013)

During a construction project, the effectiveness of top-down and bottom-up estimates will vary depending on the current phase of the project. In Figure 2.3, the change in effectiveness for both methods as the design matures is illustrated. Since top-down estimates rely solely on generic project parameters, they are most effective at the very beginning of the planning stage where there is not enough detail to use the bottom-up approach. As the design is being developed, the usefulness of the top-down method decreases since bottom-up estimates tend to be more accurate and useful for allocating and managing resources. Gransberg et al. (2013) found that bottom-up estimates for PCS were most useful right before the final design phase. At this point in the preconstruction phase, there is enough design information available to create a reliable estimate. Beyond that point, in-house departments or third-party consultants will manage the preconstruction services, so the risk of any further cost escalation is relatively low.

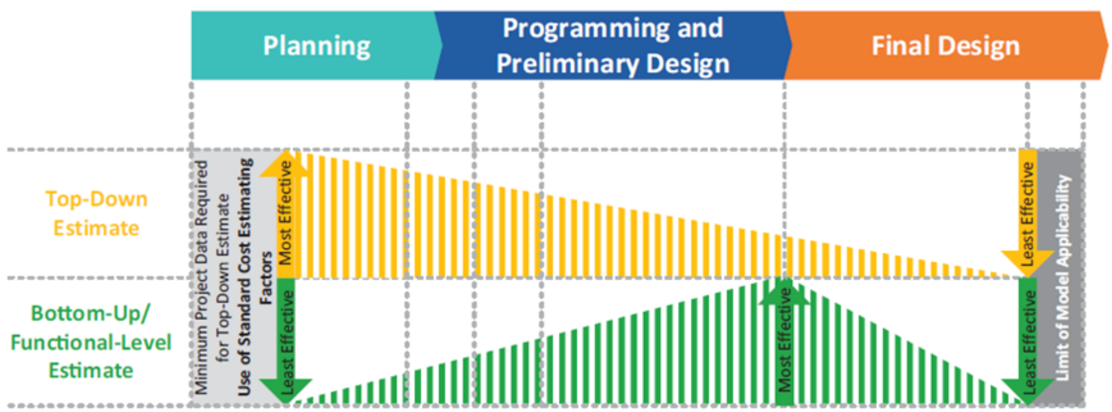

Figure 2.3: Relative effectiveness of top-down and bottom-up estimating methods (from Gransberg et al., 2013)

Gransberg et al (2016) provides a six-step framework for state agencies to follow when creating a top-down or bottom-up cost estimating model for PCS costs. This general approach can be applied to other types of construction costs that would use their own prediction models, such as construction or right-of-way costs. The framework functions as a cycle, which allows for agencies to make continuous improvements to their models. Figure 2.4 shows the six steps in the PCS cost estimating model creation framework.

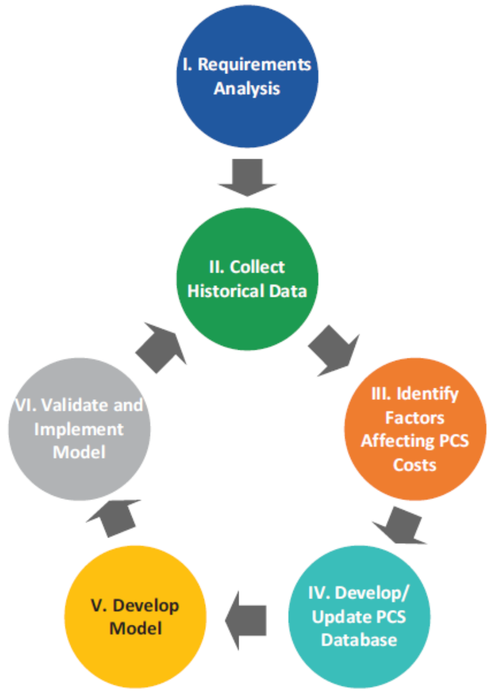

Figure 2.4: PCS cost estimating model framework (from Gransberg et al. 2016)

The first step of the model, Requirements Analysis, is where the state agency decides on whether they will be using a top-down or bottom-up estimating approach. Depending on the method chosen, the agency will also have to determine what historical data will be needed for the estimate and where the data can be sourced. In the second step, the state agencies collect historical project data from a variety of separate databases and compile it into a central database. During this stage, the estimators need to ensure that the data meets their standards of quality, quantity, and level of detail. The database should also be tailored to the end user of the data, whether it be an estimator that needs project-level historical data or a geotechnical engineer that needs specific information on soil conditions or regional geography (Gransberg et al. 2016).

After the central database has been created, the next step of the process is to identify significant factors that affect PCS costs. This can be done through engineering judgement and expertise or with a structured statistical process. Following the identification of these factors, the agency can develop or update their PCS database with respect to the significant variables. Availability of historical data can potentially limit the effectiveness of the database if the data does not meet a certain level of granularity (Gransberg et al. 2016).

The fifth step is where the model is created from the historical data. The model consists of both qualitative and quantitative aspects. The qualitative portion of the model relies on the experience and judgement of the model's creators and users to assess the quality of the data and interpret the results. Gransberg et al. (2016) discusses four different quantitative methods used in PCS cost estimating models: decision tree, multiple regression, artificial neural networks (top-down method), and three-point estimating (bottom-up method). The sixth and final step involves the validation and implementation of the new model. After the model has been validated and deemed satisfactory, it is recommended that the creators of the model set up a system to track how well the model's estimate performs and compares to actual costs. If there are discrepancies between estimated and actual costs, the creators and end users should be able to identify the causes for the deviations and create a list of "lessons learned" that can be adapted into the next development cycle of the model (Gransberg et al. 2016).

2.1.2 Sources of Data in Cost Estimates

The source of data for a cost estimate will depend on what information is available at the time of the estimate. For conceptual estimates, historical data is used to obtain typical unit costs per lane mile, interchange, or square foot of deck area. After the design has been developed to the point where specific units of work can be quantified, estimators can use the Historical Bid-Based estimating method to generate an estimate based on historic cost data. This data is often obtained from previously submitted bids for similar projects or work (WSDOT 2008). Cost estimates with the highest level of detail are created from assumed unit prices and quantity takeoffs (Foster, 1972; Peurifoy, 1975).

2.1.3 Sources of Error in Cost Estimates

The high-profile nature of most bridge projects requires that schedule and budget performance be closely monitored (Wilmot and Cheng, 2003). It is in the best interest of state transportation agencies to provide accurate estimates, or else explain publicly why the public funding was mismanaged (Wilmot and Cheng, 2003). Underestimating the cost of a project leads to delays as agencies search for additional funding, while overestimating can cause missed opportunities for projects that could have been partially or fully funded from that excess amount (Kyte et al., 2004).

According to Schexnayder et al. (2003), "an estimate is accurate if it is close to the actual final cost of the project." In describing "close to the actual final cost," Schexnayder et al. (2003) state that a good estimator will generate estimates that are reasonably close to actual costs with a reasonably small standard deviation. Acceptable confidence bounds will defend on the type of estimate and which stage of the project in which the estimate was compiled. As the project becomes more defined and more information becomes available, the confidence range becomes narrower. This is because there is less uncertainty once the design has been complete, but since uncertainty always exists it is still incorporated into the engineering estimates to some extent.

Since cost estimates are predictions, they can be wrong. Early optimism can lead to false precision, which poses problems to the schedule and scope of work (Schexnayder et al., 2003). When cost goes up, the budget must be increased or the scope reduced to keep the project cost within budget limits. As a result, the project becomes more expensive and its overall value is reduced. When the final project cost exceeds the original low bid cost, the overrun can be caused by bidding errors, poor design, constructability issues, project complexity, poor construction management, site conditions, and labor and material availabilities (Wright and Williams, 2001). Since it is difficult to anticipate the presence of these factors at the very beginning of a project, it is even more difficult to predict the magnitude to which these factors will increase the project cost.

The estimated project cost may also fall short of the actual project cost when estimators fail to apply a cost inflation factor for future year estimates. Many state agencies estimate future costs by using a construction cost index or extrapolating trends from prior years. Both methods fail to consider characteristics that have an impact on contract cost, such as contract size, duration, location, bid variance, and changes in construction practices. Wilmot and Cheng (2003) proposed a model that accounts for additional variables that have a statistically significant impact on contract costs. The new model, developed for the Louisiana Department of Transportation and Development, tends to estimate greater cost escalation, which leads to more conservative cost estimates. Even in optimal economic conditions, the model anticipates that increases in the costs of petroleum products and construction machinery will outpace the standard inflation rate. While this increase is inevitable, it can be managed and controlled by increasing contract size, reducing contract duration, minimizing plan changes, and letting fewer projects during the fourth financial quarter; all of which were shown to be significant factors in the construction cost prediction model (Wilmot and Cheng, 2003).

2.2 Overview of Cost Estimation for Bridge Replacement Projects

State transportation agencies have three alternatives to improve a deficient bridge: maintenance, rehabilitation, or replacement. Bridge replacement projects require the greatest proportion of funding (Abed-al-Rahim and Johnston, 1995). Cost predictions for replacement projects are used to estimate each bridge's present and future funding needs and create a reliable highway construction program (Behmardi et al. 2015; Wilmot and Cheng 2003).

2.2.1 Components of Total Project Cost

The cost of a bridge replacement project reflects more than just the cost of building the new structure. Costs for demolition, detour routes, surveying, design, inspections, and approach roadway improvements should also be considered when estimating the total cost of a project. Parameters such as bridge functional classification and bridge size will likely also affect the final estimated cost (Abed-al-Rahim and Johnston, 1995).

2.2.1.1 Construction Cost

The overall construction cost of a bridge involves several distinct work items. Before the new bridge can be constructed, the site must be cleared. This may involve demolition of an existing bridge structure, acquisition of right-of-way property, and relocation or removal of underground utilities (Behmardi et al., 2015; Heiner and Kockelman, 2005). Earthwork, erosion control, and construction of the bridge abutments and approach slabs are also part of the bridge construction process (Wahls, 1990). Transportation and installation of substructure and superstructure components also contribute to the overall construction cost (Saito et al., 1991).

In addition to the quantity of each material used in the bridge design, the location of the project can place additional constraints on the methods available to the contractor, which may drive up the cost of construction. For example, a bridge that crosses a waterway may have underwater substructure components that require dewatering and the installation of coffer dams to allow workers to work in dry conditions (Purvis, 1994). A shortage of fill material for the abutments may necessitate importing fill material from other areas, which can also inflate the cost of construction (Wahls, 1990).

As discussed previously, the contractor charges the owner for equipment, overhead, contingency, and profit. A more complex project may prompt the contractor to charge additional amounts for labor, specialized equipment, or greater contingency to cover the increased risk. As a result, owners will pay a greater price for construction of bridge replacement projects that are large, complex, or with less-than-optimal environmental constraints.

2.2.1.2 Roadway Cost

The need for additional capacity and mobility during a bridge replacement project often requires state agencies to purchase private or public land for the improvements. The Right-of-Way (ROW) acquisition process involves the highway agency acquiring additional land from the legal property owner while providing the property owner a reasonable compensation based on fair market value of the parcel (Chang-Albitres et al. 2014). Right-of-Way acquisition can be time consuming and costly for transportation projects (Aleithawe 2017). Under ideal circumstances, the ROW property can be acquired quickly and at fair market value. However, any delays in acquiring the property in a timely manner minimizes any potential savings for the highway agency and introduces additional risk of the project deviating from the budget and schedule (Chang-Albitres et al., 2014).

The costs for acquiring parcels includes the value of the parcel (or portion of the parcel) and any damages that must be paid out to the owner for having to relocate (Heiner and Kockelman 2005). Rising acquisition prices have prompted state DOT's to focus on minimizing ROW costs by prioritizing which parcels to purchase first (Chang-Albitres et al. 2014). These decisions are time-sensitive in nature, as land values can increase over the time that a decision is being delayed.

When performing any sort of site work, interference with existing utilities can have lasting effects on a project's schedule and budget. Utility Conflict Cost (UCC) is the combined direct and indirect estimated costs for the conflict resolution for each utility conflict (Aljadhai and Abraham 2016). If the utilities are relocated, potential costs include the relocation cost, risk to the project schedule, and impact on nearby facilities. If the utilities are left in place, cost items may include impacts on traffic, nearby facilities, and pavement service life (Aljadhai and Abraham 2016). The UCC can be estimated with different equations depending on how critical the conflict is.

2.2.1.3 Design Cost

The preliminary engineering (PE) phase of a highway project aims to accomplish two goals. The first goal is to minimize the physical, social, and human environmental impacts posed by the project. The second goal is delivery of the best solution by way of engineering design. Accurate PE estimates promote proactive allocation of funds and fiscal responsibility (Hollar et al. 2013). With tighter constraints on spending at the government level, the need for accurate and reliable cost estimates has made itself a priority (Gransberg et al. 2016).

2.2.2 Adjustment of Costs for Inflation and Productivity

When dealing with a set of historical cost data or predicting costs for future projects, it is necessary to apply a factor to the estimated amount to account for inflation and changes in productivity between years. There are several construction cost indexes that account for these factors and can be used to covert the value of a dollar from one year to another year, such as the RS Means Historical Cost Index, ENR Index, and the FHWA Price Index (Abed-al-Rahim and Johnston, 1995).

Values for the RS Means Historical Cost Index are based on a historical cost index where January 1, 1993 is equal to 100 and a current index where January 1, 2018 is estimated to be equal to 100. This allows for conversion of national average building costs between different time periods (RS Means, 2017). The ENR Index is available for 20 specific US cities based on data from 1978 to 2012 and considers material and labor costs (Engineering News Record, 2017). The National Highway Construction Cost Index, published by the FHWA, allows for conversion and prediction of construction costs for highway projects (FHWA, 2017).

The study conducted by Wright and Williams (2001) used data from 298 highway projects let by the New Jersey Department of Transportation (NJDOT) from 1989 to 1996. To make comparisons between projects let in different years, Wright and Williams applied the ENR Construction Cost Index to Formula 2.1 to convert dollar values from all the projects to their 1999-equivalent values.

$$C_2 = C_1 \times \left(\frac{I_2}{I_1}\right) \tag{2.1}$$

Where: C_1 = Cost in Year 1 dollars
C_2 = Cost in Year 2 dollars
I_1 = ENR Construction Index value for Year 1
I_2 = ENR Construction Index value for Year 2

Since cost indexes are based on data from construction projects, there are no values available for future years. To estimate the adjusted cost of a project for a future year, the index data can be extrapolated using regression techniques. For early work using the NCDOT BMS, Abed-al-Rahim and Johnston (1995) used the FHWA Structures Index to convert their bridge cost data to a common year. The first step of the conversion was used to bring the bridge project costs to a common base year, using an equation similar to Equation 2.5. The limitation to this method is that the base year must fall within the range of years from which the construction index data is sourced. To bring these common base year costs to present or future values, Abed-al-Rahim and Johnston (1995) developed a linear regression equation from the construction index data that extrapolated future year index values with a relatively good fit (R^2 = 0.84). The future year index value found in Equation 2.2 can be plugged back into Equation 2.1 to solve for the cost in Year 2 dollars.

$$IND_{(YF,YB)} = 102.21 - 3.9(YB - YF) \tag{2.2}$$

Where: $IND_{(YF,YB)}$ = Cost index for future year YF and base year YB
YB = Base year
YF = Future year

2.3 Cost Prediction Modeling Approaches

Demand for accurate cost forecasting methods for highway projects has prompted several state transportation agencies to fund research projects on cost prediction modeling. In this section, an overview of several different approaches is presented, along with background information required to develop these models. This section also covers how predictions from the models are used in the real world through implementation into a bridge management system.

2.3.1 Types of Variables

Variables can be classified by the way their data is recorded. *Continuous* or *quantitative* variables are numerical values that can be measured at any point along a range of possible values. The granularity of the data is only limited by the precision of the instrument which provided the original measurement. Many variables included in a BMS, such as daily traffic, length, and width, can be considered continuous or quantitative variables. *Discrete* variables can also be expressed as numerical values, however there is no smooth transition between values. One example of this would be the number of spans for a bridge. This field can only be expressed in whole numbers, since half-span bridges do not exist. The distinction between continuous and discrete numerical variables can become blurred whenever the precision of the continuous variable is limited or the steps between the discrete values become very small (Tabachnick and Fidell, 2006).

Discrete variables can also describe non-numerical *qualitative* data. In a BMS database, this could be deck material, functional classification of the route, or whether the bridge crosses over water or a grade change. *Dichotomous* variables have only two possible values (Tabachnick and Fidell, 2006). Categorical variables can be used to set up discrete variables into grouped categories.

2.3.2 Types of Models

The cost for bridge replacement projects can be estimated through traditional cost estimation or aggregated statistical modeling. Traditional cost estimates are calculated by listing all of the work items and multiplies their quantities by a unit price. The sum of all the costs for the work items is the estimated value for the complete project. Aggregated statistical modeling uses historical data on bridge costs and attributes to predict the project cost based on models developed through linear regression analysis (Behmardi et al. 2015).

2.3.3 Regression Analysis

Regression can be described as a statistical method that can be used to investigate the relationship between variables (Dodge and Marriott, 2003). Regression analyses are often performed to answer the following question: "How do changes in x affect the value of y?" If a relationship exists between the dependent variable (y) and the one or more independent variables (x_1, $x_2 \ldots x_n$), the value of the dependent variable can be predicted using a mathematical model (Dowdy and Wearden, 1991).

There are several different types of models to choose from. In simple linear regression, the relationship between one dependent variable and one independent variable can be modeled with a straight line, as seen in Equation 2.3. Ideally, this straight line should “fit” the actual data on a scatter plot and minimize the sum of the squares of the vertical differences between the line and the data points. The coefficient of determination (R^2) measures how well the regression model fits the data. The value of R^2 ranges from 0 to 1, with higher values indicating a better fit (Dodge and Marriott, 2003; Dowdy and Wearden, 1991).

$$Y' = A + BX \tag{2.3}$$

Where: Y' = Predicted Score
A = Value of Y when X is equal to zero
B = Slope of best-fit line
X = Value from which Y' will be predicted

To solve for the predicted score of Y', values for both A and B must be found. First, the bivariate regression coefficient (B) is calculated by using Equation 2.4. The coefficient is a ratio of the covariance of the two variables (X and Y) and the variance of X and is also the slope of the best-fit line (Tabachnik and Fidell, 2006). After B has been found, the x-intercept (A) can be calculated from Equation 2.5.

$$B = \frac{N\sum XY - (\sum X)(\sum Y)}{N\sum X^2 - (\sum X)^2} \tag{2.4}$$

Where: B = Bivariate regression coefficient
X = Independent Variable
Y = Dependent Variable

$$A = \bar{Y} - B\bar{X} \tag{2.5}$$

Where: A = X-Intercept
$\bar{X}$ = Sum of values used for the prediction
$\bar{Y}$ = Sum of values to be predicted

Multiple regression is an extension of bivariate regression in which more than one independent variable is used to predict values of a dependent variable (Tabachnik and Fidell, 2006). For example, in the case of this project, it is useful to predict the construction cost of a bridge replacement project (DV) based on the several independent variables available in the data set, such as structure length, number of spans, material, or design type. The multiple linear regression equation (Equation 2.6) is an extension of the bivariate regression equation (Equation 2.3) that is designed to be used with more than just one independent variable. Each independent variable has its own regression coefficient, which is used to bring the predicted values of Y as close as possible to the values from the data set and maximize the correlation between the predicted and obtained values for Y.

$$Y' = A + B_1X_1 + B_2X_2 + \cdots + B_kX_k \tag{2.6}$$

Where:
Y' = Predicted score for dependent variable
A = Value of Y when all X values equal zero
B_n = Regression coefficient for n-th variable
X_n = n-th independent variable
k = Number of independent variables

Collinearity is another consideration for regression equations that involve multiple independent variables. This condition exists when there is a high amount of correlation between two or more predictor variables. In layman's terms, the two variables are measuring the same thing (or highly interrelated things). In a multiple regression analysis, collinearity that is not addressed will cause variables that truly affect the dependent variable to not appear in the regression equation while the other predictor variable may have a large impact on the equation. There are several ways to deal with collinearity between variables. After the collinear variables have been identified, the two variables can be combined into one single variable by converting each of the variables into a *z* score and them using the sum of the *z* scores as the total for the new variable. Another approach is to use a factor analysis that will identify the set of factors within the collinear variables and use the factors in the regression analysis (Cramer and Howitt, 2004). Collinearity can also be addressed by removing one of the collinear variables from the regression model.

Performing multivariate regression with a bridge dataset containing hundreds of entries would be burdensome if done by hand. Fortunately, there are several computer programs that can perform the calculations automatically and provide an equation for the best fit line for a set of data. Minitab, Matlab, SPSS, and SAS are all popular computer software programs that can manage large quantities of data and perform different types of regressions.

2.3.4 Use of Cost Prediction Models in Bridge Management Systems

All state transportation agencies are required to comply with the Intermodal Surface Transportation Efficiency Act of 1991 by implementing a BMS that logs bridge data and considers the costs of repairing, rehabilitating, or replacing deficient bridges (Abed-al-Rahim and Johnston, 1995). The three alternatives are typically evaluated based on a variety of considerations, including ownership and user costs, as well as budget constraints and the preferences of state/local personnel. The decision to replace a functionally obsolete or deteriorated bridge will bring the user cost back to zero at the beginning of the new bridge's service life (Chen and Johnston, 1987).

Currently, the NCDOT BMS computes the bridge replacement cost using the bridge deck area and a unit cost based on functional classification. The deck width and length for a new bridge is calculated based on the desired level of service. Design and planning of the new structure is estimated as a fixed percentage of the base construction cost. Costs associated with roadway improvements can be added onto the subtotal as a fixed amount (Chen and Johnston, 1987).

2.4 Existing Cost Prediction Models

Conceptual cost estimating is hardly a new concept. The term "conceptual estimate" was first recognized in 1975 by a federal government publication that urged construction managers to familiarize themselves with the technique (Collier, 1984). Around this time, computerized bridge management systems were being developed to catalog bridge inspection data and prioritize bridge maintenance needs (Chen and Johnston, 1987). The ability of a BMS to estimate the cost to replace a bridge helps the system users to evaluate whether it is more feasible to repair, rehabilitate, or replace the bridge (Abed-al-Rahim and Johnston, 1995). By 1992, these systems had been implemented by several states, providing these agencies with the ability to consider user costs, owner costs, level-of-service goals, or life-cycle activity profiles to estimate replacement costs (Organization for Economic Co-operation and Development, 1992). The following sections identify and discuss some of the cost modeling approaches developed for North Carolina bridges and for bridge systems in other states.

2.4.1 Usage of Cost Prediction Models in North Carolina

At the time of the research conducted by Chen and Johnston (1987), NCDOT estimated bridge replacement costs with a fixed unit cost of $43 per square foot of deck area. This same unit cost would be applied to all bridge replacement projects without regard to project size, location, design, or traffic volume. Further research was conducted by Abed-al-Rahim and Johnston in the early 1990s to develop models that will produce a unit cost based on different project characteristics. Additional research by NCDOT focused on evaluation of PE costs and development of models (Hollar et al., 2013).

2.4.1.1 NCDOT, 1995

In 1995, researchers working on behalf of North Carolina State University (NCSU) developed a framework for the NCDOT to estimate unit costs for bridge replacement projects based on bridge-specific factors cataloged in a BMS database. Abed-al-Rahim and Johnston (1995) also developed models that would predict new bridge characteristics. The North Carolina Bridge Index (NCBI) contained the total bridge project cost for each bridge record, as well as the costs for preliminary engineering, construction, and roadway improvement. Miscellaneous items, such as right-of-way purchases, field operations, and legal fees were estimated by subtracting the three cost categories from the total project cost, as seen in Equation 2.8.

$$TOTCOST = MISCCOST + STRCOST + ROADCOST + ENGCOST \tag{2.8}$$

Where:
TOTCOST = Total project cost
MISCCOST = Miscellaneous costs
STRCOST = Bridge structure cost
ROADCOST = Roadway improvement cost
ENGCOST = Engineering cost

Abed-al-Rahim and Johnston used the FHWA Structures Index for North Carolina to convert costs to a present value. Equation 2.1 was used to convert dollar values from the year of construction (YC) to the latest available year (YL), using 1987 as a base year (YB). It was also possible to extrapolate data from the FHWA Index for future years, as shown in Equation 2.2, which was developed based upon linear regression conducted by

Abed-al-Rahim and Johnston (1995). The linear model yielded an R^2 value of 0.84. After using Equation 2.2 to determine the future year (FY) cost index, Equation 2.1 was used to calculate the future year cost.

Before a detailed bridge design is created, specific bridge characteristics such as structure length, deck width, and maximum span length are typically not known with certainty. These variables will have an impact in the overall replacement cost of a bridge, especially in cases where there is a large change in one of these characteristics for the new bridge relative to the old bridge (Abed-al-Rahim and Johnston, 1995). A set of models that can predict these new bridge characteristics based on those of the existing bridge can help estimators identify structures that would undergo a relatively large increase in size and therefore have a potentially higher cost to replace. Using the Generalized Linear Method (GLM), Abed-al-Rahim and Johnston (1995) performed regression analysis to develop an equation that could be used to predict new bridge length based on several existing bridge parameters. With new bridge length as the sole dependent variable, Abed-al-Rahim and Johnston considered several independent variables, such as existing bridge length, waterway adequacy, and under-clearance ratings. Ultimately, old bridge length was the independent variable that provided the best fit (R^2=0.9854), so the following regression equation (2.9) was developed:

$$NBLEN_{NC} = 8.45 + (1.013 \times L1) \tag{2.9}$$

Where:

$NBLEN_{NC}$: New bridge length based on NC data (in meters)
L1: Old bridge length (in meters)

Abed-al-Rahim and Johnston also utilized the FHWA Expansion Factor to predict new bridge length. These factors are based on nationwide averages of new bridge length as a function of existing bridge length. To use the expansion factor, Abed-al-Rahim and

Johnston (1995) took various original lengths from the curve (Fig. 2.5) and found their corresponding expansion factors. Multiplying the original lengths by their respective expansion factors provided the writers with a list of new lengths. A linear regression was performed with the original lengths (independent variable) and new lengths (dependent variable) to generate a regression equation (Eq. 2.10).

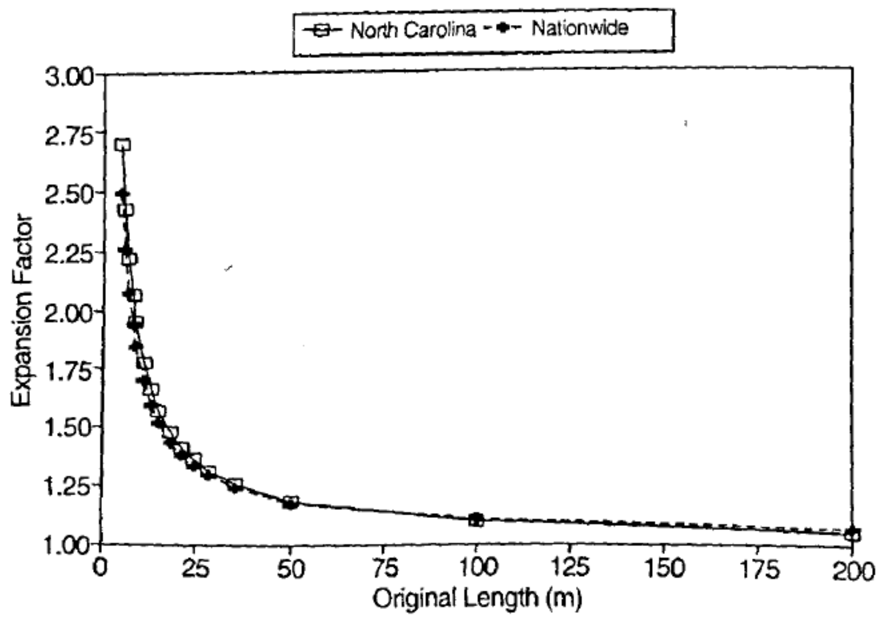

Figure 2.5: FHWA length expansion factor graph (from Abed-Al-Rahim and Johnston 1995)

$$NBLEN_{US} = 7.32 + (1.032 \times L1) \quad (2.10)$$

Where: $NBLEN_{US}$: New bridge length based on US data (in meters)
$L1$: Old bridge length (in meters)

Abed-al-Rahim and Johnston (1995) used Equation 2.11 to estimate the new bridge out-to-out deck width. Abed-al-Rahim and Johnston stated that the predicted clear deck width for the new bridge ($NBCDW_i$) was determined in OPBRIDGE by considering future level-of-service and ADT needs. OPBRIDGE was a computer program developed by Al-Subhi et al. (1989) to forecast and prioritize future bridge replacement projects. The second part of the equation exists to add the difference between current out-to-out width and current deck width to the predicted clear deck width. This assumes that the difference in width between out-to-out and clear deck widths will remain the same for the new bridge.

$$NBWID_i = NBCDW_i + (WIDTH_i - CDW_i) \tag{2.11}$$

Where: $NBWID_i$ = Predicted out-to-out width for new bridge *i*
$NBCDW_i$ = Predicted clear deck width for new bridge *i*
$WIDTH_i$ = Out-to-out width for bridge *i* that is to be replaced
CDW_i = Clear deck width of bridge *i* that is to be replaced

As one of the significant factors in predicting replacement cost, a bridge's maximum span length can also be predicted by its original maximum span length, waterway adequacy rating, structure length, and number of spans (Abed-al-Rahim and Johnston, 1995). The research team found that it was best to create two separate models for bridges over waterways and bridges over grade separations. Both models used old total length and maximum span of the bridge being replaced as independent variables and applied a logarithmic transformation to allow the models to meet the two assumptions for regression analysis:

1. Normal distribution of residuals
2. Variance is consistent along the regression line

After developing the two models, the research team was unable to prove that the coefficients in both equations were statistically different. The dataset of bridges used by Abed-al-Rahim and Johnston included 442 waterway crossings but only 39 grade separation crossings. Using a single, logarithmic-transformed model instead of two separate models yielded an R^2 value of 0.53, resulting in Equation 2.12. Both Equation 2.12 and Figure 2.6 show that new maximum span length are predicted to be shorter if the original span length is greater than 75 feet. Conversely, bridges with an original maximum span length less than 75 feet are predicted to have an increase in maximum span length for the new bridge (Abed-Al-Rahim and Johnston 1995).

$$MAXSPAN2 = 4.31 \times MAXSPAN1^{0.196} \times L1^{0.216} \tag{2.12}$$

Where: *MAXSPAN2* = Predicted maximum span length
MAXSPAN1 = Original maximum span length
L1 = Original bridge length

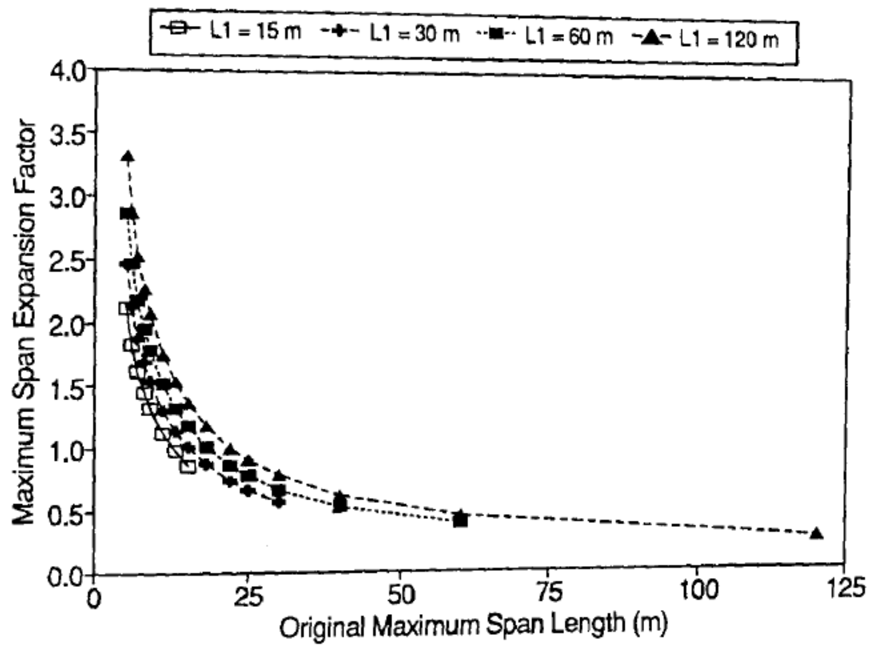

Figure 2.6: FHWA maximum span length expansion factor graph (from Abed-Al-Rahim and Johnston, 1995)

Abed-al-Rahim and Johnston (1995) developed a total cost prediction model under the assumption that the total cost would be a function of bridge length and width (unit cost) while other additional costs (engineering, roadway construction, etc.) could be added as a fixed percentage of that total cost. The resulting equation is shown in Equation 2.13.

$$TCOST_i = UREPB(NBLEN_i \times NBWID_i) \times (1 + EPC) + FIXEDC \tag{2.13}$$

Where: *TCOST*$_i$: Total cost for replacing bridge *i* in present year value
UREPB: Unit cost for bridge construction per square meter of deck area
NBLEN$_i$: Predicted length of bridge *i* in meters
EPC: Engineering cost as a ratio of structural costs
FIXEDC: Fixed cost for roadway and other incidental costs

The NCSU research team took historical cost data from 32 NCDOT bridges to find a unit structure cost based on new bridge deck area (Equation 2.14). All costs were converted to 1990 dollar-values to adjust for inflation and productivity changes. After determining which independent variables were significant (Table 2.1), the research team developed Equation 2.15 to estimate the unit structure cost for future bridges. Equation 2.14 can be rewritten in the form of Equation 2.16 to find structure cost using a predicted unit structure cost.

Table 2.1: Significant variable parameters in bridge structure cost (from Abed-Al-Rahim and Johnston, 1995)

	Parameter	Level of Significance
Grouping Parameters	Highway Functional Classification	P > 5.0%
	Rural vs. Urban	P > 5.0%
	Water vs. Grade Separation	P > 5.0%
Independent Variable Parameters	Width	P > 5.0%
	Length	P > 5.0%
	ADT	P > 5.0%
	Maximum Span Length	P = 2.6%
	Number of Spans	P > 5.0%

$$UCONST_{(YP,i)} = \frac{STRCOST_{(YP,i)}}{NBLEN_i \times NBWID_i} \tag{2.14}$$

Where: *UCONST*$_{(YP,i)}$ = Unit cost of structure construction for bridge *i* in present year dollar value
STRCOST$_{(YP,i)}$ = Structure construction cost for bridge *i* in present year dollar value
NBLEN$_i$ = Predicted length of bridge *i* in meters
NBWID$_i$ = Predicted width of bridge *i* in meters

$$UNITSTR = 919 - 40.6(MAXSPAN) + 0.927(MAXSPAN)^2 \tag{2.15}$$

Where: *UNITSTR* = Total cost for replacing bridge *i* in present year value
MAXSPAN = Unit cost for bridge construction per square meter of deck

area

$$STRCOST_i = UNITSTR \times NBLEN_i \times NBWID_i \tag{2.16}$$

Roadway improvement costs and miscellaneous costs are more difficult to predict due to the number of influencing factors. The amount of roadwork is not always necessarily linked to bridge deck area. Changing the elevation of a bridge can result in significant amounts of roadwork on one or both sides of the structure. On the other hand, miscellaneous costs (pavement markers, field office, etc.) can be calculated as the difference between the total project cost and the sum of the structure, roadway, and engineering costs (Equation 2.8).

Abed-al-Rahim and Johnston (1995) developed the following regression equations to estimate roadway improvement cost, miscellaneous cost, and engineering costs. The research team found that Equation 2.18 and Equation 2.19 tended to underestimate costs for smaller bridges and overestimate costs for larger bridges. The equation for engineering costs (Equation 2.20) had a relatively low R^2 value (0.60) but was judged by the NCSU researchers to perform rather well considering all the factors that usually affect engineering cost. The R^2 values for Equation 2.18 and 2.19 were not reported. For the regression analysis, structure cost was the only significant parameter identified for prediction of engineering cost (Equation 2.20).

$$ROADCOST = (177{,}900 \times NBWID) - 1{,}198{,}500 \tag{2.18}$$

Where: *ROADCOST* = Roadway improvement cost
NBWID = Predicted bridge width in meters

$$MISCCOST = 0.56(STRCOST) + 42{,}500(NBWID) - 364{,}000 \tag{2.19}$$

Where: *MISCCOST* = Miscellaneous costs
STRCOST = Bridge structure cost
NBWID = Predicted bridge width in meters

$$ENGCOST = 65{,}384 + 0.136(STRCOST) \quad (2.20)$$

Where: $STRCOST$ = Bridge structure cost
$ENGCOST$ = Engineering cost

2.4.1.2 NCDOT, 2013

Preliminary engineering costs for a bridge replacement project are typically estimated as being a fixed percentage of the total project cost. This technique does not address project-specific parameters that would cause PE costs to increase. According to the 2008 auditor's report for schedule and budget performance of NCDOT highway projects, PE costs for a set of 292 highway projects completed between April 1, 2004 and March 31, 2007 typically increased by 59% over the original estimated amount. This specific area had not received much attention from researchers due to the lack of reliable information available for PE costs (Hollar et al., 2013).

Hollar et al. (2013) compiled a database of 461 NCDOT bridge projects from several sources, such as online bid tabulations and construction plans, National Bridge Inventory System (NBIS) data, 12-month letting lists, meeting minutes, and funding authorizations. The bridges in the compiled database were usually three-span, two-lane concrete structures that crossed water features in rural areas. The dependent variable for this analysis was the ratio of actual PE cost to the estimated Statewide Transportation Improvement Program (STIP) construction cost. The research team used estimated costs instead of actual costs because estimators would not know the actual cost of a project during the conceptual planning stage. Using the correct PE cost ratio for a project would reduce the likelihood of cost escalation. The distribution of the PE cost ratio for the 461 NCDOT bridge projects ranged from 0.8% to 152% of estimated construction cost. The shape of the distribution was skewed to the left and needed to be transformed to improve normality to satisfy linear regression assumptions (Hollar et al., 2013).

The 461 database projects were divided into a validation set of 70 projects and a modeling set of 391 projects. The validation projects were used to test and quantify the model's performance in predicting the ratio of PE to STIP. Each candidate model was tested over the validation set by comparing the predicted PE cost values to the actual historical values for those projects. The models with lowest Mean Absolute Percentage Error (MAPE) and Average Absolute Error (AAE) were preferred over models with higher error values (Hollar et al. 2013).

The response variable (PE cost ratio) was transformed by applying an exponential power and using the Box-Cox procedure to identify the optimal transformation to get normality. In this case, the cubed root of the response variable was used to attain normality, which was then verified using a goodness-of-fit test. Since the dependent variables were normalized using the power transformation, results had to be transformed back using Equations 2.21, 2.22, and 2.23. The equations were solved using a variance value of 0.0229 for the data set (Hollar et al. 2013).

$$E.M.R. = (predicted\ cubed\ root\ of\ response)^3 \quad (2.21)$$

Where: *E.M.R.* = Estimated Median Response

$$T.C.F = 1 + \{[(var) \times (1 - {}^{1}/_{3})]/[2(predicted\ cubed\ root\ of\ response)^2]\} \quad (2.22)$$

Where: *T.C.F.* = Transformation Correction Factor
var = Variance

$$Estimated\ mean\ response = E.M.R \times T.C.F \quad (2.23)$$

Where: *E.M.R.* = Estimated Median Response
T.C.F. = Transformation Correction Factor

The one-way ANOVA technique was applied to the 16 categorical variables in the compiled database to identify those which were statistically significant. The seven significant categorical variables are listed in Table 2.2. The two categorical variables with the highest level of influence on the cubed root of the PE cost ratio were year-related. The researchers assumed that any fluctuations in STIP estimated costs over time would be mirrored by the actual PE costs, so these two variables were not used as predictor variables in the analysis (Hollar et al., 2013).

Table 2.2: Statistically significant categorical variables (Hollar et al., 2013)

Categorical Variable	R^2	F-value	p-value
Year of letting	0.3037	20.83	<0.0001
Year of environmental doc. approval	0.1220	3.47	<0.0001
Road system	0.0443	8.80	0.0002
Project construction scope	0.0322	6.45	0.0017
Geographical area of state	0.0361	4.84	0.0026
Division	0.0728	2.28	0.0068
Design live load	0.0302	3.00	0.0185

To determine which of the numerical variables should be used in the regression model, Hollar et al. (2013) used the Pearson correlation coefficients and p-values to identify which variables were statistically significant. The correlation coefficient, which ranges between -1 to +1, indicates the strength of the correlation with the cubed root of PE cost ratio. The sign of the coefficient reflects whether the independent and response variables move together (positive slope) or move apart from each other (negative slope). Table 2.3 contains the eight numerical independent variables that were determined to be statistically significant.

Table 2.3: Pearson correlation coefficients for numerical variables (Hollar et al. 2013)

Numerical Independent Variable	Pearson Coefficient	p-value
Project length	-0.3263	<0.0001

STIP-estimated construction cost	-0.3130	<0.0001
ROW cost to STIP-estimated Construction cost	+0.3089	<0.0001
Structure length	-0.1944	<0.0001
Roadway percentage of construction cost	-0.1849	<0.0001
Spans in primary unit	-0.1766	<0.0001
Horizontal clearance for loads	-0.1592	0.0006
PE duration after environmental document approval	-0.1053	0.0237

After selecting the statistically-significant categorical and numerical variables for the linear regression, the research team used the GLMSELECT procedure within SAS to create a multiple linear regression (MLR) model. Excluding all date-related variables, the completed MLR model achieved an adjusted R^2 value of 0.2745 using the following variables:

1. ROW cost to STIP-estimated construction cost (Numerical)
2. Roadway percentage of construction cost (Numerical)
3. STIP-estimated construction cost (Numerical)
4. Bypass detour length (Numerical)
5. Project construction scope (Categorical)
6. NCDOT division (Categorical)
7. Geographical area of state (Categorical)
8. Responsible party for the planning document (Categorical)

When applied to the data set of 70 projects, the MLR provided a MAPE of 0.1889. This was compared to the MAPE of 0.9137 that was achieved by a single-point estimate using the mean PE cost ratio of the remaining 391 projects. This single-parameter estimating method is commonly used by the NCDOT to estimate PE costs and also served as a baseline target to measure the MLR model's prediction capability. After obtaining the regression coefficients (Table 2.4), Equation 2.24 can be used to find the predicted cube root of the PE cost ratio to STIP construction cost (Hollar et al. 2013).

Table 2.4: Regression coefficients for MLR model (Hollar et al. 2013)

	Parameter		Coefficient
	Intercept	β_0	0.6471
x_1	NCDOT division = D12 and project construction scope = new location; 1 if true, 0 if false	β_1	-0.1657
x_2	NCDOT division = D06 and responsible party for the planning document = DOT; 1 if true, 0 if false	β_2	-0.1087
x_3	Geographical area of state = very mountainous and responsible party for the planning document = DOT; 1 if true, 0 if false	β_3	0.0701
x_4	ROW cost to STIP-estimated construction cost	β_4	0.2909
x_5	STIP-estimated construction cost if NCDOT division = D12	β_5	4.45×10^{-8}
x_6	Roadway percentage of construction cost multiplied by STIP-estimated construction cost	β_6	-1.88×10^{-7}
x_7	Bypass detour length if NCDOT division = D07	β_7	-0.0159

$$\begin{aligned} Predicted\ cubed\ root = \ & \beta_0 + \beta_1(x_1) + \beta_2(x_2) + \beta_3(x_3) + \\ & \beta_4(x_4) + \beta_5(x_5) + \beta_6(x_6) + \beta_7(x_7) \end{aligned} \tag{2.24}$$

Ideally, the MLR model would follow a 45-degree positive slope, which would mean that the predicted values would be close to the actual values. The slope of the MLR model is positive but smaller than the ideal slope. Compared to the mean-value of the PE cost ratio for the set of 391 projects, the MLR model overestimated PE cost ratios at the lesser percentages (<20%) and underestimated ratios at higher percentages (>35%). The MLR model had a MAPE value of 42.7%. Compared to the mean value's MAPE of 48.7%, the MLR had slightly better performance over the single-point estimator (Hollar et al. 2013).

Despite the relatively high prediction error percentage for the MLR model (42.7%), the results of the modeling confirmed the research team's assertions that PE costs for bridge projects were often underestimated. The historical mean reported by Hollar et al. (2013) for NCDOT bridge projects was 27.8%, which was greater than the WSDOT estimate of 10.3%, VDOT estimate range of 8-20%, and Georgia DOT estimate range of 6-12%. The data used to create the model should be readily available for most state agencies. Hollar et al. found that state agency procedures and processes could compromise the quality of PE cost data. In the case of this study, the research team found that PE costs were often charged as an overhead burden and not accurately assigned to the individual bridge projects. For this reason, it is important that databases should be expanded and updated often to create a solid data set for creating regression models (Hollar et al. 2013).

In addition to suggesting means to improve the quality of PE cost data recording procedures, Hollar et al. (2013) also recommended that future researchers analyze PE costs in terms of monetary units instead of ratios. To do this, it is necessary to convert all costs to a common year. The research team expressed a need for future research into reasons why PE costs were driven up for projects, such as the instances where the project PE cost ratio was 152% of the construction cost. An analysis of case studies may provide qualitative data on how certain factors drive up PE costs (Hollar et al. 2013).

2.4.2 Usage of Cost Prediction Models in Other State Agencies

Since bridge maintenance and replacement programs are managed by state departments of transportation, many of these agencies have funded research projects that determine the most effective way to forecast bridge replacement costs that work best for the state bridge inventories. Publications exist on the many different approaches researchers have employed to create state-specific prediction models. The techniques used by researchers to develop cost prediction models for Indiana DOT and Texas DOT are covered in more detail in the following sections.

2.4.2.1 Indiana Department of Transportation

Saito et al. (1991) developed a series of regression models for predicting costs for bridge replacement projects in Indiana. A dataset of 279 Indiana Department of Transportation (INDOT) bridges replaced between 1980 and 1985 was compiled by the researchers. Bridge attributes used for the model, such as structure length, deck width, vertical clearance, approach length, and earthwork needed, could be easily identified by inspectors and included in the database. Cost data were the dependent variables for the study, and all prices were converted to 1985 values using the FHWA construction price index. Cases where multiple bridges were included on one contract or where replacement costs were extremely high or low were removed from the data set to avoid influence from outliers.

To develop the replacement cost model, the ANOVA (analysis of variance) technique was used to determine the effect that the independent variables, such as structure length, deck width, and number of spans, had on the actual contract costs. SPSS and SAS statistical software packages were then used to take the results of the ANOVA and develop a regression model. The ANOVA was done based on three primary classification factors that were currently being used by INDOT to estimate bridge replacement cost: superstructure type, substructure type, and highway type. At the time of the study, the FHWA required state agencies to provide separate unit costs for each of the different highway types and superstructure types (Saito et al. 1991).

The ANOVA test confirmed that both superstructure and substructure types were statistically significant (5% level) in predicting unit substructure cost and that both factors should be used to generate estimates. A separate ANOVA test was performed for approach construction costs, but with total contract costs instead of unit costs. This test was based on two factors: amount of earthwork (small, medium, or large) and approach length (short, medium, or long). Results from this ANOVA test showed that amount of earthwork and approach length can and should be used as factors in predicting approach construction costs (Saito et al. 1991).

The results from both ANOVA tests were then used by the research team to develop bridge replacement cost models that required as few independent variables as possible. The models developed by Saito et al. (1991) were nonlinear and log-linear in nature, and used predictor variables that could be easily determined by engineers on the site, such as designed structure length, width, and vertical clearance. Results from ANOVA and the scatter plots showed that regressions could be done for the four cost categories (superstructure, substructure, approach, and other costs) using a multiplicative regression model, shown in Equation 2.25. Usage of the multiplicative model works under the same logic as unit costs, where structure length and width are multiplied by that unit price to determine the total cost. Since the regression coefficients are fixed values, Saito et al. (1991) cautioned users of this model (INDOT) against using it for bridges that were outside the data range used in the creation of the prediction models.

$$Y = \left(X_1^{\beta_1} X_2^{\beta_2} \ldots X_n^{\beta_n}\right) \epsilon \qquad (2.25)$$

Where: Y = Dependent Variable (Replacement Cost)
X_n = Independent Variable
B_n = Regression coefficient
ϵ = Error coefficient

With a non-linear regression, it is possible to transform the raw data to see if it is possible to perform a linear regression analysis. Equation 2.26 was transformed into Equation 2.27 using $\log_{10}$ transformation (Saito et al. 1991). The new equation could be used provided that it met the key assumptions of linear regressions (Nau 2009):

1. Linearity and additivity of relationship between independent and dependent variables
2. Statistical independence of errors
3. Heteroscedasticity (constant variance)
4. Normality of error distribution

If the transformed model met all of the four key assumptions, it was returned to the non-linear form shown in Equation 2.27. Once this was performed, the non-linear cost model and transformed log-linear model were compared to see if one model is preferable for use in estimating replacement costs. In making the comparison, the research team assumed that error terms were independent, variance was constant along the regression line, linearity of the model, and the residuals were distributed normally. When comparing the two models, residual plots were used to test the constancy of variance. Normal probability plots of the residuals were used to test normality of the error term distribution (Saito et al. 1991).

$$Y' = \beta_0' + \beta_1 X_1' + \beta_2 X_2' + \cdots + \beta_n X_n' + \epsilon' \tag{2.26}$$

Where:
$Y' = \log_{10}(Y)$
$\beta_0' = \log_{10}(\beta_0)$
$X_i' = \log_{10}(X_i)$
$\epsilon' = \log_{10}(\epsilon)$

$$Y = 10^{\beta_0'} X_1^{\beta_1'} X_2^{\beta_2'} \ldots X_n^{\beta_n'} \tag{2.27}$$

Log-linear models were developed to predict bridge replacement, superstructure, substructure, approach, and “other” costs. Separate equations were developed for significant categorical variables alongside an overall equation for all bridge types. The log-linear equations for bridge replacement total cost (BRTC) (Table 2.5), superstructure cost (Table 2.6), substructure cost (Table 2.7), and approach cost (Table 2.8) were validated with a set of bridge data for projects between January and June 1986. Out of the

37 bridges in the validation set, only 26 of the bridges had complete cost data for the other cost components while the remaining 11 bridges only had information on total project cost. After adjusting the predicted values to 1986-dollar equivalents, Saito et al. (1991) found that the models were reasonably precise. Cost values for these equations were rounded to the nearest $1,000 while bridge length (*BL*), deck width (*DW*), and vertical clearance (VC) were reported in feet.

Table 2.5: BRTC regression equations (in 1985 dollars)

Component	Type	Model	R^2	*F* Value	*n*
Other	All types	$OTHC = 0.0721(BL)^{0.696}(DW)^{0.932}$	0.524	100.60	186
Bridge Total	All types	$BRTC = 0.155(BL)^{0.903}(DW)^{0.964}$	0.951	1861.28	196
	RC Slab & Box Beam	$BRTC = 0.0781(BL)^{0.748}(DW)^{1.319}$	0.874	380.74	113
	Concrete I-Beam	$BRTC = 1.255(BL)^{0.809}(DW)^{0.534}$	0.913	205.34	42
	Steel Beam	$BRTC = 0.128(BL)^{0.785}(DW)^{1.210}$	0.971	317.50	22
	Steel Girder	$BRTC = 0.353(BL)^{1.015}(DW)^{0.603}$	0.950	150.91	19

Table 2.6: Superstructure cost regression equations (in 1985 dollars)

Type	Model	R^2	*F* Value	*n*
All types	$SUPC = 0.0107(BL)^{1.122}(DW)^{1.084}$	0.524	1861.28	196
RC Slab & Box Beam	$SUPC = 0.0137(BL)^{1.001}(DW)^{1.161}$	0.874	380.74	113
Concrete I-Beam	$SUPC = 0.0330(BL)^{0.907}(DW)^{1.043}$	0.913	205.34	42
Steel Beam	$SUPC = 0.0102(BL)^{1.120}(DW)^{1.117}$	0.971	317.50	22
Steel Girder	$SUPC = 0.8550(BL)^{0.906}(DW)^{0.747}$	0.950	150.91	19

Table 2.7: Substructure cost regression equations (in 1985 dollars)

Type	Model	R^2	*F* Value	*n*
All types	$SUBC = 0.00168(BL)^{0.906}(DW)^{1.255}(VC)^{0.487}$	0.725	168.35	196
Steel Girder	$SUBC = 0.00354(BL)^{0.744}(DW)^{1.205}(VC)^{0.515}(T)^{0.156}$	0.751	143.62	196

Note: T = 1 for solid stem piers, T = 0 for pile-type piers

Table 2.8: Approach cost regression equations (in 1985 dollars)

Models	R^2	*F* Value	*n*
$APC = 0.769(APL)^{0.823}$	0.566	248.08	192
$APC = 39.876(EW)^{0.378}$	0.633	328.20	192
$APC = 4.715(APL)^{0.403}(EW)^{0.250}$	0.696	215.93	192

Note: APL = Approach cost and EW = Earthwork (in 100CY)

2.4.2.2 Texas Department of Transportation, 2005

Chou et al. (2005) developed a probabilistic cost estimation tool for the Texas Department of Transportation (TxDOT). An analysis of TxDOT bridge data from 2001 to 2003 showed that there were 22 major work items in a bridge project that accounted for roughly 80.2% of the total cost (Table 2.9). The estimation tool was created under the assumption that estimators would be able to control at least 80.2% of the total project cost.

Table 2.9: High Cost Major Work Items for TxDOT Bridge Projects (FY 2001-FY 2003)

WORK ITEM	**COST %**	**ITEM DESCRIPTION**
100 ITEMS: EARTHWORK AND LANDSCAPE		
100	1.51%	PREPARING RIGHT-OF-WAY
110	1.67%	EXCAVATION
132	3.09%	EMBANKMENT
200 ITEMS: SUBGRADE TREATMENTS AND BASE		
247	2.62%	FLEXIBLE BASE
300 ITEMS: SURFACE COURSES AND PAVEMENT		
340	0.76%	HOT MIX ASPHALTIC CONCRETE PAVEMENT
360	1.55%	CONCRETE PAVEMENT
400 ITEMS: STRUCTURES		
409	1.21%	PRESTRESSSED CONCRETE PILING
416	11.67%	DRILLED SHAFT FOUNDATIONS
420	12.69%	CONCRETE STRUCTURES
422	7.13%	REINFORCED CONCRETE SLAB
432	0.86%	RETAINING WALL
435	9.28%	PRESTRESSED CONCRETE STRUCTURAL MEMBERS
430	2.79%	EXTENDING CONCRETE STRUCTURES
432	1.29%	RIPRAP
442	2.55%	METAL FOR STRUCTURES
450	1.65%	RAILING
462	2.65%	CONCRETE BOX CULVERTS AND SEWERS

500 ITEMS: MISCELLANEOUS CONSTRUCTION		
500	8.28%	MOBILIZATION
502	1.79%	BARRICADES, SIGNS, AND TRAFFIC HANDLING
508	1.54%	CONSTRUCTING DETOURS
534	0.73%	STRUCTURE APPROACH SLABS
SPECIAL SPECIFICATION WORK ITEM		
3146	2.91%	QA/QC OF HOT MIX ASPHALT
Total = 80.22%		

Source: Chou et al., 2005

The unit cost for each work item was expressed as a cost per lane-kilometer. Equation 2.28 was used to calculate the total project cost by adding up the unit costs for all 22 major work items. The sum of the major work item unit costs is divided by 80.2% to account for the 19.8% of the project cost covered by the minor work items. A contingency amount is added to the quotient to account for engineering costs (Chou et al., 2005).

$$Total\ Project\ Cost = \frac{\sum_{i=1}^{22} ItemCostPerLaneKm_i}{80.2\%}(1 + EngCont\%) \tag{2.28}$$

Where: $ItemCostPerLaneKm_i$ = Cost per lane-km for each of the 22 major work items

EngCont% = Engineering contingency expressed as a percentage

Chou et al. (2005) performed Monte Carlo simulations for five scenarios to create charts that can be used by estimators to determine the unit cost for a bridge project with knowledge of market conditions, need for work, location, scope changes, geological conditions, and constructability challenges. Figure 2.7 is a graph of the probability density functions (PDFs) for all five scenarios tested in the Monte Carlo simulation. Since the variables used in the Monte Carlo simulation were random and continuous, the area under each PDF curve from 0 to *x* is equal to the probability of getting a value that is less than or equal to *x*. The total area under each PDF curve is equal to one (Andrews and Phillips, 2003).

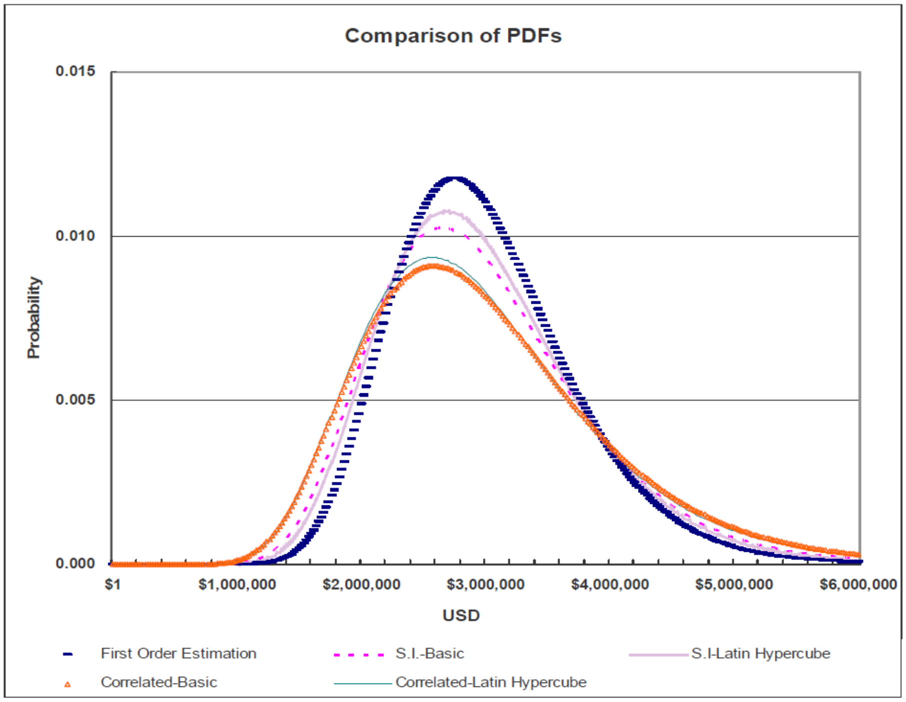

Figure 2.7: Comparison of PDF's (Chou et al. 2005)

The cumulative distribution functions (CDFs) shown in Figure 2.8 can also be used to calculate the probability of the random variable being less than or equal to *x* in real-world conditions (Chou et al., 2005). This probability is found by selecting the *y*-axis value for the chosen CDF curve at *x* (Andrews and Phillips, 2003). The total project costs are calculated from the CDFs and PDFs by multiplying the *x*-axis value ($/lane-km) by the length of the bridge (Chou et al., 2005).

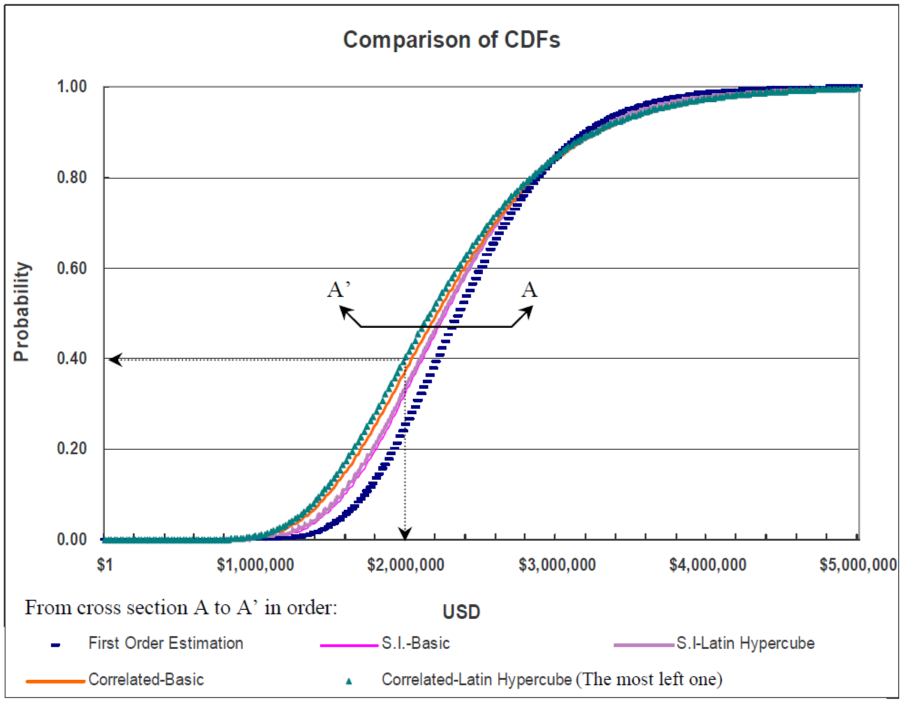

Figure 2.8: Comparison of CDFs (Chou et al., 2005)

Unlike other traditional models that are affected by untreated historical data, the probabilistic model developed by Chou et al. (2005) provides confidence bounds for an estimate, which helps control error, accounts for probability, and considers the independent or correlated relationships between the major work items. As with any other estimating method, the effectiveness of probabilistic models hinges on the quality of the data available to estimators.

2.5 Research Needs

Like most state agencies, NCDOT employs two methods of estimating project costs. The first method uses BMS-sourced data in an algorithm based solely on deck area and a unit cost allocated by type of roadway (interstate, primary, or secondary) to produce a conceptual-level estimate that can be used for prioritizing and allocating funding for future projects. This method, currently programmed into the NCDOT BMS using unit costs identified in the early 2010s, is not providing accurate estimates per NCDOT personnel.

The second method, cost-based estimation, considers labor and material costs as well as prevailing market rates to create a more detailed project-specific cost estimate. The estimated construction costs are generally close to the actual construction costs (within 2%), while the BMS estimates are less reliable, especially for small and large projects. Since the project-specific factors involved in a cost-based estimate cannot be easily integrated into the BMS, there is currently a need to identify strategies and databases used in cost-based estimates that could be used to improve the accuracy of BMS-based estimates. This can be done by analyzing the discrepancies that exist between the BMS estimates, cost-based estimates, and actual costs. This analysis should show where the BMS estimates fall short of the cost-based estimates in terms of accuracy.

Once the inaccuracies in the BMS cost estimation algorithms are found, the associated algorithms will need to be adjusted and updated. Some of refined strategies used for cost-based estimates cannot be easily integrated into an automated BMS estimating process. However, information on the appropriate contract costs can potentially be used alongside BMS data. Recent research by Hollar et al. (2013) provided insight into PE costs that can be used to support development of more comprehensive bridge replacement costs. However, more recent data on preconstruction costs (preliminary engineering, right-of-way, and utility costs) could be used to further improve current cost estimating models and improve the accuracy of their results.

Chapter 3: DATA SOURCING AND PRECONDITIONING

3.1 Data Sources

As discussed in the previous chapter, aggregated statistical models use a dataset of relevant historical project data to create regression equations (models) that predict a subset of known bridge parameters (Behmardi et al., 2015). The quality of the historical project dataset used by estimators can also influence the quality of their predictions (Gransberg et al., 2013). The dataset serves as a foundation for development of regression equations that can be used to predict costs. If data is missing or improperly recorded, the assumptions made by the regression equations are not as sound, since they do not accurately reflect the true conditions. Likewise, atypical projects included in a regression database may improperly bias prediction models design to project costs for typical projects.

The two modeling approaches utilized in this study are illustrated in Figure 3.1. One approach is to develop models that will predict the geometry of the replacement bridge based on the old bridge characteristics. The estimated dimensions for the new structure are then used to predict the costs associated with the bridge replacement. An alternative to this approach is to find a way to model the cost for the new bridge solely from information on the old bridge without needing to predict increases in bridge size.

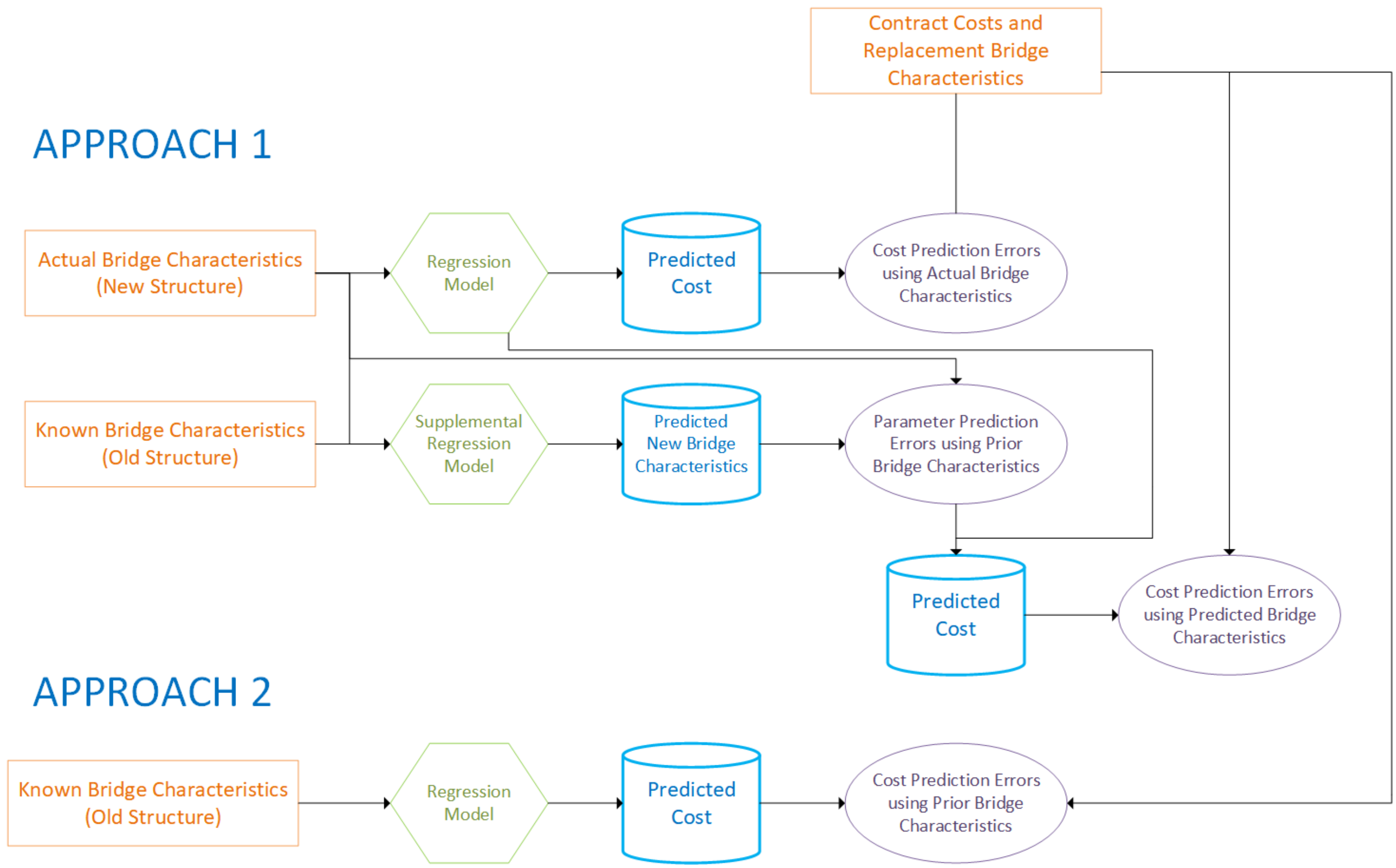

Figure 3.1: Modeling approaches estimated in study

The models for predicting cost can be developed by following one of the approaches in Figure 3.1. With the first approach, two types of prediction models are developed: models that predict new bridge characteristics from old bridge characteristics and models that predict bridge replacement contract cost from new bridge characteristics. Both models would be validated with actual data on old and new bridge characteristics and contract cost data.

The second approach omits the prediction of new bridge characteristics by modeling a relationship between bridge characteristics of the structure being replaced and replacement contract cost. This simplified approach provides the advantage of relying upon both existing data (for the original bridge) as well as the advantage of involving fewer models so, if found to be reasonably accurate, this approach would be more user-friendly and easier to incorporate into the BMS. However, prediction errors between the two approaches need to be analyzed to prevent selection of lower-quality models.

In order to support both regression modeling approaches shown in Figure 3.1, datasets needed to be assembled from various sources. The following sections provide information on these sources, the types of data imported from these sources to support this work, and preconditioning of the data prior to regression modeling.

3.1.1 NCDOT BMS Network Master

The Network Master is sourced from the NCDOT Bridge Management System (BMS). It contains information on location, structural characteristics, and performance ratings for every NCDOT-maintained bridge, culvert, and overhead sign in the state. The Performance Master is updated annually. The dataset used for this project used information sourced from the Network Master, exported in May of 2017 to an Excel spreadsheet. At the time of export for this work, the NCDOT Network Master contained 21,698 records.

3.1.2 NCDOT BMS Performance Master

The Performance Master catalogs detailed inspection data for all NCDOT-maintained structures. Condition ratings and appraisals for superstructure, substructure, deck, guardrails, and expansion joints are included for each structure. A new Performance Master annual record is updated in the BMS every year to serve as a historical record of bridge condition and status over time. Older versions of the Performance Master can be used to search for general bridge data prior to bridge replacements. The earliest version of the Performance Master used was from 2006, so old and new bridge data was available for bridge projects completed between 2007 and 2016. The 2013 Performance Master was also used to provide recent old bridge data for structures replaced between 2014 and 2016. The 2006 NCDOT Performance Master used for the central dataset contained 20,690 entries, while the 2013 Performance Master contained 22,226 entries.

The National Bridge Inventory (NBI) was used as a source for additional bridge data. The NBI includes bridge inspection information submitted to the FHWA by state and federal transportation agencies (FHWA, 1995). Maximum span length was not included for bridge entries in the 2006 and 2013 Performance Masters, so this information had to be obtained from the NBI.

3.1.3 Historical Bridge Replacement Cost Data

NCDOT provided HiCAMS-sourced data for 1,165 bridge replacement projects let between 2008 and 2016. The entries were categorized by funding source and the letting approach for the project. The 17BP projects were state-funded and were let from the NCDOT Central Office or by the NCDOT Division from where the project originated. TIP projects received federal funding and were typically centrally let. This dataset contained contract information on each replacement project, including contract cost and letting dates. A breakdown of the types of projects in this dataset is provided in Table 3.1.

Table 3.1: Breakdown of entries sourced from HiCAMS

By letting method	Count
17BP	86
TIP	235
Total	321

3.1.4 Supplementary Sources

In addition to the NCDOT BMS Performance Master, Network Master, and historical bridge replacement cost data, NCDOT also provided additional information from their Highway Construction and Materials System (HiCAMS) database. After the size of the central dataset was reduced to 334 entries, NCDOT provided cost data for each bridge entry. The cost data included contract amount, actual expenditure amount, preliminary engineering cost, right-of-way cost, and construction cost. All 17BP project entries did not have component costs for the preliminary engineering, right-of-way, and construction costs, so those entries were not used for development of regression models that used those variables.

NCDOT also provided contract information for 12 bridge replacement projects on structures with high traffic volumes. Each of these bridges had an ADT greater than 15,000 cars per day. Including these entries in the central dataset allowed for development of models that would be useful for predicting characteristics and costs of larger bridge replacement projects.

3.2 Data Conditioning

To facilitate the development of the regression models, data from five sources was merged into two central datasets. The sheer volume of bridge entries in each data source precluded the possibility of manually entering the data in the central datasets, so formulas were used to pull in specific values from the sources based on a matching bridge structure numbers. The contract cost data provided by NCDOT did not include structure numbers, so these values were determined by cross-referencing county, route name, feature crossed, and TIP number between the contract cost data and the BMS data to find a match. Typographical errors, misspellings, or empty cells would cause the formulas to return incorrect or empty values. Columns that contained descriptive data was often recorded in the source datasets in an inconsistent manner, which led to instances where a word was spelled, capitalized, or abbreviated in several different ways. In these cases, records were modified manually, changing the values to a consistent format. After all data was imported, an additional review of the dataset was performed to ensure that the data was a correct match and that the formulas were entered in properly. This issue was addressed by performing spot checks on random bridge entry rows in the central datasets and comparing values side-by-side with those in the source datasets. Entries that were missing data that could not be found in the source datasets were eliminated from the central dataset.

3.2.1 Creation of Central Datasets

Central datasets containing records useful to support this work were created by merging specific data fields from other NCDOT databases. While data on bridge structural characteristics was readily available from the BMS to support development of models predicting new bridge characteristics from old bridge characteristics, records providing data on replacement project costs was limited, providing far fewer records to support development of the cost prediction models. The new bridge characteristic prediction models did not require cost-related information, so those models could be developed using a larger dataset (the full combined dataset) than the cost prediction model dataset.

The two central datasets were created by the following process. An initial list of bridge records was created based on the dataset that would control the number of entries. For the historical cost-based central dataset, this was the historical contract cost dataset. The central dataset for the characteristic prediction models began with the 2006 Performance Master. Entries in this initial list were removed if their data was not representative of a typical bridge replacement project. Examples of non-typical bridge replacement project included bridges with more than nine spans, moveable bridges, and replaced bridges that experienced very small or large relative changes in length or width from the original bridge size. Once the entries on the initial list were conditioned, additional data was imported from supporting datasets. Table 3.2 illustrates the process that was followed to create both central datasets.

Table 3.2: Actions taken to create central datasets

Actions Taken:	**Characteristic Prediction Central Dataset:**	**Cost Prediction Central Dataset:**
Initial List:	• 2006 Performance Master	• 17BP/TIP Projects
Filter out:	• Entries that do not appear in the 2016 Network Master	• “Basket projects” (3.2.2) • Bridge rehabilitation projects • Bridge preservation projects
Import information from:	• 2013 Performance Master (ADT) • 2016 Network Master (new bridge characteristics)	• 2006 Performance Master (old bridge characteristics) • 2016 Network Master (new bridge characteristics) • Supplementary sources (project cost data)
Additional filtering:	• Entries where structure type is not a bridge • Bridges replaced prior to 2007 • Bridges with more than 7 spans	• Entries where structure type is not a bridge

3.2.1.1 Structure Numbers

Each NCDOT bridge is identified by a unique six-digit structure number. The first two numbers are for the county code (00 to 99, coding North Carolina’s 100 counties in alphabetical order) and the last four numbers of the structure number are the bridge’s number within that county (0000 to 9999). Information could be pulled from the Performance Master and Network Master by writing INDEX-MATCH formulas that would search for rows with a matching structure number. Unfortunately, the historical cost dataset did not include structure numbers for the entries, so structure numbers for each of the entries in the historical cost dataset had to be found using other methods.

The first method was to find the structure number based on the contract number, county, and route in the Network Master. These three fields were present in both the Network Master and historical cost spreadsheets. A column was created in both spreadsheets that had the concatenated county, route, and project number for each bridge. The chance of more than one bridge having the same county code, route, and contract number was perceived as being relatively low. A VLOOKUP function was then used to pull the matching structure number from the 2016 Network Master into the historical cost dataset.

This method of merging the data for each structure from the two databases was successful for 787 out of the initial 1,165 entries (67.6%) in the historical cost spreadsheet. A second method was applied to the remaining 378 bridge entries for which a matching structure number was not identified using the first method. Several entries in the historical cost spreadsheet had the four-digit bridge number listed in the comments field. Assuming that the number was entered correctly, the four-digit number could be appended to the two-digit county code to provide the bridge's structure number. This process had to be performed manually, but allowed for better quality control. This method also served as a check for the first method to verify that the VLOOKUP returned the correct structure number.

In cases where those two methods failed to identify a common field for merging of the datasets, entries with no known structure number were looked up in the Network Master by County and Route Carried. For most cases, searching with these two criteria usually returned only one entry in the Network Master, which was presumed to be the corresponding bridge from the historical cost dataset. In cases where there was more than one result, the Year Built for the Network Master entries were compared to the entry in question from the historical cost dataset. There were a few rare cases where the bridge information had to be found using sources on the internet. In total, structure numbers were found for 814 of the bridge entries in the historical cost dataset, which accounted for 69.9% of the total entries in the original dataset. These 814 entries could then be linked to physical bridge characteristics and location information from the Network Master and Performance Master. This connection allowed for further filtering based on known bridge parameters.

3.2.1.2 Linking Bridge Characteristic Information to Historical Cost Records

For both central datasets, data on bridge characteristics for the old and new structures were imported from the other datasets and paired to the corresponding entries in each central dataset. Data from the 2006 Performance Master, 2013 Performance Master, and the 2016 Network Master was imported as separate files. Once the historical cost data entries had structure numbers, the bridge characteristic data from the Network Masters and Performance Masters could be imported into a central dataset by matching the structure numbers.

The lookup formulas consisted of a combination of the INDEX and MATCH functions. The INDEX function defined a column to use as a lookup array and the MATCH function selected the row based on a structure number that matched the value on the central dataset (Table 3.3). This setup has an advantage over VLOOKUP since the reference column, which in this case is the list of structure numbers, does not have to be the leftmost column in the sheet.

Table 3.3: Typical format for data lookup formula

= INDEX('PM'!C:C,MATCH(A[x],'PM'!A:A,0))		
Where:		
'PM'!C:C	=	Column of variable to be imported from sheet 'PM'
A[x]	=	Structure number for row [x] entry receiving the data
'PM'!A:A	=	Range of structure numbers on sheet 'PM'
0	=	Return exact match

3.2.1.3 Variables

Both the Network and Performance Masters contained dozens of variables pertaining to bridge structural characteristics. Ultimately, some of the variables were judged to be more useful for prediction models than others. Decisions had to be made regarding which variables were important and should be imported into the central dataset, and which variables likely had no logical link to predicting bridge cost. Too many variables would clutter the central dataset with useless information while too few variables would defeat the purpose of introducing new variables into prediction models to create better predictions. The first step in making the decision was to compile a list of all available variables between the three datasets (Table 3.4)

Table 3.4: Fields available in NCDOT BMS databases

Field	Performance Master	Network Master	Historical Cost		Field	Performance Master	Network Master	Historical Cost
Substructure Rating		X			Structure Length	X	X	
Superstructure Rating		X			Deck Width	X	X	
Deck Condition		X			Roadway Width	X	X	
Division Number	X	X	X		Through Lanes On	X	X	
Tier ID	X	X			Min. Clearance Under		X	
County	X	X	X		Replacement Cost		X	
Structure Number	X	X	X		Maint. Resp.	X	X	
Structure Type	X	X			Owner	X	X	
Facility Carried	X	X	X		Service Type On		X	
Intersected Features	X	X			Service Type Under		X	
Maintenance History	X	X			Span Type		X	
TIP Bridge Number		X	X		Superstructure Type	X	X	
Bridge Replacement Status		X			Substructure Type	X	X	
Replacement Status (TIP)		X			Latitude		X	
Route Name		X			Longitude		X	
PRI		X			Structure Type Main	X	X	
Sufficiency Rating	X	X			Structure Type Approach	X	X	
Deficiency Points	X	X			Deck Structure Type	X	X	
Structurally Deficient?	X	X			Culvert Type		X	
Approach Roadway Width	X	X			Service Type	X	X	
Approach Trway Width	X				Brng. Assy (Girder) Grade	X		
Functionally Obsolete?	X	X			Sorting Code	X	X	
Posted SV	X	X			Scour Critical Bridge	X	X	
Posted TTST	X	X			Last Routine Insp. Date		X	
Bearing Grade	X	X			Structure Appraisal	X	X	
Posting Score (#)	X	X			Deck Geometry Appraisal	X	X	
Bridge Health Index (BHI)	X	X			Approach Alignment App.	X	X	
BHI Score (#)	X	X			Underclearance Appraisal	X	X	
Average Index (BMS)		X			Waterway Adequacy App.	X	X	
Temp. Struct. Designation		X			Culvert Condition		X	
Detour Length	X	X			Paint Condition	X	X	
Bridge Age	X	X			Min. Vert. Clearance Over		X	
Year Built	X	X			Max. Clearance Under	X	X	
Est. Remaining Life	X	X			Water Depth	X	X	
Bridge Length (NBIS)	X	X			Height Crown to Bed	X	X	
National Highway System		X			City	X	X	
Strahnet Designation		X			Bridge Name	X	X	
Deficient?	X	X			Road System	X	X	
Green Line Route	X	X			Traffic Direction		X	
Posted?	X	X			Bridge System	X	X	
No. Thru Lanes (Under)	X				Left Guardrail Mat'l Type	X		
Federal Skew	X				Rght Guardrail Mat'l Type	X		

Table 3.4: Fields available in NCDOT BMS databases (continued)

Field	Performance Master	Network Master	Historical Cost	Field	Performance Master	Network Master	Historical Cost
Functional Classification	X	X		Annual User Cost		X	
ADT	X	X		Consider Replacement?		X	
Funct. Class. (Sys. under)	X	X		Built By (Original)		X	
Funct. Class (Sys. on)	X			Location	X		
ADT Year	X	X		Project No. (Original)	X	X	
% ADT Truck	X	X		Contract No.			X
ADT (Under)	X	X		Contract Description			X
ADT Year (Under)	X	X		Contract Comments			X
Milepoint	X	X		WBS Number			X
Att.		X		Contract Bid Amount			X
Comments		X		Letting Date			X
Wearing Surface Type	X	X		Work Start Date			X
Wearing Surface Grade		X		Availability Date			X
Substructure Mat'l (Det)		X		Acceptance Date			X
No. of Main Spans	X			Operating Rating	X		
No. of Approach Spans	X			Region	X		
Exp. Jnt. Facility Carried	X			Inventory Rating	X		
Last Routine Insp. Date	X			Inventory Type	X		
Vert. Overclear. Goal	X			Lane Desirable	X		
Vert. Underclear. Goal	X	X		Project Type			X
NCB Deck Width GL ID		X		Consider Replacement?	X		
Bridge Type		X		Bridge Status	X		
Bearing Grade (#)	X	X		Channel Condition	X		
SD Calc		X		Deck Approach Slabs	X		
Undet. BHI ID		X		Deck Condition	X		
Undet. BHI Score		X		From MP	X		
Route Type			X	Deck Material Type (Det)	X		
Federal Aid Number			X	Floor Beams	X		
NCB Min. Vert. Und. Ft.	X			Superstr. Dsgn Type (Det)	X		
Safe Load Appraisal	X			Substr. Mat'l Type (Det)	X		
Scour Analyzed	X			Special Materials	X		
Substructure Condition	X			Stringer Connection	X		
Substr. Condition (#)	X			Timber Replacemnt Status	X		
Superstructure Condition	X			Total Horiz. Clearance	X		
Superstr. Condition (#)	X			County Owner	X		
Corridor	X			Statewide Owner	X		
Encroachments	X			Left Sidewalk Width	X		
Desired Cap. Goal	X			Right Sidewalk Width	X		
Deck - Concrete Grade	X			Exp Jnt – Compr. Seal Gr.	X		
Deck – Timber Grade	X			Exp Jnt – Opn Jnt Grade	X		

Deck – Stl Open Grid Grde	X				Exp Jnt – Prefab Device	X		
Deck – Stl Plank Grade					Exp. Jnt. – Std. Jnt Grade	X		

Table 3.4: Fields available in NCDOT BMS databases (continued)

Field	Performance Master	Network Master	Historical Cost		Field	Performance Master	Network Master	Historical Cost
Concrete Backwalls Grade	X				Bottom Lat Bracing Guide	X		
Concrete Int. Bent Pile Gr	X				Bottom Slab Grade	X		
Concrete Bent Ftg Grade	X				Railing Grade – Alum.	X		
Concrete Bnt Col Grade	X				Railing Grade – Timber	X		
Concr. Abut Int Bent Cap	X				Railing Grade – Steel	X		
Opening Desc 5th	X				Railing Grade - Concrete	X		
Opening Desc 4th	X				Nav Horiz. Clearance	X		
Opening Desc 3rd	X				Nav Vert Clearance	X		
Opening Desc 2nd	X				Type Floor & Wear Surf.	X		
Channel Condition (#)	X				NCB OFFICE	X		
Project No. (Reconst)	X				Bearing Grade	X		

Many of the continuous variables relating to bridge performance and condition ratings were not used in either central dataset. A cursory glance at the Network and Performance Masters showed that some of those performance-related fields were blank or not complete. Specific location information, such as city or route carried would not be useful in a regression. The supporting datasets also included qualitative variables for each bridge entry that could be used as categorical variables. Categorical variables were only useful for analysis if they met both requirements listed below:

1. The number of entries in each category was large enough to develop a model for that specific category that would apply to similar bridges outside of the analysis.

2. The categories in the variable are distinct in that each category would have a different impact on the predicted value.

Categorical variables that failed to meet the first requirement could be used in the analysis by combining two or more smaller groups to make a larger group. This was only done if that variable met the second requirement. The second requirement was utilized so that adding extra groups within a categorical variable would be statistically significant enough to warrant the added complexity. This was assessed initially through basic assumptions (i.e. deck material type may affect width change while region might not) and verified during the regression by comparing *p*-values and coefficients of the different categorical groups.

The categorical and continuous variables included in both datasets are listed in Table 3.5, along with the type of variable and the source(s). Both the characteristic prediction and cost prediction datasets included each of the variables listed in Table 3.4.

Table 3.5: Variables in central dataset

Variable	Variable Type	Source
Structure Number	Continuous	HCD, PM, NM
Functional classification	Categorical	PM, NM
Bridge system	Categorical	PM, NM
Region	Categorical	PM, NM
Division	Categorical	PM, NM
Water Depth	Continuous	PM, NM
Crown to Bed Height (CTB)	Continuous	PM, NM
Year Built	Continuous	PM
Year Replaced	Continuous	NM
Age at replacement (BRIDGEAGE)	Continuous	PM*, NM*
ADT	Continuous	NM
Old bridge length (OBLEN)	Continuous	PM
New bridge length (NBLEN)	Continuous	NM
Length expansion factor (LEF)	Continuous	PM*, NM*
Old bridge width (OBWID)	Continuous	PM
New bridge width (NBWID)	Continuous	NM
Width expansion factor (WEF)	Continuous	PM*, NM*
Superstructure type	Categorical	NM
Substructure type	Categorical	NM
Deck material type	Categorical	NM
Number of spans: original bridge	Categorical	PM
Number of spans: new bridge	Categorical	NM
Original bridge maximum span length	Continuous	NBI
New bridge maximum span length	Continuous	NM
Deck Geometry Appraisal	Discrete	PM
Underclearance Appraisal	Discrete	PM
Roadway Alignment Appraisal	Discrete	PM
Waterway Adequacy	Discrete	PM
Unit cost (total)	Continuous	Other
Unit cost (construction)	Continuous	Other
Construction cost	Continuous	Other
Preliminary engineering cost	Continuous	Other
Right of way cost	Continuous	Other

**Variable was calculated in central dataset using variables from other sources*

3.2.1.3.1 Continuous Variables

The continuous variables used in the analysis were typically measurements of physical attributes of each structure in imperial units (feet and inches). Most variables are defined in the 1995 NBI Recording and Coding Guide, which was published by the

Federal Highway Administration (FHWA, 1995). Since all of the bridge data for the central datasets were sourced from the NCDOT BMS, there was some minor variation between the quantities for the variables included in those sources.

3.2.1.3.1.1 Structure Length

The structure length for bridges is the distance measured between the backwalls of the end abutments or between expansion joints. The NBI coding uses metric units rounded to the nearest tenth (0.1) of a meter. The data in the NCDOT BMS records the structure length measurements in imperial units. This was verified by comparing the structure length field to the “Span Type” description in the Network Master. While the Network Master did not have specified units, the description of the span lengths in “Span Type” specified length measurements in feet and inches.

The terms *OBLEN* and *NBLEN* were adopted from the Abed-al-Rahim and Johnston (1995) study to represent the respective lengths of the old and new bridge structures. A unitless length expansion factor (LEF) was calculated as a ratio of *NBLEN* to *OBLEN* to identify instances where the length change was unusual. Entries with abnormal LEF values could be filtered and removed from the regression so that the prediction models would not be based upon cases with atypical increases or decreases in structure length.

3.2.1.3.1.2 Deck Width

The NBI Recording and Coding Guide calls for bridge decks to be out-to-out deck measurements, as shown as measurement 2 in Figure 3.2, rounded to the nearest tenth of a meter. The NCDOT BMS provides bridge dimensions in feet instead of meters. As with bridge structure length, *OBWID* and *NBWID* represent the old and new deck widths for a bridge that has been replaced. The width expansion factor (WEF) is the ratio of new to old bridge widths that was used to identify outlier projects that had an unusually small or large change in width.

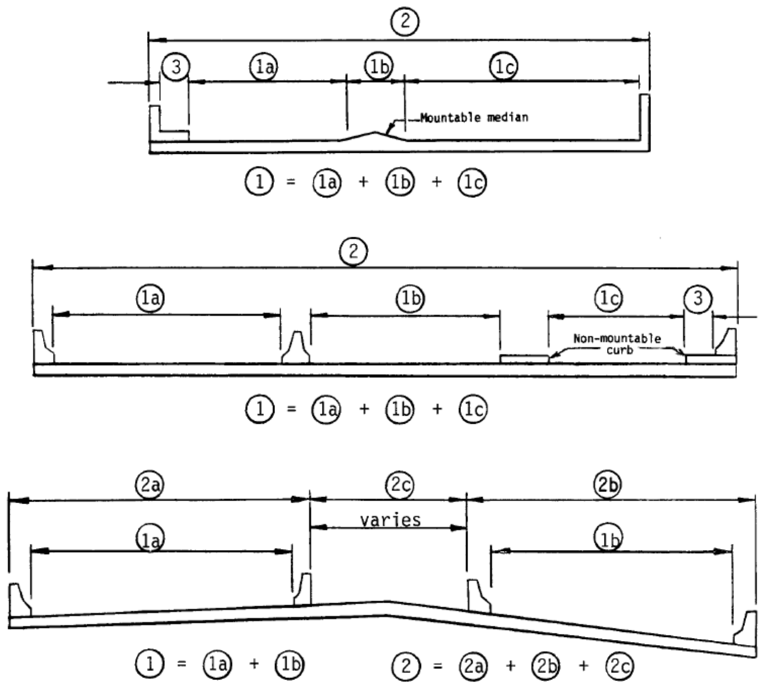

Figure 3.2: Out-to-out bridge deck measurements (from NBI Coding Guide (1995))

3.2.1.3.1.3 Maximum Span Length

According to the NBI Recording and Coding Guide (FHWA, 1995), the length of the maximum bridge span should be measured along the bridge's centerline. This measurement can either be between the centerlines of the supports or the clear distance between the supports, although the measuring points that were used should be noted. *MAXSPAN1* and *MAXSPAN2* represent the maximum span lengths of the old and new structures (respectively). Maximum span length information for the structures being replaced was not readily available in the Performance Master, so data for *MAXSPAN1* was imported from the NBI database.

3.2.1.3.1.4 Bridge Age

The age of the bridge when it was replaced was considered in the development of the prediction equations. *BRIDGEAGE* was calculated with the NCDOT BMS data as the difference between the year the structure was originally built and the year it was replaced. In theory, an older bridge could potentially require more extensive improvements when reconstructed in order to meet modern bridge performance standards than a structure that was built in a later year.

3.2.1.3.1.5 Water Depth

All 305 of the bridge records in the cost prediction dataset and 975 of the bridge records in the characteristic prediction dataset cross a body of water, which could range in size from a small creek to a deep river or bay inlet. *WATERDEPTH* was considered as a possible predictor variable because of the role that flood plains play in many bridge replacement designs. An existing bridge that crosses a river may be more prone to flooding and, when replaced, may require a height and length increase to comply with standards. Waterway adequacy (discussed in a later section) rates a structure's susceptibility to flooding. Bridges that have piers in deep water may also cost more to remove or install during a bridge replacement project as special equipment and methods need to be employed. For this analysis, *WATERDEPTH* used imperial units and was rounded to the nearest foot.

3.2.1.3.1.6 Approach Roadway Width

Based on the year replaced, the approach roadway width (*APPWID*) for the original structure was found in either the 2006 or 2013 version of the Performance

Master. *APPWID* includes the roadway width plus any useable shoulder areas on either side. “Useable shoulder area” is defined in the 1995 NBI Recording and Coding Guide as being stabilized, normally maintained, and structurally capable of handling the same traffic and weather conditions as the facility that is being carried. The 1995 NBI Coding Guide calls for approach roadway width to be measured to the nearest tenth of a meter, however the information from the Performance Master utilized imperial units rounded to the nearest foot.

3.2.1.3.1.7 Crown-To-Bed Height

The “crown” of a bridge is defined as the apex of its arch (Kassler, 1949). The crown-to-bed height for bridges is not stored in the NBI and was not included in the Coding Guide. For the purposes of this work, it was inferred that the measurement from the bed of the feature that the bridge is crossing to the top of the bridge crown represents the maximum height of the bridge structure.

3.2.1.3.2 Categorical Variables

While Abed-al-Rahim and Johnston (1995) explored the possibility of creating separate models for different bridge types, their final prediction models utilized only continuous predictor variables. Saito et al. (1991) used categorical variables indirectly by developing separate prediction models for different types of bridges. This study tested a new approach for using categorical variables to improve the accuracy of the prediction equations. Instead of creating separate equations for different bridge categories, the categorical variables were added directly into the equation during the stepwise process. The stepwise function in Minitab was used to identify and remove categorical variables that were not statistically significant. The remainder of the categorical variables were assigned numerical values in the equation based on their possible values.

Before analysis, the number of bridge entries in each group within a category was counted. For the regression equation to be reliable, each of the groups within a category needed to contain a minimum number of bridge entries. A minimum of 30 entries per classification was judged to be a reasonable number of entries to warrant a classification. Smaller groups were merged with larger groups, which in several cases resulted in binary categorical variables. The processes for binning each categorical variable is described in the following sections.

3.2.1.3.2.1 Functional Classification

Several factors are considered when assigning the functional classification for a route, such as mobility, accessibility, ADT, continuity, and system continuity (FHWA, 2013). *FUNCTCLASS* for the route carried by a bridge had six possible values: Interstate, Principal Arterial, Minor Arterial, Major Collector, Minor Collector, or Local. The characteristic prediction central dataset had a very small population of Interstate bridges, so that category was merged with Principal Arterial.

3.2.1.3.2.2 Bridge System

Bridge system (*BRIDGESYS*) is used by NCDOT to describe the highway system for the route that a bridge is carrying. In the BMS, *BRIDGESYS* could be recorded as either Interstate, Primary, or Secondary. The NCDOT defines secondary routes as DOT-maintained routes that do not carry a "US" or "NC" route number and fall outside the borders of any incorporated municipality (NCDOT, 2017).

3.2.1.3.2.3 Crossing Type

As described earlier in Section 3.2.1.3.1.5., water depth could influence changes in new bridge's size or replacement costs. Following this logic, the significance of including *CROSSINGTYPE* was to see whether the bridge crossed a body of water or a grade separation. In instances where a bridge crossed a roadway and a body of water, the bridge entry defaulted to "not a waterway crossing." This approach was utilized due to the underclearance and waterway adequacy appraisals, which are described in greater detail in sections 3.2.1.3.2.10 and 3.2.1.3.2.11.

3.2.1.3.2.4 Superstructure Material Type

The BMS uses a three-digit code that describes the superstructure material and design type for a bridge. The first digit of this code signifies the material used for the superstructure of the bridge. For this analysis, material and design type were separated into two categories. In the characteristic prediction central dataset, material type was ultimately condensed to two groups (concrete or non-concrete), making *SUPERSTRMAT* a binary variable.

3.2.1.3.2.5 Superstructure Type

The last two digits of the three-digit superstructure code mentioned in Section 3.2.1.3.2.4 indicates the design type of the superstructure. This code was also deciphered using information from the NBI coding guide. The different groups within *SUPERSTRTYPE* were consolidated as necessary to ensure that each group had a large enough population for the analysis. During the regression, further consolidation was done if the coefficients or *p*-values (discussed in Section 3.2.2) indicated that two or more groups could be grouped together to simplify the equation. Since *SUPERSTRMAT* and *SUPERSTRTYPE* were sourced from the same column in either the Network or

Performance Masters, it was expected that there might be a degree of collinearity between those two variables. In cases where this occurred, *SUPERSTRMAT* was retained for the analysis.

3.2.1.3.2.6 Substructure Material

The substructure material type is coded as a one-digit value in the Network Master and Performance Master. Both the Network Master and Performance Master recorded *SUBSTRMAT* as a text description as well as a numerical value. Timber, concrete, and steel were the most prevalent substructure material types included in the central databases.

3.2.1.3.2.7 Deck Material Type

The deck material for each bridge is coded as a single-digit numerical code in both the Network Master and Performance Master. The NBI Recording and Coding Guide was used to translate the numerical code into material types for the structures prior to replacement. Regrouping of deck material types was performed if the number of bridges within in a certain group (having a certain deck material type) was too small to use in the regression analysis. Typically, the decks on the original structures were either made of concrete, steel, or timber.

3.2.1.3.2.8 Deck Geometry Appraisal

Deck Geometry for a bridge is evaluated by the bridge's clear deck width and the minimum vertical clearance over the bridge, with the lower rating dictating the deck geometry appraisal (FHWA, 1995). The numerical rating for deck geometry was already included in the Performance Master, so the ratings were binned to create large enough groups for regressions and to also reduce equation complexity. The rating value for each bridge was imported from the 2006 or 2013 Performance Master, depending on which of the two versions was published before the year the bridge was replaced.

DECKGEOMAPP was classified as being either "Acceptable" or "Unacceptable." The cutoff for an "Acceptable" rating was a rating of 4 or higher. This number was chosen because the NBI Recording and Coding Guide dictated corrective action for a bridge with a vertical clearance rating of 3 or below. Despite the seemingly low cutoff value for acceptable *DECKGEOMAPP* ratings, there were still enough bridge entries that fell within the "Unacceptable" group to warrant the creation of the two groups.

3.2.1.3.2.9 Roadway Alignment Appraisal

The roadway alignment for a bridge is appraised by the change in speed required for vehicles due to the alignment of the approach roadway with respect to the bridge deck. The 1995 NBI Recording and Coding Guide indicates an appropriate rating of 6 for a structure that requires a minor reduction in speed and an 8 for a structure that requires no reduction in speed. In the central dataset, values in *ROADWAYALIGNAPP* were grouped as "Acceptable" (rated 6 or higher) or "Unacceptable" (rated below 6). In theory, bridges with poor approach roadway alignment will need wider bridge decks or expensive modifications to the approach roadway.

3.2.1.3.2.10 Waterway Adequacy

The NBI Recording and Coding Guide defines waterway adequacy as the appraisal of waterway openings with respect to its passage of flow through the bridge (FHWA, 1995). This rating identifies the likelihood of water "overtopping" the bridge and the extent of traffic delays caused by the flooding. The numerical adequacy rating also depends on the functional classification of the route carried by the bridge. Interstates and Principal Arterials receive a lower score for the same waterway conditions as would a Local or Minor Collector. The cutoffs for acceptable and unacceptable waterway adequacy ratings in the central dataset are dependent on *FUNCTCLASS.*

Waterway adequacy ratings only apply for bridges that cross over a waterway, and bridges that do not cross water receive an "N" rating. Another set of bridges crossed over water and roadways, so those entries had both waterway and underclearance adequacy ratings (refer to Section 3.2.1.2.2.11). The *UNDERAPP* category was created to reduce the confusion associated with using two categorical variables to describe the same attribute. *UNDERAPP* uses whichever adequacy rating (waterway or underclearance) is appropriate, and defaults to using underclearance adequacy in cases where both ratings are used.

3.2.1.3.2.11 Underclearance Adequacy

The 1995 NBI Coding Guide describes underclearance adequacy as the horizontal and vertical clearances from the bridge superstructure and substructure for the route travelling beneath the bridge. The sufficiency of the horizontal and vertical clearances are evaluated using tables in the coding guide and the lower of the two scores is used as the underclearance adequacy rating. As mentioned in the previous section, underclearance adequacy and waterway adequacy were combined to form *UNDERAPP* for the purposes of this project.

3.2.1.3.2.12 Number of spans

Multi-span bridges introduce an extra layer of complexity that may drive up cost or the dimensions of the new structure. The number of spans for each original bridge was found in the Performance Master by taking the sum of the number of approach spans and the number of main spans. A histogram of *SPAN1* for the original bridges showed that bridges with more than nine (9) spans were uncommon in the dataset. The singular structures with 17, 24, and 35 spans became a concern for the analysis, since these large and complex bridges might skew the regression. Before the regression was performed, bridges with more than nine (9) spans on the original structure were treated as atypical cases and were filtered from the dataset. The remaining bridge entries were grouped within *SPAN1* as 1, 2, 3, 4, or 5+ span bridges.

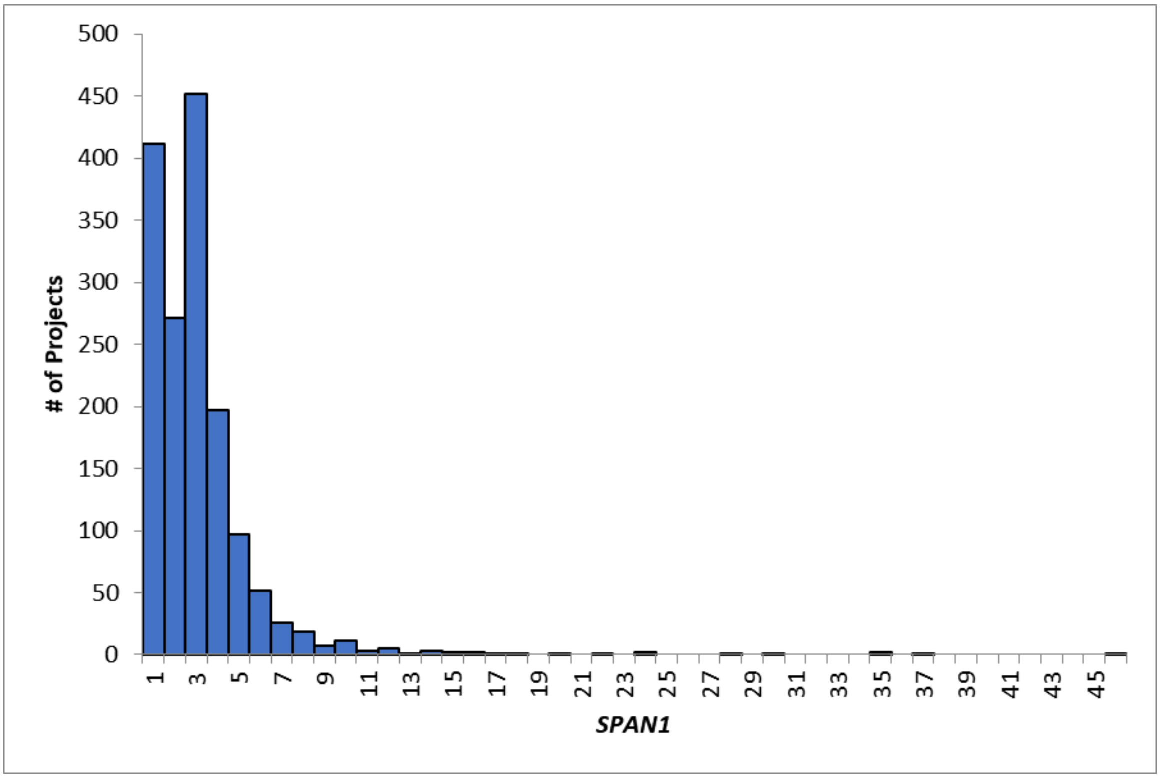

Figure 3.3: Residual of *SPAN1* values in central characteristic dataset

3.2.2 Basket Projects

The term "basket project" was used to describe larger contracts that included the replacement of multiple bridges along a route or within an area. In the historical cost dataset, the individual cost for each bridge included in the basket project was calculated by dividing the total contract amount by the number of replaced bridges in the contract.

Since the bridges included in basket projects are likely not identical, the recorded cost data for each bridge does not reflect the bridge's size, location, or design. Since the creation of cost prediction models depends on linking bridge replacement cost to bridge characteristics, the inclusion of basket project bridges would skew the results.

Basket projects were identified in the central dataset by using simple manual and formulaic data checks. Contract numbers that appeared more than once in the "Contract Number" column meant that there was more than one bridge associated with that contract. Bridges that had the same exact dollar value for "Contract Cost" were often basket projects. The dollar value would typically be a rounded amount that was could reasonably be assumed to have been calculated by division of a larger contract amount (i.e. a contract cost of $5,000,000.00 or $3,333,333.33). Before labeling a bridge as being part of a basket project, both checks were performed to eliminate false positives and to preserve the size of the dataset. Basket projects were filtered from the dataset by including a dichotomous variable that indicates whether an entry is part of a basket project. A filter was applied to hide all entries with variables equal to "Y".

Another approach utilized to find basket projects was a search of the *ConnectNCDOT.com* website. The contract or TIP number for each entry in the cost prediction central dataset was researched on that website to find a general project description for that project. Bid documents, construction drawings, and meeting minutes were reviewed to identify projects where the scope included more than just a bridge replacement. There were several instances where a bridge replacement project also included the construction of nearby roadways or interchanges. This manual approach resulted in identification of approximately 250 records associated with basket projects.

3.2.3 Culverts

There were instances in the Network and Performance Masters where a bridge was converted into a culvert or an existing culvert was replaced by another culvert. Since these structures are also maintained by NCDOT, they are included in the Network and Performance Masters. If culverts were not identified prior to regression analysis, then the regression would be skewed since culverts have no deck area and carry no traffic. Since a culvert replacement or conversion project would appear the same way as a bridge project in the historical cost spreadsheet, several checks were employed to identify the culverts. A culvert meets any of the following conditions:

1. "Culvert Type" = 1
2. "Deck Width" = 0
3. "Deck Geometry Appraisal" = N
4. "Culvert Condition" ≠ N
5. "Structure Type" = 1

As with the basket projects, culverts were filtered from the historical cost dataset. An IF statement for one of the above conditions automatically identified all culverts and assigned those entries a binary value. Entries with this value could be filtered out in Excel. The Network Master was used for identifying culverts so that any bridge-to-culvert conversion projects would also be filtered.

3.2.4 Treatment of Atypical Values

Before performing the regression analyses, Abed-al-Rahim and Johnston (1995) removed entries with unit costs outside of the 5^{th} and 95^{th} percentiles. A similar approach was tested with the central dataset for use on this project. The 5^{th} and 95^{th} percentile values for certain dependent variables were filtered out prior to the regression for that dependent variable, and then the results were reviewed to determine if this filtering approach was suitable for use with the current data and dependent variable. The intended application for the updated regression models was for typical bridge projects, so the filtering of atypical values was done to prevent fitting of the models to the outliers at the expense of fit for the rest of the typical values.

For bridge length and deck width, the 5^{th} and 95^{th} percentile filtering was used to condition the dataset. Instead of filtering the independent variable, the filtering was applied to a unitless expansion factor related to that independent variable. In theory, replacement projects with little to no change in a characteristic (low expansion factor) or an extreme change in a characteristic (high expansion factor) would be treated as abnormal cases. This also covered instances in the central dataset where a typo would cause an artificially high or low expansion factor value.

3.3 Summary of Dataset

The central cost database created for the new models had to contain bridge cost and characteristic information. Detailed geographical and physical attributes for all NCDOT highway bridges was available from the BMS. The cost information was obtained from a separate historical cost dataset and contained fewer entries, so it controlled the size of the central historical cost dataset.

Each data source was reviewed to identify atypical entries that could adversely affect the quality of the regression models. These atypical entries included basket projects, culverts and other non-bridge structures, and replacement projects with atypical costs or changes in dimensions. The result was a characteristic database with 1,506 entries and a cost database with 305 entries.

CHAPTER 4: MODELING PROCESS

4.1 Overview of Modeling Process

The new models created as part of this work are intended to be used by NCDOT personnel in the conceptual stage of a project or for budget forecasting purposes. At this stage in a project, specific characteristics of the new bridge are uncertain, whereas old bridge characteristics are known and recorded in the BMS. Models that use predictor variables that are readily accessible to estimators at the conceptual stage are likely more useful than models that require assumptions of unknown new bridge parameters. For this work, both types of models were developed and evaluated. As a reminder to the reader, this approach was shown schematically in Figure 3.1.

4.1.1 Selection of Models to be Updated

Over time, the reliability of a given prediction model can decrease. In order to keep up with changes in bridge project management and trends in the construction industry, the prediction models should be updated periodically using more current data. As part of this research effort, one of the first steps in updating prediction models was replication of the older prediction models. Using the same independent variables from an older model with a new set of bridge replacement project data shows how applicable the older models are for accurately predicting today's typical bridge replacement projects. Results of this effort also set a benchmark for prediction accuracy.

The regression models developed by Abed-al-Rahim and Johnston (1995) were excellent candidates for replication because they were also developed using data from prior NCDOT bridge replacement projects. To set a benchmark for the updated models, the NCSU NCDOT prediction models were replicated with the original predictor variables and current NCDOT bridge replacement project data. As described in Chapter 2, Abed-al-Rahim and Johnston (1995) followed Approach 1 in Figure 3.2 by developing models that predict new bridge characteristics and then developing cost prediction models that used these new bridge characteristics as predictors.

The inputs used for some of the regression models developed by Saito et al. (1991) for INDOT also allowed them to be used to predict total bridge replacement cost. These models were replicated with NCDOT bridge data using the same predictor variables. Other INDOT models published in the literature predicted costs for different components of bridge replacement projects. Since there was no NCDOT data available for the cost of those components, those INDOT models were not selected for replication.

4.1.2 Model Development Process

Each model was updated and evaluated by following a uniform modeling framework developed for this project (Figure 4.1). The first step of the modeling process was to replicate an existing prediction model using the newer project data from the NCDOT BMS. Once this step was completed and the baseline performance for the replicated model was determined, a backward elimination stepwise regression was performed on the data using Minitab. All available categorical and continuous variables were initially used as possible predictors, and as the backward elimination stepwise regression progressed, at each step the least statistically-significant variable was eliminated. Since the desirable *alpha* threshold was identified by the research team as 0.05, variables with *p*-values greater than 0.05 were removed one-at-a-time until no variables with low statistical significance remained in the equation.

If a group within a categorical variable still had a *p*-value greater than 0.05, the categorical variable would either be consolidated or removed depending on the population size, coefficients, and *p*-values of the remaining groups. Removing or consolidating a categorical variable during the third step would typically increase the *p*-value of one or more other model variables after the regression was run again. When this occurred, the variable with the highest *p*-value above 0.05 was manually removed from the equation in a manner consistent with the stepwise process in Minitab. This process continued until no variables remained with *p*-values greater than 0.05. The consolidation process for categorical variables is outlined in Section 4.1.2.1 with an example demonstrated in Table 4.1.

Adjusted-R^2 and standard error of regression (*S*) were chosen as metrics for model performance. The adjusted-R^2 statistic accounts for the degrees of freedom in models with multiple predictors and can be calculated using Equation 4.1. The standard error of regression (*S*) represents the average distance each data point was located from the regression line. For a model with relatively good fit, the *S* value should be low. This metric was important to the analysis because it can be used to estimate the 95% confidence interval for a model's prediction. This information was used to evaluate whether increasing a model's accuracy justifies the added complexity. The 95% confidence interval was calculated from *S* using Equation 4.2.

$$R^2(adj) = 1 - \left[\frac{(1-R^2)(n-1)}{(n-k-1)}\right] \tag{4.1}$$

Where: R^2 = Unadjusted R^2
n = Total sample size
k = Degrees of freedom in model

$$C.I._{95} = \pm(1.96 \times S) \tag{4.2}$$

Where: $C.I._{95}$ = 95% Confidence Interval
S = Standard error of regression

Variable coefficients were also considered during the supplemental stepwise process. For categorical variables, two groupings within that category with similar coefficients may be candidates for consolidation. Consolidation was useful for cases where one or more of the groupings within a categorical variable had high *p*-values and the grouping had to be condensed to avoid having to remove the entire variable. Comparing coefficients helped establish logical consolidations of groupings due to similar influences on the predicted variable.

4.1.2.1 Preparation of Datasets

Before bridge data from the NCDOT BMS was used for modelling, it was pre-processed (filtered, consolidated, and formatted). The processes used to import data from different sources and measure data quality are described in Chapter 3. For this project, two separate datasets were created for characteristic prediction models and cost prediction models.

In the original datasets, a categorical variable may include five or six different categories. The population within each of these categories was not always equally distributed and some categories had too low of a population to be useful in the analysis. Based on populations within each category, some were combined to simplify the categorical variable and preserve the integrity of the regression equations by eliminating statistically-insignificant categories. Categories that were merged were done so in a logical manner. For example, Interstate bridges were grouped with Principal Arterials to form a single predictor variable *Principal Arterial/Interstate*, since the two functional classifications are closest to each other in traffic capacity. It was reasonable to consolidate prestressed concrete and cast-in-place concrete material classifications, since both categories are made from the same material (concrete).

After running an initial regression, a categorical variable was sometimes further condensed to remove statistically insignificant groupings by looking at each category's model coefficients. Two categories with similar coefficients indicated that the distinction between the two categories did not have a significant effect on the final model prediction. Merging those categories simplified the model and lowered or eliminated the high *p*-values that stemmed from the unnecessary distinction. Typical groupings for categories of the categorical variables can be seen in Table 4.1.

Table 4.1: Typical grouping of categories

<table>
<tr><td rowspan="2">Categorical Variable</td><td>Original Grouping</td><td>Grouping 1</td><td>Grouping 2</td></tr>
<tr><td colspan="3">Progression of consolidation →→→</td></tr>
<tr><td rowspan="6">FUNCTCLASS</td><td>Local</td><td>Local</td><td rowspan="2">Local/Minor Collector</td></tr>
<tr><td>Minor Collector</td><td>Minor Collector</td></tr>
<tr><td>Major Collector</td><td>Major Collector</td><td>Major Collector</td></tr>
<tr><td>Minor Arterial</td><td>Minor Arterial</td><td>Minor Arterial</td></tr>
<tr><td>Principal Arterial</td><td rowspan="2">Principal Arterial/Interstate</td><td rowspan="2">Principal Arterial/Interstate</td></tr>
<tr><td>Interstate</td></tr>
<tr><td rowspan="3">REGION</td><td>Mountains</td><td>Mountains</td><td rowspan="3"></td></tr>
<tr><td>Piedmont</td><td rowspan="2">Piedmont/Coastal</td></tr>
<tr><td>Coastal</td></tr>
<tr><td rowspan="3">BRIDGESYS</td><td>Interstate</td><td rowspan="2">Interstate/Primary</td><td rowspan="3"></td></tr>
<tr><td>Primary</td></tr>
<tr><td>Secondary</td><td>Secondary</td></tr>
<tr><td rowspan="10">SUPERSTRMAT</td><td>Steel</td><td rowspan="2">Steel</td><td rowspan="2">Steel/Timber</td></tr>
<tr><td>Steel, continuous</td></tr>
<tr><td>Concrete</td><td rowspan="4">Concrete</td><td rowspan="8">Concrete</td></tr>
<tr><td>Concrete, continuous</td></tr>
<tr><td>PS Concrete</td></tr>
<tr><td>PS Concrete, cont.</td></tr>
<tr><td>Wood/Timber</td><td rowspan="4">Timber</td></tr>
<tr><td>Masonry*</td></tr>
<tr><td>Aluminum*</td></tr>
<tr><td>Other*</td></tr>
</table>

** indicates groupings that were not represented in the dataset and therefore not used in the analysis*

Table 4.1: Typical grouping of categories (continued)

Categorical Variable	Original Grouping	Grouping 1	Grouping 2
	Progression of consolidation → → →		
SUBSTRMAT	Steel	Steel	Other
	Steel, continuous		
	Concrete	Concrete	Concrete
	Concrete, continuous		
	PS Concrete		
	PS Concrete, continuous		
	Wood/Timber	Timber	Timber
	Masonry*		
	Aluminum*		
	Other	Other	
SUPERSTRTYPE	Slab	Stringer/Multi-Beam or Girder	Girder & Floor -beam System
	Stringer/Multi-Beam or Girder		
	Girder & Floorbeam System	Girder & Floorbeam System	
	Tee Beam	Other	Other
	Box Beam or Girders (mult.)*		
	Box Beam or Girders (single)*		
	Frame*		
	Orthotropic*		
	Truss (deck)*		
	Truss (thru)		
	Arch (deck)		
	Arch (thru)*		
	Suspension*		
	Stayed Girder*		
	Moveable (lift)*		
	Moveable (bascule)*		
	Moveable (swing)*		
	Tunnel*		
	Culvert*		
	Mixed types*		
	Segmental Box Girder*		
	Channel Beam	Channel Beam	
	Other		

** indicates groupings that were not represented in the dataset and therefore not used in the analysis*

Table 4.1: Typical grouping of categories (continued)

Categorical Variable	Original Grouping	Grouping 1	Grouping 2
	Progression of consolidation →→→		
DECKMAT	CIP Concrete	Concrete	
	Precast Concrete		
	Open Grating*		
	Closed Grating*	Steel	
	Steel Plate		
	Corrugated Steel		
	Aluminum*		
	Wood/Timber	Timber	
	Other		
	N/A*		
SPAN1	1 Span	1 Span	1 Span
	2 Spans	2 Spans	2+ Spans
	3 Spans	3+ Spans	
	4 Spans		
	5+ Spans		

** indicates groupings that were not represented in the dataset and therefore not used in the analysis*

Histograms were constructed for each of the independent variables to identify potential outliers. For the characteristic prediction models, length and width-related outliers were identified by expansion factors, which were ratios of the new to old dimensions. Bridge entries with length or width expansion factors outside of the 5th and 95th percentiles were flagged and not used in the regression analyses. It is important to note that the only expansion factors excluded from a regression analysis were related to the independent variable. In other words, only flagged length expansion factor values were removed from the dataset when performing regressions for new bridge length and only flagged width expansion factor values were removed when performing regressions for new bridge width. For development of models for maximum span length, an unfiltered dataset was used, since there was no reliable way to identify outliers for maximum span length since the number of spans changing for a new structure could greatly impact the relative increase or decrease in maximum span length.

The historical cost central datasets were pre-filtered based on total unit costs (dollars per square foot of deck area). The 5th and 95th percentiles were calculated separately for TIP and 17BP projects. The histograms for the unit costs (Figure 4.1 and Figure 4.2) showed that the unit costs were skewed toward the left of the x-axis with a few outliers toward the right. It did not make sense to remove the unit costs below the 5th percentile ($200.32 for 17BP and $334.67 for TIP) since those unit costs were not significantly different than the median values and followed a normal distribution. For this dataset, filtering was only applied for those outliers above the 95th percentile. The thresholds for each subset of bridges are listed in Table 4.2.

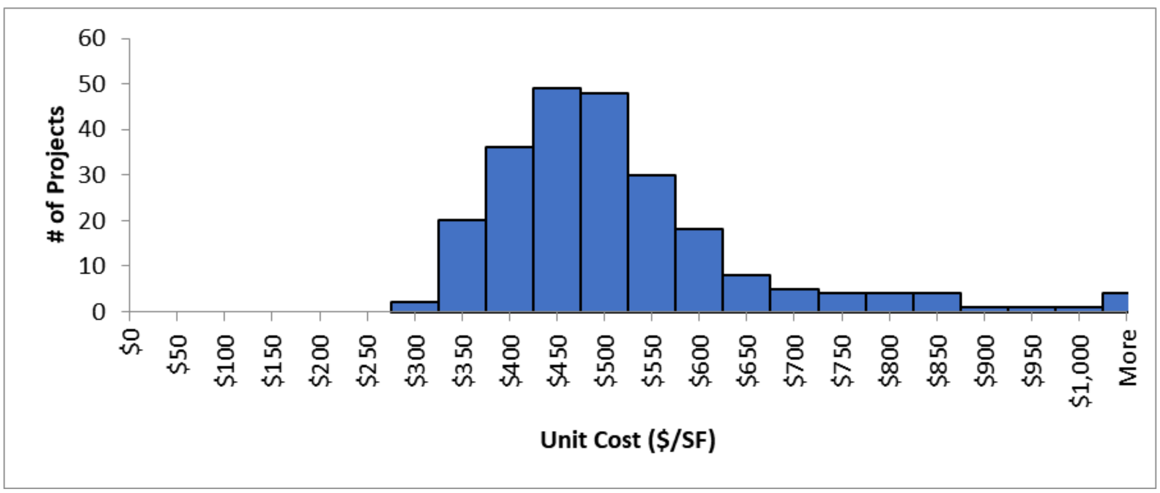

Figure 4.1: Distribution of unit costs for TIP projects prior to prefiltering

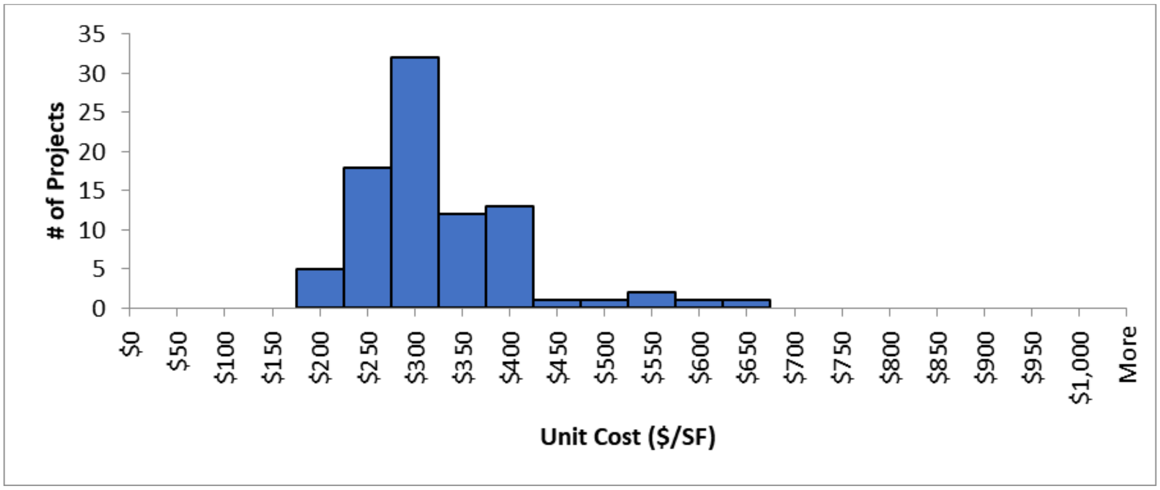

Figure 4.2: Distribution of unit costs for 17BP projects prior to prefiltering

Table 4.2: Statistics for unit costs in dataset prior to prefiltering ($/SF)

	PROJECT TYPE		
STATISTIC	17BP	TIP	ALL
Number of Entries	86	235	321
5th Percentile	\$200.32	\$334.67	\$238.03
Mean	\$298.57	\$501.33	\$447.01
Median	\$274.48	\$456.72	\$423.30
95th Percentile	\$468.80	\$792.44	\$737.44
Standard Deviation	\$84.32	\$243.67	\$231.07

4.1.2.2 Updating of Original NCSU and INDOT Models and Evaluation

Each of the original regression equations were replicated in Minitab using the bridge data from the respective central dataset. The same dependent and independent variables from the original equations were used whenever possible. In the decades following the work by Saito et al. (1991) and Abed-al-Rahim and Johnston (1995), some bridge parameters have been added, modified, or removed from the BMS. Saito et al. (1991) used cost data for specific project components (superstructure, substructure, etc.), while the central dataset only had subtotals for more general components of the project. Since the historical cost dataset provided limited component cost data, the component cost models created by Saito et al. (1991) were not replicated. The total replacement cost model created by Saito et al. (1991) was chosen instead for replication. The component cost models created by Abed-al-Rahim and Johnston (1995) were a closer match to the component cost data in the historical cost dataset, so those models were chosen for replication Some variables from Abed-al-Rahim and Johnston (1995) were substituted with variables from the historical cost dataset that were similar in nature. These substituted variables are listed in Table 4.3.

Table 4.3: Independent variable substitutions

Original Variable[1]:	Substituted with[2]:	Comments
Unit Structure Cost (UNITSTR)	Construction Cost (CONSTCOST)	• Data for structure cost was not available • Unit costs were found to be unsuitable for prediction models due to collinearity with predictors
Roadway Cost (ROADCOST)	Right-of-Way Cost (ROWCOST)	• Data for roadway improvement cost was not available
Miscellaneous Cost (MISCCOST)	Right-of-Way Cost (ROWCOST)	• Data specifically for miscellaneous costs was not available
Structure Cost (STRCOST)	Construction Cost (CONSTCOST)	• Data for structure cost (alone) was not available

[1]*: Dependent variables utilized by Abed-al-Rahim and Johnston (1995)*
[2]*: Component cost variables available from historical cost dataset*

4.1.2.3 Stepwise Regression

Each of the original prediction equations created by Saito et al. (1991) or Abed-al-Rahim and Johnston (1995) utilized fewer than three predictor variables. If a replicated model has a significantly lower adjusted-R^2 value than the original equation, adding predictor variables to the model may increase the fit of the regression equation to the data. Minitab can perform stepwise regressions by adding or removing predictor variables based on their statistical significance in the model.

The Stepwise function in Minitab can be used to perform both forward and backward stepwise regressions. With the forward pass method, the program initially begins with zero predictor variables, and then adds predictor variables one at a time until a target alpha value has been reached. Backward elimination starts with all possible terms included in the model and removes one term at a time until the alpha value has been reached. The alpha value sets the threshold for the minimum amount of statistical significance a term must have in order to be retained in the equation. For this project, the backward elimination method was used with an alpha value of 0.05. The backward elimination method was chosen in an effort to err on the side of caution (considering the potential for each variable to be significant) and allow for as many terms to be included in the model as possible. Backward elimination is recommended for making models that include as many statistically-significant independent variables as possible (Duke, 2017).

When performing a regression, Minitab can create several different models for different groups within the categorical variables provided for the analysis, or it can create one model where each categorical case is treated as a binary variable. For example, the possible deck types for a bridge can be steel, concrete, or timber. Instead of creating three separate equations for each material type, Minitab can present the model as one equation where each possible option for a categorical variable is represented as a set of binary variables represented with either a zero or non-zero coefficient depending on the configuration of the model. This method was chosen because of the number of categorical variables that were considered for the analysis.

The grouping of categorical variables (Section 4.1.2.2) and the stepwise regression were an iterative process. The stepwise selection process eliminated statistically insignificant terms from the models, but it did not remove terms with high degrees of collinearity or condense categories with extraneous groupings. These two cases were checked manually by reviewing regression coefficients and *p*-values for each of the terms in each of the stepwise-generated models. When looking at groupings within a categorical variable, two groupings with similar coefficients could be consolidated with a less significant impact to the rest of the model. When coupled with high p-values, similar coefficients between groupings suggested that the distinction between the two groupings did not have a statistically-significant impact on the rest of the model. This approach typically had minimal to no effect on adjusted-R^2 of a model and in certain cases improved the fit of the model. Reducing the number of groupings also simplified the model and made it more user-friendly. The iterative model revision process was considered complete when all *p-values* were below 0.05.

4.1.2.4 Supplemental Stepwise Process

When the stepwise process in Minitab failed to produce an initial model where all terms had *p*-values lower than 0.05, a supplemental stepwise process was employed to consolidate or remove variables with low statistical significance. This process consisted of identifying the variable with the highest *p*-value above 0.05 and either consolidating it with other groupings (if categorical) or removing it from the model. After this was done, the standard regression was performed again with the remaining variables. If any of the *p*-values from this new regression were above the *alpha* threshold of 0.05, the process was repeated. Once no more variables with high *p*-values remained in the model, the supplemental stepwise process was deemed complete.

For categorical variables with high *p*-values, several considerations factored into the decision to consolidate or remove the variable. Ideally, the decision to consolidate a categorical variable into fewer groups would improve the accuracy of the model. The nature of the consolidation was dictated in part by the coefficient of the group or groups with the highest *p*-value. If the coefficient of the group with the highest *p*-value was close to the coefficient of another group, the two groups were combined before re-running the regression. Consolidation of groups was also done in a logical manner, such as combining cast-in-place concrete structures with precast concrete structures, or placing obscure or uncommon design or material types into an "other" group. Categorical variables that became binary and still had *p*-values greater than 0.05 were removed from the equation.

4.2 Updated Bridge Characteristic Prediction Models

When creating a set of cost prediction models that considers new bridge characteristics, it is necessary to also utilize a set of new bridge characteristic prediction models since specific data on a new bridge's design is not known at the planning stage of a project. New bridge characteristic prediction models were developed to predict changes in structure length (*NBLEN*), deck width (*NBWID*), and maximum span length (*MAXSPAN2*) when a bridge is replaced. The characteristic prediction models from Abed-al-Rahim and Johnston (1995) were replicated and served as a benchmark for further improvement of the models.

The characteristic prediction models were subjected to the stepwise modeling process twice. To develop the first model, the stepwise regression was performed without any quadratic or cubic terms or variable interactions considered. Variable interactions were limited to products of the continuous variables. To develop the second model, the stepwise regression was performed again but included variable interactions. This was done to test whether adding quadratic and cubic terms and variable interactions would benefit the performance of the models enough to justify the added complexity.

Table 4.4: Continuous old bridge characteristics considered in stepwise process

Variable	Square	Cube	Products (Variable Interactions)	
OBLEN	$OBLEN^2$	$OBLEN^3$	*OBLEN*OBWID* *OBLEN*MAXSPAN1* *OBLEN*BRIDGEAGE* *OBLEN*CTB*	*OBLEN*APPWID* *OBLEN*WATERDEPTH* *OBLEN*ADTr*
OBWID	$OBWID^2$	$OBWID^3$	*OBWID*MAXSPAN1* *OBWID*BRIDGEAGE* *OBWID*CTB*	*OBWID*APPWID* *OBWID*WATERDEPTH* *OBWID*ADTr*
MAXSPAN1	$MAXSPAN1^2$	$MAXSPAN1^3$	*MAXSPAN1*BRIDGEAGE* *MAXSPAN1*CTB* *MAXSPAN1*APPWID*	*MAXSPAN1*WATERDEPTH* *MAXSPAN1*ADTr*
BRIDGEAGE	$BRIDGEAGE^2$	$BRIDGEAGE^3$	*BRIDGEAGE*CTB* *BRIDGEAGE*APPWID*	*BRIDGEAGE*WATERDEPTH* *BRIDGEAGE*ADTr*
CTB	CTB^2	CTB^3	*CTB*APPWID* *CTB*WATERDEPTH*	*CTB*ADTr*
APPWID	$APPWID^2$	$APPWID^3$	*APPWID*WATERDEPTH*	*APPWID*ADTr*
WATERDEPTH	$WATERDEPTH^2$	$WATERDEPTH^3$	*WATERDEPTH*ADTr*	
ADTr	$ADTr^2$	$ADTr^3$		

4.2.1 Predicting New Bridge Length

The new bridge length (*NBLEN*) prediction model created by Abed-al-Rahim and Johnston (1995) used original bridge length as the sole predictor variable. Before modeling, bridge entries with length expansion factors outside of the 5th and 95th percentiles were filtered from the analysis. Statistical significance of the relationship between old and new bridge length was performed in Minitab (Eq. 4.1). The adjusted-R^2 for Equation 4.1 was 86.3%. With an *S* value of 24.86, the 95% confidence interval for the predicted new bridge length in Equation 4.1 is ± 48.7 feet.

$$NBLEN = 30.95 + 1.0267(OBLEN) \tag{4.1}$$

Where: *NBLEN* = New bridge length (ft)
OBLEN = Old bridge length (ft)

The stepwise modeling process was performed for *NBLEN* with variable interactions excluded. The supplemental stepwise process required five steps to meet the desired goals (*p*-values below 0.05 as described above) (Table 4.5). The second trial of the stepwise modeling process, with variable interactions included, did not require the removal of additional variables beyond the Minitab stepwise output. The results of this stepwise regression can be seen in Table 4.6.

Table 4.5: Supplemental stepwise process for *NBLEN* (without variable interactions)

Trial	# Terms/Predictors		Action	*Adj-R^2*	*S*	95% C.I.	Max. *P*
1	12	5	*~Began SSP~*	88.0%	23.3068	± 45.68 ft	0.456
2	12	5	Consolidated *SUPERSTRTYPE* (Coded as "girder/beam system" or other type)	88.0%	23.3250	± 45.72 ft	0.107
3	12	5	Consolidated *FUNCTCLASS* (Combined Local and Minor Collector)	87.9%	23.4024	± 45.87 ft	0.074
4	11	4	Removed *ADTr*	87.9%	23.4217	± 45.91 ft	0.135
5	10	3	Removed *OBWID*	87.8%	23.4326	± 45.93 ft	0.038

Table 4.6: Supplemental stepwise process for *NBLEN* (with variable interactions)

Trial	# Terms/Predictors		Action	*Adj-R^2*	*S*	95% C.I.	Max. *P*
1	37	8	*~Began SSP~*	90.1%	21.1010	± 41.38 ft	0.044

The recommended model for predicting *NBLEN* without considering interactions between variables (Table 4.7) was generated from the fifth and final step of the supplemental stepwise process. The recommended *NBLEN* prediction model that included variable interactions was generated directly from the stepwise process in Minitab and did not require elimination or consolidation of additional variables (Table 4.8). The adjusted-R^2 for the recommended models without and with variable interactions (87.8% and 90.1% respectively) both fell short of the values reported by Abed-al-Rahim and Johnston (1995) (98.5%) but were improved, if slightly, over the replicated equation's (Eq. 4.1) adjusted-R^2.

Table 4.7: Recommended *NBLEN* prediction equation (without variable interactions)

NBLEN = 40.058 + 0.83879(OBLEN) + 0.52909(MAXSPAN1) + 1.8584(WATERDEPTH) + (REGION) + (FUNCTCLASS) + (CROSSINGTYPE) + (SUPERSTRMAT) + (SUPERSTRTYPE) + (DECKGEOMAPP) + (SPAN1)

R^2 *(adj) = 87.8%*
n = 1,356
S = 23.4326
$C.I._{95}$ = ± 45.93 ft

REGION =	0.0	*Mountains*
	9.2727	*Piedmont*
	9.8538	*Coastal*
FUNCTCLASS =	0.0	*Local/Minor Collector*
	4.2494	*Major Collector*
	7.4988	*Minor Arterial*
	12.205	*Principal Arterial/Interstate*
CROSSINGTYPE =	0.0	*Not a waterway crossing*
	- 14.446	*Waterway crossing*
SUPERSTRMAT =	0.0	*Concrete*
	- 20.860	*Steel*
	- 19.876	*Timber*
DECKGEOMAPP =	0.0	*Acceptable*
	6.6421	*Unacceptable*
SPAN1 =	0.0	*1 Span*
	7.0231	*2 Spans*
	6.8442	*3 Spans*
	9.2313	*4 Spans*
	13.890	*5+ Spans*

Table 4.8: Recommended *NBLEN* prediction equation (with variable interactions)

$$
\begin{aligned}
NBLEN = {} & 112.91 + 0.45919(OBLEN) - 4.2444(OBWID) + 1.1315(MAXSPAN1) + 4.7761(WATERDEPTH) \\
& - 1.7940(BRIDGEAGE) + 3.7412(CTB) + 0.0083855(ADTr) + 0.0026151(OBLEN^2) - 0.12119(CTB^2) \\
& - 0.062324(APPWID^2) + 0.00000018114(ADTr^2) - 0.0000017175(OBLEN^3) \\
& + 0.0027856(WATERDEPTH^3) + 0.000038388(BRIDGEAGE^3) + 0.0019823(CTB^3) \\
& - 0.0000000000015426(ADTr^3) - 0.0023589(OBLEN*BRIDGEAGE) - 0.011408(OBLEN*CTB) \\
& + 0.0093819(OBLEN*APPWID) - 0.000029465(OBLEN*ADTr) + 0.024576(OBWID*MAXSPAN1) \\
& + 0.022220(OBWID*BRIDGEAGE) + 0.078222(OBWID*APPWID) - 0.00011673(OBWID*ADTr) \\
& + 0.10987(MAXSPAN1*WATERDEPTH) - 0.037515(MAXSPAN1*CTB) \\
& - 0.050918(MAXSPAN1*APPWID) - 0.17884(WATERDEPTH*CTB) \\
& - 0.23405(WATERDEPTH*APPWID) + 0.00033224(WATERDEPTH*ADTr) \\
& + 0.031391(BRIDGEAGE*CTB) + 0.050072(BRIDGEAGE*APPWID) \\
& - 0.000081444(BRIDGEAGE*ADTr) + 0.000045257(CTB*ADTr) + (REGION) + (CROSSINGTYPE) \\
& + (SPAN1)
\end{aligned}
$$

R^2 (adj) = 90.1%
n = 1,356
S = 21.1010
$C.I._{95}$ = ± 41.36 ft

REGION =	0.0000	*Mountains*
	6.5414	*Piedmont*
	10.823	*Coastal*
CROSSINGTYPE =	0.000	*Not a waterway crossing*
	-32.716	*Waterway crossing*
SPAN1 =	0.0000	*1 Span*
	8.8601	*2 Spans*
	12.168	*3 Spans*
	15.180	*4 Spans*
	21.273	*5+ Spans*

4.2.2 Predicting New Bridge Width

In their original study, Abed-al-Rahim and Johnston (1995) did not develop a prediction model for new bridge width. For their research, new bridge width (*NBWID*) was estimated in OPBRIDGE by the NCDOT-based on level-of-service goals. The dataset of more recent bridge replacement projects included information pertaining to how the bridges were utilized, such as ADT, functional classification, bridge system classification, and the old and new bridge widths. The stepwise process was conducted for new bridge width with hopes that some of those data fields would be identified as suitable predictors via the regression analysis.

Before any regressions were performed, bridge entries with width expansion factors outside of the 5th and 95th percentiles were filtered out and excluded from the analysis. In the absence of an original new bridge width prediction equation, simple linear regression on the relationship between old and new bridge width was performed to serve as a benchmark for further model development. The simple linear model, which had an R^2 of 56.5% and an S value of 5.285, is shown below in Equation 4.2.

$$NBWID = 8.292 + 1.0454(OBWID) \tag{4.2}$$

Where: *NBWID* = New bridge width (ft)
OBWID = Old bridge width (ft)

The stepwise modeling process was performed for *NBWID* with variable interactions excluded. The supplemental stepwise process required seven steps (Table 4.9). The second run of the stepwise modeling process, with variable interactions included, required an additional four steps. The results of this stepwise regression can be seen in Table 4.10.

Table 4.9: Supplemental stepwise process for *NBWID* (without variable interactions)

Trial	# Terms/Predictors		Action	*Adj-R²*	*S*	95% C.I.	Max. *P*
1	13	6	*~Began SSP~*	73.4%	4.13603	± 8.11 ft	0.773
2	13	6	Consolidated *DECKMAT* (Coded as "steel" or other type)	73.4%	4.13460	± 8.10 ft	0.535
3	13	6	Consolidated *REGION* (Coded as "piedmont" or other region)	73.4%	4.13364	± 8.10 ft	0.089
4	13	6	Consolidated *SUPERSTRMAT* (Coded as "concrete" or other type)	73.4%	4.13821	± 8.11 ft	0.060
5	12	6	Removed *SUPERSTRMAT*	73.3%	4.14214	± 8.12 ft	0.052
6	12	6	Consolidated *SUPERSTRTYPE* (Coded as "girder/beam system" or other type)	73.1%	4.16023	± 8.15 ft	0.910
7	11	6	Removed *SUPERSTRTYPE*	73.1%	4.15869	± 8.15 ft	0.012

Table 4.10: Supplemental stepwise process for *NBWID* (with variable interactions)

Trial	# Terms/Predictors		Action	*Adj-R²*	*S*	95% C.I.	Max. *P*
1	26	8	*~Began SSP~*	78.6%	3.70872	± 7.27 ft	0.959
2	26	8	Consolidated *FUNCTCLASS* (Combined Local and Minor Collector)	78.6%	3.70892	± 7.27 ft	0.199
3	26	8	Consolidated *SUPERSTRTYPE* (Coded as "girder/beam system" or other type)	78.4%	3.72536	± 7.30 ft	0.638
4	25	8	Removed *SUPERSTRTYPE*	78.4%	3.72427	± 7.30 ft	0.061

The recommended *NBWID* prediction models without variable interactions (Table 4.11) and with variable interactions (Table 4.12) both had adjusted-R^2 values greater than that of Equation 4.2 (73.1% and 78.4% respectively).

Table 4.11: Recommended *NBWID* prediction equation (without variable interactions)

$$NBWID = -13.899 - 0.018451(OBLEN) + 0.70895(OBWID) + 0.091028(MAXSPAN1) + 0.15516(WATERDEPTH) + 0.14022(APPWID) + 0.00074050(ADTr) + (FUNCTCLASS) + (SPAN1) + (DECKMAT) + (REGION)$$

R^2 *(adj) = 73.1%*
n = 1,354
S = 4.15869
$C.I._{95}$ = ± 8.15 ft

FUNCTCLASS =	0.000	*Interstate*
	21.348	*Local*
	21.779	*Minor Collector*
	22.360	*Major Collector*
	23.336	*Minor Arterial*
	19.548	*Principal Arterial*
SPAN1 =	0.0000	*1 Span*
	1.5368	*2 Spans*
	1.2823	*3 Spans*
	2.6969	*4 Spans*
	2.5320	*5+ Spans*
DECKMAT =	0.0000	*Other material*
	- 1.1725	*Steel*
REGION =	0.00000	*Mountain/Coastal*
	0.69413	*Piedmont*

Table 4.12: Recommended *NBWID* prediction equation (with variable interactions)

<table>
<tr><td colspan="3">

$NBWID$ = 18.425 + 0.89625(CTB) + 0.0015005(ADTr) + 0.016848($OBWID^2$) - 0.0024443($MAXSPAN1^2$) - 0.050350($WATERDEPTH^2$) - 0.036797(CTB^2) - 0.000000063522($OBLEN^3$) - 0.00011639($OBWID^3$) + 0.000012774($MAXSPAN1^3$) + 0.0019365($WATERDEPTH^3$) + 0.00045756(CTB^3) - 0.000089372($APPWID^3$) - 0.00000000000034776($ADTr^3$) + 0.0017984(OBLEN*OBWID) - 0.00056837(OBLEN*BRIDGEAGE) - 0.027340(OBWID*CTB) + 0.000019513(OBWID*ADTr) + 0.0021289(MAXSPAN1*BRIDGEAGE) + 0.0069695(WATERDEPTH*BRIDGEAGE) - 0.000010663(BRIDGEAGE*ADTr) + 0.030081(CTB*APPWID) - 0.000015101(APPWID*ADTr) + (FUNCTCLASS) + (CROSSINGTYPE) + (DECKGEOMAPP)

R^2 *(adj) = 78.4%*
n = 1,354
S = 3.72427
$C.I._{95}$ = ± 7.30 ft

</td></tr>
<tr><td rowspan="4">*FUNCTCLASS =*</td><td>0.0000</td><td>*Local/Minor Collector*</td></tr>
<tr><td>7.7432</td><td>*Major Collector*</td></tr>
<tr><td>1.9887</td><td>*Minor Arterial*</td></tr>
<tr><td>- 1.7308</td><td>*Principal Arterial/Interstate*</td></tr>
<tr><td rowspan="2">*CROSSINGTYPE =*</td><td>0.0000</td><td>*Not a waterway crossing*</td></tr>
<tr><td>- 3.4204</td><td>*Waterway crossing*</td></tr>
<tr><td rowspan="2">*DECKGEOMAPP =*</td><td>0.0000</td><td>*ACCEPTABLE*</td></tr>
<tr><td>1.2151</td><td>*UNACCEPTABLE*</td></tr>
</table>

4.2.3 Predicting New Bridge Maximum Span Length

The original *MAXSPAN2* prediction equation created by Abed-al-Rahim and Johnston (1995) was replicated with the newer bridge characteristic dataset. The replicated equation (Eq. 4.3) had an adjusted-R^2 of 35.2% after being adjusted for the natural-log transformation. With an *S*-value of 21.3973, this model had a 95% confidence interval for the predictions of ±41.9 feet.

$$MAXSPAN2 = 10.3068(MAXSPAN1)^{0.2804} * (OBLEN)^{0.1977} \tag{4.3}$$

Where: *MAXSPAN2* = New bridge maximum span length (ft)
MAXSPAN1 = Old bridge maximum span length (ft)
OBLEN = Old bridge length (ft)

The variable transformation from linear to non-linear was done outside of Minitab. Key statistics for both models, such as R^2, adjusted-R^2, *S*, and the 95%

Confidence Interval were computed manually. After the coefficients were found in Minitab, the model variables and their log-transformed counterparts were exported to a separate spreadsheet. The log-transformed model was used as a formula to calculate a fitted value for each entry. Comparing the fitted value to the actual values made it possible to estimate certain model statistics. Equation 4.3 was used to calculate R^2 based on the residual sum of squares (SS_{res}) and the total sum of squares (SS_{tot}). Equation 4.1 was used to find adjusted-R^2 and Equation 4.4 was used to estimate S.

$$R^2 = 1 - \frac{SS_{res}}{SS_{tot}} \tag{4.3}$$

Where: SS_{res} = Residual sum of squares: Σ(*actual value – fitted value*)^2
SS_{tot} = Total sum of squares: Σ(*actual value – mean of actual values*)^2

$$S = \sqrt{\frac{SS_{res}}{N}} \tag{4.4}$$

Where: SS_{res} = Residual sum of squares: Σ(*actual value – fitted value*)^2
N = Number of entries

One of the limitations with using natural-log transformations for the stepwise modeling process lies in how the categorical variables are integrated into the model. One of the advantages of the updated models is the possibility of incorporating several different categorical variables into one model. The coefficients are coded so that one group within each categorical variable has a coefficient of zero. In the transformed natural-log model, the predicted value is the product of all the model's predictor variables and coefficients, so a term with a zero coefficient would cause the predicted value to also equal zero. It was presumed that the new modeling approach would allow for sufficient improvement to the *MAXSPAN2* prediction model to preclude the need for complex variable transformations.

With variable interactions considered as possible predictors, the supplemental stepwise process for *MAXSPAN2* required two additional steps beyond the Minitab stepwise regression to produce a model with acceptable *p*-values. A summary of the steps is shown in Table 4.13. The supplemental stepwise process performed with quadratic and cubic terms and variable interactions is summarized in Table 4.14.

Table 4.13: Supplemental stepwise process for *MAXSPAN2* (without variable interactions)

Trial	# Terms/Predictors		Action	*Adj-R^2*	*S*	95% C.I.	Max. *P*
1	12	5	*~Began SSP~*	48.4%	19.0707	± 37.4 ft	0.659
2	12	5	Consolidated *SUBSTRMAT*	48.4%	19.0656	± 37.4 ft	0.146
3	12	5	Consolidated *SUPERSTRTYPE*	48.4%	19.0555	± 37.3 ft	0.048

Table 4.14: Supplemental stepwise process for *MAXSPAN2* (with variable interactions)

Trial	# Terms/Predictors		Action	*Adj-R^2*	*S*	95% C.I.	Max. *P*
1	24	8	*~Began SSP~*	53.8%	18.0383	± 35.4 ft	0.292
2	24	8	Consolidated *REGION*	53.8%	18.0390	± 35.4 ft	0.206
3	24	8	Consolidated *SUBSTRMAT*	53.8%	18.0427	± 35.4 ft	0.035

Table 4.15 contains the recommended model for predicting *MAXSPAN2* without considering interactions between variables. Predictions made with variable interactions can be made with the model in Table 4.16. Both linear models had higher adjusted-R^2 and lower *S* than the transformed natural-log model, albeit with more variables.

Table 4.15: Recommended *MAXSPAN2* prediction equation (without variable interactions)

MAXSPAN2 = 73.099 + 0.53443(MAXSPAN1) + 0.42766(CTB) – 0.4458(WATERDEPTH)
+ 0.18300(BRIDGEAGE) + 0.31932(APPWID) + (FUNCTCLASS) + (CROSSINGTYPE)
+ (SPAN1) + (REGION) + (DECKMAT) + (SUBSTRMAT) + (SUPERSTRTYPE)

R^2 *(adj) = 48.4%*
n = 1,506
S = 19.0555
$C.I._{95}$ *= ± 37.3 ft*

FUNCTCLASS =	0.0	*Interstate*
	- 17.213	*Local*
	- 19.048	*Minor Collector*
	- 16.868	*Major Collector*
	- 24.385	*Minor Arterial*
	- 20.978	*Principal Arterial*
CROSSINGTYPE =	0.0	*Not a waterway*
	- 35.505	*Waterway*
SPAN1 =	0.0	*1 Span*
	15.987	*2 Spans*
	17.126	*3 Spans*
	15.555	*4 Spans*
	19.576	*5+ Spans*
REGION =	0.0	*Mountains*
	6.0209	*Piedmont*
	- 3.7617	*Coastal*
DECKMAT	0.0	*Other type*
	- 5.5290	*Steel*
	- 4.2567	*Timber*
SUBSTRMAT	0.0	*Other type*
	- 7.3787	*Concrete/Timber*
SUPERSTRTYPE	0.0	*Other type*
	- 6.0483	*Channel Beam*

Table 4.16: Recommended *MAXSPAN2* prediction equation (with variable interactions)

$$\begin{aligned}MAXSPAN2 = {} & 57.888 + 2.1620(MAXSPAN1) - 0.60896(OBLEN) + 1.6182(CTB) + 0.0014671(ADTr) \\ & + 0.002875(OBLEN^2) - 0.01772(MAXSPAN1^2) + 0.001868(BRIDGEAGE^2) \\ & + 0.000000013241(ADTr^2) - 0.0000018177(OBLEN^3) + 0.000079622(OBWID^3) \\ & + 0.000078033(MAXSPAN1^3) - 0.0033249(OBLEN*OBWID) \\ & - 0.0021610(OBLEN*MAXSPAN1) - 0.0044436(OBLEN*CTB) \\ & + 0.0027994(OBLEN*APPWID) - 0.029635(MAXSPAN1*WATERDEPTH) \\ & - 0.000020301(MAXSPAN1*ADTr) - 0.000050951(CTB*ADTr) + (FUNCTCLASS) + (REGION) \\ & + (CROSSINGTYPE) + (SUBSTRMAT) + (ROADWAYALIGNAPP) + (SPAN1)\end{aligned}$$

R^2 (adj) = 53.8%
n = 1,506
S = 18.0427
$C.I._{95}$ = ± 35.4 ft

FUNCTCLASS =	0.0	*Interstate*
	- 23.858	*Local*
	- 25.745	*Minor Collector*
	- 22.970	*Major Collector*
	- 28.917	*Minor Arterial*
	- 26.965	*Principal Arterial*
REGION =	0.0	*Mountains/Coastal*
	4.9739	*Piedmont*
CROSSINGTYPE =	0.0	*Not a waterway*
	- 33.451	*Waterway*
SUBSTRMAT =	0.0	*Other type*
	- 7.7541	*Concrete/Timber*
ROADWAYALIGNAPP =	0.0	*ACCEPTABLE*
	- 4.4703	*UNACCEPTABLE*
SPAN1 =	0.0	*1 Span*
	27.209	*2 Spans*
	36.643	*3 Spans*
	40.129	*4 Spans*
	46.464	*5+ Spans*

4.3 Updated Bridge Replacement Cost Prediction Models

As discussed previously and shown graphically in Figure 3.1, two types of cost prediction models were developed for each type of bridge cost. The first type of cost models used a combination of old and new bridge characteristics as predictors. The new bridge variables used were continuous variables estimated using the equations developed in Section 4.1. The estimated values from those equations can be used in the cost prediction models. The second type of cost model developed uses only old bridge characteristics and does not require prediction of the new bridge's physical attributes.

The cost prediction models also consider quadratic and cubic terms and interactions between continuous variables. While this approach introduced additional complexity to the models, it was also shown to improve the fit of the models. This was especially important for the cost prediction models, as a seemingly small decrease in adjusted-R^2 could translate into a much larger increase in the model's 95% confidence interval for predictions.

Possible continuous variable combinations for the first cost model type (old and new characteristics) are listed in Table 4.17. The variable combinations considered when making the second model type (old characteristics only) are listed back in Table 4.4. Both model types also included the categorical variables listed in Table 4.18 as possible predictors. One exception to this is the *PROJECTTYPE* variable. Since the component cost models (*ROWCOST*, *ENGCOST*, and *CONSTCOST*) use cost data from only TIP projects, there was no need to include this variable in those equations. The component costs values recorded for each TIP project entry did not add up to the total cost amount, which precluded using the summation of the three component cost estimates as a substitute for the *TOTCOST* prediction model. Only the total cost models used *PROJECTTYPE* as a potential predictor.

Table 4.17: Continuous new bridge characteristics considered in stepwise process

Variable	Quadratic	Cubic	Products (Variable Interactions)	
NBLEN	*NBLEN*^{2}	*NBLEN*^{3}	*NBLEN*NBWID* *NBLEN*MAXSPAN2* *NBLEN*BRIDGEAGE* *NBLEN*CTB*	*NBLEN*APPWID* *NBLEN*WATERDEPTH* *NBLEN*ADTr* *NBLEN*DECKAREA*
NBWID	*NBWID*^{2}	*NBWID*^{3}	*NBWID*MAXSPAN2* *NBWID*BRIDGEAGE* *NBWID*CTB* *NBWID*APPWID*	*NBWID*WATERDEPTH* *NBWID*ADTr* *NBWID*DECKAREA*
MAXSPAN2	*MAXSPAN2*^{2}	*MAXSPAN2*^{3}	*MAXSPAN2*BRIDGEAGE* *MAXSPAN2*CTB* *MAXSPAN2*APPWID*	*MAXSPAN2*WATERDEPTH* *MAXSPAN2*ADTr* *MAXSPAN2*DECKAREA*
BRIDGEAGE	*BRIDGEAGE*^{2}	*BRIDGEAGE*^{3}	*BRIDGEAGE*CTB* *BRIDGEAGE*APPWID* *BRIDGEAGE*WATERDEPTH*	*BRIDGEAGE*ADTr* *BRIDGEAGE*DECKAREA*
CTB	*CTB*^{2}	*CTB*^{3}	*CTB*APPWID* *CTB*WATERDEPTH*	*CTB*ADTr* *CTB*DECKAREA*
APPWID	*APPWID*^{2}	*APPWID*^{3}	*APPWID*WATERDEPTH* *APPWID*ADTr*	*APPWID*DECKAREA*
WATERDEPTH	*WATERDEPTH*^{2}	*WATERDEPTH*^{3}	*WATERDEPTH*ADTr*	*WATERDEPTH*DECKAREA*
ADTr	*ADTr*^{2}	*ADTr*^{3}	*ADTr*DECKAREA*	
DECKAREA	*DECKAREA*^{2}	*DECKAREA*^{3}		

Table 4.18: Categorical variables considered in stepwise process

REGION	(Region)	*DECKMAT*	(Deck Material)
FUNCTCLASS	(Functional Classification)	*DECKGEOMAPP*	(Deck Geometry Appraisal)
BRIDGESYS	(Bridge System)	*ROADWAYALIGNAPP*	(Roadway Align. Appraisal)
CROSSINGTYPE	(Crossing Type)	*UNDERAPP*	(Underclearance or Waterway Adequacy)
SUBSTRMAT	(Substructure Material)		
SUPERSTRMAT	(Superstructure Material)	*SPAN1*	(Original number of spans)
SUPERSTRTYPE	(Superstructure Type)	*PROJECTTYPE**	(Project Type)

** only considered for TOTCOST models*

4.3.1 Predicting Construction Cost

The structure cost prediction model created by Abed-al-Rahim and Johnston (1995) served as the foundation for the construction cost (*CONSTCOST*) prediction model. The R^2 for the original structure cost prediction model was not reported, but the replicated equation had a 95% confidence interval of ± $94 per square foot of bridge deck area (Equation 4.5).

$$UNITCONSTCOST = 199.1 + 0.31(MAXSPAN2) - 0.0007(MAXSPAN2)^2 \tag{4.5}$$

Where: *UNITCONSTCOST* = Unit Construction Cost ($\$/ft^2$)
MAXSPAN2 = Maximum span length of new bridge (ft)

For both Approach 1 and Approach 2, the final models generated from the stepwise process in Minitab did not require the removal of any additional predictors or further consolidation of categorical variables. The Minitab regression models using new bridge characteristics (Table 4.19) and old bridge characteristics (Table 4.20) are the recommended models for the approaches that they represent.

Table 4.19: Recommended *CONSTCOST* prediction equation (Approach 1)

$$\begin{aligned} CONSTCOST = {} & 15{,}902{,}000 - 351.48(ADTr) + 21{,}457(DECKAREA) + 884.64(NBLEN^2) \\ & + 1.2998(DECKAREA^2) - 71.247(DECKAREA*NBLEN) - 320.29(DECKAREA*NBWID) \\ & - 345{,}960(NBLEN) - 673{,}730\,(NBWID) - 99{,}508(MAXSPAN2) + 0.42151(NBLEN^3) \\ & + 170.63(NBWID^3) + 20.003(BRIDGEAGE^2) - 0.0000082626(DECKAREA^3) \\ & - 36.647(DECKAREA*MAXSPAN2) - 0.081224(DECKAREA*ADTr) \\ & + 1168.6(NBLEN*MAXSPAN2) + 2.9097(NBLEN*ADTr) \\ & + 3094.3(NBWID*MAXSPAN2) + 8.4513(NBWID*ADTr) + 5.7811(ADTr*CTB) \\ & + (BRIDGESYS) \end{aligned}$$

R^2 (adj) = 99.2%
n = 224
S = 95233.4
$C.I._{95}$ = ± \$186,700

BRIDGESYS =	0	*Primary/Interstate*
	- 84,962	*Secondary*

Table 4.20: Recommended *CONSTCOST* prediction equation (Approach 2)

$$\begin{aligned} CONSTCOST = {} & 278{,}480 + 17{,}595(OBWID) - 133.80(ADTr) + 1.1866(OBLEN*ADTr) \\ & + 214.97(OBLEN*CTB) - 102.48(OBWID*MAXSPAN1) - 5.9873(OBWID*ADTr) \\ & - 655.79(OBWID*APPWID) + 2.3045(MAXSPAN1*ADTr) + 10.923(ADTr*APPWID) \\ & - 0.00000017159(ADTr^3) + 28.497(BRIDGEAGE^2) - 18.234(CTB^3) + (FUNCTCLASS) \end{aligned}$$

R^2 (adj) = 98.9%
n = 224
S = 111275
$C.I._{95}$ = ± \$218,100

FUNCTCLASS =	0	*Local/Minor Collector*
	124,550	*Major Collector*
	362,950	*Minor Arterial*
	- 963,770	*Principal Arterial/Interstate*

4.3.2 Predicting Right-Of-Way Cost

The original roadway cost prediction model developed by Abed-al-Rahim and Johnston (1995) was replicated using right-of-way acquisition cost (*ROWCOST*) as the dependent variable and new bridge width (*NBWID*) as the independent variable. The replicated equation, shown in Equation 4.6, had an R^2 value of 8.4% and an S value of around 23,000. This translated into a 95% confidence interval of ± $45,100.

$$ROWCOST = -17{,}023.3 + 1{,}084.2(NBWID) \tag{4.6}$$

Where: *ROWCOST* = Right-of-way cost
NBWID = New bridge deck width (ft)

During the stepwise modeling process, an outlier was identified from among the group of 305 bridge entries. This specific entry had a *ROWCOST* value greater than $500,000, which was significantly larger than the rest of the bridge entries (Figure 4.3). Models generated with this outlier included would fit the outlying data point well at the expense of the rest of the data points. Since it was clear that this outlier heavily influenced the resulting model, this outlier was removed in the development of further *ROWCOST* prediction models. This entry was also filtered out during the initial replication of the prediction model shown in Equation 4.6.

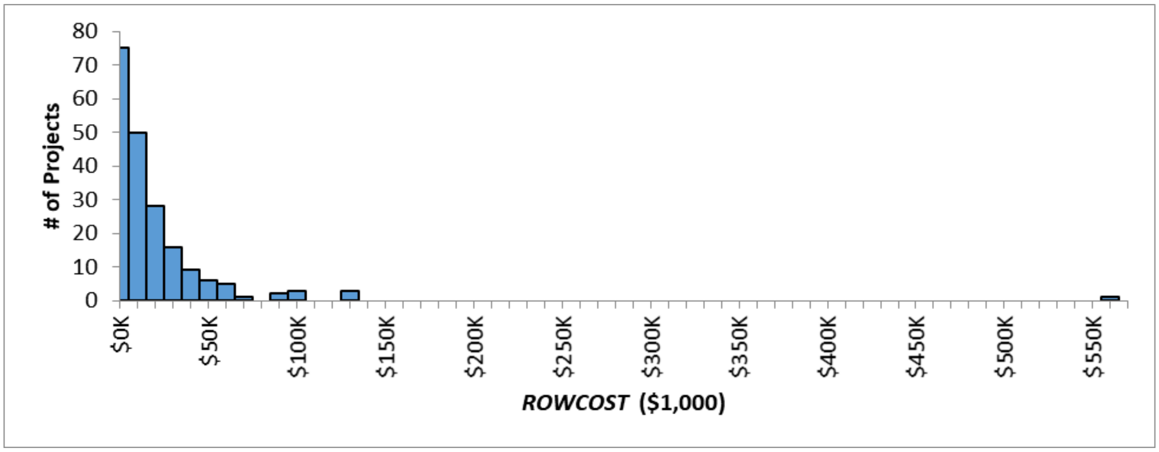

Figure 4.3: Distribution of *ROWCOST* values in central cost dataset

The two recommended *ROWCOST* models using new bridge characteristics (Approach 1, Table 4.21) and with old bridge characteristics (Approach 2, Table 4.22) had low adjusted-R^2 values (34.1% and 30.1% respectively). The original R^2 for the NCSU *ROADCOST* equation was not published, so it was not possible to make a comparison between the performance of the two generations of models.

Table 4.21: Recommended *ROWCOST* prediction equation (Approach 1)

$$\begin{aligned} ROWCOST = & - 684280 - 2564.7(DECKAREA) - 214.60(NBLEN^2) - 0.31190(DECKAREA^2) \\ & + 16.443(DECKAREA*NBLEN) + 45.554(DECKAREA*NBWID) + 36962(NBLEN) \\ & + 15145(BRIDGEAGE) - 0.087767(NBLEN^3) + 1360.3(NBWID^2) - 30.730(NBWID^3) \\ & - 0.0032239(ADTr^2) - 259.99(BRIDGEAGE^2) + 1.4136(BRIDGEAGE^3) \\ & + 0.0000033660(DECKAREA^3) - 0.15841(DECKAREA*MAXSPAN2) \\ & + 0.031507(DECKAREA*ADTr) - 0.69633(DECKAREA*APPWID) \\ & - 0.88218(NBLEN*ADTr) - 1.0757(NBWID*ADTr) - 93.985(NBWID*WATERDEPTH) \\ & + 0.55852(MAXSPAN2*ADTr) + 1.7606(ADTr*WATERDEPTH) + (DECKMAT) \end{aligned}$$

R^2 (adj) = 35.7%
$n = 198$
$S = 19258.4$
$C.I._{95} = \pm \$37,700$

DECKMAT =	0	*Steel or Timber*
	9578.9	*Other type*

Table 4.22: Recommended *ROWCOST* prediction equation (Approach 2)

$$\begin{aligned} ROWCOST = & - 32{,}103 + 6{,}609.8(APPWID) + 7.4531(ADTr) - 11.556(OBLEN*MAXSPAN1) \\ & + 0.16900(OBLEN*ADTr) - 21.804(OBLEN*APPWID) - 0.30355(OBWID*ADTr) \\ & + 83.036(MAXSPAN1*CTB) - 475.86(WATERDEPTH*CTB) \\ & - 25.408(BRIDGEAGE*CTB) + 4.0711(OBLEN^2) - 176.14(APPWID^2) \\ & + 575.51(WATERDEPTH^2) \end{aligned}$$

R^2 (adj) = 30.1%
n = 198
S = 20073.3
$C.I._{95} = \pm \$39{,}300$

4.3.3 Predicting Engineering Cost

Development of an engineering cost (*ENGCOST*) prediction model started with replication of the original *ENGCOST* model developed by Abed-al-Rahim and Johnston (1995). The replicated model (Eq. 4.6) had an adjusted-R^2 of 35.9% when predicted with *CONSTCOST*.

$$ENGCOST = 40{,}189 + 0.06875(CONSTCOST) \qquad (4.7)$$

Where: *ENGCOST* = Preliminary Engineering Cost ($)
CONSTCOST = Construction Cost ($)

The recommended model for predicting *ENGCOST* with new bridge characteristics (Approach 1) had an adjusted-R^2 of 83.6%. This value for this equation (Table 4.23) is higher than the R^2 value reported by Abed-al-Rahim and Johnston (1995) (60%) and the adjusted-R^2 of their replicated equation (35.9%), indicating a better fit of the regression line to the actual data points. When using only old bridge characteristics (Approach 2), the recommended *ENGCOST* prediction model had a lower adjusted-R^2 (Table 4.24) (79.2%) than that of the recommended equation that used new bridge characteristics. The goodness-of-fit for Table 4.24 was still an improvement over the original NCSU equation and its replicated counterpart.

Table 4.23: Recommended *ENGCOST* prediction equation (Approach 1)

$$ENGCOST = -3{,}633{,}100 + 39{,}517(APPWID) + 409.85(ADTr) - 3179.5(DECKAREA)$$
$$- 163.85(NBLEN^2) - 0.15970(DECKAREA^2) + 10.299(DECKAREA*NBLEN)$$
$$+ 29.203(DECKAREA*NBWID) + 71{,}228(NBLEN) + 94{,}908(NBWID)$$
$$+ 59{,}978(BRIDGEAGE) + 0.0073228(ADTr^2) - 0.00000050021(ADTr^3)$$
$$+ 0.10569(DECKAREA*ADTr) + 22.096(DECKAREA*BRIDGEAGE)$$
$$- 3.2093(NBLEN*ADTr) - 40.041(NBLEN*WATERDEPTH)$$
$$- 697.11(NBLEN*BRIDGEAGE) - 13.872(NBWID*ADTr)$$
$$- 1429.9(NBWID*BRIDGEAGE) - 808.29(APPWID*BRIDGEAGE)$$

R^2 *(adj) = 83.6%*
n = 224
S = 43360.5
$C.I._{95} = \pm \$85{,}000$

Table 4.24: Recommended *ENGCOST* prediction equation (Approach 2)

$$ENGCOST = 346{,}660 - 3{,}578.1(OBLEN) - 34{,}374(OBWID) - 92.278(ADTr)$$
$$+ 37.853(OBLEN*MAXSPAN1) + 233.94(OBLEN*APPWID) - 118.63(OBLEN*CTB)$$
$$- 38.327(MAXSPAN1*BRIDGEAGE) + 2.4308(ADTr*APPWID)$$
$$+ 0.94241(ADTr*BRIDGEAGE) - 329.07(APPWID*BRIDGEAGE)$$
$$+ 114.06(BRIDGEAGE*CTB) + 0.035132(OBLEN^3) + 1{,}544.1(OBWID^2)$$
$$- 20.378(OBWID^3) - 0.000000041497(ADTr^3) + 149.16(BRIDGEAGE^2)$$
$$- 1.1790(BRIDGEAGE^3) + (DECKGEOMAPP) + (SPAN1)$$

R^2 *(adj) = 79.2%*
n = 224
S = 48808.9
$C.I._{95} = \pm \$95{,}700$

DECKGEOMAPP =	0	*ACCEPTABLE*
	- 27,908	*UNACCEPTABLE*
SPAN1 =	0	*1 or 2 Spans*
	- 29,735	*3+ Spans*

4.2.4 Predicting Total Replacement Cost

The original Bridge Replacement Total Cost (BRTC) model developed by Saito et al. (1991) used bridge length and deck width as predictors. The first equation (Eq. 4.8) uses old bridge length and width while the second equation (Eq. 4.9) uses new bridge length and width. These models were originally created in log-log form, so it was necessary to adjust the coefficients, adjusted-R^2, and S values that were generated in Minitab. As with *MAXSPAN2*, the method of coding categorical variables in this modeling process is incompatible with log-transformations.

$$TOTCOST = 104{,}930 * (OBLEN)^{0.47233} * (OBWID)^{0.16097} \tag{4.8}$$

Where: *OBLEN* = Old bridge length (ft)
OBWID = Old bridge width (ft)

$$TOTCOST = 1{,}201.2 * (NBLEN)^{0.6957} * (NBWID)^{1.0775} \tag{4.9}$$

Where: *NBLEN* = New bridge length (ft)
NBWID = New bridge width (ft)

The adjusted-R^2 and S values for both transformed equations were calculated after determining the coefficients for the linear model. Equation 4.3 was used to calculate R^2, from which adjusted-R^2 was calculated by Equation 4.1. The standard error of regression (S) was found from Equation 4.4. For the replicated *TOTCOST* model using old bridge data (Equation 4.7), the adjusted-R^2 of the transformed equation was 27.6% with an S value of 890,752. This translates to a 95% confidence interval of ± $1,745,900. When using new bridge data to estimate *TOTCOST* (Equation 4.8), the adjusted-R^2 of the transformed equation was calculated to be 78.2% with an S value of 487,910. This signifies a narrower 95% confidence interval of plus or minus $956,000.

The recommended equation for predicting *TOTCOST* with new bridge characteristics is shown below in Table 4.25. When using only old bridge characteristics, the recommended model for predicting *TOTCOST* is shown in Table 4.26.

Table 4.25: Recommended *TOTCOST* prediction model (Approach 1)

$$TOTCOST = 433{,}650 - 260.32(NBLEN*BRIDGEAGE) - 3.4234(ADTr*APPWID) + 119.90(NBLEN^2) - 0.00000024817(ADTr^3) + 328.75(CTB^2) - 3.2296(DECKAREA*NBLEN) + 0.035682(DECKAREA*ADTr) + 9.7269(DECKAREA*BRIDGEAGE) + (FUNCTCLASS) + (PROJECTTYPE)$$

R^2 *(adj) = 96.1%*
n = 305
S = 207149
$C.I._{95} = \pm \$406{,}000$

FUNCTCLASS =	0	*Other classification*
	- 145900	*Local*
PROJECTTYPE =	0	*17BP*
	525570	*TIP*

Table 4.26: Recommended *TOTCOST* prediction model (Approach 2)

$$TOTCOST = 13{,}363 + 5{,}746.2(OBLEN) + 14{,}016(OBWID) + 1.3327(OBLEN*ADTr) - 1{,}363.2(OBWID*CTB) - 0.000000097673(ADTr)^3 + 56.500(BRIDGEAGE)^2 + 1{,}427.5(CTB)^2 + (FUNCTCLASS) + (PROJECTTYPE)$$

R^2 *(adj) = 94.9%*
n = 305
S = 237169
$C.I._{95} = \pm \$464{,}900$

FUNCTCLASS =	0	*Local or Minor Collector*
	250,080	*Major Collector*
	477,600	*Minor Arterial, Principal Arterial, or Interstate*
PROJECTTYPE =	0	*17BP*
	551,350	*TIP*

4.3 Summary of Findings

For each of the characteristic models developed as part of this work, the models that incorporated variable interactions as predictors had larger adjusted-R^2 values than their counterparts that did not include variable interactions, showing an improved model fit. However the increase in model fit was relatively small (between 2 to 5%) and at the expense of added model complexity. For example, the difference in adjusted-R^2 between the *NBLEN* prediction model that included interactions between variables (90.1%) and the model that excluded interactions between variables (87.8%) represents a change of only 2.3 percent. The difference between the *S* values, which was 2.33 feet, shows that the inclusion of variable interactions in the prediction model narrows the 95% confidence interval by only ± 2.33 feet. Introducing variable interactions to the *NBLEN* prediction model increased the number of predictor terms from 10 to 37. To simplify the models and make them user-friendly, the characteristic prediction models that do not include variable interactions are recommended for implementation into the BMS.

Table 4.27: Results of modeling approaches for dependent variables

	Replicated		With Variable Interactions		No Variable Interactions	
	R^2	*S*	R^2	*S*	R^2	*S*
NBLEN	86.3%	24.86	90.1%	21.10	87.8%	23.43
NBWID	56.5%	5.285	78.4%	3.724	73.1%	4.159
MAXSPAN2	35.2%		53.8%	18.04	48.4%	19.06
	Replicated		New bridge characteristics		Old bridge characteristics	
	R^2	*S*	R^2	*S*	R^2	*S*
CONSTCOST	-	48.06	99.2%	95,300	98.9%	111,300
ENGCOST	35.9%	90,600	83.6%	43,400	79.2%	48,800
ROWCOST	8.4%	23,000	35.7%	19,300	30.1%	20,100
TOTCOST	78.2%	487,900	96.1%	207,100	94.9%	237,200

During the modeling process, cost prediction was explored by project component (*ROWCOST*, *CONSTCOST, ENGCOST*) and as a total project cost (*TOTCOST*). While the recommended *CONSTCOST* and *ENGCOST* models had adjusted-R^2 values greater than 80%, the *ROWCOST* models did not. This can reduce the accuracy of a cost prediction that is made by summing up component costs. The errors for each component cost are additive, and the accumulated error could result in total cost predictions that fall outside the confidence interval for the *TOTCOST* prediction models. Component costs were also only provided for TIP bridge replacement projects, so those models were developed from those type of projects and may not accurately predict costs associated with 17BP projects. As seen with both *TOTCOST* models, *PROJECTTYPE* was a significant predictor for project cost. For that reason, the accuracy of the component cost models for 17BP projects is unclear. The models for *TOTCOST* can be applied to both project types and do not require the summation of predicted costs for project components.

The question of whether the final recommended cost models should use old bridge characteristics, or instead utilize new bridge characteristics predicted from the old bridge characteristics will be addressed in the following chapter, after additional analysis of model residuals and validation using the existing dataset is performed. Before selecting an approach, each pair of models should be compared to quantify the effects that error from the characteristic prediction models have on the cost models that use new bridge data.

CHAPTER 5: MODEL VALIDATION

5.1 Overview

The recommended prediction models from the previous chapter can be used with two different approaches. The first approach (Approach 1) involves predicting cost with predicted new bridge characteristics, which are predicted with their own set of models from known old bridge characteristics. The second approach (Approach 2) is used to predict costs directly from old bridge characteristics, thereby eliminating the need for new bridge characteristic prediction models. After developing the models using actual old and new bridge characteristics, the prediction models that incorporated actual new bridge characteristics were stronger models for predicting cost in terms of goodness-of-fit and having a narrower prediction confidence interval. This is because replacement costs are more closely tied to the characteristics of the new structure that is being built. To use these models in actual practice, however, the new bridge characteristics need to be estimated by NCDOT since the actual characteristics are not known prior to the development of the design. When utilizing predicted characteristics from a first prediction model in a second prediction model, there is potential for prediction error from the first stage of forecasting to be compounded into the final estimate. For these reasons, the results from Approach 1 and Approach 2 need to be computed and compared, in order to provide a recommendation regarding the best means of estimating bridge replacement costs and replacement cost components.

Error within a model can be classified as being either systematic or stochastic. The values of stochastic (random) errors in a model will vary for each trial, while systematic errors remain the same. Model validation as a process involves quantifying systematic errors within a model and deciding whether that amount of error in the model is acceptable for its intended application (Blischke and Murthy, 2000).

The goal of model validation for this project was to determine whether there is an acceptable tradeoff between the higher accuracy of the cost prediction models that use new characteristics and the potential prediction error from those estimated new characteristics. Using the same datasets, predicted values for predicted *NBLEN*, *NBWID*, and *MAXSPAN2* were calculated using their respective recommended prediction equations. Those predicted values took the place of the actual values in the cost prediction models that used new bridge characteristics.

Residual analysis is an important step in a regression analysis, during which the differences between the observed and estimated values, or residuals, are evaluated to identify underlying trends or outliers in the regression model (Toit et al., 1986). Losses in model prediction quality are assessed with histograms of residuals and the standard error of regression (*S*). In this text, model prediction quality encompasses *fidelity*, which is how well the model represents what it is simulating (Cross, 1999); *fit*, which is how close the model estimates are to actual data points (Karen, 2017); and *predictive accuracy*, which is measured by the amount of prediction error in the model (Penn State Eberly College of Science, 2017). Histograms can be compared to look for changes in residual error distribution between the two modeling approaches for a dependent variable. While adjusted-R^2 can be used as a metric to compare goodness-of-fit between models, standard error of regression (*S*) is quantified in the same terms as the dependent variable. This makes *S* more useful for validating models and determining the best prediction approach for each application.

The residuals and histograms were computed outside of Minitab to allow for more flexibility in how the data was binned and displayed. To accomplish this, the fitted values for were calculated using coefficients in scientific notation out to the fourth decimal place. Since these coefficients used more significant figures, fitted values and residuals varied slightly from the original Minitab values, but reflect residuals associated with implementation of the prediction models with regression coefficients truncated to precisions suitable for use in the BMS.

5.2 Comparison of Approaches

In this analysis, three sets of residual error values were compared for each predicted cost variable. The cost prediction models for Approach 1 were developed from a set of known new bridge characteristics sourced from a historical dataset. Determining the spread of the residual errors for those models was done to verify that the mean residual error values were close to zero, with a relatively normal distribution. This would have meant that the errors were “random” in nature and would, for the most part, cancel each other out.

While evaluating the distribution of residual errors for Approach 1 was useful, using actual new bridge characteristic values as predictors would not be an accurate reflection of how the Approach 1 models are to be used outside of this project. The point of using Approach 1 is to be able to estimate new bridge characteristics that would not be known before the design is complete and use those estimated values to predict cost. Aside from predicting cost, the characteristic predictions from Approach 1 can be used for forecasting and design purposes. Residual error distributions for Approach 1 were also found after substituting actual (known) new bridge characteristic values for predicted values generated from the recommended and updated characteristic prediction models from Chapter 4 of this thesis. This set of histograms is a more accurate reflection of how residual error would be distributed when using Approach 1 as it was intended. The residual error distribution histograms for Approach 2 serve the same purpose as with Approach 1 and allow for fair comparisons to be made in deciding on a preferred approach for a particular cost variable.

Histograms were created of the residual error distributions for predicted *NBLEN*, *NBWID*, and *MAXSPAN2* values. These predicted values were calculated from the recommended characteristic equations from Chapter 4: Table 4.7 (*NBLEN*), Table 4.11 (*NBWID*), and Table 4.15 (*MAXSPAN2*). The residual error distributions for these predicted characteristics are shown in Figure 5.1 (*NBLEN*), Figure 5.2 (*NBWID*), and Figure 5.3 (*MAXSPAN2*).

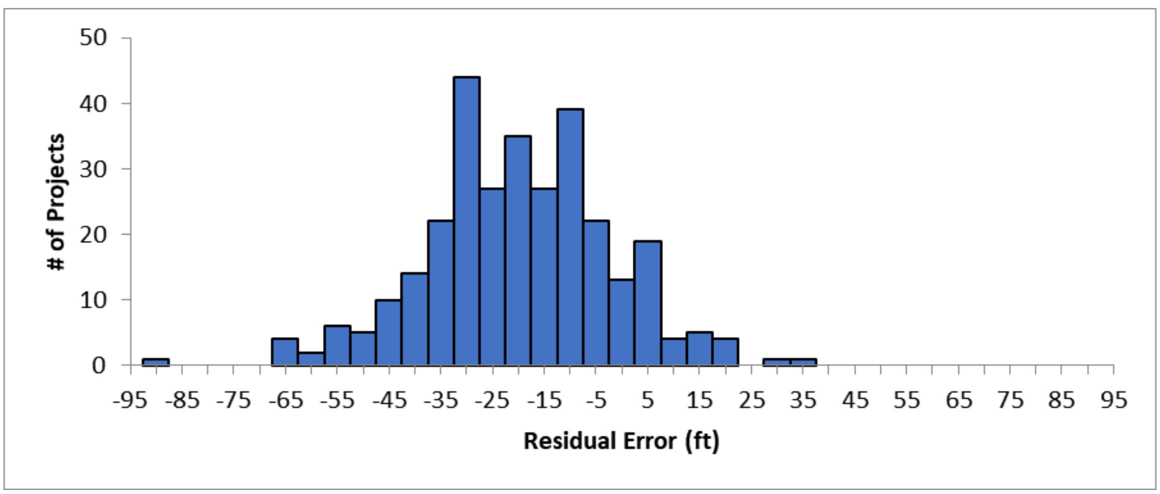

Figure 5.1: Residual error for predicted *NBLEN*

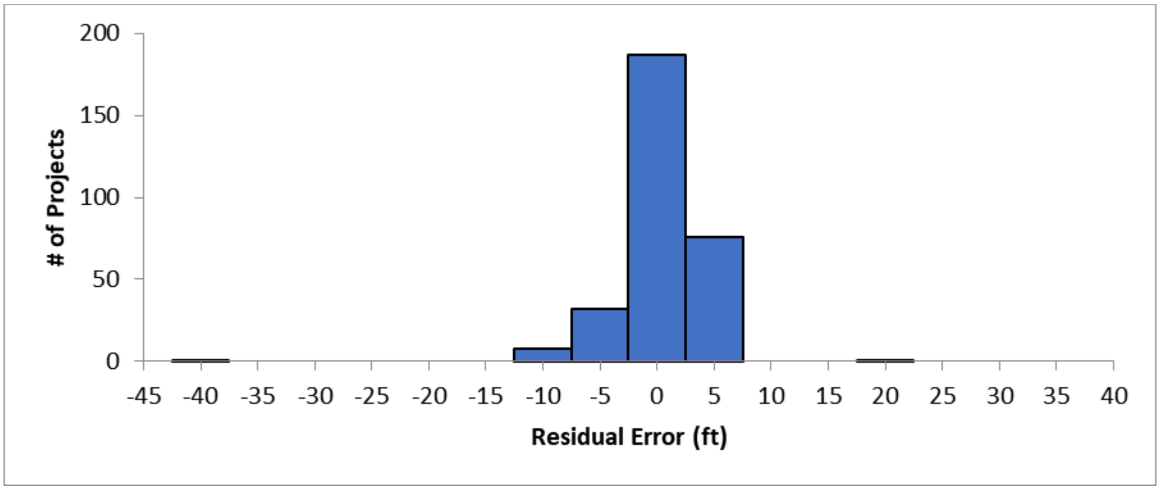

Figure 5.2: Residual error for predicted *NBWID*

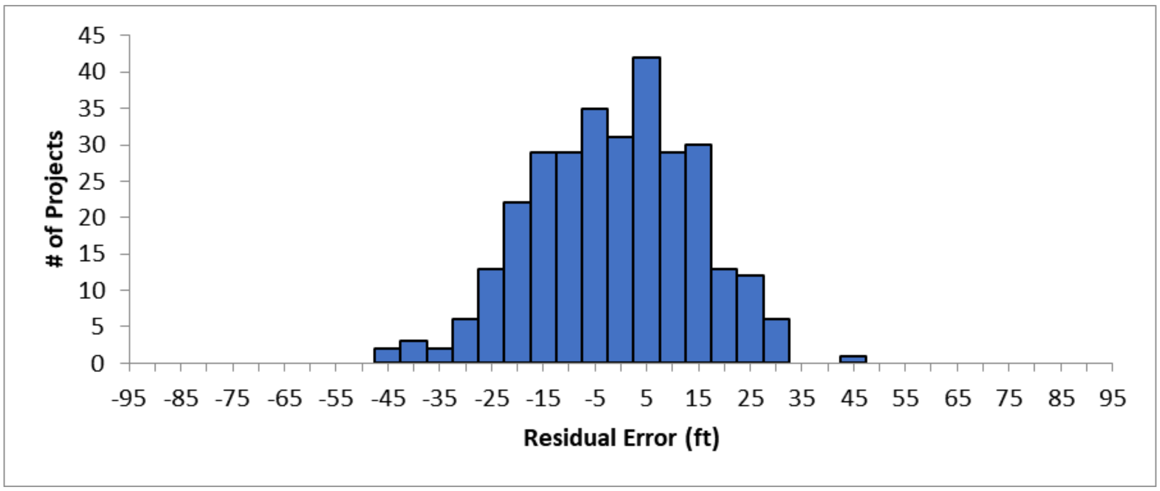

Figure 5.3: Residual error for predicted *MAXSPAN2*

Using the recommended characteristic prediction equations and the recommended cost prediction equations with new bridge characteristics for Approach 1 and the recommended cost prediction equations with old bridge characteristics for Approach 2, the fitted values and residuals were calculated for each final *ROWCOST* prediction. Bin sizes were determined by dividing the difference between the minimum and maximum values by the number of bins. The number of bins required was calculated by taking the square root of the number of data points ($\sqrt{N}$) and rounding up to the nearest whole number. This is the default method used by Microsoft Excel to set up histogram bins (Cameron, 2009).

For some histograms, an extreme outlier would influence the bin sizing. Having an artificially wide range of residual error values meant that the range in which most values lay would be represented by a very small number of bins. This was expected in the histograms for cost predictions with predicted characteristics (Approach 1) since the compounded errors would cause a shift in the mean residual values and a larger standard error (*S*). In cases where outliers were removed to increase the number of bins representing the majority of the data, the quantity and approximate values of the removed entries were recorded.

The mean residual error was found by taking the average of the residuals for each approach. Residuals were calculated by subtracting the actual value from the predicted value, so a positive mean residual error indicates a tendency of a model to over-estimate its predictions for the given dataset. Conversely, a negative mean residual error shows that the approach tends to under-estimate its predictions for the given dataset. For the dataset used for the creation of the models, the mean residual error will be close to zero. The mean and relative distribution of the residual errors will shift if applied to a different dataset or even with different rounding of model coefficients.

5.2.1 Construction Cost

Histograms of residual error for predicting *CONSTCOST* through Approach 1 with actual new bridge characteristics (Figure 5.4) and predicted new bridge characteristics (Figure 5.5) are shown below. Using old bridge characteristics to predict cost as part of Approach 2 resulted the distribution of residual error shown in Figure 5.6.

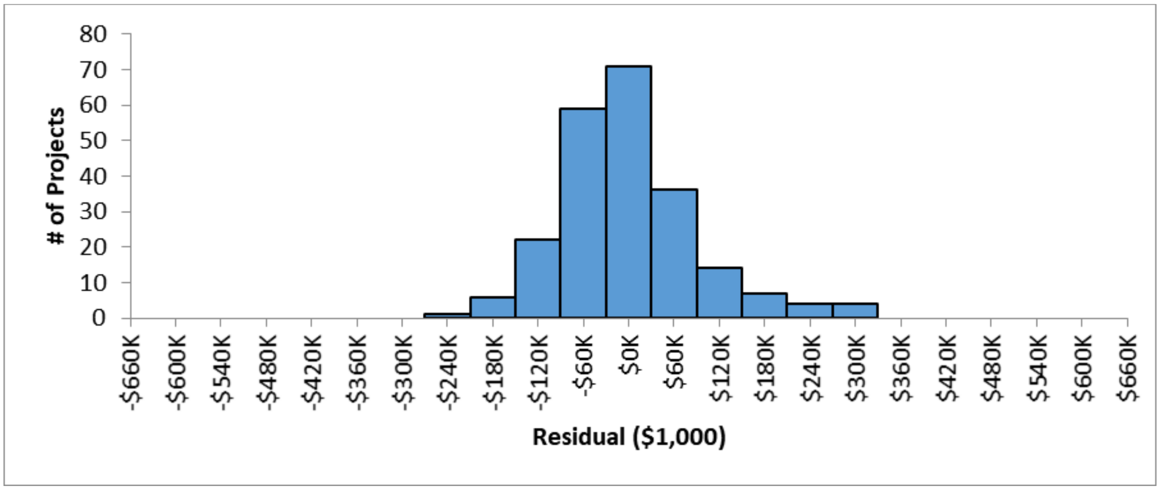

Figure 5.4: Residual error for *CONSTCOST* (with actual new bridge characteristics)

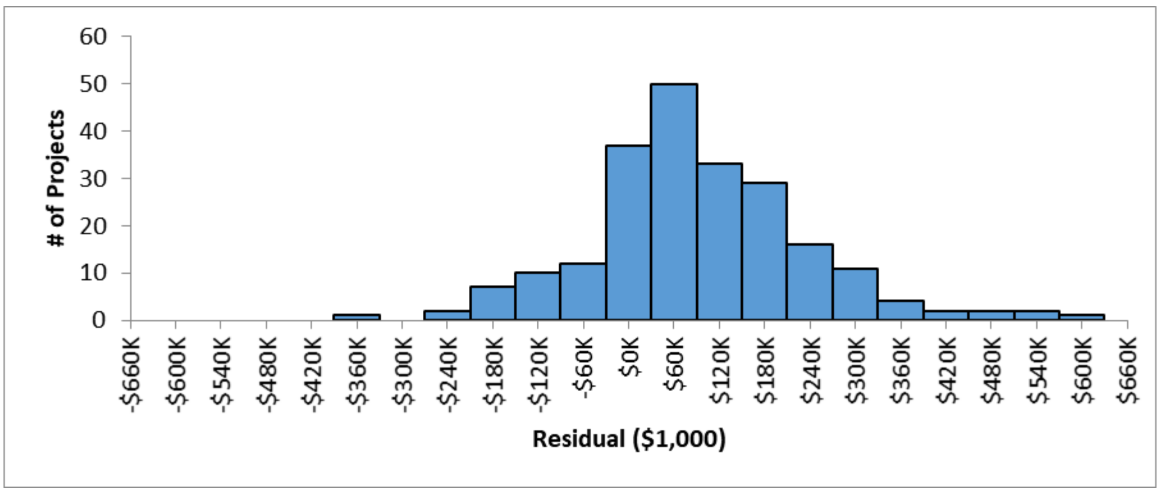

Figure 5.5: Residual error for *CONSTCOST* (with predicted new bridge characteristics)

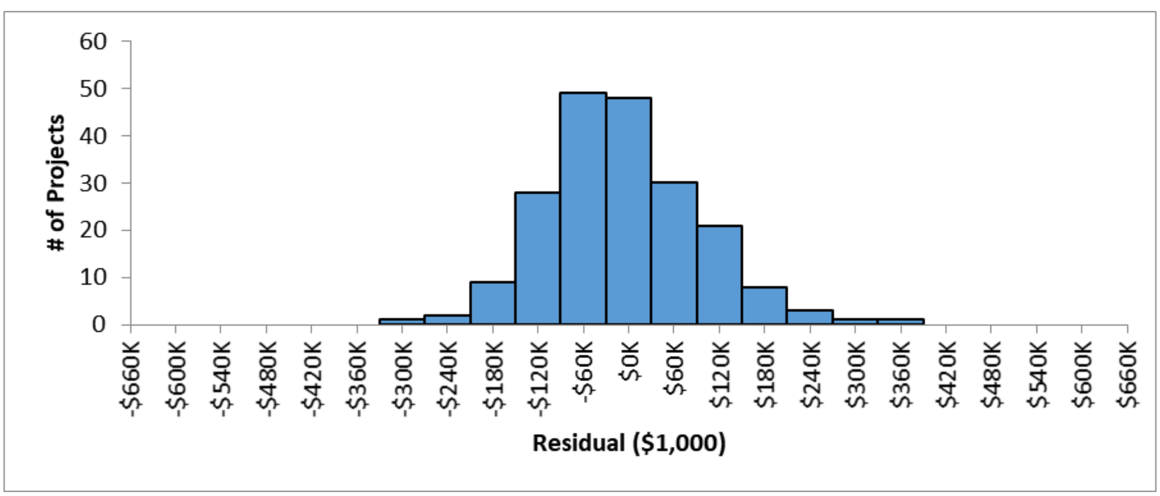

Figure 5.6: Residual error for *CONSTCOST* (with old bridge characteristics)

Table 5.1: Comparison between approaches (*CONSTCOST*)

	Approach 1 *actual* (Figure 5.4)	Approach 1 *predicted* (Figure 5.5)	Approach 2 (Figure 5.6)
Mean Residual Error	- $901.43	$130,564.19	- $22.04
Standard Deviation	$90,400	$1,298,800	$107,200
95% Confidence Interval	± $177,300	± $2,545,600	± $210,200

Table 5.2: Residual error outliers not displayed in histograms (*CONSTCOST*)

<table>
<tr><th></th><th>Below</th><th>Range Displayed</th><th>Above</th></tr>
<tr><td>Approach 1 (w/ actual) (5.4)</td><td>0</td><td rowspan="3">- $660,000 to $660,000</td><td>0</td></tr>
<tr><td>Approach 1 (w/ predicted) (5.5)</td><td>2</td><td>3</td></tr>
<tr><td>Approach 2 (5.6)</td><td>0</td><td>0</td></tr>
</table>

Based on Table 5.1, Approach 2 is the recommended approach for predicting *CONSTCOST*, indicating that although replacement bridge characteristics can be reasonably predicted from existing bridge characteristics, construction costs for the replacement bridge are best predicted from old bridge characteristics. Initially, Approach 1 appeared to have a narrower confidence interval and a lower standard deviation. Figure 5.4 shows residual error for Approach 1 when using actual new bridge characteristics. Using known new bridge characteristics for Approach 1 was helpful in seeing how well the cost prediction models were developed to fit the data, however estimates from Approach 1 use predicted new bridge characteristics. Compounding error from the characteristic predictions into the final cost prediction was evident in an increase in standard deviation and the widening of the confidence interval. The shifting of the mean residual error in the positive direction states that this approach will tend to generate cost estimates that are higher than the actual amount. The presence of the residual error outliers that are beyond the extents of the x-axis display range influenced the shifting of the mean error and widening of the confidence interval for this approach. Approach 2 had a standard deviation and confidence interval close to Approach 1 with actual old bridge characteristics. Approach 2 is also simpler to use since there is no need to estimate any bridge characteristics.

5.2.2 Right-Of-Way Cost

Histograms of residual error for predicting *ROWCOST* through Approach 1 with actual new bridge characteristics (Figure 5.7) and predicted new bridge characteristics (Figure 5.8) are shown below. Using old bridge characteristics to predict cost as part of Approach 2 created the distribution of residual error shown in Figure 5.9.

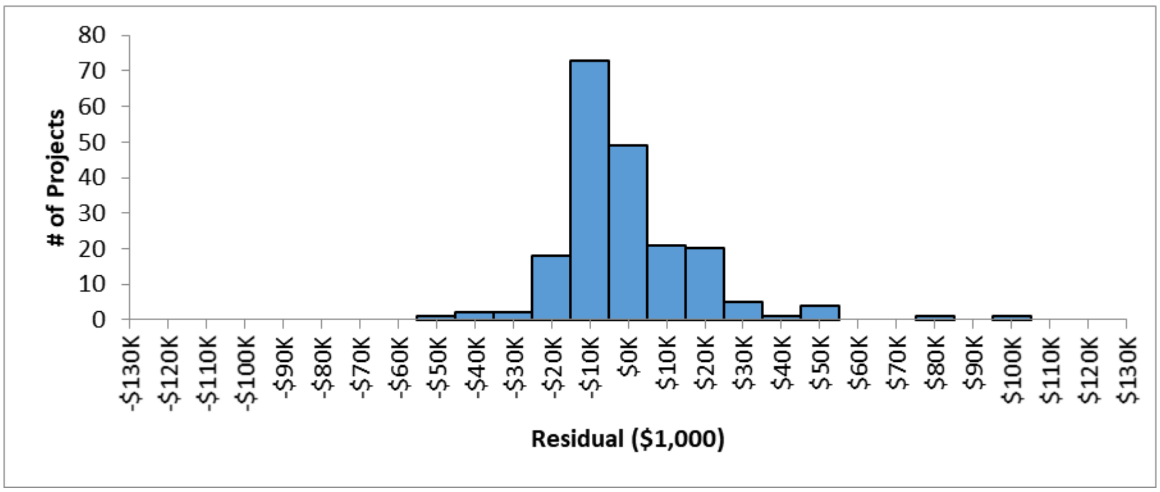

Figure 5.7: Residual error for *ROWCOST* (with actual new bridge characteristics)

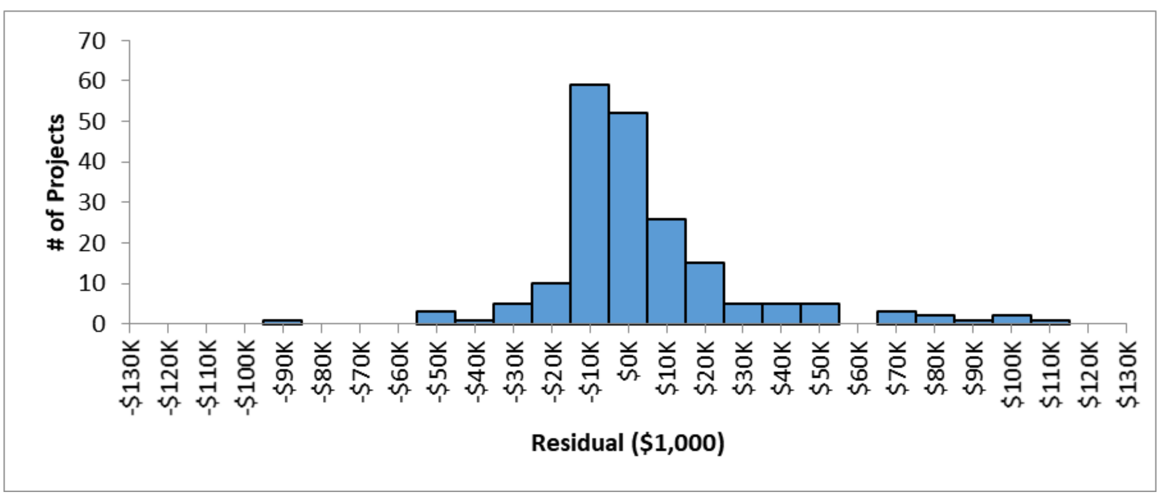

Figure 5.8: Residual error for *ROWCOST* (with predicted new bridge characteristics)

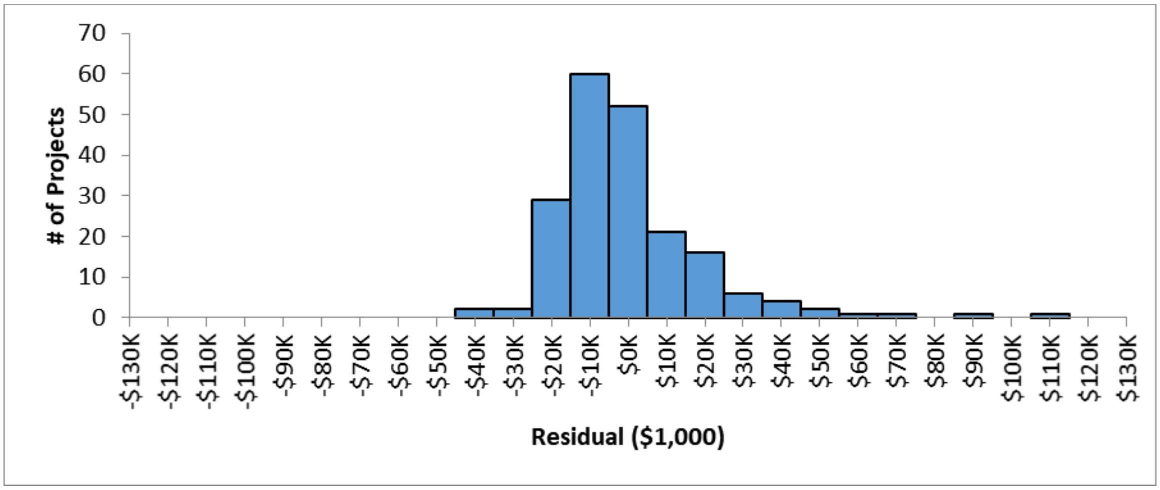

Figure 5.9: Residual error for *ROWCOST* (with old bridge characteristics)

Table 5.3: Comparison between approaches (*ROWCOST*)

	Approach 1 *actual* (Figure 5.7)	Approach 1 *predicted* (Figure 5.8)	Approach 2 (Figure 5.9)
Mean Residual Error	- $137.85	$22,612.51	- $0.78
Standard Deviation	$18,100	$197,100	$19,400
95% Confidence Interval	± $35,400	± $386,300	± $38,000

Table 5.4: Residual error outliers not displayed in histograms (*ROWCOST*)

	Below	Range Displayed	Above
Approach 1 (w/ actual) (5.7)	0	- $130,000 to $130,000	0
Approach 1 (w/ predicted) (5.8)	0		2
Approach 2 (5.9)	0		0

Despite Approach 1 having a slightly narrower confidence interval than Approach 2, the usage of predicted new bridge characteristics in Approach 1 widened the confidence interval by reasonably a large amount (approximately 10x). The shifting of the mean residual error for Approach 1 with predicted characteristics appeared to be driven by the two error outliers in Table 5.4 that had magnitudes greater than the display range of ±$130,000. These two outliers also influenced the widening of the confidence interval.

5.2.3 Engineering Cost

Histograms of residual error for predicting *ENGCOST* through Approach 1 with actual new bridge characteristics (Figure 5.10) and predicted new bridge characteristics (Figure 5.11) are shown below. Using old bridge characteristics to predict cost as part of Approach 2 created the distribution of residual error shown in Figure 5.12.

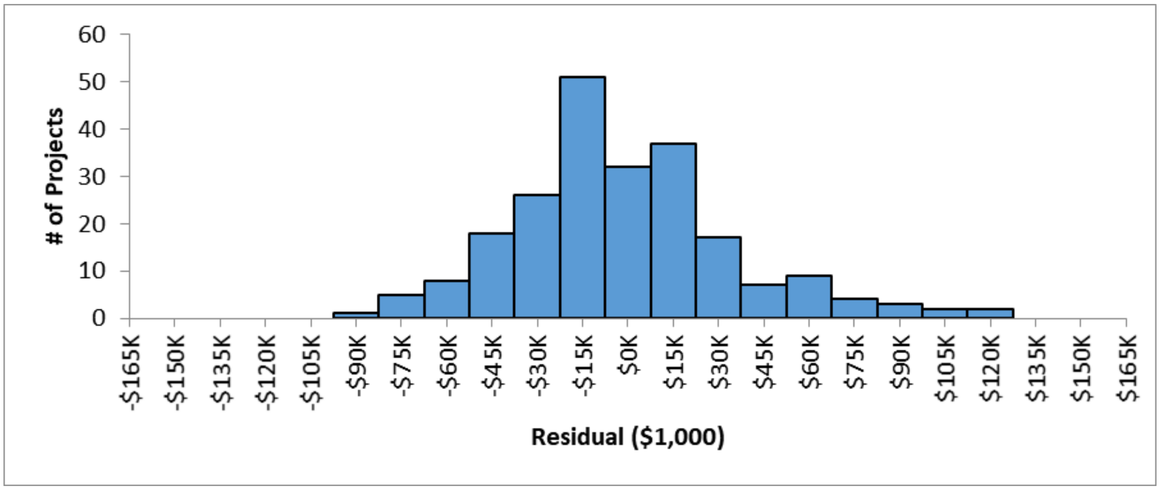

Figure 5.10: Residual error for *ENGCOST* (with actual new bridge characteristics)

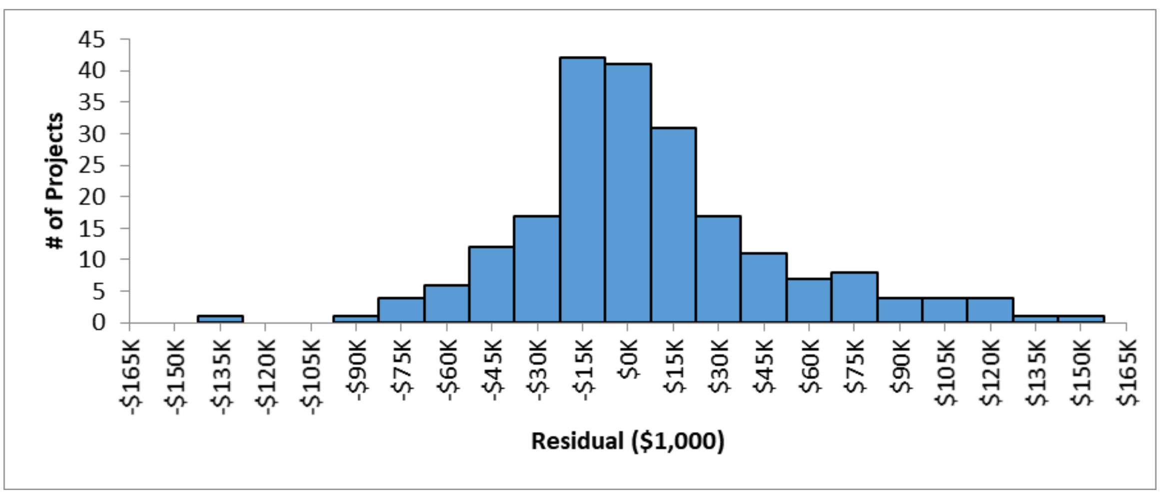

Figure 5.11: Residual error for *ENGCOST* (with predicted new bridge characteristics)

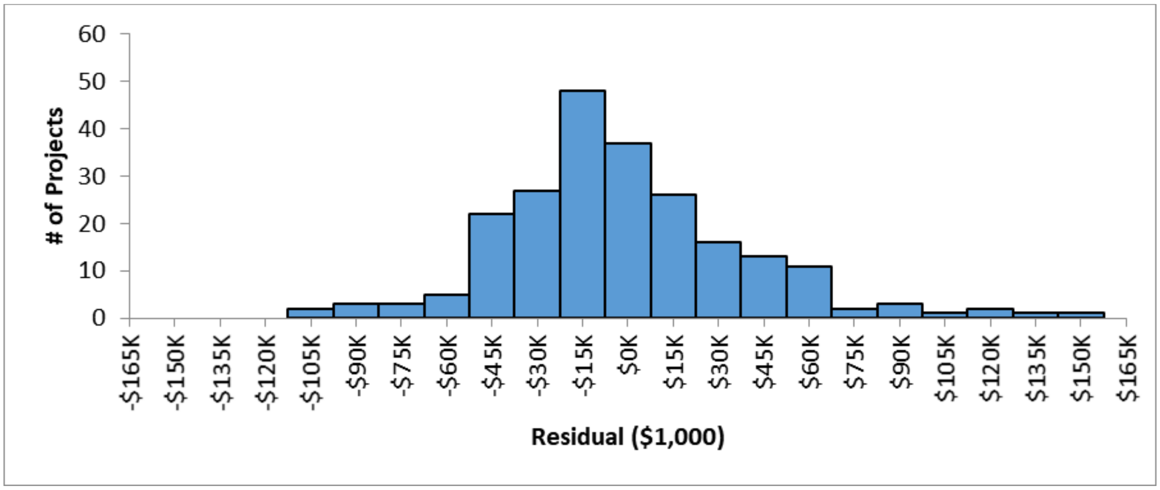

Figure 5.12: Residual error for *ENGCOST* (with old bridge characteristics)

Table 5.5: Comparison between approaches (*ENGCOST*)

	Approach 1 *actual* (Figure 5.10)	Approach 1 *predicted* (Figure 5.11)	Approach 2 (Figure 5.12)
Mean Residual Error	$324.44	$44,230.60	$14.10
Standard Deviation	$41,300	$590,200	$46,600
95% Confidence Interval	± $80,900	± $1,156,800	± $91,300

Table 5.6: Residual error outliers not displayed in histograms (*ENGCOST*)

	Below	Range Displayed	Above
Approach 1 (w/ actual) (5.10)	0	- $165,000 to $165,000	2
Approach 1 (w/ predicted (5.11)	2		10
Approach 2 (5.12)	0		1

Based on the results in Table 5.5, Approach 2 is the recommended method for predicting *ENGCOST*. Initially Approach 1 appeared to have the narrowest confidence interval when using actual new bridge characteristics. While actual characteristics were used to create the models for Approach 1, the Approach 1 cost models are intended to be used with estimated characteristics. Substituting predicted new bridge characteristics for the actual characteristics used to create the model resulted in a drastic widening of the 95% confidence interval and a shift in the mean residual error. The presence of twelve residual error values with magnitudes larger than $165,000 indicate that the shift in mean residual error and widening of the confidence interval were influenced by these atypical residual values. The positive value for Approach 1's mean residual error ($44,230.60) indicates that this approach, when used with predicted characteristics, will tend to generate under-predictions of *ENGCOST*. Approach 2 has a confidence interval that is only slightly wider than that of Approach 1 (when using actual new bridge characteristics) and uses a more streamlined modeling process that eliminates the need to predict new bridge characteristics.

5.2.4 Total Replacement Cost

Histograms of residual error for predicting *TOTCOST* through Approach 1 with actual new bridge characteristics (Figure 5.13) and predicted new bridge characteristics (Figure 5.14) are shown below. Using old bridge characteristics to predict cost as part of Approach 2 created the distribution of residual error shown in Figure 5.15.

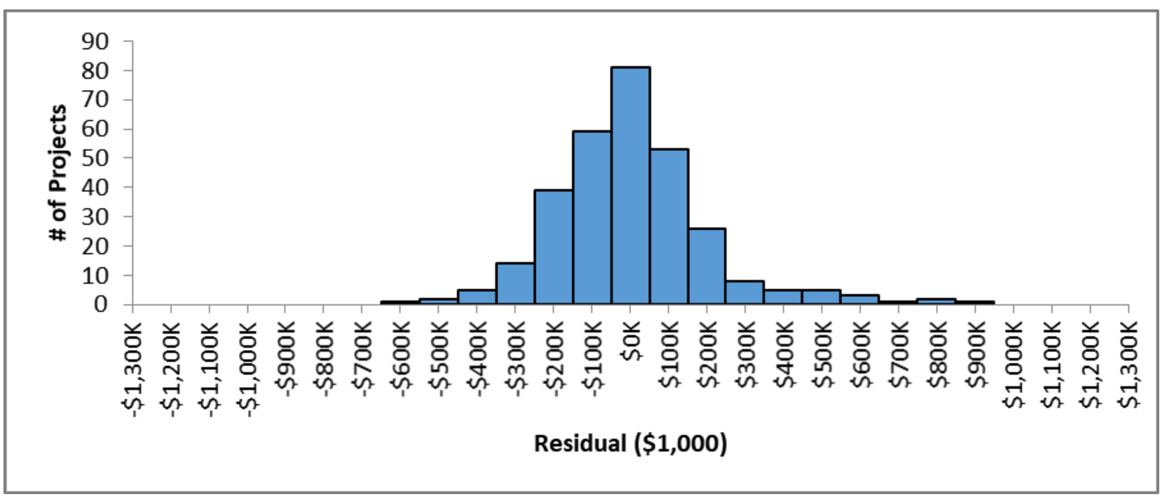

Figure 5.13: Residual error for *TOTCOST* (with actual new bridge characteristics)

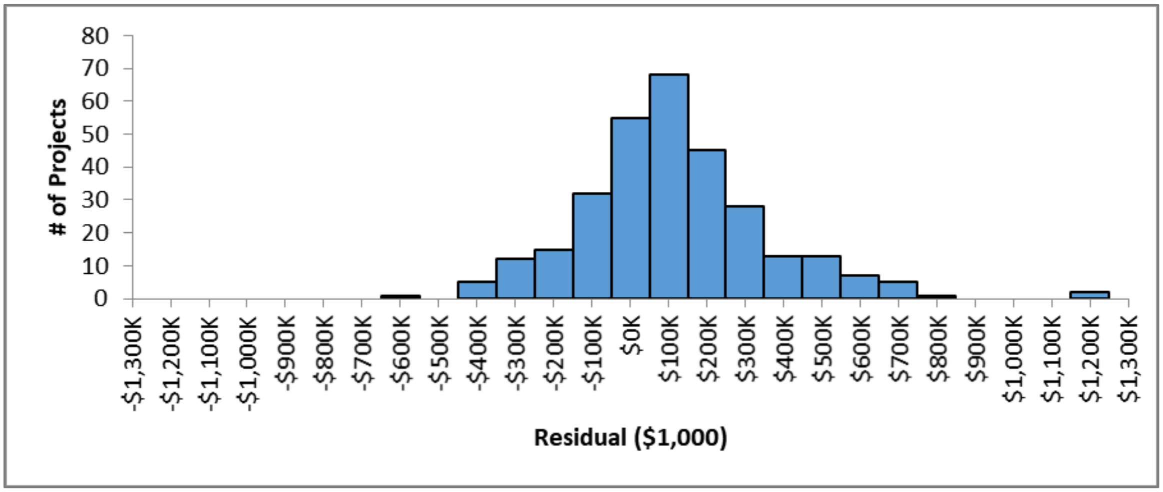

Figure 5.14: Residual error for *TOTCOST* (with predicted new bridge characteristics)

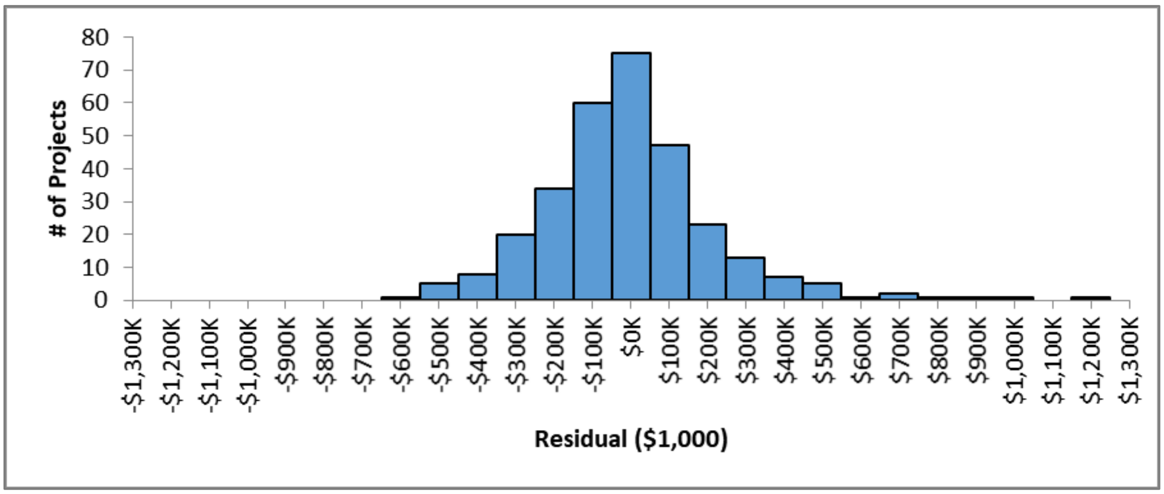

Figure 5.15: Residual error for *TOTCOST* (with old bridge characteristics)

Table 5.7: Comparison between approaches (*TOTCOST*)

	Approach 1 *actual* (Figure 5.13)	Approach 1 *predicted* (Figure 5.14)	Approach 2 (Figure 5.15)
Mean Residual Error	$26.68	$142,771.49	$20.00
Standard Deviation	$203,400	$777,200	$232,900
95% Confidence Interval	± $398,600	± $1,523,300	± $456,400

Table 5.8: Residual error outliers not displayed in histograms (*TOTCOST*)

	Below	Range Displayed	Above
Approach 1 (w/ actual) (5.13)	0	- $1,300,000 to $1,300,000	0
Approach 1 (w/ predicted) (5.14)	2		1
Approach 2 (5.15)	0		0

Based on the results in Table 5.7, Approach 2 is again the best prediction approach for *TOTCOST* when considering the 95% confidence interval. If actual new bridge characteristics were to be known for a future bridge project, using those values in Approach 1 would still generate predictions with a slightly wider confidence interval than with Approach 2. As seen with the other dependent cost variables, introducing predicted new bridge characteristics compounds further error into the final cost estimates, resulting in a relatively wide confidence interval range (increased by approximately 3x). The changes in mean residual error and confidence interval when using predicted bridge characteristics appeared to be influenced by the three residual error outliers seen in Table 5.8. Each of these outliers had a magnitude larger than the display range of ±$1,300,000, which could affect the mean residual error and confidence interval. Using Approach 1 as intended (with predicted new bridge characteristics) will tend to over-predict *TOTCOST*, as seen in the positive values for its mean residual error. The mean residual error values for Approach 1 (with actual characteristics) and Approach 2, while non-zero, are significantly smaller in magnitude in comparison to their respective standard deviations. Approach 2 for *TOTCOST* also possesses the inherent advantage of its simplicity, since there is no need for an intermediate prediction step as with Approach 1.

A comparison of the *TOTCOST* predictions using the current BMS unit cost method (Table 1.1) and the regression model developed from old bridge characteristics (Table 4.25) shows how the two methods predict costs differently for the same replacement projects (Figure 5.16 and Figure 5.17). Points on the red-dashed line in Figure 5.17 would denote equal predictions from both estimation methods for the same project. Since most of these points appear to be above the red line towards the positive yaxis, this confirms that the predictions from the regression model tend to be higher than the unit cost-based estimates for the same projects.

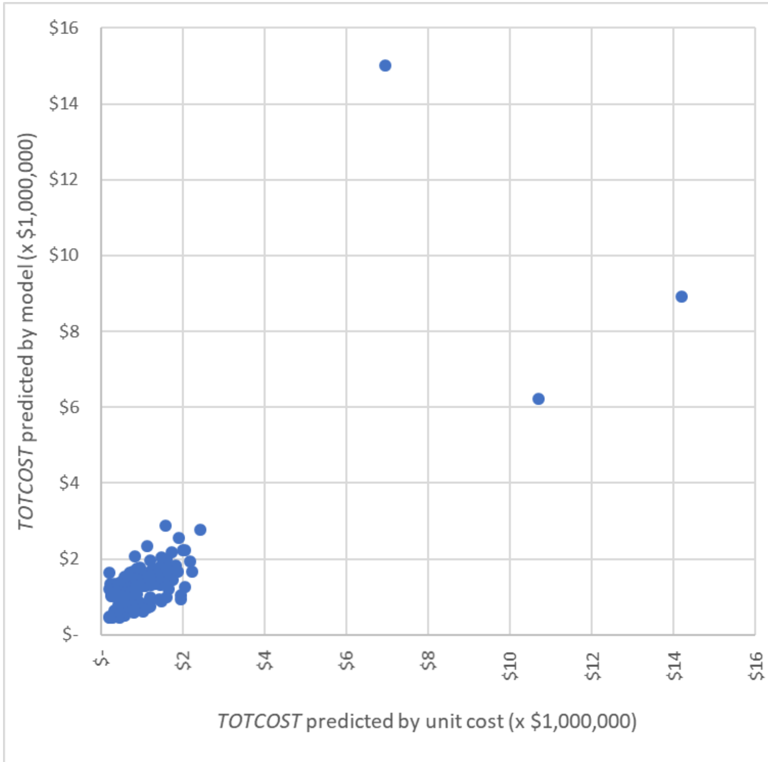

Figure 5.16: Plot of *TOTCOST* predictions by current versus updated methods

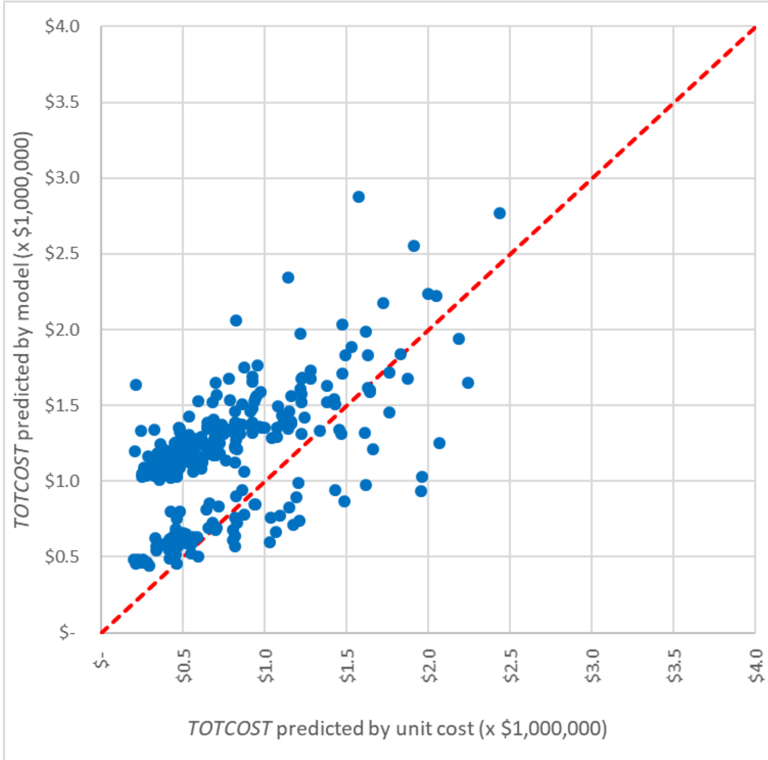

Figure 5.17: Plot of *TOTCOST* predictions by current and updated methods (zoomed-in)

5.3 Summary of Findings

The comparison between the new bridge characteristic-based cost predictions from Approach 1 and the old bridge characteristic-based cost predictions from Approach 2 showed that the Approach 1 models had a slight advantage over the Approach 2 models. The Approach 1 residuals were also calculated with the predicted new bridge characteristics instead of the actual values. The residuals from those histograms showed a significant widening of the confidence interval and a shift in the mean residual error away from zero. This shift was, in most cases, toward the positive direction on the x-axis, which signified a tendency for the model to over-predict costs and create positive residual values. It is speculated that error in each characteristic prediction model prediction from Approach 1 was compounded into the final cost estimate. The three predicted characteristics (*NBLEN*, *NBWID*, and *MAXSPAN2*) each appear several times in the

Approach 1 cost models as solitary variables, squares, cubes, or products with other predictors. The effects of any biases and error from the characteristic models would be additive and influence the final cost estimate.

The Approach 2 models for each of the cost variables had reasonable values for mean residual error and confidence intervals. The limited availability of component costs for all bridge replacement projects used in the project meant that the universal applicability of the component cost models is limited to TIP projects. The component cost values for each TIP bridge replacement project did not add up to the recorded total project cost, so the component cost prediction models cannot be used for aggregated total cost estimates. In terms of simplicity, the Approach 2 total cost model showed good mean residual error and a reasonable confidence interval for both 17BP and TIP bridge replacement projects.

With the models in Approach 2, there is the inherent possibility that future data will follow different trends than what is modeled in the equations presented in this chapter. This condition can be checked by performing additional residual analyses with newer datasets to observe shifts in mean residual error or widening of the confidence interval. Changing relationships between old bridge characteristics and new bridge costs may require that the Approach 2 models be periodically updated to retain model fidelity.

CHAPTER 6: CONCLUSIONS AND RECOMMENDATIONS

6.1 Conclusions

The current cost estimation method used by NCDOT considers deck area and bridge system classification for predicting bridge replacement cost. While conceptual-level estimates are expected to have degree of variability, NCDOT reported that their model typically failed to make accurate predictions for bridges on the high and low ends of the cost scale, possibly indicating that more location or bridge design factors need to be considered for the cost estimates. This work aimed to provide improved bridge replacement models for NCDOT by updating existing models with new data and by developing new models that consider additional predictor variables and more complex forms (such as models with quadratic and cubic terms and variable interactions). Direct approaches to model replacement costs using old bridge characteristics were explored, as well as approaches using an intermediate step – predicting new bridge characteristics from old bridge characteristics, then predicting replacement costs from these predicted new characteristics.

In replicating approaches proposed by Saito et al. (1991) and Abed-al-Rahim and Johnston (1995), the original models were updated in Minitab using the same predictor variables and a new bridge dataset. As presented in Chapter 3, two different datasets were used in this project: one for the characteristic prediction models and one for the cost prediction models, and both datasets were preconditioned to identify and filter out entries that did not represent typical bridge replacement projects. The original cost and characteristic prediction models created for NCDOT and INDOT were replicated by preserving the original predictor variables and using the newer datasets of NCDOT bridge replacement projects to determine the coefficient values and assess model performance. With the newer NCDOT bridge replacement data, the predictor variables utilized in the existing NCDOT and INDOT prediction models were not able to provide consistently accurate predictions for the dependent variables, as evidenced by reduced adjusted-R^2 and increased S values compared to the initial values reported in their parent literature.

It was presumed that the poorer performance of the original NCDOT and INDOT prediction models with the newer bridge data was due to (1) changes over time with respect to the design of new structures and with how replacement projects are managed and (2) the limited number of predictor variables utilized for each prediction model. To address these shortcomings, new models were developed. The process involved performing a stepwise regression for each dependent variable using the newer NCDOT bridge replacement datasets and a more robust list of possible predictor variables. From this list, the stepwise process automatically rejected predictor variables that did not have a strong enough statistical significant link to the dependent variable.

The characteristic prediction models that included quadratic and cubic terms and interactions between predictor variables had higher adjusted-R^2 and S values than their counterparts without variable interactions. These improvements were typically between 2.3-5.4% for goodness-of-fit (adjusted-R^2) and 0.43-2.33 for standard error (S). However, between each pair of models (with and without the quadratic and cubic terms and variable interactions) the difference between the goodness-of-fit and standard error statistics was judged to be not enough to justify introducing more complexity into the models with the variable interactions. The models without quadratic and cubic terms and variable interactions are recommended.

The four dependent variables for cost (right-of-way cost, engineering cost, construction cost, and total cost) were modeled with new bridge characteristics as predictors (Approach 1) or with old bridge characteristics as predictors (Approach 2). A comparison of the residual errors from each approach was done to observe trends in mean residual error and changes in the confidence interval. Since Approach 1 utilizes new bridge characteristics as predictors, the histogram of residuals was created twice for this approach: once with actual new bridge characteristics and once with predicted new bridge characteristics that were calculated from the recommended characteristic prediction models. Outside of this project, actual new bridge characteristics will not be available for use in cost prediction models. However, since actual new bridge characteristic data was available within the dataset, those known values were used to develop models which provided a performance baseline to help evaluate and validate the models produced using the old bridge characteristics.

While Approach 1 with the actual new bridge characteristics initially had a narrower confidence interval compared to Approach 2, the confidence interval became much wider with the predicted characteristics as predictors for cost. It is believed that error in the characteristic models compounded into the final estimate from Approach 1, sometimes increased the magnitude of the confidence interval up to 15 times, as seen with *CONSTCOST*. The increases in mean residual error magnitude and confidence interval range was driven by a handful of extreme residual error outliers in each analysis. Approach 2 offered a confidence interval comparable to Approach 1 with the actual variables but with less complexity since intermediate characteristic predictions were not required as part of this approach. The mean error for Approach 2 did not shift far from zero, indicating that the models did not have a strong tendency to under or over-predict.

Data on component costs was provided for a limited set of projects and the sums of the component costs did not appear to account for the entire total cost. As a result, the component cost prediction models developed from this dataset cannot be used to provide an accurate aggregated estimate for total cost. Further research into the unaccounted costs not captured in the HiCAMS-based historical cost dataset would be required before developing aggregated cost models.

In summary, the central datasets created from NCDOT bridge data were successfully used to develop models for predicting new bridge characteristics and replacement project costs for typical bridge replacement projects. One limitation of these models is that they were not designed for bridge replacement projects that would be considered atypical. Having a large set of potential predictor variables, considering interactions between the variables as additional predictors, and the handling of multiple categorical variables within a single model allowed for improved R^2 and narrower confidence intervals for the predictions. The recommended models for each predicted variable can be easily implemented into the NCDOT BMS and can be used to provide improved bridge replacement cost forecasting.

6.2 Recommendations

To retain prediction accuracy, the models recommended in this study should be updated periodically using data from recent bridge replacement projects. The recommendations outlined in this section, if implemented, could streamline the modeling process for any future model updates.

6.2.1 Inclusion of Bridge Structure Number in Contract Databases

When creating or updating a regression model, a dataset of the highest quality reasonably obtainable is essential for ensuring that the model can produce reliable and accurate estimates. The dataset creation process used in this project involved several additional steps because of the lack of a structure number in the provided contract database from HiCAMS. The various methods used to find structure numbers was successful for about 70% of the provided entries. If structure numbers had been provided with the HiCAMS-sourced dataset, a large number of the remaining 30% of entries could have been retained for use in the central cost dataset, potentially improving the models.

6.2.2 Identification of Basket Projects

The existence of basket projects (multiple bridge replacements let under the same contract) in the HiCAMS-sourced dataset presented an issue during the data conditioning phase. Basket project entries did not provide reliable cost information for those bridges and had to be filtered from the central cost dataset. Many of these basket projects were identified during data pre-processing. However there were cases of other basket projects being found in the central cost dataset later on in the project. Identifying contracts that cover more than one bridge replacement would make it easier to flag and remove these entries for future model updates.

6.2.3 Consistent Terminology for Text Entries

To prepare the datasets for use in Minitab, some qualitative variables had to be converted to quantitative scores. In other cases, groups within categorical variables had to be consolidated based on group size or perceived similarities between the groups. During this process, it was noted that many of these qualitative entries in the Network and Performance Masters were not phrased in a consistent manner. One single value would have several different spellings, capitalizations, or abbreviations. This complicated the data conditioning process and, during future model updates, could potentially cause extra entries to be erroneously filtered from the dataset. Using a consistent terminology or abbreviation convention for qualitative text fields in the Network Master and Performance Master would streamline this process.

6.2.4 Recording Project Cost Components in Databases

Bridge replacement projects that included component cost information could not be used to develop aggregated total cost models. This was because the values for each component cost did not add up to the recorded total cost amount. This is possibly due to additional project costs (demolition, inspection, miscellaneous) that were not recorded in the historical cost dataset. Structuring the component cost data so that it sums up to the total project cost value could allow for future development of aggregated total cost models as an alternative to predicting total cost with one model.

REFERENCES

Abed-Al-Rahim, I., and Johnston, D. (1995). “Bridge Replacement Cost Analysis Procedures.” Transportation Research Record: Journal of the Transportation Research Board, 1490, 23–31.

Aleithawe, I. (2017). “Right of Way Acquisition Workflow Model to Reduce Acquisition Duration in Mississippi.” Journal of Legal Affairs and Dispute Resolution in Engineering and Construction. 9(3).

Aljadhai, S., and Abraham, D. (2016). “Evaluation of Proposed Highway Routes Based on the Existing Utilities: A Cost Estimate Framework.” Construction Research Congress 2016, American Society of Civil Engineers, 2060–2069.

Al-Subhi, K., Johnston, D., and Farid, F. (1989). “Optimizing System-Level Bridge Maintenance, Rehabilitation, and Replacement Decisions.” Center for Transportation Studies, Department of Civil Engineering, North Carolina State University.

Andrews, L. and Phillips, R. (2003). “Mathematical techniques for engineers and scientists.” Bellingham, Wash: SPIE.

Behmardi, B., Doolen, T., and Winston, H. (2015). “Comparison of Predictive Cost Models for Bridge Replacement Projects.” Journal of Management in Engineering, 31(4).

Blischke, W. R. and Murthy, D. N. P. (2000). “Reliability: modeling, prediction, and optimization.” New York: Wiley.

Cameron, A. (2009). “Excel 2007: Histogram.” Available at: http://cameron.econ.ucdavis.edu/excel/ex11histogram.html. Accessed on: November 24, 2017.

Chang-Albitres, C., Feldman, R., Krugler, P. E., and Ibarra, I. (2014). “Simulation Model to Prioritize Right-of-Way Acquisitions.” Journal of Infrastructure Systems, 20(1).

Chen, C. and Johnston, D. W. (1987). “Bridge management under a level of service concept providing optimum improvement action, time, and budget prediction.” Raleigh, N.C.: Center for Transportation Engineering Studies, Dept. of Civil Engineering, N.C. State University.

Chou, J.-S., Wang, L., Chong, W. K., and O’Connor, J. (2005). “Preliminary cost estimates using probabilistic simulation for highway bridge replacement projects.” American Society of Civil Engineers, San Diego, CA.

Collier, K. (1984). “Estimating construction costs: A conceptual approach.” Reston, VA: Reston Pub. Co.

Cramer, D. and Howitt, D. (2004). “The SAGE Dictionary of Statistics.” SAGE Publications.

Cross, D. (1999). “Report from the Fidelity Implementation Study Group.” Spring Simulation Interoperability Workshop. Orlando, FL.

Dodge, Y. and Marriott, F. (2003). “The Oxford dictionary of statistical terms.” Oxford: Oxford University Press.

Dowdy, S. and Wearden, S. (1991). “Statistics for research.” New York: Wiley.

Du Toit, S. H. C., A. G. W. Steyn, and R. H. Stumpf. (1986). “Graphical exploratory data analysis.” New York: Springer-Verlag.

Engineering News Record. (2017). “Construction Economics.” Available at: https://enr.com/economics. Date Accessed: December 4, 2017.

FHWA. (1995). “Recording and Coding Guide for the Structure Inventory and Appraisal of the Nation’s Bridges.”

FHWA. (2017). “National Highway Construction Cost Index.” Available at: https://fhwa.dot.gov/policy/otps/nhcci. Date Accessed: December 4, 2017.
Foster, N. (1972). “Construction estimates from take-off to bid.” New York: McGraw-Hill.

Gould, F. (2005). “Managing the construction process: estimating, scheduling, and project control.” Upper Saddle River, N.J.: Pearson/Prentice Hall.

Gransberg, D., Jeong, H. D., Craigie, E., Rueda-Benavides, J. A., and Shrestha, K. J. (2016). Estimating Highway Preconstruction Service Costs, Volume 1: Guidebook. Transportation Research Board.

Hearn, G. (2012). “Deterioration and Cost Information for Bridge Management.”

Heiner, J. D., and Kockelman, K. M. (2005). “Costs of Right-of-Way Acquisition: Methods and Models for Estimation.” Journal of Transportation Engineering, 131(3), 193–204.

Hollar, D. A., Rasdorf, W., Liu, M., Hummer, J. E., Arocho, I., and Hsiang, S. M. (2013). “Preliminary Engineering Cost Estimation Model for Bridge Projects.” Journal of Construction Engineering and Management, 139(9), 1259–1267.

Karen. (2017). “Assessing the fit of regression models.” Available at: http://www.theanalysisfactor.com/assessing-the-fit-of-regression-models/. Date accessed: November 24, 2017.

Kassler, E. B. (1949). “The architecture of bridges.” New York: Museum of Modern Art.

Kyte, C., Perfater, M., Haynes, S. and Lee, H. (2004). “Developing and Validating a Highway Construction Project Cost Estimation Tool.” Virginia Transportation Research Council.

Levy, S. (2006). “Project Management in Construction.” McGraw-Hill.

Nau, R. (2009). “Testing the assumptions of linear regression.” Available at: http://people.duke.edu/~rnau/testing.htm#homoscedasticity. Date Accessed: March 27, 2017.

NCDOT. (2017). “North Carolina Bridge Information.” Available at: https://www.ncdot.gov/projects/ncbridges/. Date Accessed: November 24, 2017.

Organization for Economic Co-operation and Development. (1992). “Bridge management: report.” Paris: Organization for Economic Co-operation and Development.

Penn State Eberly College of Science. (2017). “2-1 Prediction Accuracy.” Available at: https://onlinecourses.science.psu.edu/stat857/node/160. Date Accessed: November 13, 2017.

Peurifoy, R. L. (1975). *Estimating construction costs*. New York: McGraw-Hill.

Purvis, R. (1994). *Underwater bridge maintenance and repair*. Washington, D.C.: National Academy Press.

RS Means. (2017). “Historical Cost Indexes.” Available at: https://rsmeansonline.com/references/unit/refpdf/hci.pdf. Date Accessed: December 4, 2017.

Saito, M., Sinha, K. C., and Anderson, V. L. (1991). “Statistical models for the estimation of bridge replacement costs.” Transportation Research Part A: General, 25(6), 339–350.

Schexnayder, C., Weber, S., and Fiori, C. (2003). “Project Cost Estimating: A synthesis of highway practice.” American Association of State Highway and Transportation Officials.

Tabachnik, B. and Fidell, L. (2006). “Using Multivariate Statistics.” Pearson Education.

Wahls, H. (1990). “Design and construction of bridge approaches.” Washington, D.C.: Transportation Research Board, National Research Council.

Wilmot, C. G., and Cheng, G. (2003). “Estimating Future Highway Construction Costs.” Journal of Construction Engineering and Management, 129(3), 272–279.

Wright, M., and Williams, T. (2001). "Using bidding statistics to predict completed construction cost". *The Engineering Economist.* 46 (2): 114-128.

WSDOT. (2008). “Cost Estimating Manual for WSDOT Projects.”

APPENDIX A: SUPPLEMENTAL MODEL INFORMATION

	A	B	C	D	E	F	G	H	I	J	K
1	Substructure Condition	Deck Condition	Superstructure Condition	Divisio	Tier ID	County	Structure Type	Structure Type	Facility Carried	Intersected Features	Structure No.
2	N	N	N	12	Sub-Regio	35 - GASTC	C	1 - Reinforced con	SR1103	MCGILL BRANCH	3503
3	6	5	6	4	Statewide	95 - WAYN	0	0 - Bridge	US70 WBL	WALNUT CREEK	9500
4	N	N	N	5	Sub-Regio	91 - WAKE	C	1 - Reinforced con	SR3009	RICHLAND CREEK	9109
5	N	N	N	5	Sub-Regio	91 - WAKE	C	1 - Reinforced con	SR1830	CREEK	9106
6	N	N	N	10	Statewide	59 - MECK	S	14 - Overhead Sig	OVERHEAD SIGN	I485 NBL	5911
7	N	N	N	7	Statewide	40 - GUILF	L	8 - Cantilever Sign	CANTILEVER SIGN	I40 EBL	4010
8	N	N	N	10	Statewide	59 - MECK	S	14 - Overhead Sig	OVERHEAD SIGN	I85 SBL	5912
9	N	N	N	10	Statewide	59 - MECK	L	8 - Cantilever Sign	CANTILEVER SIGN	I77 NBL	5912
10	7	7	7	7	Statewide	40 - GUILF	0	0 - Bridge	I85 NBL	I74, US311	4009
11	N	N	N	7	Statewide	40 - GUILF	S	14 - Overhead Sig	OVERHEAD SIGN	I85 SBL	4009
12	N	N	N	5	Sub-Regio	38 - GRAN	P	12 - Pipe culvert,	SR1147	CREEK OFF TAR RIVER	3802
13	7	7	7	14	Sub-Regio	87 - TRANS	0	0 - Bridge	SR1535	LITTLE RIVER	8700
14	N	N	N	6	Statewide	25 - CUMB	C	1 - Reinforced con	I95	CREEK	2502
15	N	N	N	6	Regional	23 - COLU	C	1 - Reinforced con	NC87	WAYMANS CREEK	2303
16	N	N	N	9	Sub-Regio	29 - DAVIE	P	12 - Pipe culvert,	SR1147	LITTLE CREEK	2900
17	N	N	N	7	Sub-Regio	0 - ALAMA	P	12 - Pipe culvert,	SR1005	PRONG CANE CREEK	0001
18	N	N	N	7	Sub-Regio	0 - ALAMA	P	12 - Pipe culvert,	SR1615	TOMS CREEK	0001
19	8	8	8	14	Sub-Regio	43 - HAYW	0	0 - Bridge	BOYD AVENUE	RICHLAND CREEK	4302
20	5	5	5	3	Sub-Regio	81 - SAMP	0	0 - Bridge	SR1811	LITTLE COHARIE CR	8103
21	5	7	6	13	Sub-Regio	10 - BUNC	0	0 - Bridge	SR2802	BROAD RIVER	1001
22	N	N	N	8	Regional	61 - MONT	R	13 - Structure owr	A,C & WEST RR	NC109	6100
23	N	N	N	3	Sub-Regio	9 - BRUNS	P	12 - Pipe culvert,	SR1335	WET ASH SWAMP S. PR.	0900
24	N	N	N	7	Regional	40 - GUILF	R	13 - Structure owr	RAILROAD OVERP	SR4762	4001
25	7	7	7	10	Regional	59 - MECK	0	0 - Bridge	SR1625	I85	5900

Figure A.1: Screenshot of 2017 NCDOT Network Master

	A	B	C	D	E	F	G	H	I	J
1	Contract Number	Str. No.	TIP Number	Physical	County	County	County	Contrac	Contract Description	Contract Comments
2	C202781	990031		13	Yancey	99	99 - Yance	Other	BRIDGE MAINTENANCE AND REPAIR	Grading, drainage, paving, a
3	C202912	100322		13	Buncombe	10	10 - Buncc	Other	BRIDGE PRESERVATION	
4	C202915	#N/A		1	Bertie	7	7 - Bertie	Other	BRIDGE PRESERVATION	
5	C202915	#N/A		1	Bertie	7	7 - Bertie	Other	BRIDGE PRESERVATION	
6	C202915	#N/A		1	Bertie	7	7 - Bertie	Other	BRIDGE PRESERVATION	
7	C202915	#N/A		1	Bertie	7	7 - Bertie	Other	BRIDGE PRESERVATION	
8	C202915	#N/A		1	Bertie	7	7 - Bertie	Other	BRIDGE PRESERVATION	
9	C202915	570014		1	Martin	57	57 - Marti	Other	BRIDGE PRESERVATION	
10	C202915	570012		1	Martin	57	57 - Marti	Other	BRIDGE PRESERVATION	
11	C202915	570042		1	Martin	57	57 - Marti	Other	BRIDGE PRESERVATION	
12	C202915	#N/A		1	Martin	57	57 - Marti	Other	BRIDGE PRESERVATION	
13	C202915	#N/A		1	Martin	57	57 - Marti	Other	BRIDGE PRESERVATION	
14	C202921	570011		1	Martin	57	57 - Marti	Design Bu	DESIGN BUILD	
15	C202923	160102		7	Caswell	16	16 - Caswe	Design Bu	DESIGN BUILD	
16	C202923	#N/A		7	Guilford	40	40 - Guilfo	Design Bu	DESIGN BUILD	
17	C202923	#N/A		7	Orange	67	67 - Orang	Design Bu	DESIGN BUILD	
18	C202923	#N/A		7	Rockingha	78	78 - Rocki	Design Bu	DESIGN BUILD	
19	C202924	190018		14	Cherokee	19	19 - Chero	Design Bu	DESIGN BUILD	
20	C202924	#N/A		14	Jackson	49	49 - Jacksc	Design Bu	DESIGN BUILD	
21	C202924	#N/A		14	Macon	55	55 - Maco	Design Bu	DESIGN BUILD	
22	C202924	860134		14	Swain	86	86 - Swain	Design Bu	DESIGN BUILD	

Figure A.2: Screenshot of 17BP/TIP project information from HiCAMS. (Structure number field was added for this project)

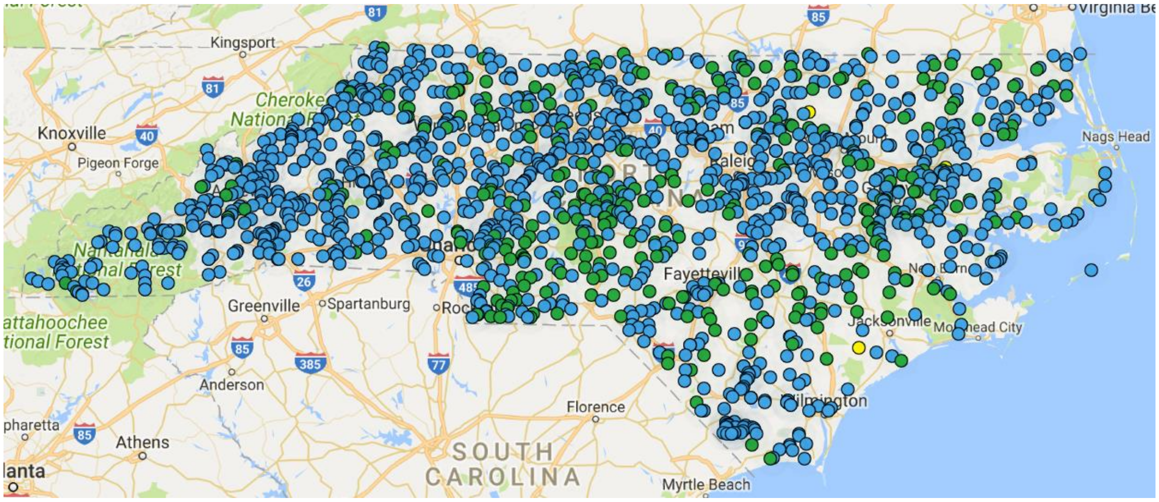

Figure A.3: Locations of replaced bridges in cost dataset (yellow), characteristic dataset (blue) or in both datasets (green).

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
21.1010	90.44%	90.14%	80.84%

Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	112.9	21.4	5.28	0.000	
OBLEN	0.459	0.128	3.59	0.000	183.60
OBWID	-4.244	0.789	-5.38	0.000	88.16
MAXSPAN1	1.131	0.260	4.35	0.000	45.03
WATERDEPTH	4.78	1.52	3.14	0.002	57.36
BRIDGEAGE	-1.794	0.330	-5.43	0.000	40.99
CTB	3.741	0.829	4.51	0.000	94.32
ADTr	0.00839	0.00117	7.19	0.000	128.87
OBLEN^2	0.002615	0.000317	8.24	0.000	262.51
CTB^2	-0.1212	0.0272	-4.45	0.000	262.75
APPWID^2	-0.0623	0.0260	-2.39	0.017	223.69
ADTr^2	0.000000	0.000000	6.80	0.000	259.12
OBLEN^3	-0.000002	0.000000	-6.14	0.000	106.83
WATERDEPTH^3	0.00279	0.00138	2.01	0.044	4.90
BRIDGEAGE^3	0.000038	0.000016	2.36	0.018	14.33
CTB^3	0.001982	0.000260	7.61	0.000	120.02
ADTr^3	-0.000000	0.000000	-6.72	0.000	126.33
OBLEN*BRIDGEAGE	-0.00236	0.00105	-2.24	0.025	48.93
OBLEN*CTB	-0.01141	0.00243	-4.69	0.000	138.15
OBLEN*APPWID	0.00938	0.00306	3.06	0.002	104.65
OBLEN*ADTr	-0.000029	0.000003	-9.28	0.000	39.53
OBWID*MAXSPAN1	0.02458	0.00543	4.52	0.000	35.68
OBWID*BRIDGEAGE	0.02222	0.00932	2.38	0.017	61.46
OBWID*APPWID	0.0782	0.0288	2.72	0.007	380.76
OBWID*ADTr	-0.000117	0.000027	-4.40	0.000	126.65
MAXSPAN1*WATERDEPTH	0.1099	0.0246	4.47	0.000	28.97
MAXSPAN1*CTB	-0.0375	0.0105	-3.57	0.000	136.10
MAXSPAN1*APPWID	-0.0509	0.0106	-4.80	0.000	73.60
WATERDEPTH*CTB	-0.1788	0.0514	-3.48	0.001	40.33
WATERDEPTH*APPWID	-0.2341	0.0710	-3.29	0.001	56.90
WATERDEPTH*ADTr	0.000332	0.000070	4.73	0.000	5.87
BRIDGEAGE*CTB	0.03139	0.00923	3.40	0.001	50.36
BRIDGEAGE*APPWID	0.0501	0.0106	4.74	0.000	51.12
BRIDGEAGE*ADTr	-0.000081	0.000017	-4.68	0.000	76.54
CTB*ADTr	0.000045	0.000012	3.93	0.000	8.88
REGION					
Piedmont	6.54	1.71	3.81	0.000	2.12
Coastal	10.82	2.26	4.80	0.000	3.64
CROSSINGTYPE					
Waterway	-32.72	7.64	-4.28	0.000	6.19
SPAN1					
2	8.86	2.78	3.18	0.001	3.58
3	12.17	3.87	3.14	0.002	9.82
4	15.18	5.13	2.96	0.003	8.91
5	21.27	6.08	3.50	0.000	11.98

Figure A.4: Minitab model summary for *NBLEN*, with variable interactions

```
NBLEN  =  112.9 + 0.459 OBLEN - 4.244 OBWID + 1.131 MAXSPAN1 + 4.78 WATERDEPTH
          - 1.794 BRIDGEAGE + 3.741 CTB + 0.00839 ADTr + 0.002615 OBLEN^2
          - 0.1212 CTB^2 - 0.0623 APPWID^2 + 0.000000 ADTr^2
          - 0.000002 OBLEN^3 + 0.00279 WATERDEPTH^3 + 0.000038 BRIDGEAGE^3
          + 0.001982 CTB^3 - 0.000000 ADTr^3 - 0.00236 OBLEN*BRIDGEAGE
          - 0.01141 OBLEN*CTB + 0.00938 OBLEN*APPWID - 0.000029 OBLEN*ADTr
          + 0.02458 OBWID*MAXSPAN1 + 0.02222 OBWID*BRIDGEAGE
          + 0.0782 OBWID*APPWID - 0.000117 OBWID*ADTr
          + 0.1099 MAXSPAN1*WATERDEPTH - 0.0375 MAXSPAN1*CTB
          - 0.0509 MAXSPAN1*APPWID - 0.1788 WATERDEPTH*CTB
          - 0.2341 WATERDEPTH*APPWID + 0.000332 WATERDEPTH*ADTr
          + 0.03139 BRIDGEAGE*CTB + 0.0501 BRIDGEAGE*APPWID
          - 0.000081 BRIDGEAGE*ADTr + 0.000045 CTB*ADTr + 0.0 REGION_Mountains
          + 6.54 REGION_Piedmont + 10.82 REGION_Coastal
          + 0.0 CROSSINGTYPE_Non-waterway - 32.72 CROSSINGTYPE_Waterway
          + 0.0 SPAN1_1 + 8.86 SPAN1_2 + 12.17 SPAN1_3 + 15.18 SPAN1_4
          + 21.27 SPAN1_5
```

Figure A.4: Minitab model summary for *NBLEN*, with variable interactions (continued)

```
Model Summary

      S    R-sq  R-sq(adj)  R-sq(pred)
23.4326  87.99%     87.84%      87.34%

Coefficients

Term                                  Coef   SE Coef  T-Value  P-Value    VIF
Constant                         4.0058E+1      6.56     6.11    0.000
OBLEN                            8.3879E-1    0.0289    29.07    0.000   7.60
MAXSPAN1                         5.2909E-1    0.0955     5.54    0.000   4.91
WATERDEPTH                       1.8584E+0     0.273     6.81    0.000   1.50
REGION
  Piedmont                       9.2727E+0      1.77     5.23    0.000   1.84
  Coastal                        9.8538E+0      2.23     4.43    0.000   2.87
FUNCTCLASS
  Major Collector                4.2494E+0      2.04     2.08    0.038   1.10
  Minor Arterial                 7.4988E+0      2.99     2.51    0.012   1.18
  Principal Arterial/Inter.      1.2205E+1      4.40     2.78    0.006   1.23
CROSSINGTYPE
  Waterway                      -1.4446E+1      4.04    -3.57    0.000   1.41
SUPERSTRMAT
  Steel                         -2.0860E+1      5.78    -3.61    0.000  19.66
  Timber                        -1.9876E+1      5.77    -3.44    0.001  15.08
SUPERSTRTYPE
  Other (Not Girder/beam sys)   -2.1740E+1      5.68    -3.83    0.000  10.22
DECKGEOMAPP
  UNACCEPTABLE                   6.6421E+0      1.42     4.68    0.000   1.23
SPAN1
  2                              7.0231E+0      2.30     3.05    0.002   1.98
  3                              6.8442E+0      2.53     2.71    0.007   3.39
  4                              9.2313E+0      3.42     2.70    0.007   3.22
  5                              1.3890E+1      4.34     3.20    0.001   4.94



Regression Equation

 NBLEN   =  40.058 + 0.83879 OBLEN + 0.52909 MAXSPAN1 + 1.8584 WATERDEPTH
            + 0.0 REGION_Mountains + 9.2727 REGION_Piedmont
            + 9.8538 REGION_Coastal + 0.0 FUNCTCLASS_Local/Minor Collector
            + 4.2494 FUNCTCLASS_Major Collector
            + 7.4988 FUNCTCLASS_Minor Arterial
            + 12.205 FUNCTCLASS_Principal Arterial/Interstate
            + 0.0 CROSSINGTYPE_Non-waterway - 14.446 CROSSINGTYPE_Waterway
            + 0.0 SUPERSTRMAT_Concrete - 20.860 SUPERSTRMAT_Steel
            - 19.876 SUPERSTRMAT_Timber + 0.0 SUPERSTRTYPE_Girder/Beam System
            - 21.740 SUPERSTRTYPE_Other type + 0.0 DECKGEOMAPP_ACCEPTABLE
            + 6.6421 DECKGEOMAPP_UNACCEPTABLE + 0.0 SPAN1_1 + 7.0231 SPAN1_2
            + 6.8442 SPAN1_3 + 9.2313 SPAN1_4 + 13.890 SPAN1_5
```

Figure A.5: Minitab model summary for *NBLEN*, without variable interactions

```
Model Summary
      S    R-sq  R-sq(adj)  R-sq(pred)
 3.72427  78.84%     78.41%      75.92%
Coefficients
 Term                                   Coef   SE Coef  T-Value  P-Value     VIF
 Constant                              18.42      1.27    14.52    0.000
 CTB                                   0.896     0.165     5.45    0.000   99.78
 ADTr                               0.001500  0.000208     7.20    0.000   59.07
 OBWID^2                             0.01685   0.00181     9.32    0.000   60.67
 MAXSPAN1^2                        -0.002444  0.000601    -4.06    0.000   73.98
 WATERDEPTH^2                        -0.0503    0.0142    -3.54    0.000   36.78
 CTB^2                              -0.03680   0.00688    -5.35    0.000  265.87
 OBLEN^3                           -0.000000  0.000000    -4.57    0.000    6.66
 OBWID^3                           -0.000116  0.000027    -4.29    0.000   75.70
 MAXSPAN1^3                         0.000013  0.000004     3.43    0.001   40.57
 WATERDEPTH^3                       0.001937  0.000425     4.56    0.000   14.83
 CTB^3                              0.000458  0.000103     4.46    0.000  123.52
 APPWID^3                          -0.000089  0.000032    -2.81    0.005   37.33
 ADTr^3                            -0.000000  0.000000    -5.83    0.000    5.51
 OBLEN*OBWID                        0.001798  0.000270     6.65    0.000   33.08
 OBLEN*BRIDGEAGE                   -0.000568  0.000124    -4.57    0.000   22.20
 OBWID*CTB                          -0.02734   0.00327    -8.35    0.000   61.80
 OBWID*ADTr                         0.000020  0.000007     2.94    0.003  117.78
 MAXSPAN1*BRIDGEAGE                 0.002129  0.000472     4.51    0.000   16.86
 WATERDEPTH*BRIDGEAGE                0.00697   0.00194     3.59    0.000   11.74
 BRIDGEAGE*ADTr                    -0.000011  0.000003    -3.82    0.000   38.11
 CTB*APPWID                          0.03008   0.00437     6.88    0.000   78.38
 APPWID*ADTr                       -0.000015  0.000007    -2.20    0.028   95.95
 FUNCTCLASS
   Major Collector                     0.774     0.354     2.19    0.029    1.31
   Minor Arterial                      1.989     0.578     3.44    0.001    1.66
   Principal Arterial/Interstate      -1.731     0.923    -1.88    0.061    2.21
 CROSSINGTYPE
   Waterway                            -3.42      1.08    -3.16    0.002    4.16
 DECKGEOMAPP
   UNACCEPTABLE                        1.215     0.273     4.45    0.000    1.79
Regression Equation

 NBWID  =  18.42 + 0.896 CTB + 0.001500 ADTr + 0.01685 OBWID^2
           - 0.002444 MAXSPAN1^2 - 0.0503 WATERDEPTH^2 - 0.03680 CTB^2
           - 0.000000 OBLEN^3 - 0.000116 OBWID^3 + 0.000013 MAXSPAN1^3
           + 0.001937 WATERDEPTH^3 + 0.000458 CTB^3 - 0.000089 APPWID^3
           - 0.000000 ADTr^3 + 0.001798 OBLEN*OBWID - 0.000568 OBLEN*BRIDGEAGE
           - 0.02734 OBWID*CTB + 0.000020 OBWID*ADTr
           + 0.002129 MAXSPAN1*BRIDGEAGE + 0.00697 WATERDEPTH*BRIDGEAGE
           - 0.000011 BRIDGEAGE*ADTr + 0.03008 CTB*APPWID
           - 0.000015 APPWID*ADTr + 0.0 FUNCTCLASS_Local/Minor Collector
           + 0.774 FUNCTCLASS_Major Collector + 1.989 FUNCTCLASS_Minor Arterial
           - 1.731 FUNCTCLASS_Principal Arterial/Interstate
           + 0.0 CROSSINGTYPE_Non-waterway - 3.42 CROSSINGTYPE_Waterway
           + 0.0 DECKGEOMAPP_ACCEPTABLE + 1.215 DECKGEOMAPP_UNACCEPTABLE
```

Figure A.6: Minitab model summary for *NBWID* (with variable interactions)

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
4.15869	73.44%	73.08%	*

Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	-1.3899E+1	4.72	-2.95	0.003	
OBLEN	-1.8451E-2	0.00509	-3.62	0.000	7.10
OBWID	7.0895E-1	0.0385	18.39	0.000	3.86
MAXSPAN1	9.1028E-2	0.0158	5.75	0.000	3.82
WATERDEPTH	1.5516E-1	0.0447	3.47	0.001	1.26
APPWID	1.4022E-1	0.0451	3.11	0.002	3.20
ADTr	7.4050E-4	0.000051	14.39	0.000	2.89
FUNCTCLASS					
Local	2.1348E+1	4.70	4.55	0.000	393.63
Major Collector	2.2360E+1	4.67	4.78	0.000	182.97
Minor Arterial	2.3336E+1	4.65	5.02	0.000	86.32
Minor Collector	2.1779E+1	4.70	4.64	0.000	218.43
Principal Arterial	1.9548E+1	4.53	4.31	0.000	41.59
DECKGEOMAPP					
UNACCEPTABLE	2.2631E+0	0.281	8.05	0.000	1.51
SPAN1					
2	1.5368E+0	0.379	4.05	0.000	1.70
3	1.2823E+0	0.411	3.12	0.002	2.77
4	2.6969E+0	0.568	4.75	0.000	2.72
5	2.5320E+0	0.722	3.51	0.000	4.27
DECKMAT					
Steel	-1.1725E+0	0.468	-2.50	0.012	1.07
REGION					
Piedmont	6.9413E-1	0.251	2.77	0.006	1.18

Regression Equation

```
NBWID  =  -13.899 - 0.018451 OBLEN + 0.70895 OBWID + 0.091028 MAXSPAN1
          + 0.15516 WATERDEPTH + 0.14022 APPWID + 0.00074050 ADTr
          + 0.0 FUNCTCLASS_Interstate + 21.348 FUNCTCLASS_Local
          + 22.360 FUNCTCLASS_Major Collector
          + 23.336 FUNCTCLASS_Minor Arterial
          + 21.779 FUNCTCLASS_Minor Collector
          + 19.548 FUNCTCLASS_Principal Arterial + 0.0 DECKGEOMAPP_ACCEPTABLE
          + 2.2631 DECKGEOMAPP_UNACCEPTABLE + 0.0 SPAN1_1 + 1.5368 SPAN1_2
          + 1.2823 SPAN1_3 + 2.6969 SPAN1_4 + 2.5320 SPAN1_5
          + 0.0 STEELDECKMAT_0 - 1.1725 DECKMAT_Steel
          + 0.0 REGION_Mountains/Coastal + 0.69413 REGION_Piedmont
```

Figure A.7: Minitab model summary for *NBWID*, without variable interactions

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
18.0427	54.72%	53.77%	49.13%

Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	57.9	10.6	5.45	0.000	
MAXSPAN1	2.162	0.316	6.85	0.000	111.38
OBLEN	-0.6090	0.0818	-7.45	0.000	125.71
CTB	1.618	0.188	8.62	0.000	7.51
ADTr	0.001467	0.000337	4.36	0.000	19.02
OBLEN^2	0.002875	0.000331	8.68	0.000	429.68
MAXSPAN1^2	-0.01772	0.00508	-3.49	0.000	374.86
BRIDGEAGE^2	0.001868	0.000378	4.95	0.000	1.30
ADTr^2	0.000000	0.000000	3.48	0.001	8.22
OBLEN^3	-0.000002	0.000000	-7.95	0.000	99.55
OBWID^3	0.000080	0.000024	3.33	0.001	5.72
MAXSPAN1^3	0.000078	0.000026	3.06	0.002	167.44
OBLEN*OBWID	-0.00332	0.00118	-2.82	0.005	41.01
OBLEN*MAXSPAN1	-0.002161	0.000980	-2.21	0.028	180.84
OBLEN*CTB	-0.004444	0.000909	-4.89	0.000	27.00
OBLEN*APPWID	0.002799	0.000927	3.02	0.003	16.31
MAXSPAN1*WATERDEPTH	-0.02963	0.00631	-4.69	0.000	2.69
MAXSPAN1*ADTr	-0.000020	0.000006	-3.50	0.000	24.88
CTB*ADTr	-0.000051	0.000007	-7.22	0.000	4.63
REGION					
Piedmont	4.97	1.08	4.59	0.000	1.30
FUNCTCLASS					
Local	-23.86	9.63	-2.48	0.013	98.57
Major Collector	-22.97	9.53	-2.41	0.016	45.29
Minor Arterial	-28.92	9.36	-3.09	0.002	22.52
Minor Collector	-25.75	9.66	-2.67	0.008	51.38
Principal Arterial	-26.97	9.22	-2.93	0.003	11.40
CROSSINGTYPE					
Waterway	-33.45	3.97	-8.43	0.000	3.79
SUBSTRMAT					
Concrete/Timber	-7.75	2.70	-2.87	0.004	7.16
ROADWAYALIGNAPP					
UNACCEPTABLE	-4.47	2.11	-2.12	0.035	1.06
SPAN1					
2	27.21	3.48	7.82	0.000	8.26
3	36.64	4.25	8.63	0.000	17.54
4	40.13	5.13	7.82	0.000	13.85
5	46.46	5.85	7.95	0.000	16.23

Figure A.8: Minitab model summary for *MAXSPAN2*, with variable interactions

```
Regression Equation

 MAXSPAN2  =  57.9 + 2.162 MAXSPAN1 - 0.6090 OBLEN + 1.618 CTB + 0.001467 ADTr
              + 0.002875 OBLEN^2 - 0.01772 MAXSPAN1^2 + 0.001868 BRIDGEAGE^2
              + 0.000000 ADTr^2 - 0.000002 OBLEN^3 + 0.000080 OBWID^3
              + 0.000078 MAXSPAN1^3 - 0.00332 OBLEN*OBWID
              - 0.002161 OBLEN*MAXSPAN1 - 0.004444 OBLEN*CTB
              + 0.002799 OBLEN*APPWID - 0.02963 MAXSPAN1*WATERDEPTH
              - 0.000020 MAXSPAN1*ADTr - 0.000051 CTB*ADTr
              + 0.0 REGION_Mountains/Coastal + 4.97 REGION_Piedmont
              + 0.0 FUNCTCLASS_Interstate - 23.86 FUNCTCLASS_Local
              - 22.97 FUNCTCLASS_Major Collector - 28.92 FUNCTCLASS_Minor
              Arterial - 25.75 FUNCTCLASS_Minor Collector
              - 26.97 FUNCTCLASS_Principal Arterial

              + 0.0 CROSSINGTYPE_Non-waterway - 33.45 CROSSINGTYPE_Waterway
              + 0.0 SUBSTRMAT_Other - 7.75 SUBSTRMAT_Concrete/Timber
              + 0.0 ROADWAYALIGNAPP_ACCEPTABLE
              - 4.47 ROADWAYALIGNAPP_UNACCEPTABLE + 0.0 SPAN1_1
              + 27.21 SPAN1_2 + 36.64 SPAN1_3 + 40.13 SPAN1_4 + 46.46 SPAN1_5
```

Figure A.8: Minitab model summary for *MAXSPAN2*, with variable interactions (continued)

```
Model Summary
 S         R-sq    R-sq(adj)  R-sq(pred)
 19.0555  49.15%    48.43%       46.60%

Coefficients
 Term                    Coef         SE Coef  T-Value  P-Value  VIF
 Constant                 7.3099E+1  8.44     8.66     0.000
 MAXSPAN1                 5.3443E-1  0.0466   11.46    0.000    2.18
 CTB                      4.2766E-1  0.124    3.44     0.001    2.96
 FUNCTCLASS
   Local                 -1.7213E+1  6.28     -2.74    0.006    37.63
   Major Collector       -1.6868E+1  6.26     -2.69    0.007    17.52
   Minor Arterial        -2.4385E+1  6.29     -3.88    0.000    9.13
   Minor Collector       -1.9048E+1  6.35     -3.00    0.003    19.90
   Principal Arterial    -2.0978E+1  6.37     -3.29    0.001    4.89
 CROSSINGTYPE
   Waterway              -3.5505E+1  3.95     -9.00    0.000    3.36
 SPAN1
   2                      1.5987E+1  3.09     5.17     0.000    5.85
   3                      1.7126E+1  3.09     5.54     0.000    8.33
   4                      1.5555E+1  3.36     4.62     0.000    5.33
   5                      1.9576E+1  3.52     5.56     0.000    5.28
 WATERDEPTH              -4.4458E-1  0.225    -1.98    0.048    1.62
 BRIDGEAGE                1.8300E-1  0.0517   3.54     0.000    1.43
 APPWID                   3.1932E-1  0.131    2.44     0.015    1.80
 REGION
   Piedmont               6.0209E+0  1.34     4.49     0.000    1.78
   Coastal               -3.7617E+0  1.81     -2.08    0.038    3.10
 DECKMAT
   Steel                 -5.5290E+0  2.36     -2.35    0.019    1.41
   Timber                -4.2567E+0  1.63     -2.61    0.009    2.75
 SUBSTRMAT
   Concrete/Timber       -7.3787E+0  2.84     -2.60    0.009    7.10
 SUPERSTRTYPE
   Channel Beam          -6.0483E+0  2.16     -2.80    0.005    1.38

Regression Equation
 MAXSPAN2  =  73.099 + 0.53443 MAXSPAN1 + 0.42766 CTB
              + 0.0 FUNCTCLASS_Interstate - 17.213 FUNCTCLASS_Local
              - 16.87 FUNCTCLASS_Major Collector
              - 24.385 FUNCTCLASS_Minor Arterial
              - 19.048 FUNCTCLASS_Minor Collector
              - 20.978 FUNCTCLASS_Principal Arterial
              + 0.0 CROSSINGTYPE_Non-waterway - 35.505 CROSSINGTYPE_Waterway
              + 0.0 SPAN1_1 + 15.987 SPAN1_2 + 17.126 SPAN1_3 + 15.555 SPAN1_4
              + 19.576 SPAN1_5 - 0.4458 WATERDEPTH + 0.18300 BRIDGEAGE
              + 0.31932 APPWID + 0.0 REGION_Mountains + 6.0209 REGION_Piedmont
              - 3.7617 REGION_Coastal + 0.0 DECKMAT_Concrete
              - 5.5290 DECKMAT_Steel - 4.2567 DECKMAT_Timber
              + 0.0 SUBSTRMAT_Other type - 7.3787 SUBSTRMAT_Concrete/Timber
              + 0.0 SUPERSTRTYPE_Other type - 6.0483 SUPERSTRTYPE_Channel beam
```

Figure A.9: Minitab model summary for *MAXSPAN2*, without variable interactions

```
Model Summary

      S    R-sq  R-sq(adj)  R-sq(pred)
19258.4  43.19%     35.68%       0.00%


Coefficients

Term                     Coef   SE Coef  T-Value  P-Value          VIF
Constant          -6.8428E+05    236636    -2.89    0.004
DECKAREA          -2.5647E+03       732    -3.50    0.001   1592748.34
NBLEN^2           -2.1460E+02      51.8    -4.14    0.000    150464.77
DECKAREA^2        -3.1190E-01    0.0860    -3.63    0.000  13778279.67
DECKAREA*NBLEN     1.6443E+01      4.32     3.81    0.000   4892427.16
DECKAREA*NBWID     4.5554E+01      12.9     3.53    0.001   3631803.61
NBLEN              3.6962E+04     10558     3.50    0.001     75727.72
BRIDGEAGE          1.5145E+04      7079     2.14    0.034      1958.82
NBLEN^3           -8.7767E-02    0.0358    -2.45    0.015      5379.37
NBWID^2            1.3603E+03       523     2.60    0.010     71622.02
NBWID^3           -3.0730E+01      11.1    -2.77    0.006    267912.48
ADTr^2            -3.2239E-03  0.000954    -3.38    0.001      5670.39
BRIDGEAGE^2       -2.5999E+02       120    -2.16    0.032      7801.71
BRIDGEAGE^3        1.4136E+00     0.659     2.15    0.033      2070.43
DECKAREA^3         3.3660E-06  0.000001     3.22    0.002   1352380.24
DECKAREA*MAXSPAN2-1.5841E-01    0.0477    -3.32    0.001        77.18
DECKAREA*ADTr      3.1507E-02   0.00704     4.47    0.000    114659.80
DECKAREA*APPWID   -6.9633E-01     0.268    -2.60    0.010      1004.71
NBLEN*ADTr        -8.8218E-01     0.211    -4.18    0.000     13709.37
NBWID*ADTr        -1.0757E+00     0.353    -3.05    0.003      4720.05
NBWID*WATERDEPTH  -9.3985E+01      30.8    -3.05    0.003         3.36
MAXSPAN2*ADTr      5.5852E-01     0.139     4.02    0.000       740.29
ADTr*WATERDEPTH    1.7606E+00     0.602     2.93    0.004         8.32
DECKMAT
  STEEL/TIMBER     9.5789E+03      3567     2.69    0.008         1.60


Regression Equation

ROWCOST = -684280 - 2564.7 DECKAREA - 214.60 NBLEN^2 - 0.31190 DECKAREA^2
          + 16.443 DECKAREA*NBLEN + 45.554 DECKAREA*NBWID + 36962 NBLEN
          + 15145 BRIDGEAGE - 0.087767 NBLEN^3 + 1360.3 NBWID^2
          - 30.730 NBWID^3 - 0.0032239 ADTr^2 - 259.99 BRIDGEAGE^2
          + 1.4136 BRIDGEAGE^3 + 0.0000033660 DECKAREA^3
          - 0.15841 DECKAREA*MAXSPAN2 + 0.031507 DECKAREA*ADTr
          - 0.69633 DECKAREA*APPWID - 0.88218 NBLEN*ADTr - 1.0757 NBWID*ADTr
          - 93.985 NBWID*WATERDEPTH + 0.55852 MAXSPAN2*ADTr
          + 1.7606 ADTr*WATERDEPTH + 0.0 DECKMAT_Concrete
          + 9578.9 DECKMAT_Steel/Timber
```

Figure A.10: Minitab model summary for *ROWCOST*, Approach 1

```
Model Summary

      S    R-sq  R-sq(adj)  R-sq(pred)
20073.3  34.38%     30.13%      14.90%


Coefficients

Term               Coef  SE Coef  T-Value  P-Value     VIF
Constant    -3.2103E+04    26923    -1.19    0.235
APPWID       6.6098E+03     2503     2.64    0.009   92.95
ADTr         7.4531E+00     2.20     3.39    0.001   24.18
OBLEN*MAXSPAN1
            -1.1556E+01     4.80    -2.41    0.017   31.97
OBLEN*ADTr   1.6900E-01   0.0493     3.42    0.001  563.19
OBLEN*APPWID-2.1804E+01     9.80    -2.23    0.027  152.02
OBWID*ADTr  -3.0355E-01   0.0855    -3.55    0.000  220.36
MAXSPAN1*CTB 8.3036E+01     21.7     3.83    0.000   15.83
WATERDEPTH*CTB
            -4.7586E+02      156    -3.05    0.003   19.40
BRIDGEAGE*CTB
            -2.5408E+01     10.5    -2.42    0.017    5.21
OBLEN^2      4.0711E+00     1.22     3.34    0.001   53.08
APPWID^2    -1.7614E+02     59.0    -2.99    0.003  347.11
WATERDEPTH^2 5.7551E+02      202     2.85    0.005   15.60


Regression Equation

ROWCOST = -32103 + 6609.8 APPWID + 7.4531 ADTr - 11.556 OBLEN*MAXSPAN1
          + 0.16900 OBLEN*ADTr - 21.804 OBLEN*APPWID - 0.30355 OBWID*ADTr
          + 83.036 MAXSPAN1*CTB – 475.86 WATERDEPTH*CTB
          - 25.408 BRIDGEAGE*CTB + 4.0711 OBLEN^2 - 176.14 APPWID^2
          + 575.51 WATERDEPTH^2
```

Figure A.11: Minitab model summary for *ROWCOST*, Approach 2

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
43360.5	85.04%	83.56%	0.00%

Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	-3.6331E+06	806777	-4.50	0.000	
APPWID	3.9517E+04	10518	3.76	0.000	380.43
ADTr	4.0985E+02	66.3	6.18	0.000	8273.87
NBLEN*NBWID	-3.1795E+03	496	-6.42	0.000	243244.24
NBLEN^2	-1.6385E+02	47.1	-3.48	0.001	36916.65
DECKAREA^2	-1.5970E-01	0.0266	-6.00	0.000	573598.31
DECKAREA*NBLEN	1.0299E+01	2.24	4.60	0.000	511419.31
DECKAREA*NBWID	2.9203E+01	4.68	6.24	0.000	179016.14
NBLEN	7.1228E+04	12364	5.76	0.000	26362.26
NBWID	9.4908E+04	21429	4.43	0.000	2985.59
BRIDGEAGE	5.9978E+04	13764	4.36	0.000	1677.35
ADTr^2	7.3228E-03	0.00222	3.29	0.001	12708.10
ADTr^3	-5.0021E-07	0.000000	-6.72	0.000	21589.89
DECKAREA*ADTr	1.0569E-01	0.0177	5.98	0.000	385117.31
DECKAREA*BRIDGEAGE	2.2096E+01	4.34	5.09	0.000	60577.17
NBLEN*ADTr	-3.2093E+00	0.661	-4.86	0.000	66176.31
NBLEN*WATERDEPTH	-4.0041E+01	12.1	-3.30	0.001	1.82
NBLEN*BRIDGEAGE	-6.9711E+02	143	-4.87	0.000	12832.21
NBWID*ADTr	-1.3872E+01	1.86	-7.44	0.000	50322.07
NBWID*BRIDGEAGE	-1.4299E+03	350	-4.09	0.000	4051.74
APPWID*BRIDGEAGE	-8.0829E+02	187	-4.31	0.000	605.00

Regression Equation

```
ENGCOST = -3633100 + 39517 APPWID + 409.85 ADTr - 3179.5 NBLEN*NBWID
          - 163.85 NBLEN^2 - 0.15970 DECKAREA^2 + 10.299 DECKAREA*NBLEN
          + 29.203 DECKAREA*NBWID + 71228 NBLEN + 94908 NBWID + 59978 BRIDGEAGE
          + 0.0073228 ADTr^2 - 0.00000050021 ADTr^3 + 0.10569 DECKAREA*ADTr
          + 22.096 DECKAREA*BRIDGEAGE - 3.2093 NBLEN*ADTr
          - 40.041 NBLEN*WATERDEPTH - 697.11 NBLEN*BRIDGEAGE
          - 13.872 NBWID*ADTr - 1429.9 NBWID*BRIDGEAGE
          - 808.29 APPWID*BRIDGEAGE
```

Figure A.12: Minitab model summary for *ENGCOST*, Approach 1

```
Model Summary

      S    R-sq  R-sq(adj)  R-sq(pred)
48808.9  80.94%     79.17%      32.87%


Coefficients

Term                      Coef   SE Coef  T-Value  P-Value      VIF
Constant             3.4666E+05   108946     3.18    0.002
OBLEN               -3.5781E+03      955    -3.75    0.000   126.72
OBWID               -3.4374E+04    10692    -3.21    0.002   524.10
ADTr                -9.2278E+01     16.4    -5.62    0.000   400.07
OBLEN*MAXSPAN1       3.7853E+01     13.7     2.77    0.006    76.99
OBLEN*APPWID         2.3394E+02     48.1     4.86    0.000   733.35
OBLEN*CTB           -1.1863E+02     31.7    -3.74    0.000   223.67
MAXSPAN1*BRIDGEAGE
                    -3.8327E+01     16.5    -2.32    0.021    10.42
ADTr*APPWID          2.4308E+00    0.815     2.98    0.003  2996.64
ADTr*BRIDGEAGE       9.4241E-01    0.309     3.05    0.003   466.65
APPWID*BRIDGEAGE    -3.2907E+02     71.3    -4.61    0.000    69.09
BRIDGEAGE*CTB        1.1406E+02     44.4     2.57    0.011    18.80
OBLEN^3              3.5132E-02   0.0133     2.64    0.009   140.83
OBWID^2              1.5441E+03      379     4.08    0.000  5492.87
OBWID^3             -2.0378E+01     4.61    -4.42    0.000  6034.16
ADTr^3              -4.1497E-08 0.000000    -2.88    0.004   639.45
BRIDGEAGE^2          1.4916E+02     34.9     4.28    0.000   113.16
BRIDGEAGE^3         -1.1790E+00    0.295    -3.99    0.000    69.98
DECKGEOMAPP
  UNACCEPTABLE       2.7908E+04    10290     2.71    0.007     2.48
SPAN1
  3+                -2.9735E+04    11850    -2.51    0.013     3.26


Regression Equation

ENGCOST = 346660 - 3578.1 OBLEN - 34374 OBWID - 92.278 ADTr
          + 37.853 OBLEN*MAXSPAN1 + 233.94 OBLEN*APPWID - 118.63 OBLEN*CTB
          - 38.327 MAXSPAN1*BRIDGEAGE + 2.4308 ADTr*APPWID
          + 0.94241 ADTr*BRIDGEAGE - 329.07 APPWID*BRIDGEAGE
          + 114.06 BRIDGEAGE*CTB + 0.035132 OBLEN^3 + 15441 OBWID^2
          - 20.378 OBWID^3 - 0.000000041497 ADTr^3 + 149.16 BRIDGEAGE^2
          - 1.1790 BRIDGEAGE^3 + 0.0 DECKGEOMAPP_ACCEPTABLE
          + 27908 DECKGEOMAPP_UNACCEPTABLE + 0.0 SPAN1_1or2
          - 29735 SPAN1_3+
```

Figure A.13: Minitab model summary for *ENGCOST*, Approach 2

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
95233.4	99.27%	99.19%	7.87%

Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	1.5902E+07	2938571	5.41	0.000	
ADTr	-3.5148E+02	101	-3.49	0.001	3966.23
NBLEN*NBWID	2.1457E+04	4360	4.92	0.000	3903336.18
NBLEN^2	8.8464E+02	208	4.25	0.000	149385.79
DECKAREA^2	1.2998E+00	0.267	4.87	0.000	11986750.24
DECKAREA*NBLEN	-7.1247E+01	15.2	-4.69	0.000	4878338.09
DECKAREA*NBWID	-3.2029E+02	67.9	-4.72	0.000	7807562.36
NBLEN	-3.4596E+05	68885	-5.02	0.000	169646.02
NBWID	-6.7373E+05	128813	-5.23	0.000	22363.85
MAXSPAN2	-9.9508E+04	28276	-3.52	0.001	4167.37
NBLEN^3	4.2151E-01	0.112	3.76	0.000	3864.61
NBWID^3	1.7063E+02	36.8	4.63	0.000	203612.27
BRIDGEAGE^2	2.0003E+01	7.27	2.75	0.007	1.29
DECKAREA^3	-8.2626E-06	0.000002	-4.34	0.000	497501.33
DECKAREA*MAXSPAN2	-3.6647E+01	9.05	-4.05	0.000	172316.42
DECKAREA*ADTr	-8.1224E-02	0.0162	-5.02	0.000	66975.11
NBLEN*MAXSPAN2	1.1686E+03	293	3.99	0.000	28874.31
NBLEN*ADTr	2.9097E+00	0.618	4.71	0.000	12002.50
NBWID*MAXSPAN2	3.0943E+03	878	3.52	0.001	14920.87
NBWID*ADTr	8.4513E+00	2.69	3.15	0.002	21639.17
ADTr*CTB	5.7811E+00	1.84	3.15	0.002	2667.50
BRIDGESYS					
Secondary	-8.4962E+04	23927	-3.55	0.000	1.30

Regression Equation

```
CONSTCOST = 15902000 - 351.48 ADTr + 21457 NBLEN*NBWID + 884.64 NBLEN^2
            + 1.2998 DECKAREA^2 - 71.247 DECKAREA*NBLEN - 320.29 DECKAREA*NBWID
            - 345960 NBLEN - 673730 NBWID - 99508 MAXSPAN2 + 0.42151 NBLEN^3
            + 170.63 NBWID^3 + 20.003 BRIDGEAGE^2 - 0.0000082626 DECKAREA^3
            - 36.647 DECKAREA*MAXSPAN2 - 0.081224 DECKAREA*ADTr
            + 11686 NBLEN*MAXSPAN2 + 2.9097 NBLEN*ADTr + 3094.3 NBWID*MAXSPAN2
            + 8.4513 NBWID*ADTr + 5.7811 ADTr*CTB + 0.0 BRIDGESYS_Not Secondary
            - 84962 BRIDGESYS_Secondary
```

Figure A.14: Minitab model summary for *CONSTCOST*, Approach 1

```
Model Summary

     S    R-sq  R-sq(adj)  R-sq(pred)
111275  98.97%     98.89%           *


Coefficients

Term                         Coef   SE Coef  T-Value  P-Value       VIF
Constant               2.7848E+05     68183     4.08    0.000
OBWID                  1.7595E+04      7484     2.35    0.020     49.40
ADTr                  -1.3380E+02      39.5    -3.38    0.001    446.66
OBLEN*ADTr             1.1866E+00     0.123     9.65    0.000    299.76
OBLEN*CTB              2.1497E+02      24.8     8.68    0.000     26.20
OBWID*MAXSPAN1        -1.0248E+02      49.1    -2.09    0.038      7.17
OBWID*ADTr            -5.9873E+00      2.13    -2.81    0.005   5093.96
OBWID*APPWID          -6.5579E+02       302    -2.17    0.031    444.03
MAXSPAN1*ADTr          2.3045E+00     0.786     2.93    0.004    566.22
ADTr*APPWID            1.0923E+01      3.89     2.81    0.005  13147.21
ADTr^3                -1.7159E-07  0.000000    -4.48    0.000    866.38
BRIDGEAGE^2            2.8497E+01      8.43     3.38    0.001      1.27
CTB^3                 -1.8234E+01      5.90    -3.09    0.002     79.46
FUNCTCLASS
  Major Collector      1.2455E+05     32195     3.87    0.000      1.66
  Minor Arterial       3.6295E+05    124216     2.92    0.004      1.24
  Principal Arterial/Int.
                      -9.6377E+05    287305    -3.35    0.001     26.19


Regression Equation

CONSTCOST = 278480 + 17595 OBWID - 133.80 ADTr + 1.1866 OBLEN*ADTr
            + 214.97 OBLEN*CTB - 102.48 OBWID*MAXSPAN1 - 5.9873 OBWID*ADTr
            - 655.79 OBWID*APPWID + 2.3045 MAXSPAN1*ADTr
            + 10.923 ADTr*APPWID - 0.00000017159 ADTr^3 + 28.497 BRIDGEAGE^2
            - 18.234 CTB^3 + 0.0 FUNCTCLASS4_Local/Minor Collector
            + 124550 FUNCTCLASS4_Major Collector
            + 362950 FUNCTCLASS_Minor Arterial
            - 963770 FUNCTCLASS_Principal Arterial/Interstate
```

Figure A.15: Minitab model summary for *CONSTCOST*, Approach 2

```
Model Summary

     S    R-sq  R-sq(adj)  R-sq(pred)
207149  96.25%     96.12%      68.86%


Coefficients

Term                      Coef   SE Coef  T-Value  P-Value     VIF
Constant             4.3365E+5    64329     6.74    0.000
NBLEN*BRIDGEAGE     -2.6032E+2     51.3    -5.08    0.000   87.43
ADTr*APPWID         -3.4234E+0    0.398    -8.60    0.000   40.02
NBLEN^2              1.1990E+2     15.1     7.93    0.000  183.52
ADTr^3              -2.4817E-7 0.000000    -6.50    0.000  249.44
CTB^2                3.2875E+2      149     2.21    0.028    7.38
DECKAREA*NBLEN      -3.2296E+0    0.446    -7.24    0.000  911.98
DECKAREA*ADTr        3.5682E-2  0.00380     9.39    0.000  783.04
DECKAREA*BRIDGEAGE   9.7269E+0     1.48     6.59    0.000  324.08
FUNCTCLASS
  Local/Minor Collector
                    -1.4590E+5    47074    -3.10    0.002    1.56
PROJECTTYPE
  TIP                5.2557E+5    27107    19.39    0.000    1.02


Regression Equation

TOTCOST = 433650 - 260.32 NBLEN*BRIDGEAGE - 3.4234 ADTr*APPWID + 119.90 NBLEN^2
          - 0.00000024817 ADTr^3 + 328.75 CTB^2 - 3.2296 DECKAREA*NBLEN
          + 0.035682 DECKAREA*ADTr + 9.7269 DECKAREA*BRIDGEAGE
          + 0.0 FUNCTCLASS_Other Classification
          - 145900 FUNCTCLASS_Local/Minor Collector
          + 0.0 PROJECTTYPE_17BP + 525570 PROJECTTYPE_TIP
```

Figure A.16: Minitab model summary for *TOTCOST*, Approach 1

```
Model Summary

     S    R-sq  R-sq(adj)  R-sq(pred)
237169  95.09%     94.92%      93.70%


Coefficients

Term                                  Coef   SE Coef  T-Value  P-Value    VIF
Constant                         1.3363E+4    135842     0.10    0.922
OBLEN                            5.7462E+3       609     9.44    0.000   2.69
OBWID                            1.4016E+4      5825     2.41    0.017   7.20
OBLEN*ADTr                       1.3327E+0    0.0658    20.25    0.000  19.01
OBWID*CTB                       -1.3632E-3       199    -6.84    0.000  19.65
ADTr^3                          -9.7673E-8  0.000000   -11.28    0.000   9.79
BRIDGEAGE^2                      5.6500E+1      14.6     3.87    0.000   1.15
CTB^2                            1.4275E+3       274     5.20    0.000  19.18
FUNCTCLASS
  Major Collector                2.5008E+5     49666     5.04    0.000   1.15
  Minor Arterial/Principal Arterial/Int.
                                 4.7760E+5    184811     2.58    0.010   2.99
PROJECTTYPE
  TIP                            5.5135E+5     31790    17.34    0.000   1.07


Regression Equation

TOTCOST = 13363 + 5746.2 OBLEN + 14016 OBWID + 1.3327 OBLEN*ADTr
          - 1363.2 OBWID*CTB - 0.000000097673 ADTr^3 + 56.500 BRIDGEAGE^2
          + 1427.5 CTB^2 + 0.0 FUNCTCLASS_Local/Minor Collector
          + 250080 FUNCTCLASS_Major Collector
          + 477600 FUNCTCLASS_Minor Arterial/Principal Arterial/Interstate
          + 0.0 PROJECTTYPE_17BP + 551350 PROJECTTYPE_TIP
```

Figure A.17: Minitab model summary for *TOTCOST*, Approach 2

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
1	STRNUI	OBLEN	OBWID	MAXSPAN1	NBLEN	NBWID	MAXSPAI	WATERDEPT	BRIDGEAG	CTB	APPWII	ADTr	REGION	DIVISIO	FUNCTCLASS	BRIDGESY
2	000007	70	20	35	100	33	96	2	54	14	20	1911	2	7	Local	Secondary
3	000032	70	20	35	77	30	76	1	57	13	20	290	2	7	Local	Secondary
4	000037	52	20	26	92	30	89	5	59	15	19	310	2	7	Local	Secondary
5	000066	91	25	31	93	30	89	3	50	15	20	350	2	7	Local	Secondary
6	000096	75	20	25	145	39	59	1	54	14	20	869	2	7	Local	Secondary
7	000097	26	24	26	62	30	59	1	60	11	20	710	2	7	Local	Secondary
8	000102	181	24	40	180	30	97	2	42	24	19	525	2	7	Local	Secondary
9	000134	121	26	30	138	30	68	2	52	22	20	420	2	7	Local	Secondary
10	000137	36	20	18	82	30	82	1	61	11	19	160	2	7	Local	Primary
11	000144	37	20	18	62	27	62	3	59	10	18	360	2	7	Local	Secondary
12	000149	31	20	31	72	33	67	4	54	11	18	770	2	7	Local	Secondary
13	000162	81	24	40	120	33	67	5	57	15	18	1500	2	7	Major Collector	Secondary
14	000267	19	20	19	62	30	62	2	58	8	19	190	2	7	Local	Secondary
15	010008	113	22	38	120	35	114	1	57	20	18	330	2	12	Local	Secondary
16	010053	76	26	26	92	30	56	1	40	17	20	2000	2	12	Local	Secondary
17	010070	106	26	36	122	36	118	1	59	20	19	710	2	12	Minor Collector	Secondary
18	010079	121	25	40	132	30	50	1	52	16	18	100	2	12	Minor Collector	Secondary
19	010081	91	26	30	102	30	71	2	46	15	20	380	2	12	Local	Secondary
20	010087	81	24	30	102	30	102	1	38	24	20	100	2	12	Local	Secondary
21	010092	71	20	36	87	30	56	1	52	11	16	260	2	12	Local	Secondary
22	010093	50	20	26	82	27	82	1	54	18	18	270	2	12	Local	Secondary

Figure A.18: Screenshot of central characteristic dataset

Table A.1: Category counts for both central datasets

Categorical Variable	Categories	Cost Dataset	Characteristic Dataset
FUNCTCLASS	Local	224	967
	Minor Collector	47	208
	Major Collector	29	185
	Minor Arterial	1	89
	Principal Arterial	3	45
	Interstate	1	12
REGION	Mountains	31	378
	Piedmont	152	599
	Coastal	122	529
BRIDGESYS	Interstate	3	19
	Primary	34	286
	Secondary	268	1201
SUPERSTRMAT	Steel	188	922
	Concrete	49	231
	Timber	68	353
SUBSTRMAT	Steel	111	460
	Concrete	42	237
	Timber	152	809
SUPERSTRTYPE	Stringer/Multibeam or Girder	236	1194
	Girder & Floorbeam System	21	84
	Tee Beam	25	111
	Channel Beam	23	117
DECKMAT	Concrete	122	678
	Steel	25	99
	Timber	158	729
SPAN1	1	99	411
	2	72	271
	3	78	452
	4	30	197
	5+	23	175