A Stochastic Three-Dimensional Cost Estimation System for Hot Mix Asphalt in the State of Alabama

by

Keren Xu

A thesis submitted to the Graduate Faculty of Auburn University
in partial fulfillment of the
requirements for the Degree of
Master of Science in Civil
Engineering

Auburn, Alabama
May 05, 2018

Keywords: bid-based cost estimation, location cost index, moving-window cross validation, risk-based cost estimation

Copyright 2018 by Keren Xu

Approved by

Jorge Rueda-Benavides, Chair, Assistant Professor of Civil Engineering
Jeffrey LaMondia, Associate Professor of Civil Engineering
Wesley Zech, Professor of Civil Engineering

ABSTRACT

Asphalt is one of the most critical commodities for the American infrastructure. It is used as a paving material on the majority of the U.S. roadways. The National Asphalt Paving Association (NAPA) estimates that about 94 percent of all paved roads in the U.S. are surfaced with asphalt (NAPA 2018). The relevance of this paving material is not different in the state of Alabama, where this material is used in about 98 percent of the paved roads (AAPA 2016). It has resulted in the purchase of Hot Mix Asphalt (HMA) consuming over 40 percent of the Alabama Department of Transportation (ALDOT) annual construction budget, becoming an item that requires special attention during resource allocation and cost estimating procedures. Therefore, any improvements in ALDOT's HMA cost estimating procedures is expected to be reflected in better budget control and a more effective use of ALDOT's limited available resources. ALDOT's available resources have become increasingly scarce in recent decades while there is an increasing demand for transportation infrastructure rehabilitation and improvements. This is what motivated the research presented in this thesis.

To improve the effectiveness of ALDOT's HMA cost estimating practices, this research proposes a three-dimensional HMA cost estimating system based on historical bid data from previous projects awarded by ALDOT. This system was the result of the assessment of five factors and the modeling of their impacts on HMA prices in Alabama. The five factors are: 1) project scale, 2) time, 3) geographic location, 4) estimating uncertainty, and 5) level of competition. The first three factors correspond to the "three dimensions" of the system, while the fourth factor was considered by designing the system to generate stochastic HMA cost estimates in lieu of the traditional deterministic estimates. The impact of level of competition on HMA prices was evaluated in this study, but no significant variations on this factor were found across the state, so this factor was not incorporated into the proposed cost estimating system.

The proposed HMA cost estimating system was developed, and its effectiveness demonstrated, through its application to a case study item and using historical bid data from 3,661 projects awarded by ALDOT between 2006 and 2016. The system integrates a number of elements and quantitative procedures including non-linear regression, a construction cost indexing system (CCIS), a location cost index (LCI), and the stochastic analysis of the expected performance of the system.

Finally, the effectiveness of the system is demonstrated using an innovative Moving-Window Cross Validation (MWCV) approach and via statistical testing to assess the level of significance of the estimating improvement offered by each of the three dimensions. The results of the validation process revealed that each dimension would significantly improve cost estimating effectiveness for the case study item, with an overall improvement of 47.7% and 43.6% in accuracy and reliability, respectively.

ACKNOWLEDGEMENT

The author would like to express a heartfelt appreciation to Dr. Jorge Rueda-Benavides for his generous devotion and mentorship while studying at Auburn University. Without his enlightening guidance and instruction, this research could not have been completed. The author would like to thank the committee members, Dr. Jeffrey LaMondia and Wesley Zech, for their time and instruction during this period of conducting research.

The author would like to express gratitude to the Graduate Research Assistants within Dr. Rueda's team: Kaylyn Cardinal, Karthik Chowadary Pakalapati, and Mohamed Khalafalla. Their support, friendship, and academic communication throughout this time helped to make the author successful in writing the thesis.

Lastly, the author would like to express a special thanks to her father, Xiuhui Xu and her mother, Tongcui Zhu. Since without their unconditional love, support, and sacrifice, the author could not have completed the studies and research level required during this educational endeavor.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

AAPA	Alabama Asphalt Pavement Association
AASHTO	American Association of State Highway and Transportation Officials
ALDOT	Alabama Department of Transportation
APA	Asphalt Pavement Alliance
CCIS	Construction Cost Indexing System
DOT	Department of Transportation
EAD	Exploratory Data Analysis
EC	East Central Region in Alabama
FHWA	Federal Highway Administration
HMA	Hot Mix Asphalt
LCI	Location Cost Index
MAPE	Mean Absolute Percentage Error
MWCV	Moving-Window Cross Validation
NAPA	National Asphalt Pavement Association
SE	South East Region in Alabama
STA	State Transportation Agency
SW	South West Region in Alabama
UP	Unit Price
WC	West Central Region in Alabama

CHAPTER ONE:

INTRODUCTION

Historical records show that asphalt was first introduced into the paving industry in Europe in 1838, and it was not until 30 years later that this practice was first used in the United States (U.S.) (Holley 2003). Since then, and after almost 150 years of history in the U.S., asphalt paving has become a mature method and a key player in socio-economic growth and development. Asphalt is currently the most popular form of road surfacing in the U.S. with about 94 percent of all paved roads and highways surfaced with asphalt (NAPA 2018). The National Asphalt Pavement Association (NAPA) estimates that there are around 3,500 asphalt plants across the country, producing about 400 million tons of asphalt every year, worth over $30 billion. Likewise, the asphalt paving industry supports more than 400,000 jobs in the asphalt production, aggregate production, and road construction sectors (NAPA 2018).

A review of the existing literature on paving materials revealed two main reasons that motivate the use of asphalt in paving projects. First, asphalt is considered to be a cost-effective material due to its ability to be rejuvenated and reused. “Asphalt pavements are America’s most recycled product. About 100 million tons of asphalt pavement is reclaimed each year, and over 95 percent of that total is reused or recycled” (APA 2010). It is estimated that reclaimed asphalt saves U.S. taxpayers about $1.8 billion per year (in 2010 dollars) (APA 2010). The second reason is that asphalt pavements provide a smoother ride in comparison with other paving materials (Nicholls, McHale and Griffiths 2008; Marks 2009). “In general, the smoother a pavement is built, the smoother it stays over time, resulting in lower maintenance costs, and more comfort and safety for the traveling public” (Kelly, Leslie, and Lynn 2002). The NCHRP Project 1–31 (Smith et al. 1997) found that a 50 percent increase in smoothness could lead to a 15 percent increase in pavement life.

Every year, state transportation agencies (STAs) spend billions of dollars to maintain, repair, and expand the U.S. highway network; and most of those funds are invested on asphalt and asphalt paving activities (Rueda 2016). Thus, the ability of STAs to effectively estimate costs for asphalt related activities is critical for planning and resource allocation purposes. The study presented in this thesis is intended to support the efforts of the Alabama Department of Transportation (ALDOT) to improve construction cost estimating. This thesis proposes a stochastic data-driven methodology to estimate unit prices for Hot Mix Asphalt (HMA) using bid data from all previous projects awarded by ALDOT during the previous two years. This methodology uses non-linear regression, cost indexing techniques, and other data processing methods to factor three key variables (hereby referred to as *dimensions*) into the cost estimating process. Theses variables are: project scale, time, and location. A fourth factor, estimating uncertainty, was considered by designing the system to generate stochastic HMA prices in lieu of the traditional deterministic estimates. The development of the proposed system was illustrated through its application to a case study item and using historical bid data provided by ALDOT for 3,661 projects awarded between 2006 and 2016. The system was successfully validated using an innovative Moving-Window Cross Validation (MWCV) approach and statistical significance testing techniques. The validation process was systematically applied to demonstrate how estimating effectiveness is improved by each of the three dimensions.

1.1 Motivation Background

1.1.1 Hot Mix Asphalt in the State of Alabama

"Roadways form the backbone of Alabama's economy by getting residents to work, transferring goods and services to market, and connecting residents and visitors to recreational and tourist destinations" (ASCE 2015). The state of Alabama has over 102,000 miles of public roads, including 11,000 miles maintained by ALDOT (ASCE 2015). Approximately 98 percent of all paved roads in Alabama are surfaced with asphalt (AAPA 2016). Asphalt paving is the most common type of work done by ALDOT. On average, the annual amount paid by ALDOT for HMA between 2012 and 2016 corresponds to over 40 percent of its annual construction budget. This percentage refers to "superpave bituminous concrete base, binder, and wearing surface layers," defined under Section 424 of ALDOT 2018 Standard Specifications for Highway Constructions (ALDOT 2018) as "a hot or warm bituminous plant mixed pavement layer placed on a prepared surface" (ALDOT 2018). HMA unit prices are calculated in this study on a tonnage basis and included "all materials, procurement, handling, hauling, and processing cost, and includes all equipment, tools, labor, and incidentals required to complete the work" (ALDOT 2018).

1.1.2 Importance of Effective Construction Cost Estimating

In project management, a project is defined as a "temporary endeavor undertaken to create a unique product, service, or result" (PMI 2013), and these endeavors usually demand the consumption of different types of resources (i.e., money, time, materials, and labor/equipment hours). Under this definition, cost estimating is the process to predict the approximate amount of money required to complete a project. The final amount depends on the required quantities for the other resources. Higher costs are expected from larger projects that require a significant consumption of materials and labor/equipment hours. Cost estimating processes are used in all industries and businesses, not only on construction projects. However, unlike other industries, a single construction owner or contractor may need to manage a highly diversified project portfolio in terms of project-specific scopes, designs, and requirements. Each construction project is characterized by a unique combination of several factors, including project objectives, deliverables, location, environmental requirements, technical complexity, etc. This uniqueness, and the fact that it is virtually impossible to accurately quantify the impact of all these factors on a project, makes construction cost estimating a particularly challenging process.

From an owner's perspective, cost estimates are commonly used to define the project scope, to determine whether or not a project should proceed, and to allocate the required funds for its completion (AASHTO 2013). On the other hand, a construction contractor uses cost estimates to assess its financial capacity to undertake a given project and to prepare the bid for an owner. In both cases, cost estimates are basically used for risk assessment purposes, to support business decisions, and to maximize returns from project portfolios. Therefore, effective cost estimating could be translated into more effective decision-making and greater returns for owners or contractors (Fakültesi and Zeynep 2004; Arafa and Alqedra 2011; Byrnes 2002).

The selection of projects for funding is becoming even more challenging for STAs due to the increasing gap between available resources and those actually required to maintain the national transportation infrastructure in optimal conditions. This situation is demanding more effective resource allocation and cost estimating practices by transportation public owners. The 2013 Report Card for America's Infrastructure (ASCE 2013), published by the U.S. Society of Civil Engineers (ASCE), estimates that "32 percent of America's major roads are in poor or mediocre condition, costing U.S. motorists more than $67 billion a year […] in additional repairs and operating costs." The same study has found that about $170 billion should be invested annually to improve the current conditions and performance of all roads and bridges across the country by 2028. However, only $91 billion is being currently assigned for this purpose (ASCE 2013). The evident growth in this gap during the last two decades led to the enactment of the Moving Ahead for Progress in the 21st Century Act (MAP-21), through which the federal government allocated over $105 billion to surface transportation programs (U.S. Congress 2012). To ensure the best value for taxpayers' money, and recognizing the limitations of traditional deterministic cost estimating practices, MAP-21 also requires STAs to develop and implement better cost estimating and risk control strategies (U.S. Congress 2012), as those presented in this thesis.

1.1.3 Challenges of Cost Estimating

There are several internal and external factors affecting the accuracy of construction cost estimates and it is virtually impossible to identify all of them, as well as to exactly quantify their impact on estimating accuracy. Thus, effective cost estimating should not be conditioned to a 100 percent accuracy, since that would be unrealistic. In this study, effective cost estimating is defined as the capacity of STAs to maximize estimating accuracy and reliability. Accuracy refers to the level of validity of the system. Based on Golafshani (2003), the definition for validity is the degree to which the system truly measured what it is intended to measure, which is usually assessed with measures of central tendency such as mean, median, and mode values. In this study, the level of accuracy in a given HMA price estimate is determined by the absolute percentage error (APE), using Equation 1-1; while the overall accuracy of the system is determined by averaging the APEs of all projects on which the system was applied during the validation process. This is called the mean absolute percentage error (MAPE) and is calculated as shown in Equation 1-2. MAPE values are commonly used to measure and compare accuracy between cost estimating models (Gardner 2015).

$$APE = \frac{|A_i - E_i|}{A_i} \times 100\% \qquad \textbf{Eq. 1-1}$$

$$MAPE = \frac{\sum_{i=1}^{n} \frac{|A_i - E_i|}{A_i}}{n} \times 100\% \qquad \textbf{Eq.1-2}$$

Where: APE = Absolute Percentage Error

MAPE = Mean Absolute Percentage Error

A_i = Actual unit price for HMA in project i

E_i = Estimated unit price for HMA in project i

n = Number of projects using HMA during period under consideration for validation

On the other hand, estimating reliability refers to the degree of consistency in the outputs of quantitative models (Golafshani 2003). Under the context of this study, and using the terms introduced in the previous paragraph, reliability is the degree to which the proposed cost estimating system consistently yield similar APEs every time that it is used. Thus, reliability is measured in this thesis using variance and standard deviation values, which indicate the level of dispersion of the APEs produced by the system.

As will be discussed later in Section 2.2, the concepts of "accuracy" and "reliability" are relative to the type of estimate and project phase in which is it produced. While conceptual estimates performed early in the planning phase of construction projects are expected to have an accuracy between -50% and +200%, detailed estimates at design completion tend to be significantly more accurate with errors ranging between -5% and +10% (AASHTO 2013). It must be noted that the HMA cost estimating approach presented in this thesis is intended to be applied at design completion or as soon as the expected amount of asphalt (in tons) to be used in the project becomes available.

The main challenge faced by STAs in their efforts to produce effective cost estimates are associated with their capacity to identify, understand, and model the impact of the main cost-influencing factors on transportation construction projects. Therefore, one of the first steps in this study was to conduct a literature search to identify the main factors affecting the effectiveness of cost estimating practices in the transportation construction industry. Project scale, time, geographic location, estimating uncertainty, and level of competition have been identified by the ASCE as the major cost-influencing factors. The first three factors have been directly incorporated into the proposed methodology (the three model dimensions), while the uncertainty factor is included in the methodology in the form of stochastic estimates as a byproduct of the validation process. Finally, the level of competition factor was assessed in this study, but not considered in the system because the assessment revealed that this factor is not substantially impacting HMA bid prices in the state of Alabama. Each of the five cost-driven factors are described below.

1.1.3.1 Scale

Project scale refers to the size of the project, which is expected to be reflected in the project budget, schedule, physical dimensions, and overall consumption of resources (Odeck 2004). The amount of HMA to be placed on a given paving project is a great indicator of the project size and may also provide a good idea of the expected project cost. It is usually assumed that larger projects have higher costs. However, the relationship between these two parameters is not linear. To better understand the relationship between project scale and cost, it is necessary to talk about economies of scale (Akintoye 2000).

"Economies of scale refers to a reduction in total cost per unit as output increases" (Betts 2007). In other words, the higher the quantities of work, the lower the unit price (Zhang and Sun 2007; Akintoye 2000). Figure 1.1 uses HMA unit prices (for case study item) received by ALDOT between 2012 and 2016 to illustrate the concept of economies of scales and how it impacts unit prices for this commodity. As shown in this figure, the quantity-unit price relationship for HMA can be modeled using non-linear regression techniques. That is actually how this study incorporates project scale into the proposed three-dimensional estimating system, with scale being the first of the three dimensions. A detailed description of the non-linear regression models developed in this study is presented in Chapter 3 of this thesis.

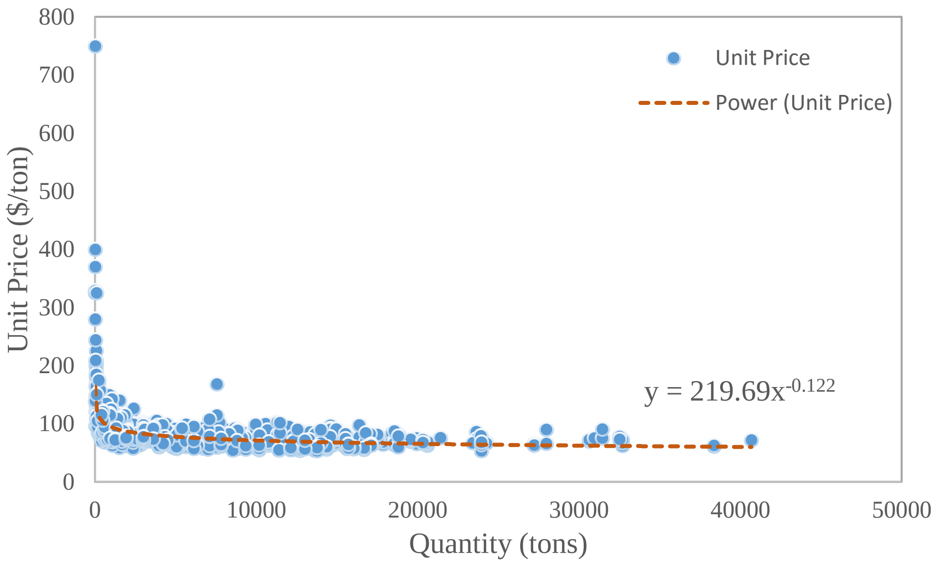


Figure 1.1: Quantity and Unit Price Relationship for HMA (Pay Item 424A360).

1.1.3.2 Time

The ability to track changes in the construction market over time is one of the main challenges faced by STA estimators (Shane et al. 2009). The total cost of a given project today is not expected to be equal to the cost of the same project one year ago or next year. Fluctuations of construction prices may occur due to inflationary trends at industry, or commodity-level changes in demand conditions of labor and materials, changes in interest rates, or seasonal effects (Xu and Moon 2011; Zou, Zhang, and Wang 2007). In the context of this study, the time dimension refers to the fact that HMA prices change over time. It does not refer to the expected project duration. Even though the author recognizes that the duration of a project has some impact on its cost, it is assumed that project duration is directly proportional to project scale. Therefore, it can be said that this factor, project duration, has already been accounted for in the project scale dimension.

The existing literature on construction cost estimating includes several examples of data-drive cost estimating models, using known prices from previous projects to estimate current or future prices. However, a question that has not been answered yet is how many years' historical data (look-back period) should be used in bid-based estimating. There is a misconception that larger datasets always produce better data-drive models, so the larger the look-back period the better. That could be true in many other contexts, but this is not always a valid assumption to determine the length of look-back periods for data retrieval in construction cost estimating. The look-back period should be long enough to include sufficient data to model price fluctuation trends, but not too long, so that, older (and now irrelevant) trends do not affect estimating accuracy.

The use of historical data to estimate current prices, and the fact that construction prices fluctuate over time, poses another issue: a data-driven model based on past data might effectively estimate prices for projects in the past, but might not be effective for current projects (old data = old estimates, even for a short look-back period). To address this issue, and to incorporate the time dimension into the estimating process, this thesis relays on the results of a parallel study conducted by Pakalapati (2018). In this concurrent study, Pakalapati proposed a methodology to determine optimal look-back periods for specific cost items or commodities, as well as a Construction Cost Indexing System (CCIS) to adjust the old estimates produced with historical data. Pakalapati's methodology is used in this thesis to determine the amount of previous data required to create a non-linear regression model similar to the one shown in Figure 1.1. Then, the HMA unit price obtained from the model is brought into present dollars using the CCIS developed by Pakalapati. According to Pakalapati, a two-year look-back period and the use of a quarterly CCIS would maximize estimating accuracy for the case study item.

1.1.3.3 Location

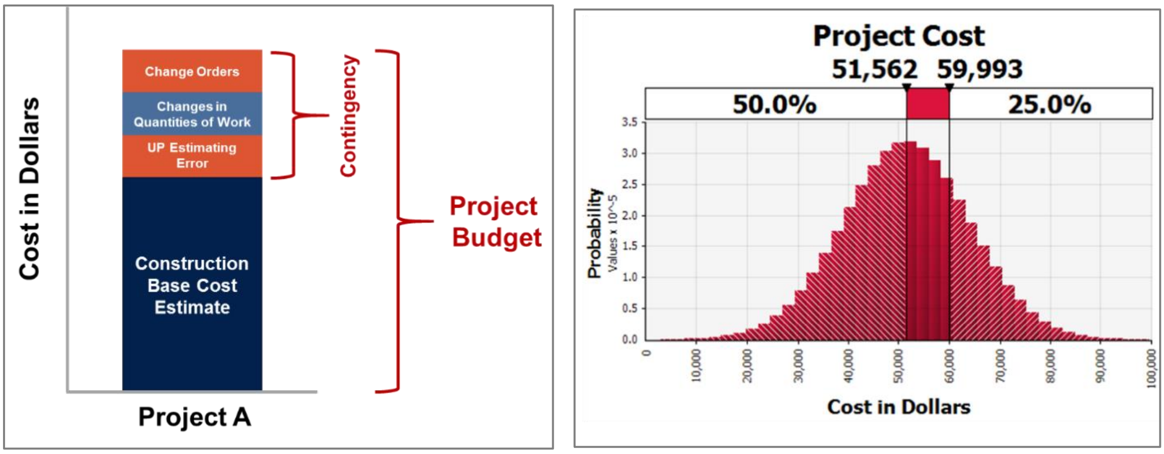
Different geographic locations bring different challenges and project requirements. Therefore, different prices could be expected for the same type of work or commodity on different locations. Price variability across the country, or even at the state level, depends on multiple factors including: local climate and geological conditions; availability of qualified local labor, suppliers, and subcontractors; and local applicable regulations (Akanni, Oke, and Akpomiemie 2015; Cuervo and Sui 2003; Kaming et al. 1997). Traffic characteristics at the jobsite are also a key factor to be considered when estimating costs for transportation construction projects since those dictate the strictness or laxity of traffic control requirements increasing or reducing construction costs.

The location dimension has been addressed in this study by developing a HMA Location Cost Index (LCI). This index was developed by dividing the state of Alabama into three different regions: north, central, and south. The process to develop the LCI is described in Chapters 3 of this thesis.

1.1.3.4 Uncertainty

Traditional deterministic cost estimating practices have shown to have a limited capacity to cope with the current needs of the transportation construction industry. They fail to objectively account for the unavoidable and increasing uncertainty associated with the development of cost estimates. In an attempt to quantify this uncertainty, some STAs have already developed and implemented data-driven systems to produce stochastic cost estimates on a per project basis, also called risk-based estimates. These agencies include the STAs in Florida, Colorado, Washington State, Nevada, New Jersey, and Texas. A risk-based cost estimate is a range of possible values of final project costs with their probability of occurrence (ASCE 2013). These are probability distribution functions that allow agencies to make better estimating decisions based on specific confidence levels.

Figure 1.2 (a) shows an example of a risk-based estimate for a given project. Using the probability distribution from Figure 1.2 (b), an agency could decide to set a base cost estimate of $51,600 for this project plus a contingency of $8,400 ($59,993 - $51,562 ≈ $8,400) if a 75% confidence level is desired. It means that the agency would be 75% confident of having enough funds to complete the project. The base cost estimate in this example corresponds to the 50% confidence level. The budget contingency is intended to account for the uncertainty associated with unit price estimating errors, changes in quantities of work, and other variations in project costs due to potential change orders issued by the agency. Risk-based estimates provide STAs with the flexibility of establishing risk tolerance levels based on the specifics of each project.



(a) (b)

Figure 1.2: Risk-Based Cost Estimate – Example.

Uncertainty is also a factor considered in the proposed cost estimating methodology; however, unlike the three dimensions mentioned above (i.e. scale, time, and location), uncertainty is not an input to the estimating system. It is actually the main output of the validation process. As explained in more detail in Chapter 3, a probability distribution, similar to the one shown in Figure 1.2 (b), is developed using the deterministic HMA unit price, which is estimated with the non-linear regression function and adjusted for time and location. The distribution of percentage errors is obtained during the validation process.

1.1.3.5 Level of Competition

"The larger the number of competitors, the higher the level of competition intensity, and the lower the bid will be" (Ye, Shen, and Lu 2014; Lo, Lin, and Yan 2007). The potential participation of a large number of contractors on a bidding process is expected to increase the level of competition perceived by contractors, forcing them to reduce their profits to ensure the submission of competitive price proposals (Dikmen, Birgonul, and Gur 2007). The Federal Highway Administration (FHWA) requires, to the maximum extent possible, the use of competitive procurement processes by STAs (Delaney and Mohan 2016). An appropriate level of competition is the one that would bring reasonably low construction prices without affecting the quality of the final products. This thesis has also assessed the impact of the "level of competition" factor on HMA prices in the state of Alabama, however, there are no significant variations on this factor were found across the state Therefore, it was not considered in the proposed cost estimating system.

1.2 HMA Stochastic Three-Dimensional Cost Estimating System

Figure 1.3 is a simplified representation of the three-dimensional cost estimating system proposed in this study. The three-dimensional system by itself has only the capacity to produce deterministic HMA cost estimates. The stochastic capability of the system has been introduced as shown in Figure 1.4, which illustrates the process to develop and implement the proposed cost estimating system, from the construction of the three dimensions to the final use of the system by STA estimators to produce stochastic HMA cost estimates.

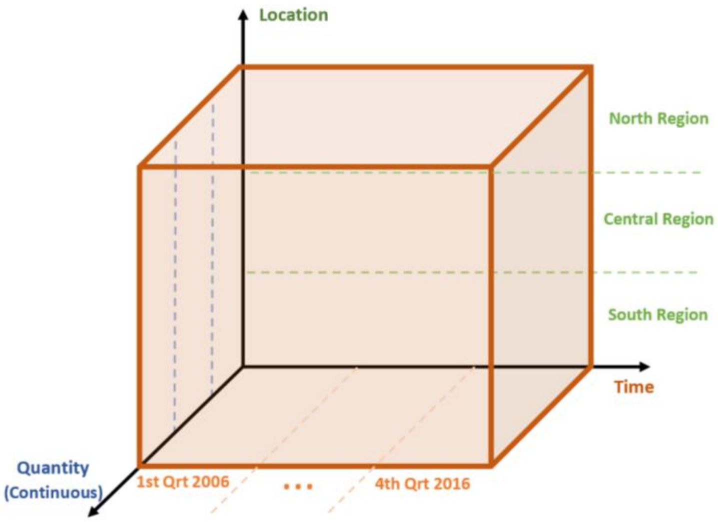


Figure 1.3: Three-Dimensional Cost Estimating Model for ALDOT.

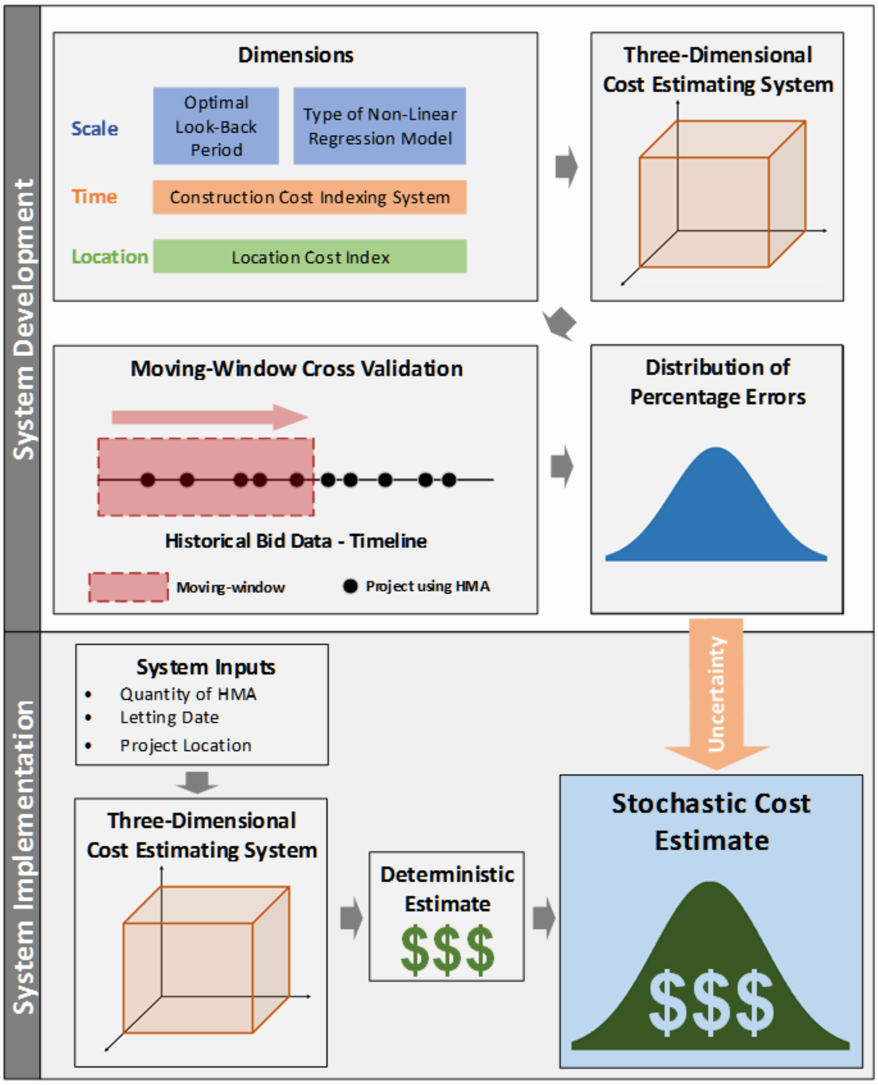


Figure 1.4: Three-Dimensional Cost Estimating Model – Development and Implementation.

The system development process shown in the top of Figure 1.4 starts by indicating the primary elements used to model each of the dimensions. To build the scale dimension it is required to know the optimal look-back period, as well as the type of non-linear regression to be used to model the quantity-unit price relationship with the data comprised in the look-back period. On the other hand, the CCIS (previously developed by Pakalapati [2018]) and LCI (developed in this study) are used to shape the time and location dimensions, respectively.

Once the three-dimensional cost estimating system has been assembled, it is validated with the MWCV. This validation process allows for the application of the system to several previous projects, providing sufficient APEs to determine the overall accuracy of the system (through the MAPE), as well as the level of reliability (standard deviation of APEs). All percentage errors, calculated with Equation 3 (not absolute values), are used to create a probability distribution of percentage errors. This distribution is then saved for future use during the actual implementation of the system to include the uncertainty factor into the cost estimating process. It should be noted that the errors in this distribution are calculated as percentages of the estimated values, unlike APEs, which are percentages of the actual values. It is due to the fact that the distribution of percentage errors is intended to indicate how far actual prices are from the estimated prices.

$$PE = \frac{A_i - E_i}{E_i} \times 100\% \qquad \textbf{Eq. 1-3}$$

Where: PE = Percentage Error

A_i = Actual unit price for HMA in project i

E_i= Estimated unit prioce for HMA in project i

For example, if the estimated HMA unit price for a given project is $71/ton and the actual price in project is $67/ton, the APE for this estimate, according to Equation 1-1, would be 6.0%, meaning that the percentage difference with respect to the actual price is 6.0%. It is irrelevant if it is greater or lower since the APE is an absolute value, but, in this case, the estimated price is 6.0% greater than the actual price. If the same percentage difference were calculated with respect to the estimated value, as in the values used to create the distribution of percentage errors, it would be -5.6%, meaning that the actual price is 5.6% lower than the estimated price.

As shown in Figure 1.4, the utilization of the proposed estimating system only requires three inputs: 1) the quantity of HMA (in tons); 2) letting date; and 3) project location. For the purposes of this study, the letting date is assumed to be the current date since the proposed system is intended to be used at design completion shortly before advertising the project. The letting date is used to define the beginning and end of the look-back period, as well as to identify the base and current index values required to adjust unit prices over time. On the other hand, the location input is a categorical variable that can take on three possible values, based on the geographical location of the project: north, central, or south region.

Once the three inputs are entered into the three-dimensional cost estimating system, they are processed into a deterministic estimate for the expected unit price to be paid by ALDOT for the stated quantity of HMA, in the current date, and in the specified location. Recognizing the unavoidable uncertainty inherent in construction cost estimates, the deterministic output of the system is multiplied by the distribution of percentage errors that resulted from the validation of the system during its development. The product of this multiplication is a stochastic estimate that represents all possible HMA unit prices and their probability of occurrence in the project under consideration.

The processes described in this section are explained in more detail throughout this thesis, as they are applied to the HMA pay item most commonly used in ALDOT's construction contracts (hereinafter referred to as the case study item): "*Superpave Bituminous Concrete Wearing Surface Layer, 1/2" Maximum Aggregate Size Mix, ESAL Range C/D – Item ID 424A360.*"

1.3 Research Objectives

The main objective of this study is to develop a HMA stochastic cost estimating system that accounts for the main factors influencing the estimating process. The cost influencing factors initially considered in this study are project scale, time (inflationary effects and changes in the construction market over time), geographic location, estimating uncertainty, and level of competition. The following sub-objectives have been identified as necessary steps to accomplish the main research objective:

- Assess and model the relationship between each of the cost-influencing factors and HMA prices paid by ALDOT. This sub-objective led to the development of the LCI and the methodology to produce stochastic estimates out of a distribution of percentage errors.
- Design and develop a system that integrates the impacts of all cost-influencing factors to maximize accuracy and reliability in cost estimates for HMA in projects executed by ALDOT.
- Develop and implement a reliable research validation approach to demonstrate the effectiveness of the proposed HMA cost estimating system. This sub-objective led to the development of the proposed innovative MWCV methodology.

The objectives and sub-objectives outlined above were achieved following the research plan presented in Chapter 3 Methodology, which can be summarized in the following seven tasks:

1. Conduct an extensive literature review on construction cost estimating.

2. Collect, clean, and explore ALDOT historical bid data from all projects awarded between 2006 and 2016.

3. Identify a relevant HMA case study item to demonstrate and validate the effectiveness of the proposed cost estimating system.

4. Identify and assess the main cost-influencing factors affecting HMA cost estimating procedures, and model the relationship between these factors and HMA unit prices.

5. Develop a framework that integrates all the elements and quantitative methods used to model the relationships established in Task 4.

6. Apply the proposed MWCV approach to measure the performance of the proposed three-dimensional cost validation approach.

7. Analyze the results from the MWCV approach and formulate conclusions and recommendations.

1.4 Expected Outcome

The proposed methodology not only increases accuracy in the estimation of costs associated with the most relevant paving material used by ALDOT, but also provides ALDOT with a better understanding of the main factors influencing cost estimating in asphalt paving projects. A better understanding of these factors is expected to be reflected in better decision-making at different management levels, as well as to improve ALDOT's budget control capabilities and resource allocation procedures. Moreover, a more effective and efficient use of the limited available resources would improve ALDOT's capacity to handle the financial challenges brought by the unavoidable increasing funding gap that has been affecting the transportation construction industry during the last two decades.

1.5 Organization of Thesis

This thesis has been divided into five chapters in an attempt to present the research efforts associated with this study in a logical and organized manner. Following this introductory chapter,

Chapter Two - Literature Review, summarizes the existing literature on cost estimating, including previous studies, research reports, and industry manuals. Specifically, this chapter explains the cost estimating process across project development phases and describes the different cost estimating approaches currently used by STAs, paying special attention to bid-based and risk-based cost estimating, which are the primary concern of this thesis. ***Chapter Three - Methodology***, describes the research plan that led to the development of the HMA cost estimating methodology proposed in this study. The research plan is presented in detail, describing all research procedures and tools used to develop and validate the stochastic three-dimensional cost estimating methodology. ***Chapter Four - Results and Discussion***, presents the main results of this study and a critical analysis of these findings. ***Chapter Five - Conclusions and Recommendations***, summarizes the main conclusions, findings, and contributions to the body of knowledge made by this thesis. Finally, this chapter outlines some research topics that should be considered to take this study's findings and contributions to the next level.

CHAPTER TWO:

LITERATURE REVIEW

2.1 Introduction

Cost estimating is critical at any project phase, from conceptual design and planning to the operation and maintenance of the built assets. The estimation of construction costs may be a complex process involving different types of challenges. It seems that it was not until the mid-1960s that these challenges started to be addressed in the U.S. through formal research. From 1965 and onwards, there has been an exponential increase in research conducted on how to develop better and more effective cost estimating systems at all project phases (Trost and Oberlender 2003). Even though there is a great variety of project life cycle configurations among STAs, a construction project can be represented, in a simplified manner, as a sequence of three generic phases: conceptual; engineering/design; and execution phase (Phaobunjong 2002). Each phase poses different estimating challenges and requirements. With growing information provided from the infancy to project completion, the estimator's understanding of the project increases as the project moves across development phases allowing for better estimating accuracy (Manfredonia 2016). Figure 2.1 shows the classification of construction cost estimates based on the three generic project phases.

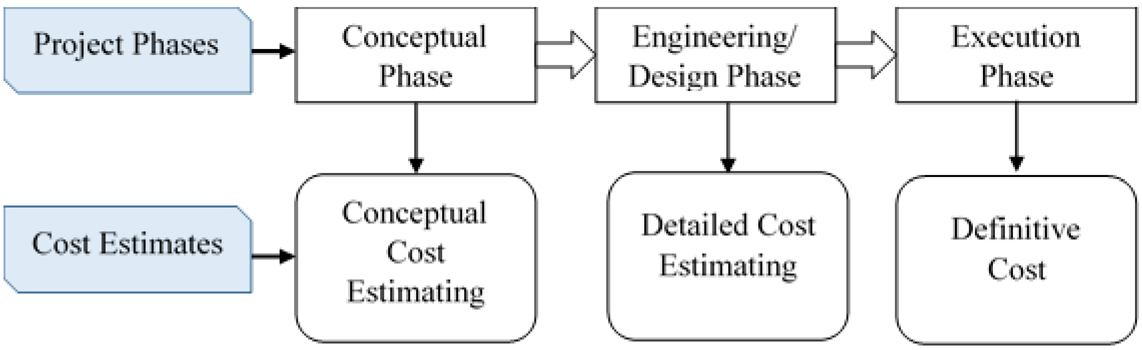


Figure 2.1: Construction Cost Estimating Phases.

First, a conceptual cost estimate is developed on an early stage of project development. Since this is an early estimate, it is usually calculated with little information and with a roughly defined scope of work; therefore, conceptual estimates are expected to be the most inaccurate among all the types of estimates developed along the life cycle of construction projects (Phaobunjong 2002). However, in cost estimating, the level of accuracy should not always be considered as directly proportional to the value of the estimates from a planning and management perspective (Liu and Zhu 2007). Regardless to the high uncertainty inherent in conceptual estimates, they are highly necessary to determine the financial viability of candidate projects, define feasible scopes of work, and to support other strategic business decisions. These estimates are used to design project portfolios that match the financial capabilities of both construction owners and contractors, as well as to compare the cost implications associated with different construction methods and materials (Manfredonia 2016). The lack of project-specific information at such an early stage makes it hard for estimators to produce effective conceptual estimates (Kim, An, and Kang 2004); however, the existing literature provides a number of tools and methodologies to aid STAs on this matter (Fragkakis et al. 2010; Asmar et al. 2011; Juszczyk 2013; Sonmez 2004; Phaobunjong 2002).

Second, a detailed cost estimate is developed by the end of the engineering/design phase or at completion of design work. This estimate is based on a detailed design for a well-defined scope of work, and with information on the type of contract to be awarded, construction methods and materials to be used in the project, and schedule constraints and required milestones and deliverables –among other project-specific characteristics. Due to the greater amount of project related information available at this phase, detailed cost estimates are expected to be considerably more accurate than conceptual estimates (Phaobunjong 2002; Manfredonia 2016). This is the last cost estimating attempt made by owners before awarding a contract. It is used to verify if the intended project is still feasible and to make any needed adjustments to the project budget (AASHTO 2013). These estimates are also intended to serve as a point of reference during potential price negotiations with contractors. The HMA cost estimates to be developed with the system proposed in this thesis full within this second category. The stochastic three-dimensional cost estimating system is aimed to be applied at design completion or by the end of the design phase after estimating the amount of HMA (in tons) to be used in the intended project.

Finally, the execution phase begins with the signature of the contract. The definitive cost estimate at this phase is the result of the price proposal submitted by the selected contractor and/or any price negotiations held between the owner and the contractor before signing the contract. It is also referred to as the actual contracted price for a construction project. This estimate is compared against the detailed estimate previously developed by the STA to determine if further budget adjustments are required (Manfredonia 2016; Phaobunjong 2002). The contracted price is used for payment purposes and to monitor the financial status of the project throughout the construction period (Hendrickson and Au 1989).

After the conceptual estimate, all subsequent estimates are built using the cost estimate from the previous phase as the starting point (Hendrickson and Au 1989). Thus, each subsequent estimate can be considered as a revised version of the previous estimate, with the revision occurring in the light of the additional information that became available in between phases. It must be noted that all three types of estimates described in section are equally important at their respective phases.

2.2 Current Cost Estimating Approaches used in Transportation Projects

This section discusses the four primary cost estimating approaches outlined in the Practical Guide to Cost Estimating, published in 2013 by the American Association of State Highway Transportation Officials (AASHTO): 1) parametric; 2) bid-based; 3) cost-based; and 4) risk-based estimating. In this guidebook, each of these estimating approaches is associated with different project development phases, as shown in Table 2.1. The four project development phases in Table 2.1 correspond to a more detailed phased configuration of the conceptual and engineering/design phases discussed in the previous section and showed above in Figure 2.1. This table also shows the level of project maturity at each project development phase (as a percentage of all the planning and design work required to successfully award the contract), as well as the most suitable cost estimating approach at each phase with its respective expected accuracy. Table 2.2 summarizes the main activities contained within each project development phase.

Table 2.1 Cost Estimating Classification (Adapted from AASHTO 2013)

<table>
<tr><th>Project Development Phase</th><th>Project Maturity (% project definition completed)</th><th colspan="2">Estimating Approach</th><th>Estimating Accuracy</th></tr>
<tr><td rowspan="2">Planning</td><td>0% to 2%</td><td>Parametric</td><td rowspan="5">Risk-Based (optional combination with other approaches)</td><td>-50% to +200%</td></tr>
<tr><td>1% to 15%</td><td>Parametric or Historical Bid-Based</td><td>-40% to +100%</td></tr>
<tr><td>Scoping</td><td>10% to 30%</td><td rowspan="3">Historical Bid-Based or Cost-Based</td><td>-30% to +50%</td></tr>
<tr><td>Design</td><td>30% to 90%</td><td>-10% to +25%</td></tr>
<tr><td>Final Design</td><td>90% to 100%</td><td>-5% to +10%</td></tr>
</table>

Table 2.2 Project Development Phases and Typical Activities (Adapted from AASHTO 2013)

Project Development Phase	Typical Activities
Planning	Purpose and need; improvement or requirement studies; environmental considerations; right-of-way considerations; schematic development; project benefit/cost feasibility; public involvement/participation; interagency conditions.
Scoping	Environmental analysis; alternative analysis; preferred alternative selection; public hearings; right-of-way impact; environmental clearance; design criteria and parameters; funding authorization (programming).
Design	Right-of-way development and acquisition; preliminary plans for geometric alignments; preliminary bridge layouts; surveys/utility locations/drainage.
Final Design	Plans, specifications, and estimate (PS&E) development—final right-of-way acquisition; final pavement and bridge design; traffic control plans; utility drawings; hydraulics studies/final drainage design; final cost estimates.

As mentioned before, the greater the level of project development, the more accurate the cost estimates. That explains the reduction in the estimating accuracy ranges in the last column of Table 2.1. The risk-based estimating is presented in the AASHTO guidebook as an optional version of the other three approaches. The following section present a more detailed description of each of the four estimating approaches.

2.2.1 Parametric Estimating

Parametric cost estimating techniques are usually applied during early project development in the planning phase and at a conceptual level. "Parametric estimating techniques are primarily used to support development of planning or early scoping phase estimates when minimal project definition is available. Statistical relationships or non-statistical ratios, or both, between historical data and other project parameters are used to calculate the cost of various items of work (i.e., center lane miles or square foot of bridge deck area)" (AASHTO 2013). The use of parametric estimating models have surge as an alternative to the traditional (and almost discontinued by STAs) expert-based approach, in which conceptual cost estimates are solely the result of the estimator's subjective opinions built from previous construction experiences (Phaobunjong 2002). Thus, the expert-based approach may not appropriately account for all the important factors influencing the cost estimating process if the estimators' experience does not match the scope of the current project. As a data-driven approach, parametric cost estimating methods have added some needed objectivity to the development of conceptual estimates and has proven to improve estimating accuracy during early project phases (Trost and Oberlender 2003; Kwak and Watson 2005).

Parametric estimating models are usually straightforward user-friendly models (Bajaj, Gransberg, and Grenz 2002). However, these models have a limited capacity to handle the high levels of cost uncertainty during early project development phases. According to Harbuck (2002), the main sources of uncertainty in construction cost estimating are 1) changes in the scope of work; 2) potential design changes; 3) errors in the calculation of quantities of work and unit costs; and 4) unforeseen site conditions. When considering the nature of these four factors, it is easy to understand how difficult it could be for a STA to quantify their potential impact on early construction cost estimates to be developed only from a preliminary scope of work.

2.2.2 Historical Bid-Based Estimates

As shown in Table 1.2, historical bid-based cost estimates are used at all project development phases. Likewise, this is recognized as the most common cost estimating approach currently used by STAs, being used to some extend by all STAs (Anderson et al. 2009; Schexnayder et al. 2003). The AASHTO guidebook defines bid-based estimating as an approach that "uses data from recently let contracts as the basis for determining estimated unit prices for a future project" (AASHTO 2013). As per AASHTO guidelines, bid-based estimates are usually developed with data from projects awarded during the last one or two years, which also seems to be the policy adopted by ALDOT. Longer look-back periods may be considered when the most recent two years of historical data do not provide sufficient relevant data. However, STAs does not count with a mechanism to make objectives decisions on the length of look-back periods for bid-based estimating. The decision of whether to use one, two, or more years of data is mainly based on estimators' subjective judgments.

Besides the determination of optimal look-back periods for data retrieval, another challenge identified in this study with respect to the implementation of bid-based estimating techniques is related to the estimation of current prices using historical data. Historical data may be more suitable to estimate prices for past projects than to estimate current prices, unless the estimates are adjusted for observed inflationary effects and price fluctuations during the look-back period. Since the methodology proposed in this thesis is based on bid-based estimating techniques, the author was required to find the way to overcome these two major challenges (determination of optimal loo-back periods and price adjustment for time effects). As mentioned before in Chapter 1, these two issues were addressed in this thesis by using findings from a concurrent study conducted by Pakalapati (2018).

One of the reasons why STAs prefer to use a bid-based approach over other estimating approaches, is that historical bid-based estimating does not require an extensive amount of previous experience from estimators since most of the required experience is replaced by trends and relevant statistic parameters extracted from the bid data. This is an important observation considering the increasing retirement rates that STAs have been experiencing during the last decade, which are leaving them with fewer, and overall, less experience staff.

2.2.3 Cost-Based Estimates

Cost-based estimates are composite estimates resulting from the aggregation of seven cost elements: time, equipment, labor, subcontractor, material, overhead, and profit (AASHTO 2013). In comparison to bid-based cost estimating, a cost-based approach demands greater estimating efforts to apply quantitative procedures at a deeper level of detail. Bid-based techniques are frequently used to support cost-based estimates by providing prices for one or more of the seven elements mentioned above (AASHTO 2013). Even though STAs tend to prefer the use of bid-based cost estimating techniques, a cost-based approach should be favored if the estimator perceives high levels of uncertainty in the bid-based estimates (AASHTO 2013).

Although cost-based estimates are more complex and require greater estimating efforts, if well developed, they could facilitate better planning and design by forcing the project staff to better understand the project, allowing for the timely identification of issues that otherwise would be discovered during construction. Problems found during construction, and resulting from poor planning and design errors, usually have a negative impact on projects in terms of increased costs or extended project durations, or both. A better understanding of construction projects and timely identification of design and constructability issues through the appropriate implementation of cost-based estimating techniques have allowed the Utah Department of Transportation to save about $11 million due to a reduced number of change orders issued during construction (Utah Construction & Design 2013).

2.2.4 Risk-Based Estimates

"Risk-based cost estimation entails developing probable cost for project components, and the project, based on identified known quantities and costs and contingency developed from a list of identified uncertainties from both opportunities and threats and their potential impact on the project" (Shane et al. 2015). In a simpler definition provided by AASHTO (2013), risk-based estimating basically refers to the combination of risk analysis techniques with any of the estimating approaches presented above to convert the typical deterministic outputs of those approaches into stochastic cost estimates in the form of probability distributions. "This approach is used to establish the range of total project cost and to define how contingency should be allocated among the critical project elements" (AASHTO 2013). The concept of risk-based estimating has been further explained in Chapter 1, Section 1.3.4 Uncertainty.

Even though the use of risk-based estimating techniques has not been adopted by most STAs, it is becoming increasingly popular among these transportation agencies. Some STAs, including the Washington State (WSDOT) and Montana (MDT), have already developed their own risk-based estimating systems (Molenaar 2005; Gardner 2015). The estimating system for WSDOT was developed by Molenaar (2005) and was intended for projects greater than $100 million, referred to as "Highway Megaprojects." Molenaar found that the main benefit provided by WSDOT risk-based cost estimating system was better budget control and resource allocation, as well as an increase in public confidence due to a more transparent communication. Molenaar estimating system was found to be effective; however, it was estimated that its implementation would have a cost of about $3 million for WSDOT. Likewise, its implementation required the employment of a risk-analyst expert, unlike the methodology proposed in this thesis.

On the other hand, MDT's risk-based estimating system was developed by Gardner (2015) using 189 paving projects previously awarded by this agency. Gardner's used multiple regression and artificial neural network models to produce stochastic conceptual cost estimates using the 14 input variables listed below. Even though some of the input variables used in Gardner's systems should be considered for further improvement of the estimating methodology proposed in this thesis, these 14 inputs do not take into consideration adjustments required to counteract the impact of inflation and construction market volatility over time. Likewise, Gardner's system considers specific topographic and geotechnical characteristics of the jobsite, but fail to consider potential regional market conditions that might influence construction pricing, which this thesis has found to significantly affect HMA prices in Alabama.

2.2.5 Classification of the Proposed Cost Estimating System

Based on the information presented above the proposed system can be classified as a combination of the bid-based and risk-based cost estimating approaches since the system uses historical data from previous projects awarded by ALDOT to produce a stochastic HMA cost estimate. Likewise, it is important to remember that the proposed stochastic three-dimensional cost estimating system is to be used during the final design phase at design completion (see Table 2.1).

2.3 Validation of Data-Driven Cost Estimating Models

Cross validation (CV) is a common research approach to assess the performance of data-driven models, such as cost estimating models. A typical CV process is performed in four general steps: 1) first, the available data is split into a training and a testing dataset; 2) the training dataset is then used to develop the model; next, 3) the model is applied to the testing dataset to estimate the values of the independent variable(s) on each observation; and finally, 4) the estimated values are compared against the actual values of the independent variable(s) in the testing dataset, and the result of this comparison is analyzed to assess the performance of the model. The CV process is intended to simulate the actual implementation of the model, so that, the CV results are assumed to reflect the level of performance that should be expected by the final users. However, the literature review has revealed some major issues with traditional CV approaches that may compromise the integrity of the validation results.

A review conducted by Gardner (2015) on several data-driven construction cost estimating models presented in the literature revealed that in spite of the fact that some of this models showed high effectiveness when validated by their respective authors, they have never been implemented by STAs due to their questionable validation processes. Some positive CV results are the result of testing the performance of the models using samples of projects with narrow scopes of work (e.g. all projects in the testing dataset have similar characteristics and are for the same type of work) or are performed with very small testing samples, in some cases, using only two projects for validation (Gardner 2015).

Rueda (2016) identified another problem associated with the traditional CV procedures used in cost estimating modeling in the transportation industry. Observations for the training and testing datasets are usually assigned in a random manner. However, as suggested by Rueda (2016), a random partition of the data might not be appropriate for a bid-based cost estimating model since model developers might end using recent projects to estimate older prices, which would be an impossible scenario during actual implementation. This simulated scenario would be a case in which an estimator uses future prices (which are obviously unknown) to estimate current prices. According to Rueda (2016), model developers should intentionally (instead of randomly) place older projects in the training dataset and the more recent projects in the testing dataset for the validation of bid-based cost estimating models. In that way, the CV process would not only "determine the accuracy of the construction cost estimating models, but also the ability of bid data from previous projects to estimate current construction costs" (Rueda 2016).

Although Rueda's approach is a more accurate representation of the actual model implementation, this thesis has identified another opportunity to improve CV procedures in bid-based estimating, allowing for an even more accurate representation of the eventual utilization of the estimating models. In the implementation of bid-based estimating models, estimators use historical bid data from the most recent projects contained in the look-back period, so that, the look-back period for data retrieval ends just before the current date. Therefore, an effective CV process for bid-based estimating should allowed for the adjustment of the look-back period to match the period of time immediately preceding the current date. That is not possible with a fixed training dataset as the one typically used in CV. The proposed MWCV approach, described in detail in Section 3.6 and Chapter 4, is intended to address this CV limitation, by allowing for a dynamic partition for training and testing observations.

CHAPTER THREE:
DATA COLLECTION AND CLEANING

3.1 Introduction

After concluding the literature review efforts summarized in Chapter 2, the author proceeded to collect, clean, and explore ALDOT's historical bid data for all projects awarded between 2006 and 2016, a total of 3,661 projects. This chapter describes the data collection process and the tools and statistical methods used to identify and remove outliers, as well as to reformat the available data into a tidy format, which is more suitable for data analysis purposes. Finally, the chapter describes the Exploratory Data Analysis (EDA) that led the author to identify and better understand the main cost-influencing factors affecting HMA prices in Alabama, to identify the most relevant HMA item to be used as the case study item, and to detect and fix errors and inconsistencies in the tidy dataset.

3.2 Data Collecting and Cleaning

After gaining a better understanding of the research problem through the comprehensive literature review, and after designing a thoughtful research plan to address the problem, the author proceeded with the data collection process, which consisted in mining historical pricing data from the bid tabulations for all projects awarded by ALDOT between 2006 and 2016. A total of 3,661 projects were awarded during this 11-year period. Data was extracted from ALDOT's Bid Tabulations website. The bid data published on this website is available in Portable Document Format (PDF), which is not a suitable format for data manipulation and processing. Figure 3.1 is a screen capture from one of the PDF files. This figure shows a few unit prices submitted by three bidders for a bridge replacement project in Etowah County, Alabama. Although this study is mainly focused on asphalt paving projects, the proposed methodology intended to analyze the behavior of HMA prices over time, and paving activities are required, to some extent, in most ALDOT's projects. Therefore, relevant data could be provided by non-paving projects, so that, bid data was collected and cleaned for all types of projects awarded by ALDOT between 2006 and 2016.

Alabama Department of Transportation
Tabulation of Bids

DATE: 1/29/2016
Page: 1

Call Order: 001 **Contract ID:** 20160129001 **County:** ETOWAH
Letting Date: January 29, 2016 **Area/District:** 0105 **Contract Time:** 310 Working Days
Contract Description: **Project(s):** BR-0001(565)

REPLACEMENT OF DUAL BRIDGES(GRADE, DRAIN BASE, PAVE,SIGNALS) ON SR-1 (US-431) OVER BLACK CREEK TO INCLUDE THE REMOVAL OF A PEDESTRIAN BRIDGE AND RETAINING WALL IN GADSDEN

Line No / Item ID Item Description Alt Set / Alt Member	Quantity and Units	(1) BELL & ASSOCIATES CONSTRUCTION, L.P. Unit Price	Ext Amount	(2) IKAROS, LLC Unit Price	Ext Amount	(3) WRIGHT BROTHERS CONSTRUCTION COMPANY, INC. Unit Price	Ext Amount
SECTION: 0001 Total							
0010 206A000 Removal Of Old Bridge, Station 120+58.85, NBR	(1) LS	200,000.00	200,000.00	150,000.00	150,000.00	115,000.00	115,000.00
0020 206A001 Removal Of Old Bridge, Station 120+42.16, SBR	(1) LS	200,000.00	200,000.00	100,000.00	100,000.00	115,000.00	115,000.00
0030 206A002 Removal Of Old Bridge, Station 123+75.00, Pedestrian Bridge	(1) LS	270,000.00	270,000.00	600,000.00	600,000.00	100,000.00	100,000.00

Figure 3.1: ALDOT Bid Tabulations - PDF Format.

A web-based free format conversion application was used to reformat the data into a more suitable format compatible with Microsoft Excel (hereinafter referred to as Excel). After the conversion process, the data arrangement in the Excel spreadsheets was still similar to the arrangement provided in the PDF file, which is still not ideal for data processing purposes. Additionally, the converted data presented several critical format inconsistencies such as discontinued or shifted columns and unintended combined cells. Given the large amount of collected data, manual reformatting and correction inconsistencies was not an option. Part of the data cleaning efforts in this study were aimed to develop an Excel data cleaning spreadsheet carefully designed to identify and correct all formatting inconsistencies while rearranging the data into a tidy format. A screen capture of the tidy dataset is shown in Figure 3.2.

	A	B	F	H	K	CH	CI	CJ	CK	CL	CM	C
1	PostingDate	ContractID	County	Year	Quarter	Item ID	Item Description	Units	Quantity	Unit Price 1	EXT1	Unit F
2	1/17/2006	20060106001	CULLMAN	2006	1	405A000	Tack Coat	GAL	15046	1.35	20312.10	1.24
3	1/17/2006	20060106001	CULLMAN	2006	1	408A053	Planing Existing Pavement (Approximately 2.10" Thru 3.0" Thick)	SQYD	250765	1.53	383670.45	3.9
4	1/17/2006	20060106001	CULLMAN	2006	1	408A054	Planing Existing Pavement (Approximately 3.10" Thru 4.0" Thick)	SQYD	300	3	900.00	18.84
5	1/17/2006	20060106001	CULLMAN	2006	1	410C000	Contractor Retained Profilograph	EACH	1	15000	15000.00	13600
6	1/17/2006	20060106001	CULLMAN	2006	1	410H000	Material Remixing Device	EACH	1	90000	90000.00	12690
7	1/17/2006	20060106001	CULLMAN	2006	1	420A015	Polymer Modified Open Graded Friction Course	TON	11285	61.11	689626.35	58.98
8	1/17/2006	20060106001	CULLMAN	2006	1	423A002	Stone Matrix Asphalt Wearing Layer, 3/4" Maximum Aggregate Size	TON	25077	64.11	1607686.47	46.39
9	1/17/2006	20060106001	CULLMAN	2006	1	424B659	Superpave Bituminous Concrete Upper Binder Layer, Leveling, 1" Maximum Aggre	TON	50	200	10000.00	150
10	1/17/2006	20060106001	CULLMAN	2006	1	600A000	Mobilization	LUMP	1	286995.78	286995.78	28699
11	1/17/2006	20060106001	CULLMAN	2006	1	701A028	Solid White, Class 2, Type A Traffic Stripe (0.06" Thick) (6" Wide)	MILE	17	1737.2	29532.40	1825
12	1/17/2006	20060106001	CULLMAN	2006	1	701A032	Solid Yellow, Class 2, Type A Traffic Stripe (0.06" Thick) (6" Wide)	MILE	17	1737.2	29532.40	1825
13	1/17/2006	20060106001	CULLMAN	2006	1	701A041	Broken White, Class 2, Type A Traffic Stripe (0.09" Thick) (6" Wide)	MILE	17	1060.5	18028.50	1115
14	1/17/2006	20060106001	CULLMAN	2006	1	701B009	Dotted Class 2, Type A Traffic Stripe (0.09" Thick)(6" Wide)	LF	500	1.01	505.00	1.06
15	1/17/2006	20060106001	CULLMAN	2006	1	701C000	Broken Temporary Traffic Stripe	MILE	33	626.2	20664.60	657.5
16	1/17/2006	20060106001	CULLMAN	2006	1	701C001	Solid Temporary Traffic Stripe	MILE	66	686.8	45328.80	721.2
17	1/17/2006	20060106001	CULLMAN	2006	1	703A002	Traffic Control Markings, Class 2, Type A	SQFT	50	5.05	252.50	5.3
18	1/17/2006	20060106001	CULLMAN	2006	1	705A030	Pavement Markers, Class A-H, Type 2-C	EACH	66	5.56	366.96	5.83
19	1/17/2006	20060106001	CULLMAN	2006	1	705A031	Pavement Markers, Class A-H, Type 1-A	EACH	1086	5.56	6038.16	5.83
20	1/17/2006	20060106001	CULLMAN	2006	1	731A010	Traffic Counting Units, Type K	EACH	1	7000	7000.00	4410
21	1/17/2006	20060106001	CULLMAN	2006	1	731A014	Traffic Counting Units, Type O	EACH	1	7500	7500.00	5040
22	1/17/2006	20060106001	CULLMAN	2006	1	740B000	Construction Signs	SQFT	979	7.95	7783.05	8.91
23	1/17/2006	20060106001	CULLMAN	2006	1	740C000	Special Construction Signs	SQFT	20	7.95	159.00	9.47
24	1/17/2006	20060106001	CULLMAN	2006	1	740D000	Channelizing Drums	EACH	200	60.8	12160.00	55.68
25	1/17/2006	20060106001	CULLMAN	2006	1	741C010	Portable Sequential Arrow And Chevron Sign Unit	EACH	2	2500	5000.00	1350
26	1/17/2006	20060106001	CULLMAN	2006	1	742A001	Portable Changeable Message Sign, Type	EACH	2	15000	30000.00	5600
27	1/17/2006	20060106001	CULLMAN	2006	1	998A000	Construction Fuel (Maximum Bid Limited To $ 145,000.00)	LUMP	1	0.00	0	
28	1/17/2006	20060113002	MONTGOMERY	2006	1	206D001	Removing Guardrail	LF	2978	2.5	7445.00	1.25
29	1/17/2006	20060113002	MONTGOMERY	2006	1	206E008	Removing Guardrail End Anchor (All Type	EACH	34	150	5100.00	130
30	1/17/2006	20060113002	MONTGOMERY	2006	1	210D001	Borrow Excavation (Loose Truckbed Measurement)	CUYD	1380	25	34500.00	31

Figure 3.2: Tidy Format.

"Tidy datasets are easy to manipulate, model and visualize, and have a specific structure: each variable is a column, each observation is a row, and each type of observational unit is a table" (Wickham 2014). There is only one type of observational unit in this study: pay items included in contracts awarded by ALDOT between 2006 and 2016. Thus, there is only one table, with each row referring to a single pay item used in a given project, and with the columns presenting all the available attributes and information associated with each pay item in each contract. Information provided for each item on each row includes, but it is not limited to, item identification number; item description; awarded quantity; unit of measurement; number of the contract in which the item was used; project location (county[ies]); number of bidders proposing on the project, names of bidders, and unit price submitted by each bidder for the same item.

The tidy dataset has 169,947 rows and 131 columns. It should be noted that ALDOT has a standard list of pay items and some of them are frequently used in different projects. It means that the same item may appear several times in the dataset, but each time that it appears, it corresponds to its use in a different project. A total of 5,246 different pay items have been used by ALDOT in the 3,661 projects contained in this dataset.

Initially, all 11 years of collected data were considered in this study. However, while analyzing the data, the author found an apparent trend change in HMA prices in 2010, suggesting that something may have changed the conditions of the paving construction market in the state of Alabama during that year. Since this study is intended to demonstrate the applicability of the proposed methodology in the current construction market, the author decided to limit the validation process to the most recent six years of data, from 2011 to 2016.

3.2.1 Outlier Detection and Removal

A critical part of data cleaning processes consists of removing those observations that do not appear to belong with the rest of the data, generally called outliers. "Usually, the presence of an outlier indicates some sort of problem. This can be a case that does not fit the model under study, or an error in measurement" (Cho et al. 2010). In this study, the author used two main outlier identification methods strategically selected and applied to serve different purposes.

The first outlier detection approach used in this thesis was the modified Z-score method, which was applied following the guidelines provided by Iglewicz and Hoaglin (1993). This method was applied at the pay item level (to each row) in order to identify outliers among the unit prices received by ALDOT for the same item under the same contract, meaning that the unit prices were estimated for the same amount of HMA. While some of these errors may be due to typographical mistakes or the misinterpretation of the scope contained within the unit price, many of them are the result of unbalanced bids (Rueda 2016). "A bid is considered unbalanced if the unit rates are substantially higher or lower, in relation to the estimate and the rates quoted by other bidders" (JICA 2000). A contractor unbalance a bid to either protect its intended profit and fixed costs (which could be partially lost if actual quantities of work are less than the original expected quantities), to maximize profits by taking advantage of errors in the quantities of work in the solicitation documents, or to inflate prices for early activities to reduce financial costs (the cost of borrowing money) (FHWA 1988). Regardless of the possible ethical implications typically associated with unbalanced bids, it is fact that this practice is currently used by construction contractors, and it is also a fact that unbalanced bids may affect the performance of bid-based estimating models.

The modified Z-score method was applied using Equation 3-1. The reason behind the use of this method is that outliers are identified using the sample median ($\tilde{x}$) and the median absolute deviation

(MAD) making it more suitable for small samples. It should be noted that this method was applied at the pay item to compare bids submitted by different contractors under the same contract, so that, it was applied to relatively small samples. The average number bids received by ALDOT for a single contract is between 3 and 4. The maximum number of vendors competing for a contract between 2006 and 2016 was 21, and it only occurred in a single contract. Other more commonly used outlier detection methods rely on the sample mean and standard deviation to identify outliers. However, these two statistics are more sensitive to extreme values in small samples, increasing the risk of not detecting outliers that should be discarded (Iglewicz and Hoaglin, 1993). Based on Iglewicz and Hoaglin guidelines, all unit prices with absolute modified Z-score greater than 3.5 ($|M_i| > 3.5$) were removed from the dataset.

$$M_i = \frac{0.6745(X_i - \tilde{x})}{MAD} \qquad \textbf{Eq. 3-1}$$

Where: M_i = Modiafied Z-Score for Observation i

MAD = Median Absolute Deviation = {|X_i – Median|}

x_i = Value of Observation i

$\tilde{x}$ = Median of All Observations

The second outlier detection approach used in this thesis is the Robust Regression and Outlier Removal method (ROUT). This method was developed and proposed by Motulsky and Brown in 2006. In this method, Motulsky and Brown combine robust regression and non-linear regression techniques to identify values that could be significantly apart from the regression equation. This second filter was used to detect outliers during the development of the non-linear regression models to correlate quantity and HMA unit prices. These are outliers not detected by the modified Z-score method, including those resulting from unusual project requirements that may force all contractors to bid outside the typical unit price ranges, this is actually the main reason to use this method. Since the modified Z-score method compares unit prices for the same item under a given contract, this method may find no outliers if all bidders are forced to submit unit prices substantially higher (or lower) than those typically pay by ALDOT for the same pay item in other projects. The ROUT method was applied using GraphPad Prims 7, a statistical software equipped with a ROUT function that can be activated during the development of non-linear regression models. Figure 3.3 shows an example of the output yielded by this software. This is a non-linear regression model for the case study item developed using all unit prices recorded between 2006 and 2016. All red data points are outliers detected by the ROUT method and excluded from the regression analysis.

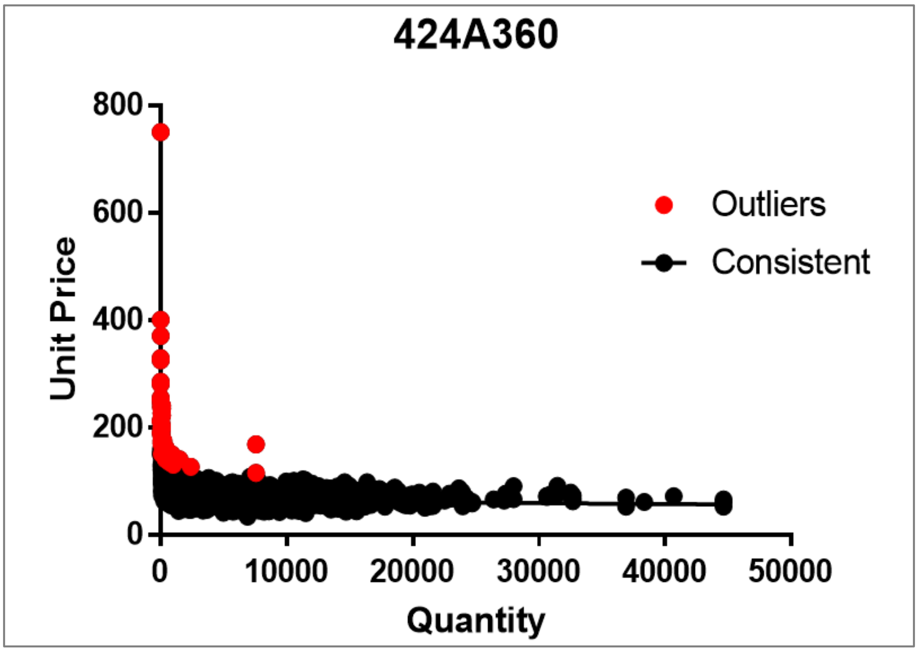


Figure 3.3: GraphPad Prims 7 Output – Example.

3.3 Exploratory Data Analysis

Exploratory data analysis is a common step in data-driven research. It is intended to provide researchers with a better understanding of each of the variables contained in the available data and the relationships among them. In this particular study, EDA facilitated the further identification of inconsistencies and errors in the data, allowing the author to take the necessary measures to correct them before proceeding with data processing. The EDA also helped with the selection of the case study item and with the identification of potential elements that could be used to model the relationship between each of the cost-influencing factors described in Section 1.3 and HMA prices paid by ALDOT.

With regard to the case study item, the author was looking for the most relevant asphalt paving item used in ALDOT contracts. The EDA showed that there is a single HMA pay item that be clearly identified as the most relevant in terms of frequency of use and dollar expenditure: "Superpave Bituminous Concrete Wearing Surface Layer, 1/2" Maximum Aggregate Size Mix, ESAL Range C/D

– Item ID 424A360.” This item corresponds to the second highest dollar expenditure in ALDOT’s annual construction program. It is only outranked by mobilizations expenses, which are paid by ALDOT in almost all construction contracts, including non-paving projects, which may explain why this is the top ranked item.

The better understanding of the data gained through the EDA was also used to make some final adjustments to the research plan. For example, in order to model the relationship between geographic location and HMA prices, it was necessary to split the historical bid data into comparable regions. Each region was required to provide sufficient data to allow for a reliable analysis, and at the same time, they could not be too large, so that, they would become meaningless geographic-wise. The study initially considered the five geographic regions used by ALDOT to organize its operations: north (N), east-central (EC), west-central (WC), south-central (SC), and south-west region (SW). Figure 3.4 shows the partition of the state of Alabama according to these five regions.

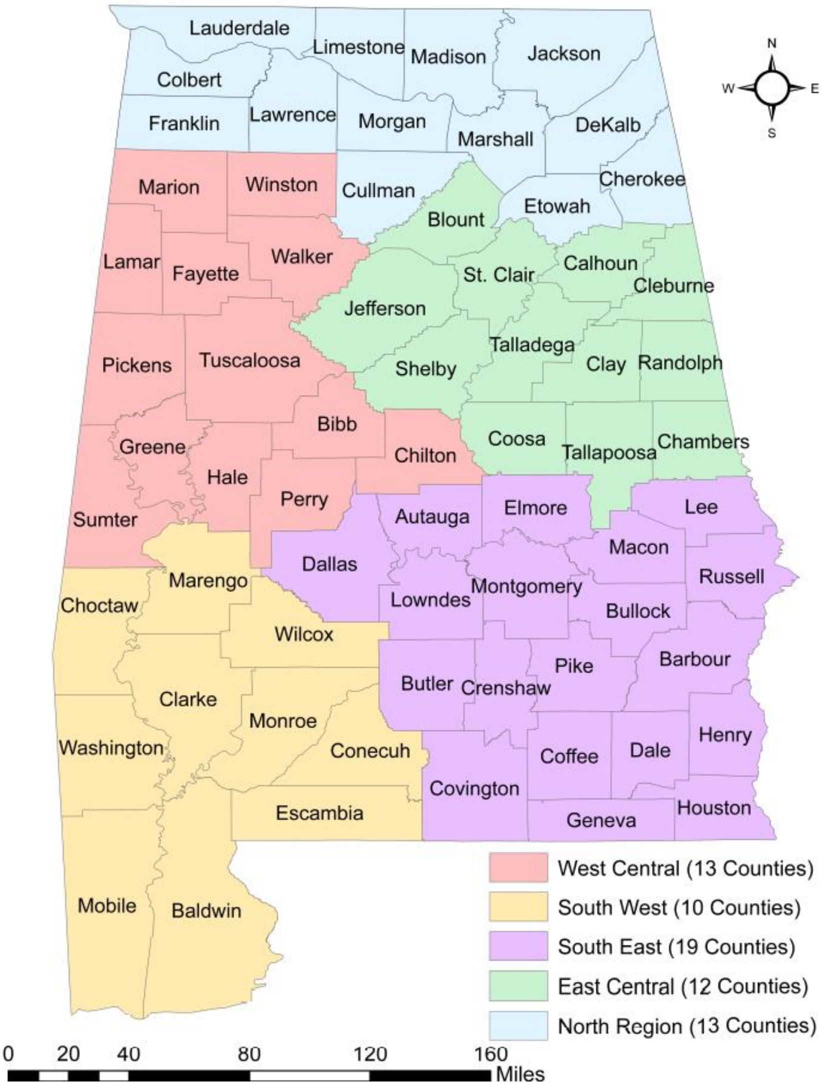


Figure 3.4: ALDOT Geographic Regions: Five Regions Classification.

The decision to use this partition had to be reevaluated when the EDA revealed that some of these regions were not providing a constant stream of HMA pricing data along the period of time considered in this study. More specifically, the regions providing the lowest count of paving projects per quarter are the WC and SW regions. This issue was solved by rearranging this partition into three regions: north, central, and south region. The final partition is shown in Figure 3.5.

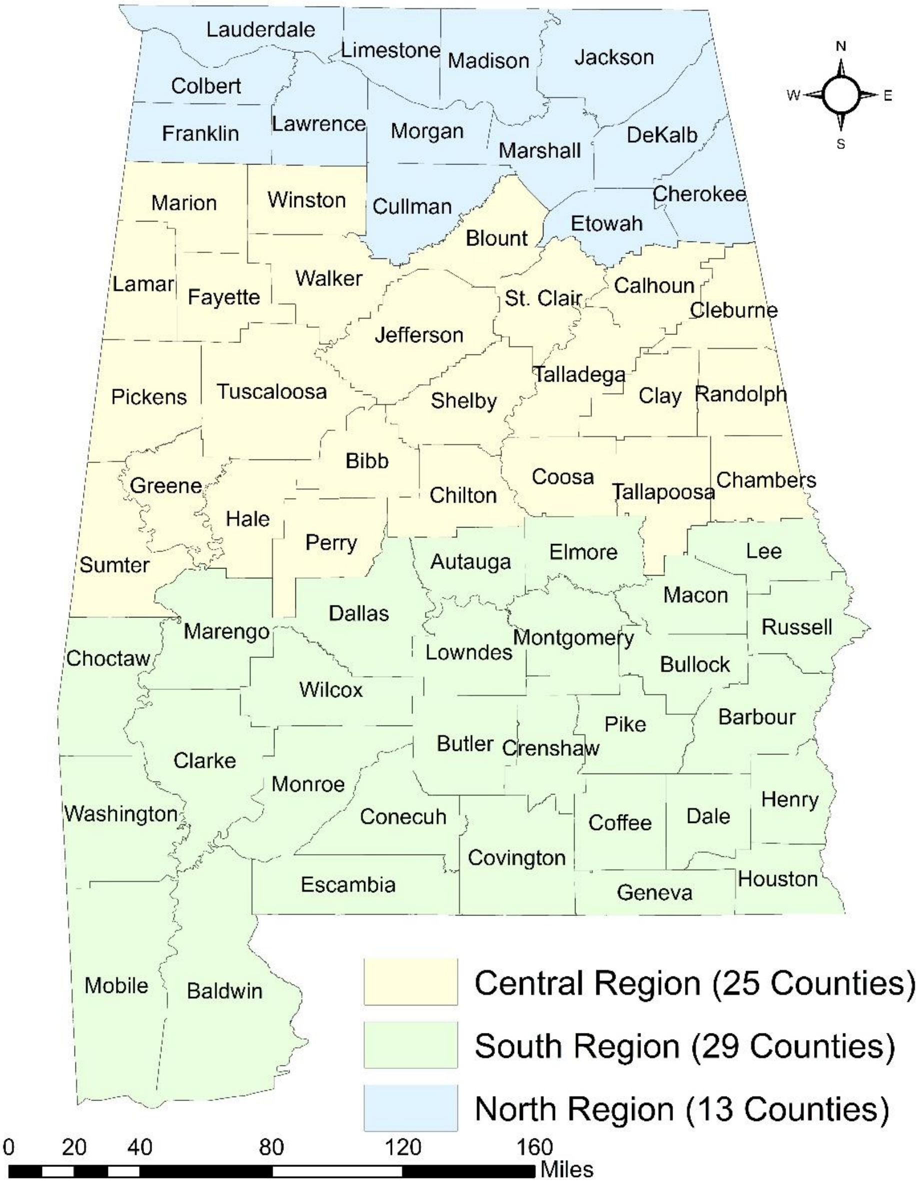


Figure 3.5: Final Geographic Regions: Three Regions Classification.

CHAPTER FOUR:

METHODOLOGY

4.1 Introduction

After gaining a better understanding of the research problem through the literature review summarized in Chapter 2, and after conducting the EDA described in Chapter 3, the author proceeded to design an appropriate research plan. The flow chart in Figure 4.1 illustrates the research plan followed in this study for the development of the stochastic three-dimensional HMA cost estimating system for ALDOT. The process and research activities performed at each step in the research plan are described throughout this chapter following the sequence of work show in Figure 4.1. It should be noted that findings and information collected during the literature review were summarized in Chapter 2, and the data collection/cleaning efforts and the EDA are presented in Chapter 3; therefore, this chapter starts with the description of the quantitative modeling process for each of the cost-influencing factors. Likewise, the Analysis of Results and Conclusions and Recommendations (last two boxes in the flow chart in Figure 4.1) are presented in Chapters 5 and 6, respectively.

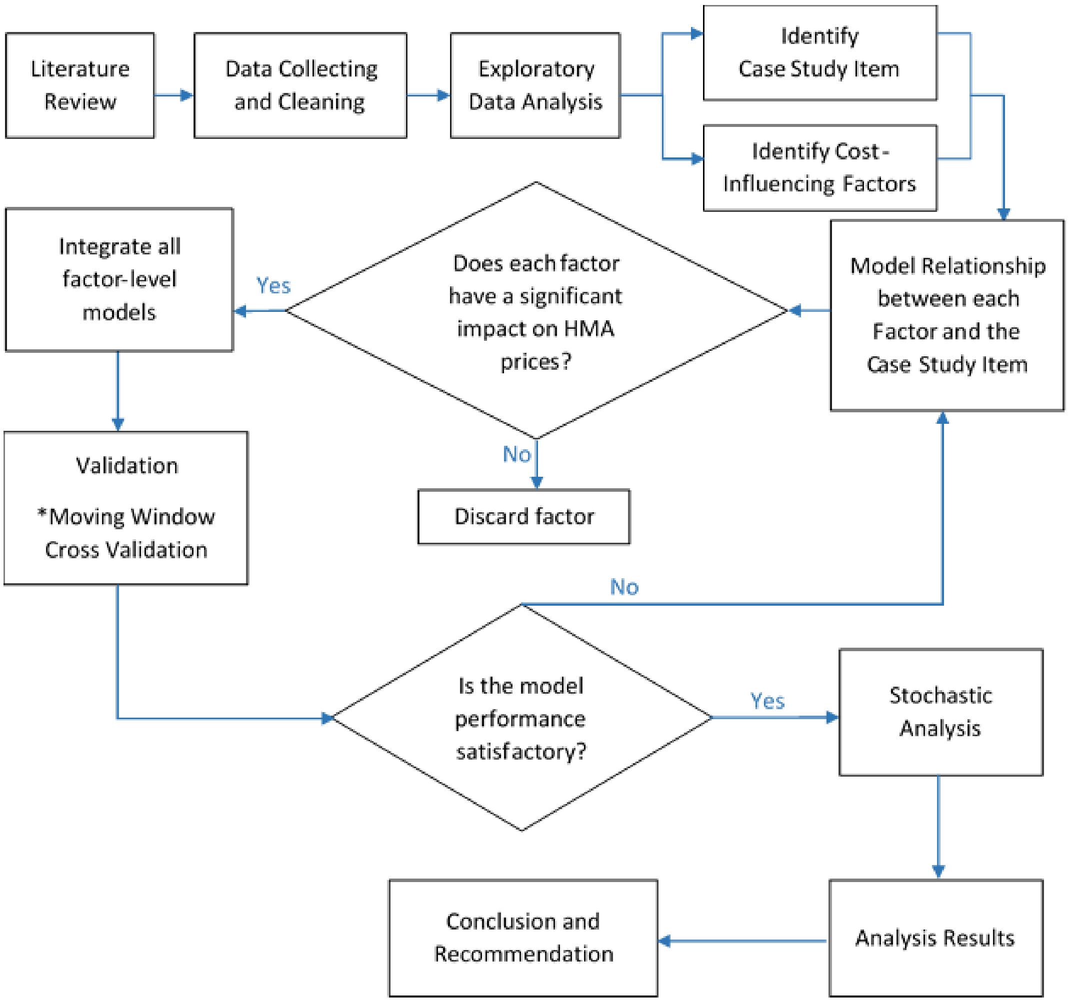


Figure 4.1: Methodology Flow Chart.

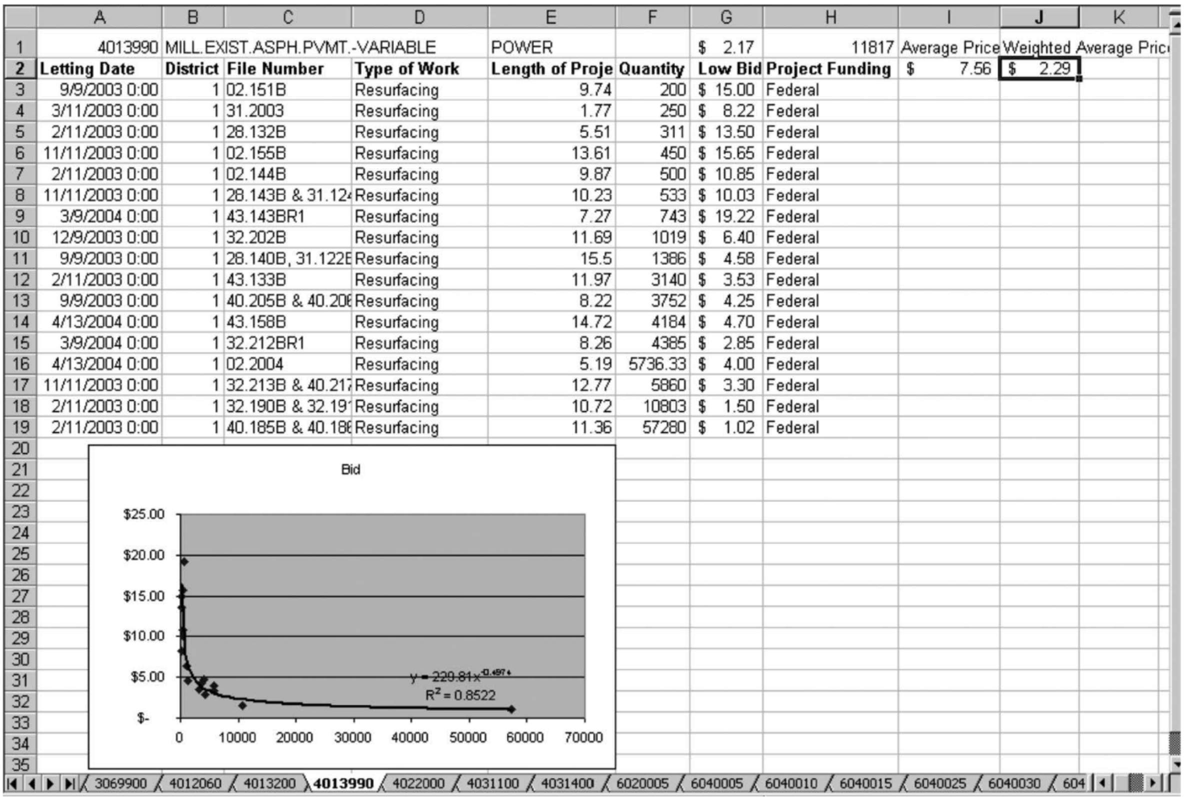
4.2 Quantitative Modeling

This section refers to the process to quantitatively model the relationship between each of the five cost-influencing factors and unit prices for the case study item. The assessment and modeling process for each of the factors is presented in following sections.

4.2.1 Scale

Base on the concept of economies of scale discussed in Section 1.3.1, it was necessary to model the reduction in price as the purchased amount of HMA increases. The relation between the quantities of work for the case study item and the unit prices paid by ALDOT was modeled using non-linear regression techniques. Specifically, the author used power regression models. This type of regression equation has been successfully used by Rueda (2016) to model HMA prices for the Minnesota

Department of Transportation. It has also proven to be a suitable regression approach in the concurrent study conducted by Pakalapati (2018). Moreover, it seems to be a widely accepted approach in the transportation construction industry as suggested by Figure 4.2, which was taken from the AASHTO Practical Guide to Cost Estimating (2013). Power regression models are defined by Equation 4-1, where 'A' and 'B' are constant values determined for each set of observations to be modeled.



	A	B	C	D	E	F	G	H	I	J	K
1	4013990	MILL.EXIST.ASPH.PVMT.-VARIABLE			POWER		$ 2.17	11817	Average Price	Weighted Average Pric	
2	Letting Date	District	File Number	Type of Work	Length of Proje	Quantity	Low Bid	Project Funding	$ 7.56	$ 2.29	
3	9/9/2003 0:00	1	02.151B	Resurfacing	9.74	200	$ 15.00	Federal			
4	3/11/2003 0:00	1	31.2003	Resurfacing	1.77	250	$ 8.22	Federal			
5	2/11/2003 0:00	1	28.132B	Resurfacing	5.51	311	$ 13.50	Federal			
6	11/11/2003 0:00	1	02.155B	Resurfacing	13.61	450	$ 15.65	Federal			
7	2/11/2003 0:00	1	02.144B	Resurfacing	9.87	500	$ 10.85	Federal			
8	11/11/2003 0:00	1	28.143B & 31.12	Resurfacing	10.23	533	$ 10.03	Federal			
9	3/9/2004 0:00	1	43.143BR1	Resurfacing	7.27	743	$ 19.22	Federal			
10	12/9/2003 0:00	1	32.202B	Resurfacing	11.69	1019	$ 6.40	Federal			
11	9/9/2003 0:00	1	28.140B, 31.122E	Resurfacing	15.5	1386	$ 4.58	Federal			
12	2/11/2003 0:00	1	43.133B	Resurfacing	11.97	3140	$ 3.53	Federal			
13	9/9/2003 0:00	1	40.205B & 40.20	Resurfacing	8.22	3752	$ 4.25	Federal			
14	4/13/2004 0:00	1	43.158B	Resurfacing	14.72	4184	$ 4.70	Federal			
15	3/9/2004 0:00	1	32.212BR1	Resurfacing	8.26	4385	$ 2.85	Federal			
16	4/13/2004 0:00	1	02.2004	Resurfacing	5.19	5736.33	$ 4.00	Federal			
17	11/11/2003 0:00	1	32.213B & 40.21	Resurfacing	12.77	5860	$ 3.30	Federal			
18	2/11/2003 0:00	1	32.190B & 32.19	Resurfacing	10.72	10803	$ 1.50	Federal			
19	2/11/2003 0:00	1	40.185B & 40.18	Resurfacing	11.36	57280	$ 1.02	Federal			

Figure 4.2: Historical Bid Analysis using Power Regression Modeling (AASHTO, 2013).

$$Unit\ Price = A * (Quantity)^B \qquad \textbf{Eq. 4-1}$$

Where: A and B are constant values.

4.2.2 Time

Two elements must be defined to incorporate the time dimension into the proposed bid-based cost estimating system: an optimal look-back period for data retrieval and a CCIS to adjust cost estimates for inflation and price fluctuations over time. As previously mentioned in this thesis, these two elements will be taken from a parallel study conducted by Pakalapati (2018). In that study, Pakalapati used the same case study item to demonstrate a look-back determination protocol and to develop a CCIS to be applied to bid-based estimating. After assessing 5 possible look-back periods (ranging from 1 to 5 years) and 20 indexing approaches, Pakalapati concluded that 2 years of historical bid data and a quarterly CCIS facilitate better accuracy in cost estimates for the case estimate item, which is the recommendation that has been followed in this study.

4.2.3 Location

The assessment of the location factor started by determining if significantly differences in HMA prices should be expected for the same amount of HMA across the three geographic locations. The study identified an actual typical paving project awarded by ALDOT, took the quantity of HMA estimated for that project (8,715 tons), and used the collected bid data to determine the annual average unit price actually paid by ALDOT for the same amount of HMA in each region between 2006 and 2016. Figure 4.3 shows how the unit price for 8,715 tons of HMA changed across these 11 years in each region.

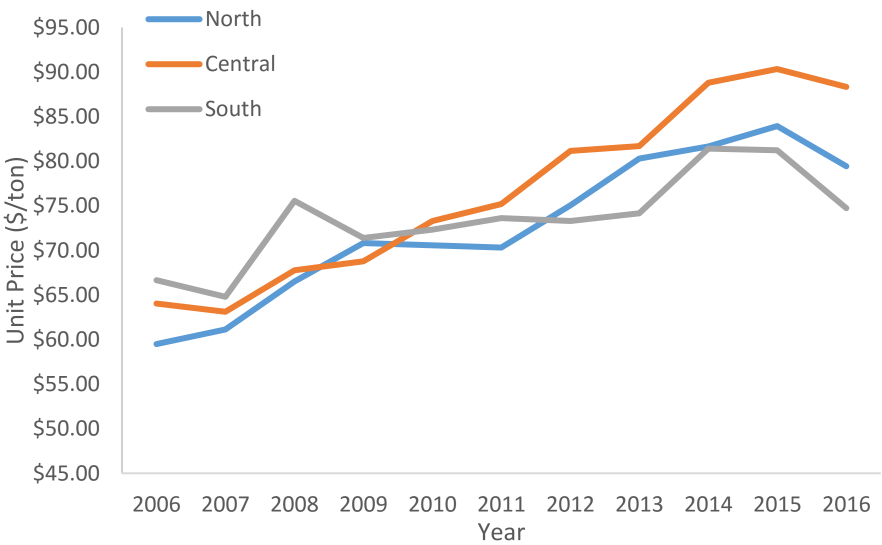


Figure 4.3: Unit Price for HMA per Region 2006-2016.

The next step was to determine if there is a significant difference between the three-time series in Figure 4.3. The location dimension would not be necessary if no significant difference is found. A visual inspection of this figure seems to show that HMA prices across the three regions started to increasingly spread out after 2010. A series of ANOVA tests applied to different time frames were used to validate this statement. The results of these tests is presented in Table 4.1. The first test was conducted to compare the 11-year average unit price (2006-2016) between the three regions and it revealed no significant difference with a significance level of 5%. The time frame was then reduced by one year (2007-2016) and the ANOVA test was run again with the same results. The process was repeated a number of times, reducing the time frame by one year at a time, showing that significant differences between the regions started to appear after 2010.

Table 4.1 ANOVA Test Results to Compare Average Unit Prices per Region

Years	Time Frame	Region	Average Unit Price	P-value	Conclusion
11	2006-2016	North	67.05	0.607	No enough information to prove a significant difference with a significance level of 5%
		Central	74.74		
		South	73.00		
10	2007-2016	North	68.41	0.549	
		Central	76.15		
		South	74.12		
9	2008-2016	North	69.90	0.394	
		Central	77.87		
		South	75.31		
8	2009-2016	North	71.36	0.215	
		Central	79.19		
		South	74.96		
7	2010-2016	North	72.43	0.076	
		Central	81.27		
		South	75.66		
6	2011-2016	North	73.72	0.034	Significant difference with a significance level of 5%
		Central	82.86		
		South	76.44		
5	2012-2016	North	75.48	0.005	
		Central	84.41		
		South	77.37		
4	2013-2016	North	76.59	0.007	
		Central	84.89		
		South	78.57		
3	2014-2016	North	76.42	0.018	
		Central	86.29		
		South	79.58		

Two important findings were derived from the statistical analysis described above. First, it can be assumed that current HMA prices may change significantly between regions. The second finding is that some sort of event(s) may have happened in 2010 affecting the paving construction market in

Alabama. Therefore, the author decided to continue developing the proposed system and to validate it using only data from projects awarded between 2011 and 2016 because old HMA pricing trends may affect the results of this study, misleading ALDOT on the expected performance of the proposed system in today's construction industry. Thus, the author proceeded to develop a LCI to quantity the difference between these regions at each year, starting in 2011.

The LCI was developed following a similar approach as the one adopted by the RSMeans for the calculation of its City Cost Index (RSMeans 2018). The RSMeans City Cost Index compares average construction costs among 731 U.S. and Canadian Cities. The index values for all U.S. cities are calculated using the U.S. national average as a reference. Every year, the U.S. national average is assigned an index value of 100, and index values at the city level are calculated in a proportional manner around the national index. Thus, if for example, the index value for a given city is 95, that would mean that average construction costs in that city are 5% lower than the national average. Likewise, an index value of 102 would mean that average costs for that city are expected to be 2% above the national average.

Figure 4.4 shows the same time series from Figure 4.3, by adding one more series for the state average HMA unit price for 8,715 tons of HMA. The values plotted in this figure for each region were compared against the state average of their respective years. The results of these comparisons were then translated into index values in a similar fashion as in the RSMeans City Cost Index. The resulting LCI is shown in Table 4.2.

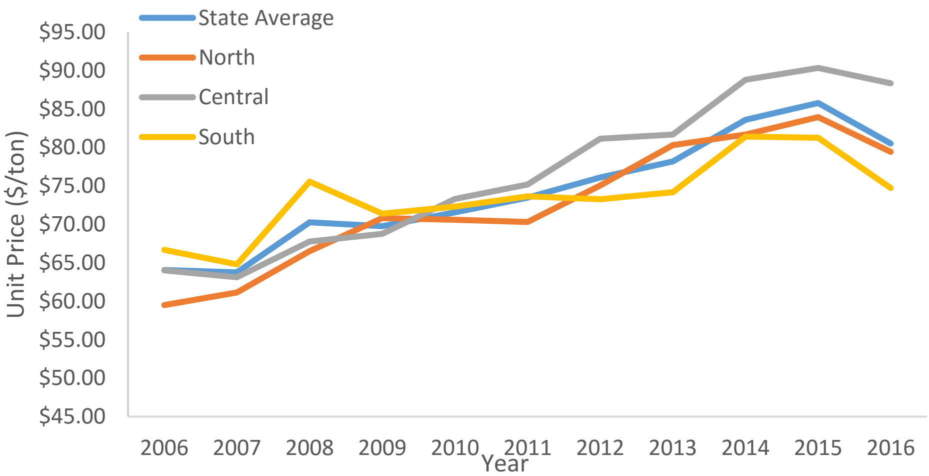


Figure 4.4: Unit Price for 8,715 Tons of HMA per Region & State Average 2006-2016.

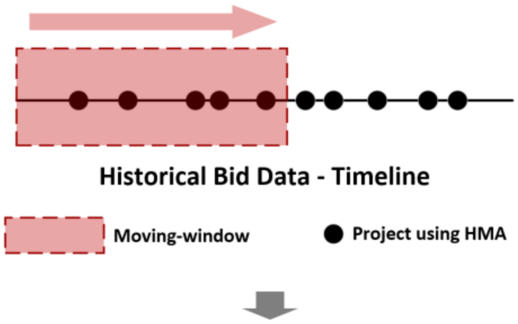
Table 4. 2 HMA Location Cost Index in Alabama

Year	State	North	Central	South
2011	100.00	95.15	102.56	100.21
2012	100.00	97.53	106.98	96.35
2013	100.00	102.02	104.51	94.93
2014	100.00	97.12	106.75	97.37
2015	100.00	97.80	105.70	94.22
2016	100.00	98.46	110.25	92.42

While it does not seem to be a clear pattern to define the difference in HMA prices between the north and south regions, Figure 4.2 and the LCI show a clear trend of higher HMA prices in the central region in comparison with the other two regions. On average among the six years shown in Table 4.2, HMA prices in the central region are 8.3% and 10.8% higher than in the north and south regions. Further research is required to attempt to explain the reason behind the price difference between regions.

4.2.4 Uncertainty

Unlike the scale, time, and location factors described above, the estimating uncertainty factor is not incorporated as an input to the proposed estimating system. It is essentially a byproduct of the validation process, as illustrated in Figure 4.5. This factor actually connects the system development and system implementation phases. Before the initial implementation of the system, ALDOT's estimators must performed the same process presented in this research with the most recent historical bid data. After using this data to develop the three-dimensional cost estimating system and to run the MWCV, the distribution of percentage errors obtained from the MWCV must be saved for future use during the actual implementation of the system in intended projects to be executed by ALDOT. The same process should be follow for every pay item frequently used by ALDOT in asphalt paving projects. Therefore, each item should have its own distribution of percentage errors.



2016 Projects	Actual HMA Unit price	Estimated HMA Unit Price	Percentage Error
Project 1	$89.06	$84.56	**5.3%**
Project 2	$67.10	$71.06	**-5.6%**
Project 3	$87.77	$93.60	**-6.2%**
...	...	...	**...**
Project n	$78.10	$72.85	**67.2%**

n = number of 2016 projects including the case study item

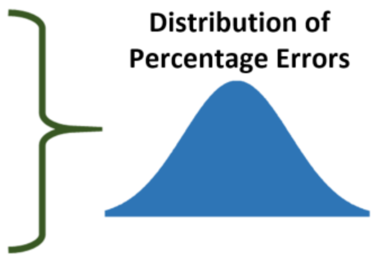


Figure 4.5: Generation of Distribution of Estimating Errors.

4.2.5 Level of Competition

A review of the average number of asphalt paving contractors competing per contract in each region revealed no significant difference in this number between regions. The average values per region are shown in Table 4.3. Assuming that the number of vendors competing per project is the main measure of the level of competition, it could be assumed from Table 4.3 that there is a uniform level of competition across the state. The results of an ANOVA test shown in Table 4.4 revealed significant differences in the average number of vendors per project across the three regions. However, as shown in Table 4.3, that significant difference could be around 0.2 vendors per project, which has been disregarded by the author given that such a small difference in a discrete variable like “number of vendors” would not have any practical meaning. Therefore, it has been assumed that the level of competition would have the same impact on all paving projects across the state, making it unnecessary the inclusion of this factor in the proposed cost estimating system. This was the only of the five cost-influencing factor not included in the system. The author recognizes that a more exhaustive review of the level of competition factor is required and should be considered in future research.

Table 4. 3 Level of Competition Analysis

Regions	Number of Vendors
North	3.2
Central	3.0
South	3.1

Table 4. 4 Level of Competition ANOVA Test

Groups	Count	Sum	Average	Variance	P-Value
North	830	2731	3.2	3.24	0.001348
Central	1230	3686	3.0	3.05	
South	1687	5315	3.1	3.54	

4.3 Integration of Cost-Influencing Factors

The previous section has presented all the elements and models used to assess the impact of each of each cost-influencing factors on bid-based cost estimating. Now it is necessary to establish a framework to integrate and facilitate the use of all these elements and models. Figure 4.6 summarize this framework in four steps. This is the step-by-step implementation of the system by ALDOT's estimators. The first three steps correspond to three-dimensional system, while the fourth and last step refers to the generation of the stochastic estimate using the distribution of percentage errors obtained during the development of the system. All the steps in Figure 4.6 are described below.

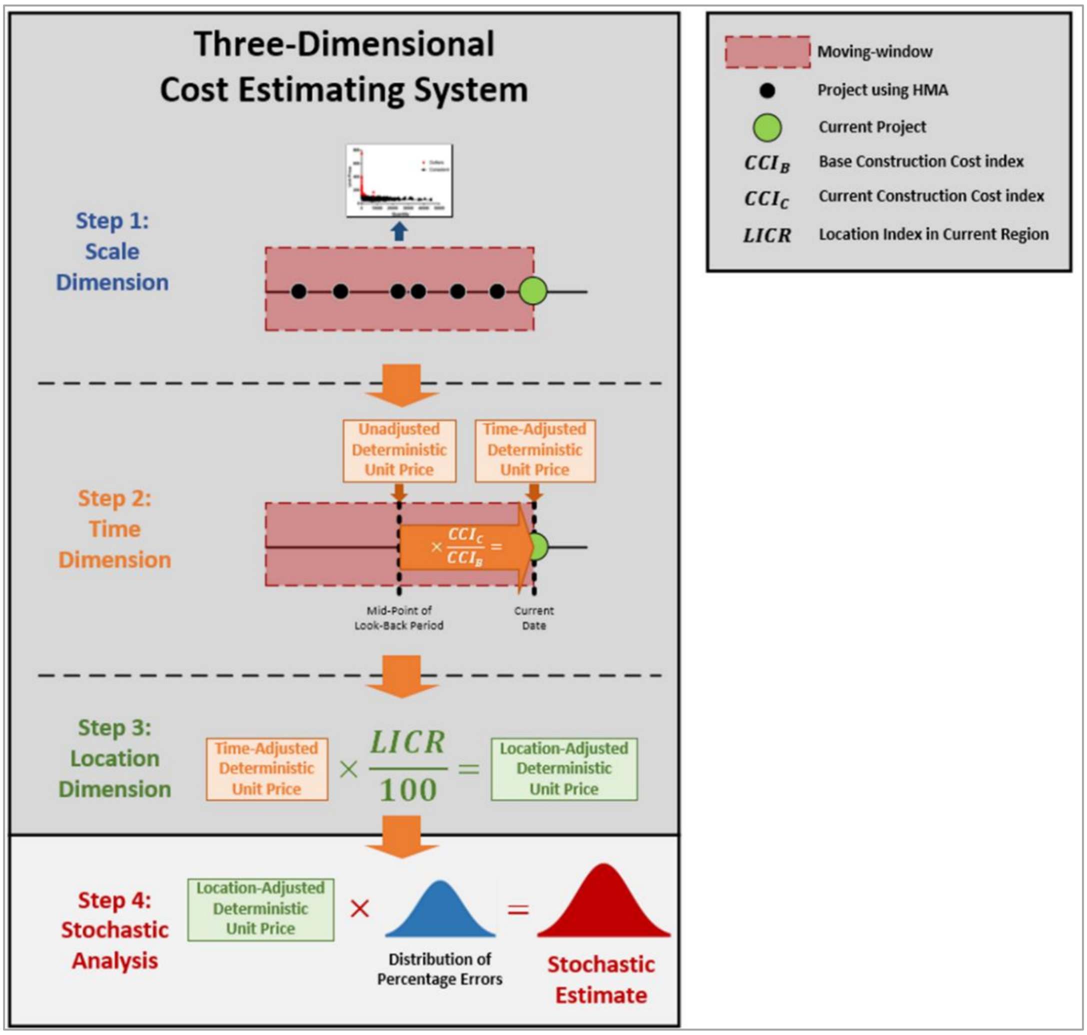


Figure 4.6: Integration of Factor for Implementation.

1. **Scale Dimension:** Develop a power regression model using HMA unit prices from all previous projects contained in the two-year look-back period.
2. **Time dimension:** Use the power regression model to estimate a deterministic unit price (unadjusted) for the amount of HMA to be used in the current project. Assuming that the unadjusted deterministic unit price corresponds to the mid-point of the look-back period, use the quarterly CCIS to adjust the deterministic estimate for time using equation 4-2.

$$TADUP = UADUP \times \frac{Current\ CCI}{Base\ CCI} \qquad \textbf{Eq. 4-2}$$

Where: TADUP = Time-Adjusted Deterministic Unit Price

UADUP = Unadjusted Deterministic Unit Price

Current CCI = Last Construction Cost Index Known at Current Date

Base CCI = Last Construction Cost Index Known at Mid-Point of Look-Back Period

3. **Location Dimension:** The time-adjusted deterministic unit price from Step 2 was estimated with projects awarded across the state. Therefore, it is assumed to be a state average unit price. Use the LCI and the equation 4-3 to adjust this unit price for the region in which the current project is to be built:

$$LADUP = TADUP \times \frac{Location\ Index\ in\ Current\ Region}{100} \qquad \textbf{Eq. 4-3}$$

Where: LADUP= Location-Ajusted Deterministic Unit Price

TADUP= Time-Ajusted Deterministic Unit Price

4. **Stochastic Analysis:** Develop the final stochastic HMA unit price by multiply the location-adjusted deterministic unit price by the distribution of percentage errors obtained during the development of the system.

4.4 Validation and Stochastic Analysis

Given that the stochastic analysis is to be performed during the actual implementation of the proposed estimating system, the stochastic analysis in the fourth step of previous section is not considered in the validation process. Thus, in this thesis, the estimating accuracy and reliability of the system is demonstrated at the deterministic level. However, this thesis still shows the process to define the distribution of the percentage errors after applying the system and the MWCV to the case study item (see Chapter 5). The proposed MWCV process is illustrated in Figure 4.7.

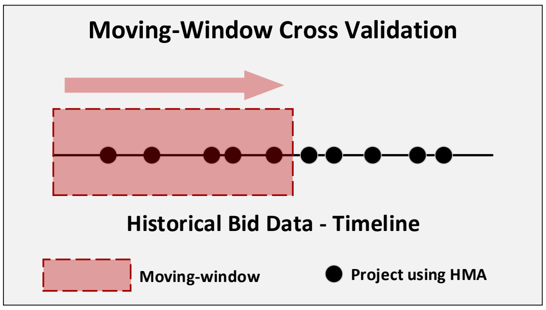


Figure 4.7: Moving-Window Cross Validation Approach.

In this study, the moving-window refers to the two-year look-back period moving across the testing projects, which in this case, consist of all 2016 projects using the case study item. The MWCV process places the end of the two-year window at the beginning of the testing period, and then starts moving it towards the end of the testing timeline. Every time that the right-end of the moving-window finds a project, it stops, the HMA unit price for that project is estimated, the APE is calculated, and then, the moving-window continues its way until finding the next project. At the end of the MWCV process, the MAPE and standard deviation of all APEs are calculated to determine the overall accuracy and reliability of the system. The MWCV was actually applied four times to the case study item to quantify the improvement in estimating accuracy and reliability offered by each of the three dimensions. A MAPE and standard deviation value was calculated at the end of each MWCV calculation and compared against the values from the previous run of the MWCV to demonstrate the progressive estimating improvement as including the dimensions one-by-one. Each of the four MWCV calculations is described below. Likewise, Chapter 5 is mainly aimed to present and analyze the results obtained from the four runs of the MWCV.

1. First MWCV – Dimensionless: The first time, the MWCV was performed without considering any of the three dimensions. The estimated HMA unit price for each project participating in the validation process was calculated as the arithmetic average of all HMA unit prices contained in the look-back period.
2. Second MWCV – Scale Dimension: The estimated HMA unit price for each project participating in the validation process was calculated using only the non-linear regression model (only the scale dimension – Step 1 in Table 5.4).
3. Third MWCV – Scale plus Time Dimension: The estimated HMA unit price for each project participating in the validation process was calculated using the non-linear regression model and the quarterly CCIS (the scale and time dimensions – Step 1 and 2 in Table 5.4).
4. Four MWCV – Scale, Time, and Location Dimensions: The estimated HMA unit price for each project participating in the validation process was calculated using the non-linear regression model, the quarterly CCIS, and the location index (all three dimensions – Steps 1 to 3 in Table 5.4).

4.4 Summary

The chapter has explained in detail each step of the research plan followed by the author for the development of the stochastic three-dimensional cost estimating system for HMA in paving projects executed by ALDOT. The chapter started with a description of the data collection and cleaning efforts, including details on the two outlier identification methods (modified Z-score and ROUT) used to remove HMA unit prices deemed to be outside the normal price ranges, preventing them from affecting the performance of the system. The chapter then described the assessment of the five HMA cost-influencing factors initially considered in this study and explained the rationale that led the author to discard one of them; level of competition. Subsequently, the chapter presented a four-step framework to integrate all the elements and quantitative methods used to model each of the four remaining factors: scale, time, location, and uncertainty. Finally, the chapter described how four different calculations of the proposed MWCV process were used to quantify the improvement in estimating accuracy and reliability offered by each factor and by the system as a whole. The results from the MWCV process are presented and statistically analyzed in the following chapter.

CHAPTER FIVE:
MOVING-WINDOW CROSS VALIDATION:
ANALYSIS OF RESULTS

5.1 Introduction

This chapter presents and analyzes the results obtained from the application of the MWCV process described in Chapter 4. The four MWCV calculations used to measure the effectiveness of the proposed system were applied to all 2016 projects using the case study item, as described in Section 4.2. The comparison of MAPE and standard deviation values between MWCV calculations showed a statistically significant improvement in estimating accuracy and reliability with the incorporation of each dimension. The following sections discuss in more detail the results from the MWCV process.

5.2 Moving-Window Cross Validation – Summary of Results

Table 5.1 summarizes the results of the four MWCV calculations. Each calculation was applied to 97 projects using the case study item during 2016. The table shows the actual unit price submitted by the lowest bidder on each project, the four unit prices obtained from the four MWCV calculations, and the APE for each of the unit price estimates calculated using Equation 1-1 (see Chapter 1). The MAPE and standard deviation values for each MWCV calculation are shown at the end of the table. These results are analyzed in detail in the following sections.

Table 5.1 Moving-Cross Validation Calculations – Summary of Results

Project	Actual UP ($/Ton)	Estimated Unit Price				Average Percentage Error			
		MWCV 1	MWCV 2	MWCV 3	MWCV 4	MWCV 1	MWCV 2	MWCV 3	MWCV 4
1	$115.00	$83.29	$75.28	$67.90	$71.76	27.6%	34.5%	41.0%	37.6%
2	$108.98	$83.45	$120.57	$108.75	$106.36	23.4%	10.6%	0.2%	2.4%
3	$60.00	$83.45	$75.54	$68.13	$64.19	39.1%	25.9%	13.6%	7.0%
4	$70.00	$83.45	$79.18	$71.42	$69.85	19.2%	13.1%	2.0%	0.2%
5	$87.77	$83.45	$103.86	$93.68	$88.26	4.9%	18.3%	6.7%	0.6%
6	$67.10	$83.45	$70.24	$63.36	$59.69	24.4%	4.7%	5.6%	11.0%
7	$88.75	$83.45	$79.96	$72.12	$76.23	6.0%	9.9%	18.7%	14.1%
8	$89.06	$83.45	$77.37	$69.79	$68.26	6.3%	13.1%	21.6%	23.4%
9	$64.34	$83.45	$77.09	$69.53	$68.01	29.7%	19.8%	8.1%	5.7%
10	$80.00	$83.45	$85.70	$77.29	$72.83	4.3%	7.1%	3.4%	9.0%
11	$53.89	$83.45	$70.74	$63.81	$60.12	54.8%	31.3%	18.4%	11.6%
12	$80.00	$83.45	$103.69	$93.53	$98.85	4.3%	29.6%	16.9%	23.6%
13	$79.16	$83.44	$80.89	$72.96	$68.74	5.4%	2.2%	7.8%	13.2%
14	$87.38	$83.44	$93.01	$83.89	$82.05	4.5%	6.4%	4.0%	6.1%
15	$72.75	$83.44	$70.89	$63.94	$60.24	14.7%	2.6%	12.1%	17.2%
16	$69.85	$83.44	$82.02	$73.97	$69.70	19.5%	17.4%	5.9%	0.2%
17	$59.46	$83.44	$76.55	$69.04	$67.53	40.3%	28.7%	16.1%	13.6%
18	$71.06	$83.44	$71.20	$64.22	$67.87	17.4%	0.2%	9.6%	4.5%
19	$78.10	$83.44	$80.41	$72.52	$68.33	6.8%	3.0%	7.1%	12.5%
20	$62.77	$83.93	$84.59	$76.30	$80.64	33.7%	34.8%	21.6%	28.5%
21	$78.77	$83.93	$77.40	$69.81	$65.78	6.6%	1.7%	11.4%	16.5%
22	$62.90	$83.93	$74.24	$66.96	$63.09	33.4%	18.0%	6.5%	0.3%
23	$66.00	$83.93	$79.41	$71.62	$70.05	27.2%	20.3%	8.5%	6.1%
24	$65.87	$83.93	$80.80	$72.88	$71.28	27.4%	22.7%	10.6%	8.2%
25	$58.00	$83.93	$85.64	$77.24	$75.55	44.7%	47.7%	33.2%	30.3%
26	$72.15	$83.93	$99.25	$89.52	$84.34	16.3%	37.6%	24.1%	16.9%
27	$79.15	$83.93	$71.30	$64.31	$60.59	6.0%	9.9%	18.7%	23.4%
28	$68.85	$83.93	$68.89	$62.14	$58.55	21.9%	0.1%	9.7%	15.0%
29	$64.99	$83.93	$68.75	$62.01	$65.54	29.1%	5.8%	4.6%	0.8%
30	$75.75	$83.93	$75.25	$67.87	$71.74	10.8%	0.7%	10.4%	5.3%
31	$82.50	$83.93	$74.47	$67.17	$71.00	1.7%	9.7%	18.6%	13.9%
32	$54.25	$83.93	$71.06	$64.09	$60.39	54.7%	31.0%	18.1%	11.3%
33	$95.00	$83.93	$115.14	$103.85	$97.85	11.6%	21.2%	9.3%	3.0%
34	$57.07	$83.83	$78.51	$71.36	$67.24	46.9%	37.6%	25.0%	17.8%
35	$67.23	$83.83	$81.30	$73.90	$78.10	24.7%	20.9%	9.9%	16.2%
36	$70.06	$83.83	$93.71	$85.18	$90.03	19.7%	33.8%	21.6%	28.5%
37	$66.01	$83.83	$65.38	$59.43	$62.82	27.0%	1.0%	10.0%	4.8%
38	$80.25	$83.83	$72.83	$66.20	$69.97	4.5%	9.2%	17.5%	12.8%
39	$56.00	$83.83	$70.21	$63.82	$62.42	49.7%	25.4%	14.0%	11.5%
40	$66.86	$83.83	$70.40	$63.99	$62.59	25.4%	5.3%	4.3%	6.4%
41	$63.00	$83.83	$72.41	$65.82	$62.01	33.1%	14.9%	4.5%	1.6%
42	$69.75	$83.83	$75.10	$68.26	$72.15	20.2%	7.7%	2.1%	3.4%
43	$53.25	$83.83	$66.48	$60.43	$56.93	57.4%	24.8%	13.5%	6.9%
44	$55.82	$83.83	$69.10	$62.81	$66.39	50.2%	23.8%	12.5%	18.9%
45	$55.79	$83.83	$71.38	$64.89	$68.58	50.3%	27.9%	16.3%	22.9%
46	$90.25	$83.83	$71.23	$64.75	$63.32	7.1%	21.1%	28.3%	29.8%
47	$83.92	$83.83	$77.67	$70.60	$74.62	0.1%	7.4%	15.9%	11.1%
48	$55.00	$83.83	$72.84	$66.21	$64.76	52.4%	32.4%	20.4%	17.7%
49	$85.00	$83.83	$102.09	$92.80	$87.43	1.4%	20.1%	9.2%	2.9%
50	$60.58	$82.52	$69.64	$63.30	$61.91	36.2%	15.0%	4.5%	2.2%
51	$81.25	$82.52	$82.22	$74.73	$73.09	1.6%	1.2%	8.0%	10.0%
52	$80.46	$82.52	$86.63	$78.74	$83.23	2.6%	7.7%	2.1%	3.4%
53	$115.00	$82.47	$86.75	$78.86	$74.30	28.3%	24.6%	31.4%	35.4%
54	$79.20	$82.47	$95.46	$86.77	$81.75	4.1%	20.5%	9.6%	3.2%
55	$74.90	$82.47	$94.08	$85.51	$90.38	10.1%	25.6%	14.2%	20.7%
56	$82.60	$82.47	$85.88	$78.06	$82.51	0.2%	4.0%	5.5%	0.1%

Table 5.1 Moving-Cross Validation Calculations – Summary of Results (Cont.)

Project	Actual UP ($/Ton)	Estimated Unit Price				Average Percentage Error			
		MWCV 1	MWCV 2	MWCV 3	MWCV 4	MWCV 1	MWCV 2	MWCV 3	MWCV 4
57	$110.92	$82.47	$88.16	$80.13	$75.50	25.7%	20.5%	27.8%	31.9%
58	$77.50	$82.47	$79.01	$71.82	$75.91	6.4%	2.0%	7.3%	2.0%
59	$81.08	$82.47	$77.63	$70.57	$74.59	1.7%	4.2%	13.0%	8.0%
60	$78.86	$82.47	$76.18	$69.25	$73.19	4.6%	3.4%	12.2%	7.2%
61	$88.11	$82.47	$76.72	$69.74	$73.71	6.4%	12.9%	20.9%	16.3%
62	$56.50	$82.47	$69.98	$63.61	$59.94	46.0%	23.9%	12.6%	6.1%
63	$72.62	$81.65	$86.97	$79.42	$74.83	12.4%	19.8%	9.4%	3.0%
64	$64.00	$81.65	$72.96	$66.63	$65.16	27.6%	14.0%	4.1%	1.8%
65	$65.19	$81.65	$68.65	$62.69	$66.26	25.3%	5.3%	3.8%	1.6%
66	$80.02	$81.65	$89.03	$81.29	$85.92	2.0%	11.3%	1.6%	7.4%
67	$63.68	$81.65	$67.25	$61.41	$60.06	28.2%	5.6%	3.6%	5.7%
68	$54.45	$81.65	$70.96	$64.80	$61.05	50.0%	30.3%	19.0%	12.1%
69	$60.00	$81.65	$74.13	$67.69	$66.21	36.1%	23.6%	12.8%	10.3%
70	$85.65	$81.65	$97.38	$88.92	$83.78	4.7%	13.7%	3.8%	2.2%
71	$59.75	$81.65	$70.49	$64.37	$60.65	36.7%	18.0%	7.7%	1.5%
72	$56.14	$81.65	$72.50	$66.21	$64.75	45.4%	29.1%	17.9%	15.3%
73	$56.00	$81.65	$74.78	$68.29	$72.18	45.8%	33.5%	21.9%	28.9%
74	$60.00	$81.65	$79.51	$72.60	$68.40	36.1%	32.5%	21.0%	14.0%
75	$84.33	$80.66	$70.04	$63.96	$67.60	4.4%	16.9%	24.2%	19.8%
76	$62.41	$80.66	$73.80	$67.39	$71.23	29.2%	18.2%	8.0%	14.1%
77	$72.25	$80.66	$77.59	$70.85	$66.75	11.6%	7.4%	1.9%	7.6%
78	$87.25	$80.66	$74.72	$68.23	$66.73	7.6%	14.4%	21.8%	23.5%
79	$68.86	$80.66	$80.89	$73.86	$69.59	17.1%	17.5%	7.3%	1.1%
80	$70.13	$80.12	$86.50	$79.06	$74.49	14.2%	23.3%	12.7%	6.2%
81	$88.33	$80.12	$89.15	$81.48	$86.12	9.3%	0.9%	7.8%	2.5%
82	$75.00	$80.12	$92.58	$84.61	$82.75	6.8%	23.4%	12.8%	10.3%
83	$78.00	$80.12	$80.32	$73.40	$77.59	2.7%	3.0%	5.9%	0.5%
84	$56.85	$80.12	$77.82	$71.13	$69.56	40.9%	36.9%	25.1%	22.4%
85	$106.26	$80.12	$100.39	$91.75	$96.98	24.6%	5.5%	13.7%	8.7%
86	$70.64	$80.12	$75.28	$68.80	$72.72	13.4%	6.6%	2.6%	2.9%
87	$92.00	$80.12	$79.26	$72.44	$68.25	12.9%	13.8%	21.3%	25.8%
88	$80.30	$80.12	$89.60	$81.89	$77.15	0.2%	11.6%	2.0%	3.9%
89	$66.00	$80.12	$82.24	$75.16	$73.51	21.4%	24.6%	13.9%	11.4%
90	$57.24	$80.12	$74.55	$68.13	$66.64	40.0%	30.2%	19.0%	16.4%
91	$61.23	$80.12	$72.48	$66.24	$70.02	30.9%	18.4%	8.2%	14.3%
92	$56.95	$80.12	$66.70	$60.96	$57.43	40.7%	17.1%	7.0%	0.9%
93	$115.00	$80.12	$102.15	$93.36	$98.68	30.3%	11.2%	18.8%	14.2%
94	$73.11	$80.12	$85.21	$77.87	$73.37	9.6%	16.5%	6.5%	0.4%
95	$73.23	$79.85	$80.33	$71.74	$67.60	9.0%	9.7%	2.0%	7.7%
96	$58.50	$79.85	$66.34	$59.25	$55.83	36.5%	13.4%	1.3%	4.6%
97	$90.00	$79.85	$75.06	$67.03	$70.85	11.3%	16.6%	25.5%	21.3%
Mean Absolute Percentage Error (MAPE)						21.6%	16.6%	12.5%	11.3%
Standard Deviation of Absolute Percentage Errors						16.2%	10.9%	8.21%	9.1%

Note: MWCV 1 = Dimensionless;
MWCV 2 = Scale Dimension;
MWCV 3 = Scale plus Time Dimension;
MWCV 4 = Scale, Time, plus Location Dimension.

5.3 Analysis of Scale Dimension

Table 5.2 shows the MAPE and standard deviation values for the first two MWCV calculations, the dimensionless calculation and the calculation that consider the scale dimension. This table also shows the results of the statistical analysis conducted to compare the results between these two calculations. Two different statistical significance tests were used to compare these results. The paired two-sample t-test was first used to determine the level of significance in accuracy improvement (reduction of MAPE), while the F-test was used to assess the improvement in reliability. The paired two-sample t-test was appropriate in this case because both MWCV calculations were applied to each of the 97 testing project.

The results of this analysis revealed that both the 23.10% and 32.94% improvements in estimating accuracy and reliability, respective, can be considered statistically significant with a 1% significance level (paired t-test p-value = 1.38×10^{-4}; F-test p-value = 5.77×10^{-5}). It allows to strongly conclude that the use of non-linear models in the scale dimension significantly improves the estimating effectiveness for the case study item. Equation 5.1 and 5.2 show the mathematical calculation used between MWCV calculations to measure the improvement in accuracy and reliability using the MAPE and standard deviation values of APEs, respectively.

$$\text{Improvement on Accuracy} = \frac{MAPE_{i+1} - MAPE_i}{MAPE_i} \times 100\% \qquad \textbf{Eq. 5-1}$$

$$\text{Improvement on Reliability} = \frac{SD_{i+1} - SD_i}{SD_i} \times 100\% \qquad \textbf{Eq. 5-2}$$

Where: MAPE_{i+1} =Mean Absolute Percentage Error-MWCV Calculation i+1

$MAPE_i$= Mean Absolute Percentage Error-MWCV Calculation i

SD_{i+1}= Standard Deviation of Absolute Percentage Errors-MWCV Calculation i+1

SD_i= Standard Deviation of Absolute Percentage Errors-MWCV Calculation i

Table 5.2 Dimensionless vs. Scale Dimension – MWCV Results

	MAPE	**Standard Deviation of APEs**
MWCV 1 - Dimensionless	21.62%	16.18%
MWCV 2 - Scale Dimension	16.62%	10.85%
Reduction	23.10%	32.94%
Significance of Improvement at 1% Significance Level	Statistically Significant	Statistically Significant
p-value	<1%	<1%

5.4 Analysis of Time Dimension

This section presents a comparative analysis between the second and third MWCV calculations following a similar approach and the same statistical techniques described in the Section 4.3. More specifically, the analysis in this section was intended to determine if the time adjustments made with the quarterly CCIS significantly improved cost estimating effectiveness for the case study item. The result of this analysis are show in Table 5.3. Once again, the statistical analysis showed statistically significant improvements in accuracy and reliability as projects move forward in the three-dimensional cost estimating system. This time, the significant estimating improvement was achieved after the incorporation of the time dimension, with a MAPE and standard deviation improvement of 24.70% and 23.33%, respectively (Equation 5-1 and Equation 5-2). These improvements are deemed significant at a significance level of 1% (paired t-test p-value = 6.15×10^{-6}; F-test p-value = 3.39×10^{-3}), confirming the importance of adjusting bid-based estimates to counteract the impact of inflation and construction market fluctuations.

Table 5.3 Scale Dimension vs. Time Dimension – MWCV Results

	MAPE	**Standard Deviation of APEs**
MWCV 2 – Scale Dimension	16.62%	10.85%
MWCV 3 - Scale + time Dimension	12.52%	8.21%
Reduction	24.70%	24.33%
Significance of Improvement at 1% Significance Level	Statistically Significant	Statistically Significant
p-value	<1%	<1%

5.5 Analysis of Location Dimension

This section presents the statistical comparative analyses between the third and fourth MWCV calculations. The fourth and last calculation was intended to determine the improvement in cost estimating effectiveness offered by the LCI. The results of this analysis are show in Table 5.4. The paired two-sample t-test revealed a further improvement in cost estimating accuracy due to the adjustments for location. There was a statistical significance reduction of 9.61% in the MAPE value (Equation 5-1 and Equation 5-2). On the other hand, the results showed an increase in the standard deviation of the APEs, which would suppose a reduction of 11.08% (Equation 5-1 and Equation 5-2) in the level of reliability after using the LCI. However, the F-test did not show this reduction as statistically significant. Therefore, this study can conclude that the implementation of the LCI would have a significant positive impact on ALDOT's construction cost estimating accuracy with affecting estimating reliability.

Table 5.4 Time Dimension vs. Location Dimension – MWCV Results

	MAPE	**Standard Deviation of APEs**
MWCV 3 - Scale + time Dimension	12.52%	8.21%
MWCV 4 – Scale + Time + Dimension	11.31%	9.12%
Reduction	9.61%	-11.08%
Significance of Improvement at 1% Significance Level	Statistically Significant	Not Statistically Significant
p-value	<1%	>1%

5.6 Distribution of Percentage Errors

Finally, after the development and successful validation of the system, ALDOT should develop the distribution of percentage errors to be used during the eventual implementation of the system to produce stochastic estimates for actual intended projects, as explained in Chapter 4. Table 5.5 shows the percentage errors calculated for each of the testing projects using Equation 1-3. The percentage errors are obtained using the estimated unit prices produced by the three-dimensional cost estimating system, which in the case of this thesis, would be the unit prices yielded by the last calculation of the MWCV process. Figure 5.1 corresponds to the empirical probability distribution built with the percentage errors listed in Table 5.5.

Table 5.5 Time Dimension vs. Location Dimension – MWCV Results

Project	Actual Unit Price	Estimated UP MWCV 4	Percentage Error	Project	Actual Unit Price	Estimated UP MWCV 4	Percentage Error
1	$115.00	$71.76	60.2%	50	$60.58	$61.91	-2.2%
2	$108.98	$106.36	2.5%	51	$81.25	$73.09	11.2%
3	$60.00	$64.19	-6.5%	52	$80.46	$83.23	-3.3%
4	$70.00	$69.85	0.2%	53	$115.00	$74.30	54.8%
5	$87.77	$88.26	-0.6%	54	$79.20	$81.75	-3.1%
6	$67.10	$59.69	12.4%	55	$74.90	$90.38	-17.1%
7	$88.75	$76.23	16.4%	56	$82.60	$82.51	0.1%
8	$89.06	$68.26	30.5%	57	$110.92	$75.50	46.9%
9	$64.34	$68.01	-5.4%	58	$77.50	$75.91	2.1%
10	$80.00	$72.83	9.9%	59	$81.08	$74.59	8.7%
11	$53.89	$60.12	-10.4%	60	$78.86	$73.19	7.7%
12	$80.00	$98.85	-19.1%	61	$88.11	$73.71	19.5%
13	$79.16	$68.74	15.2%	62	$56.50	$59.94	-5.7%
14	$87.38	$82.05	6.5%	63	$72.62	$74.83	-3.0%
15	$72.75	$60.24	20.8%	64	$64.00	$65.16	-1.8%
16	$69.85	$69.70	0.2%	65	$65.19	$66.26	-1.6%
17	$59.46	$67.53	-11.9%	66	$80.02	$85.92	-6.9%
18	$71.06	$67.87	4.7%	67	$63.68	$60.06	6.0%
19	$78.10	$68.33	14.3%	68	$54.45	$61.05	-10.8%
20	$62.77	$80.64	-22.2%	69	$60.00	$66.21	-9.4%
21	$78.77	$65.78	19.8%	70	$85.65	$83.78	2.2%
22	$62.90	$63.09	-0.3%	71	$59.75	$60.65	-1.5%
23	$66.00	$70.05	-5.8%	72	$56.14	$64.75	-13.3%
24	$65.87	$71.28	-7.6%	73	$56.00	$72.18	-22.4%
25	$58.00	$75.55	-23.2%	74	$60.00	$68.40	-12.3%
26	$72.15	$84.34	-14.5%	75	$84.33	$67.60	24.8%
27	$79.15	$60.59	30.6%	76	$62.41	$71.23	-12.4%
28	$68.85	$58.55	17.6%	77	$72.25	$66.75	8.2%
29	$64.99	$65.54	-0.8%	78	$87.25	$66.73	30.7%
30	$75.75	$71.74	5.6%	79	$68.86	$69.59	-1.1%
31	$82.50	$71.00	16.2%	80	$70.13	$74.49	-5.9%
32	$54.25	$60.39	-10.2%	81	$88.33	$86.12	2.6%
33	$95.00	$97.85	-2.9%	82	$75.00	$82.75	-9.4%
34	$57.07	$67.24	-15.1%	83	$78.00	$77.59	0.5%
35	$67.23	$78.10	-13.9%	84	$56.85	$69.56	-18.3%
36	$70.06	$90.03	-22.2%	85	$106.26	$96.98	9.6%
37	$66.01	$62.82	5.1%	86	$70.64	$72.72	-2.9%
38	$80.25	$69.97	14.7%	87	$92.00	$68.25	34.8%
39	$56.00	$62.42	-10.3%	88	$80.30	$77.15	4.1%
40	$66.86	$62.59	6.8%	89	$66.00	$73.51	-10.2%
41	$63.00	$62.01	1.6%	90	$57.24	$66.64	-14.1%
42	$69.75	$72.15	-3.3%	91	$61.23	$70.02	-12.5%
43	$53.25	$56.93	-6.5%	92	$56.95	$57.43	-0.8%
44	$55.82	$66.39	-15.9%	93	$115.00	$98.68	16.5%
45	$55.79	$68.58	-18.7%	94	$73.11	$73.37	-0.4%
46	$90.25	$63.32	42.5%	95	$73.23	$67.60	8.3%
47	$83.92	$74.62	12.5%	96	$58.50	$55.83	4.8%
48	$55.00	$64.76	-15.1%	97	$90.00	$70.85	27.0%
49	$85.00	$87.43	-2.8%	-	-	-	-

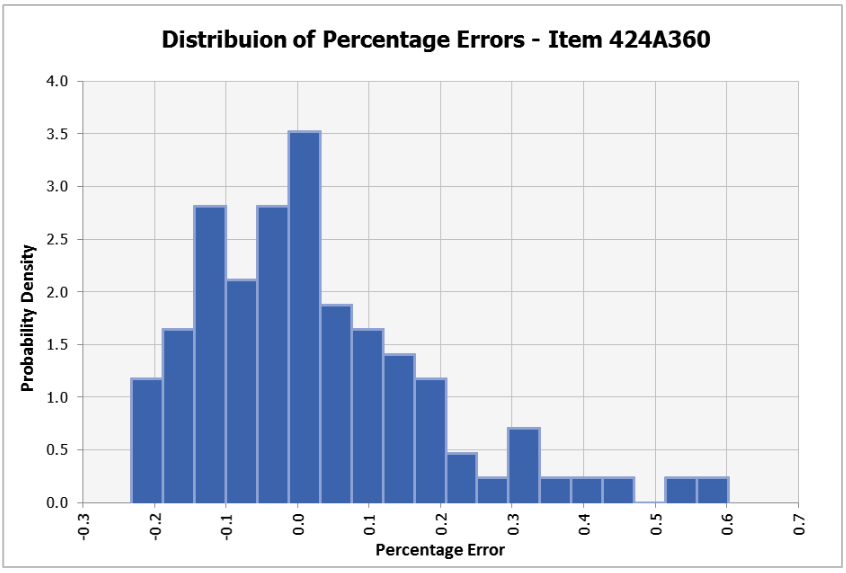


Figure 5.1: Probability Distribution of Percentage Errors – Case Study Item.

Even though this empirical probability distribution could be perfectly used to produce stochastic estimates by multiplying it by the deterministic output of the three-dimensional system, it is important to consider that this distribution is to be saved by ALDOT and distributed among estimators across the state. Sharing an empirical probability distribution is usually more complicated than when using a standard distribution with two or three parameters. For example, if ALDOT uses a normal distribution instead of the empirical distribution in Figure 5.1, all the information to be saved for future use and to be shared among the estimators would be the mean and standard deviation values of the distribution.

Thus, in the last step of the research presented in this thesis, the author used the chi-square goodness of fit statistical test to infer the most suitable standard probability distribution for Figure 5.1. This test was conducted using @Risk, a statistical software package that facilitates the performance of this test to find the most suitable distribution among several possible options. This test found that the distribution of percentage errors for the case study item most probably follows an extreme value distribution, which is defined by two parameters: alpha (location parameter; α) and beta (scale parameter; β). The extreme value distribution that best fits the empirical distribution of percentage errors is illustrated in Figure 5.6. The alpha and beta parameters for this distributions are -0.049 and 0.125, respectively.

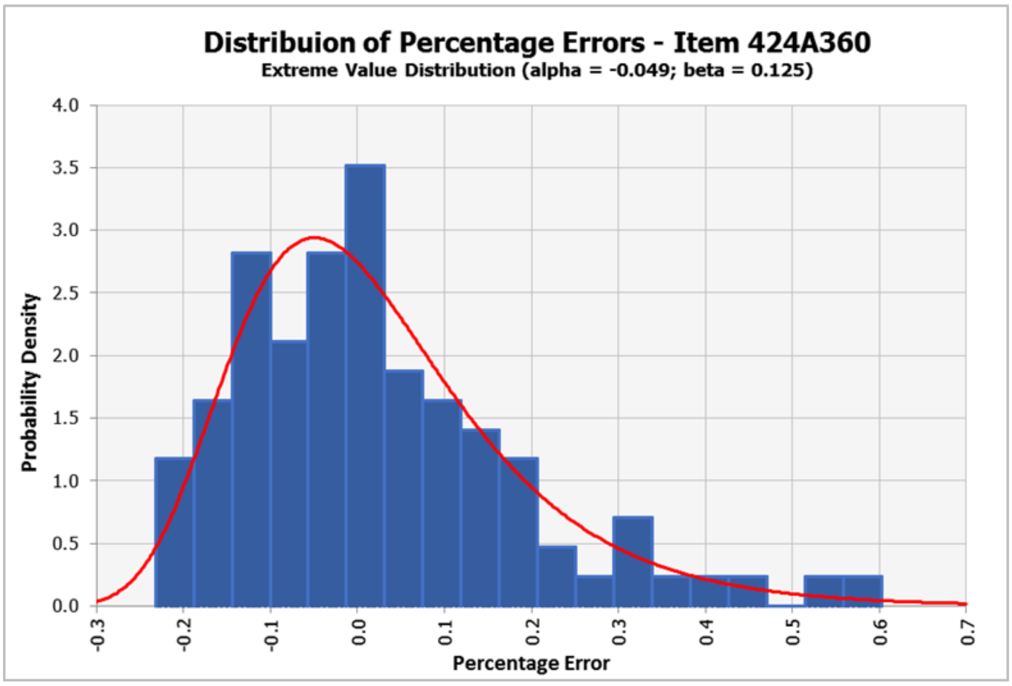


Figure 5.2: Extreme Value Distribution – Case Study Item.

At this point, ALDOT could start using the proposed three-dimensional cost estimating system to estimate unit costs for the case study item, and the extreme value distribution in Figure 5.2 to convert the deterministic outputs of the system into stochastic estimates by multiplying them by this distribution. Different pay items may have different types of percentage error distributions, and it must be noted that the rules for arithmetic operations between constants and probability distribution may vary among them. Equation 5-3 shows how to use the deterministic estimate and the extreme value distribution of the percentage errors to produce an extreme value distribution for the final stochastic estimate for the case study item.

$$SE = DE \times \left(1 + EV(\alpha_{DPE};\beta_{DPE})\right) = EV(\alpha_{SE};\beta_{SE}) = EV(DE \times (1+\alpha); DE \times \beta) \qquad \textbf{Eq. 5-3}$$

Where: SE= Stochastic Estimate

DE= Deterministic Estimate

EV($\alpha_{DPE};\beta_{DPE}$)= Extreme Value Distribution of Probability Errors

EV($\alpha_{SE};\beta_{SE}$)= Extreme Value Distribution for Stochastic Estimate

If for example, ALDOT wants to develop a stochastic estimate for the case study item in a project for which the three-dimensional system has estimated a deterministic unit price of $80/ton, the alpha and beta values of the stochastic estimate would be 76.08 $\left(\alpha_{SE} = 80 \times \left[1 + [-0.049]\right]\right)$ and 10.00 ($\beta_{SE} = 80 \times 0.125$), respectively. This stochastic estimate is illustrated in Figure 5.3, which also shows that ALDOT can be 75% sure that the unit price for the case study item in this contract will not be more than \$88.5/ton. Similarly, ALDOT could use this stochastic estimate to make management decision at difference confidence levels.

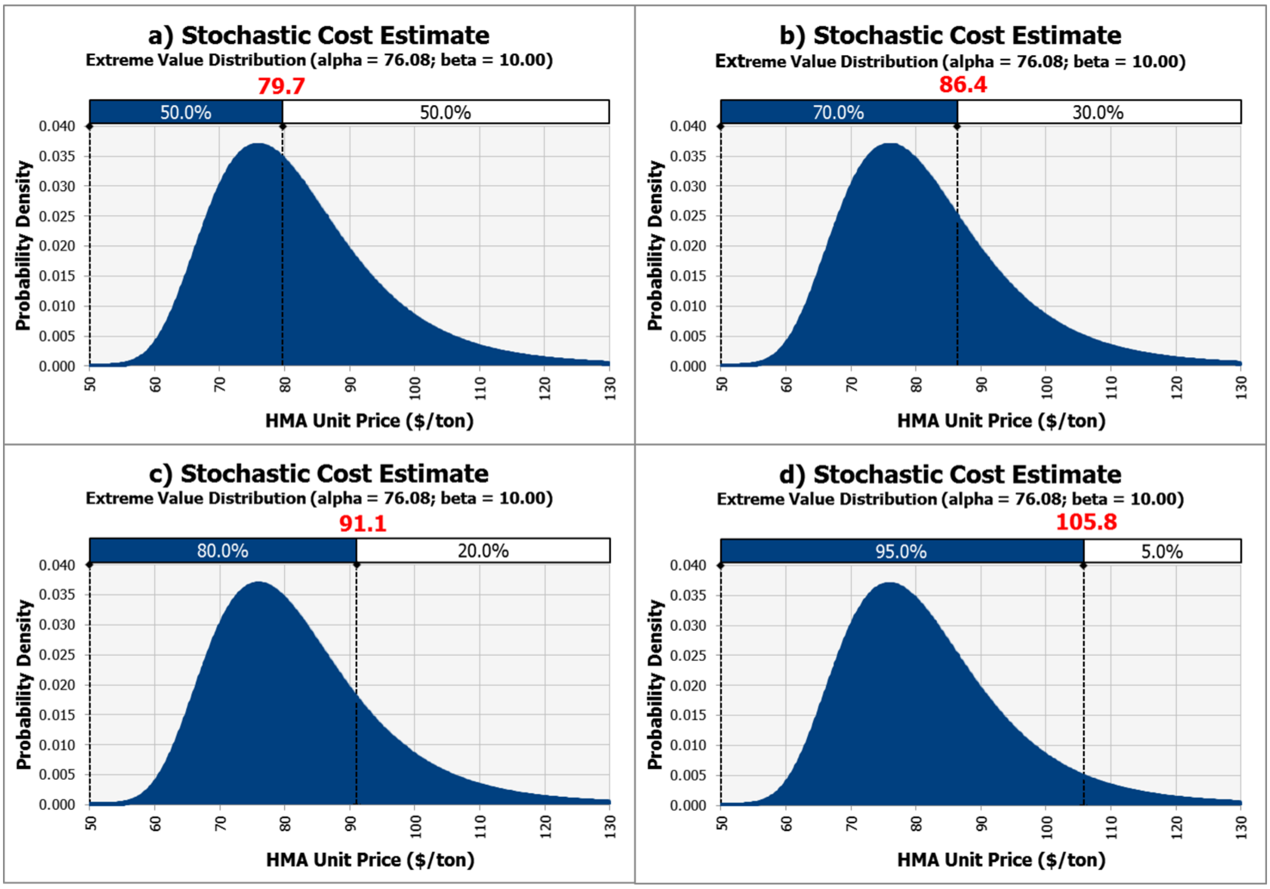


Figure 5.3: Stochastic Cost Estimate – Example.

The confidence levels in Figure 4.3 have been arbitrarily selected for illustration purposes. It is reasonable to assume that estimators would not use confidence levels below 50% for any project given that the probability of underfunding the project would be higher. Other than that, there is no objective way to establish the optimal confidence level for a given project. This is usually a subjective decision made by estimators based on their previous experiences and on their level of understanding of the specific project requirements. Further research efforts should be aimed to develop mechanisms to guide STA estimators in the effective selection of confidence level in stochastic estimates.

5.7 Summary

This chapter presented a critical analysis of the results obtained from the four MWCV calculations conducted to validate the effectiveness of the proposed three-dimensional cost estimating system. The MAPE and standard deviation of APEs were calculated after running each MWCV calculation on all 2016 projects using the case study item. The results of the validation process are summarized in Figure 5.4. The comparison of MAPE and standard deviation values between MWCV calculations showed a significant improvement in estimating accuracy and reliability with the incorporation of each dimension. The negative improvement in reliability between the third and fourth MWCV calculations (MWCV 3 and 4) indicates that there was no actual improvement in reliability after including the location factor in the system. On the contrary, the level of reliability decreased after considering this dimension. However, a statistical analysis revealed that the increase in the standard deviation between these two calculations was not statistically significant, unlike the improvement in accuracy observed between the same two calculations. It allows this study to conclude that the location dimension also has a positive impact on cost estimating effectiveness, by significantly improving accuracy without affecting estimating reliability. It should be noted that all other reductions in MAPE and standard deviation values have proven to be statistically significant, resulting in an overall significant improvement of 47.7% and 43.6% in accuracy and reliability, respectively.

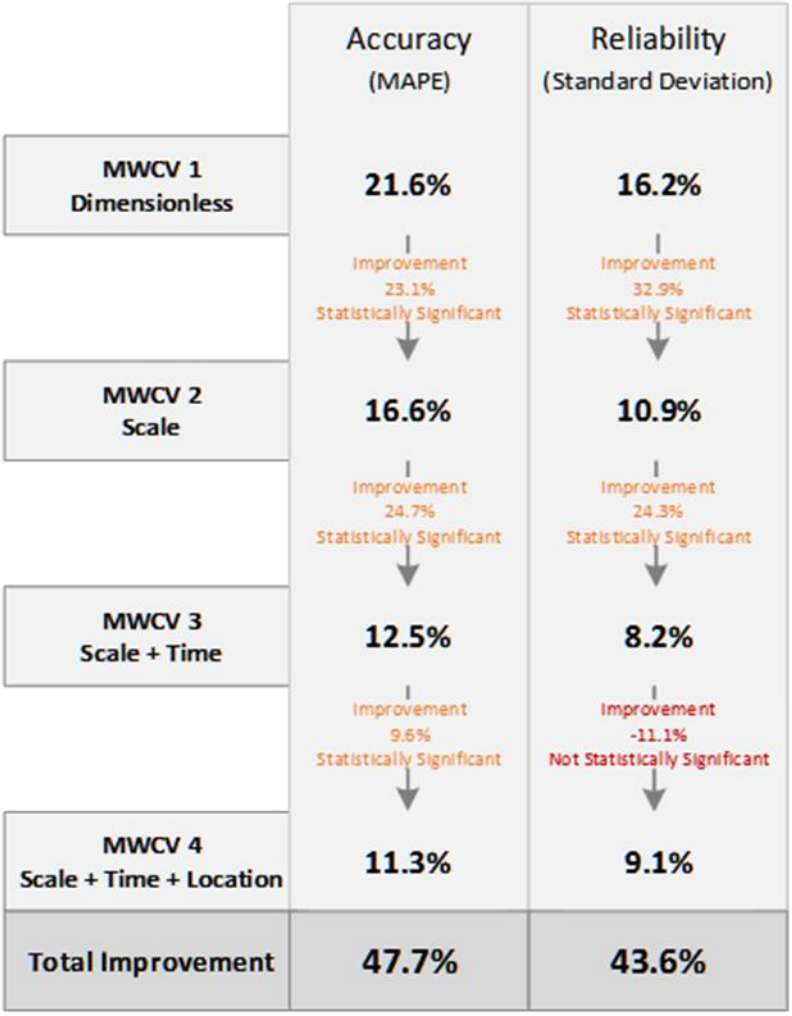


Figure 5.4:MAPE and Standard Deviation of APEs for each MWCV.

This chapter finished with a calculation of a stochastic HMA unit price estimate for a hypothetical project, including some examples on how to use the stochastic estimate under various confidence levels. This chapter has also concluded the presentation of the research efforts undertaken by the author before compiling in the next chapter all major findings and contributions made in this study.

CHAPTER SIX:

CONCLUSION AND RECOMMENDATION

This thesis has presented the exhaustive research efforts undertaken for the development of a stochastic three-dimensional cost estimating system for ALDOT. The three dimensions in this system are associated with three major factors influencing the estimation of construction costs in the transportation industry: 1) scale; 2) time; and 3) location. The three-dimensional system is actually part of a larger bid-based cost estimating framework that incorporates another important cost-influencing factor, estimating uncertainty. This additional factor was considered by developing a distribution of percentage errors with the results from the validation process during system development. This distribution is intended to be used during the actual implementation of the system to quantify the uncertainty associated with each deterministic estimate produced by the three-dimensional system. Level of competition was also considered as one of the cost-driven factors in the proposed methodology, but the assessment of this factor revealed no substantial changes in levels of competition across the state of Alabama; therefore, this factor was discarded. All cost-influencing factors were assessed using bid data from all projects awarded by ALDOT between 2006 and 2016 (3,661 projects).

The relationship between HMA prices across the state of Alabama and the scale, time, and location dimensions was modeled using non-linear regression techniques, a quarterly indexing approach, and a location cost index, respectively, to estimate HMA unit prices at the pay item level. To facilitate the implementation of the proposed methodology by ALDOT, all the elements of the system were integrated through the four-step framework outlined below. ALDOT's estimators could apply this four-step framework after having calculated the expected amount of HMA required for an intended project.

- **Step 1 – Scale Dimension:** Define an optimal look-back period for data retrieval following the protocol proposed by Pakalapati (2018) and develop a non-linear regression model to correlate the historical unit prices contained in the look-back period with their respective quantities of work. Use this regression model to generate a deterministic HMA cost estimate for the expected amount of HMA to be placed in the current project.

- **Step 2 – Time Dimension:** Assuming that the deterministic estimate obtained from Step 1 is valid at the mid-point of the look-back period; use a quarterly CCIS to adjust this estimate for inflation and any other price fluctuations that may have happened between the mid-point of the look-back period and the current date. This step provides deterministic time-adjusted estimates.
- **Step 3 – Location Dimension:** The deterministic time-adjusted estimate from Step 2 corresponds to a state average value since the historical projects in the look-back period were awarded all across the state and the quarterly CCIS has been designed to be applied at the state-level. Thus, the proposed LCI is then applied to adjust the estimate form Step 2 to the specific region in which the current project is intended to be executed. For location indexing purposes, this study has divided the state of Alabama into three geographic regions: north, central, and south.

Validation was systematically performed through multiple calculations of an innovative MWCV process intended to demonstrate and quantify the improvement in estimating accuracy and reliability offered by each of the three dimensions. The improvement offered by each dimension refers to the increase in estimating accuracy and reliability measured after incorporating the dimension into a system that was not considering it before. Four MWCV calculations were performed for a case study item, which is the most relevant item in ALDOT's asphalt paving projects (HMA pay item 424A360). The following are the results for each MWCV calculation for the case study item:

- **MWCV 1 – Dimensionless:**
    - MAPE = 21.6%
    - Standard Deviation of APEs = 16.2%
- **MWCV 2 – Scale Dimension:**
    - MAPE = 16.6%. Statistically significant improvement in accuracy in comparison to the dimensionless MWCV.
    - Standard Deviation of APEs = 10.9%. Statistically significant improvement in reliability in comparison to the dimensionless MWCV.
- **MWCV 3 – Scale plus Time Dimension:**

- MAPE = 12.5%. Statistically significant improvement in accuracy in comparison to the MWCV calculation for the scale dimension.
- Standard Deviation of APEs = 8.2%. Statistically significant improvement in reliability in comparison to the MWCV calculation for the scale dimension.

* **MWCV 4 – Scale plus Time plus Location Dimension (Three-Dimensional Cost Estimating System):**
    - MAPE = 11.3%. Statistically significant improvement in accuracy in comparison to the MWCV calculation for the scale plus time dimension.
    - Standard Deviation of APEs = 9.1%. There is no enough evidence to prove a statistically significant reduction in reliability in comparison to the MWCV calculation for the scale plus time dimension.

The results summarized before allowed the author to conclude that each of the three dimensions would be expected to significantly improve HMA cost estimating effectiveness in terms of accuracy and reliability. Even though the fourth MWCV calculation, with all three dimensions, revealed an increase in the standard deviation of APEs after incorporating the location dimension, a statistical F-test found that there is no evidence to prove that this increase corresponds to a significant reduction in the level of reliability. Therefore, it would be reasonable to conclude that the use of the proposed LCI would have a significant positive impact on HMA cost estimating accuracy without significantly affecting estimating reliability.

A comparison between the estimating performance of the first and fourth MWCV calculations showed an overall statistically significant improvement of 47.7% and 43.6% in accuracy and reliability, respectively, as a result of the integration of the scale, time, and location dimensions.

The thesis also presents an example of the calculation of a stochastic estimate for a hypothetical project using the case study item by combining its deterministic estimate from the three-dimensional cost estimating system with the distribution of percentage errors obtained from the validation process presented in this thesis. Finally, it should be noted that besides proving the importance of considering scale, time, location, and uncertainty impacts on HMA cost estimating, this thesis has also proven the effectiveness of the tools used to model these impacts, as well as the effectiveness of the four-step framework used to combine these factors. The consideration of these factors only improves cost estimating effectives if each factor is properly assessed, modeled, and integrated into the estimating process, as demonstrated in this thesis.

6.1 Study Limitations

The following are the main limitations associated with this study, which must be considered when interpreting the results presented in this thesis and during the implementation of the proposed cost estimating methodology:

- The cost estimating system presented in this thesis is intended to maximize accuracy and reliability on the estimation of original contract amounts. It is not aimed to estimate construction costs at project completion, which may differ from original contract amounts due to scope creep, design errors or omissions, differing site conditions, and/or change orders issued by ALDOT during the project construction phase.
- The specific quantitative results presented in this study are only applicable to the case study item in contracts to be awarded by ALDOT. However, every step of the study is presented in great detail to make it possible for ALDOT to develop and implement this methodology for other pay items.

6.2 Recommendations for Future Research

During the study, the author identified a number of research questions that should be considered in future studies. These questions are outlined below:

- What factors are causing the HMA price difference between geographic regions in Alabama?
    - The LCI developed in this study revealed significant differences in HMA prices across the state of Alabama. However, the factors causing price differences among geographic

regions have not been clearly defined at this writing. A better understanding of these factors could help to improve the performance of the cost estimating system proposed in this thesis.

- What factors are causing the HMA price difference between geographic regions in Alabama?
  - The LCI developed in this study revealed significant differences in HMA prices across the state of Alabama. However, the factors causing these differences among geographic regions have not been clearly defined at this writing. A better understanding of these factors could help to improve the performance of the cost estimating system proposed in this thesis.
- Does "level of competition" actually impact HMA prices in Alabama?
  - In this study, the impact of level of competition on HMA prices was assumed to be related to the number of vendors competing on ALDOT's projects on each of the three geographic regions. No substantial differences were found on the average number of vendors per project per region; therefore, this factor was not considered in the proposed cost estimating system since it was assumed to equally impact all ALDOT's projects. However, the author recognizes that further research is needed on this area given that the level of competition is not only a result of the number of vendors competing on a project. In fact, the level of competition is relative to the perception of each bidder. Even though an anticipated large number of bidders may drive vendors to submit lower price proposal, there are other factors that may have a similar effect. Some of those factors could be anticipated competitors with access to cheaper suppliers/subcontractors or to more cost-effective materials or construction methods.
- How the proposed cost estimating system would perform for other pay items different from the case study item?
  - The proposed MCVW approach has positively proven the ability of the proposed cost estimating methodology to improve ALDOT's HMA cost estimating in terms of accuracy and reliability. However, the validation process was only conducted on a

single relevant HMA item. Further research is needed to determine if the proposed system would have a similar effect on cost estimating procedures for other HMA pay items used in ALDOT paving projects.

- How the proposed cost estimating system would perform for other pay items different from the case study item?
    - The proposed MCVW approach has positively proven the ability of the proposed cost estimating methodology to improve ALDOT's HMA cost estimating in terms of accuracy and reliability. However, the validation process was only conducted on a single relevant HMA item. Further research is needed to determine if the proposed system would have a similar effect on cost estimating procedures for other HMA pay items used in ALDOT paving projects.
- Is the proposed system more effective than ALDOT's current HMA cost estimating system?
    - This study measures the improvement in accuracy and reliability offered by each dimension by comparing the estimating performance of the system before and after incorporating the dimensions one-by-one. The results from these comparisons have allowed the author to conclude that the factors considered in this study has the potential to improve ALDOT's cost estimating practices. However, the effectiveness of the proposed system has not been compared against the accuracy and reliability of ALDOT's current cost estimating practices given that the author had no access to ALDOT's estimates. Future research should compare the estimating performance of the proposed system against HMA cost estimates produced by ALDOT's estimators in other to better determine the level of improvement to be expected by ALDOT if the three-dimensional cost estimating system presented in this thesis is implemented.
- How can an estimator determine the optimal confidence level to set a cost estimate for a given project using a stochastic estimate?
    - By the end of the thesis, the author presents some general scenarios that may lead to higher or lower confidence levels during the interpretation of stochastic estimates.

However, further research is required towards the development of objective confidence levels according to the specifics of each project.

REFERENCES

Alabama Asphalt Pavement Association (AAPA). (2016). *Asphalt Facts*. Retrieved from AAPA: http://alasphalt.com/asphalt-facts/

Alabama Department of Transportation (ALDOT). (2018). *2018 Standard Specification for Highway Construction.* ALDOT.

American Association of State Highway and Transportation Officials (AASHTO). (2013). *Practical Guide to Cost Estimating.* AASHTO

American Society of Civil Engineers (ASCE). (2013). *2013 Report Card for America's Infrastructure*. Retrieved from http://www.infrastructurereportcard.org/

American Society of Civil Engineers (ASCE). (2015). *Report Card for Alabama's Infrastructure - 2016*. ASCE

Anderson, S., Damnjanovic, I., Nejat, A., & Ramesh, S. (2009). Synthesis on construction unit cost development: Technical report.

Akanni, P. O., Oke, A. E., & Akpomiemie, O. A. (2015). Impact of environmental factors on building project performance in Delta State, Nigeria. *HBRC Journal*, *11*(1), 91-97.

Akintoye, A. (2000). Analysis of factors influencing project cost estimating practice. *Construction Management & Economics*, *18*(1), 77-89.

Arafa, M., & Alqedra, M. (2011). Early stage cost estimation of buildings construction projects using artificial neural networks. *Journal of Artificial Intelligence*, *4*(1), 63-75.

Asmar, M. E., Hanna, A. S., & Whited, G. C. (2011). New approach to developing conceptual cost estimates for highway projects. *Journal of construction engineering and management*, *137*(11), 942-949.

Asphalt Pavement Alliance (APA). (2010). *The Role of Asphalt in Livable Communities.* APA

Bajaj, A., Gransberg, D. D., & Grenz, M. D. (2002). Parametric estimating for design costs. *AACE International Transactions*, ES81.

Byrnes, J. E. (2002). *Best practices for highway project cost estimating*. ARIZONA STATE UNIV TEMPE.

Betts, M. (2007). Standardization in Information Technology Industries: Emerging Issues under Section Two of the Sherman Antitrust Act. *Oklahoma Journal of Law and Technology*, *3*(1), 6.

Cho, Y. K., Youn, J. H., & Martinez, D. (2010). Error modeling for an untethered ultra-wideband system for construction indoor asset tracking. *Automation in Construction*, *19*(1), 43-54.

Congress, U. S. (2012). Moving ahead for progress in the 21 st century act. *Washington, DC*.

Cuervo, J. C., & Sui Pheng, L. (2003). Significance of location factors for Singapore transnational construction corporations. *Engineering, Construction and Architectural Management*, *10*(5), 342-353.

Delaney, J. W., & Mohan, Satish. (2016). *The Effect of the Level of Competition on Construction Bid Quality.* State University of New York at Buffalo.

Dikmen, I., Birgonul, M. T., & Gur, A. K. (2007). A case-based decision support tool for bid mark-up estimation of international construction projects. *Automation in Construction*, *17*(1), 30-44.

Fakültesi, M., & Zeynep, D. (2004). A Neural Network Approach for Early Cost Estimation of Structural Systems of Buildings. *International Journal of Project Management*, 595-602.

Federal Highway Administration FHWA (1988). Bid Analysis and Unbalanced Bids – Memorandum HHO-32. *U.S. Department of Transportation*, Washington, DC.

Fragkakis, N., Lambropoulos, S., & Tsiambaos, G. (2010). Parametric model for conceptual cost estimation of concrete bridge foundations. *Journal of Infrastructure Systems*, 17(2), 66-74.

Gardner, B. J. (2015). *Applying artificial neural networks to top-down construction cost estimating of highway projects at the conceptual stage*. Iowa State University.

Golafshani, N. (2003). Understanding reliability and validity in qualitative research. *The qualitative report*, *8*(4), 597-606.

Harbuck, R. H. (2002). Using models in parametric estimating for transportation projects. *AACE International Transactions*, ES51.

Hendrickson, C., & Au, T. (1989). *Project management for construction: Fundamental concepts for owners, engineers, architects, and builders*. Chris Hendrickson.

Holley, I. B. (2003). Blacktop: How asphalt paving came to the urban United States. *Technology and Culture*, *44*(4), 703-733.

Iglewicz, B., & Hoaglin, D. C. (1993). *How to detect and handle outliers* (Vol. 16). Milwaukee, WI: ASQC Quality Press.

JICA (Japan International Cooperation Agency). (2010). Evaluation guide for prequalification and bidding under Japanese ODA loans: Procurement of goods and services.

Juszczyk, M. (2013, October). The use of artificial neural networks for residential buildings conceptual cost estimation. In *AIP Conference Proceedings* (Vol. 1558, No. 1, pp. 1302-1306). AIP.

Kaming, P. F., Olomolaiye, P. O., Holt, G. D., & Harris, F. C. (1997). Factors influencing construction time and cost overruns on high-rise projects in Indonesia. *Construction Management & Economics*, *15*(1), 83-94.

Kelly, L. S., Leslie, T. G., & Lynn, D. E. (2002). Pavement Smoothness Index Relationships: Final Report. *Publication No. FHWA-RD-02-057, U.S. Department of Transportation, Federal Highway Administration, Research, Development, and Technology, Turner-Fairbank Highway Research Center*.

Kim, G. H., An, S. H., & Kang, K. I. (2004). Comparison of construction cost estimating models based on regression analysis, neural networks, and case-based reasoning. *Building and environment*, *39*(10), 1235-1242.

Kwak, Y. H., & Watson, R. J. (2005). Conceptual estimating tool for technology-driven projects: exploring parametric estimating technique. *Technovation*, *25*(12), 1430-1436.

Liu, L., & Zhu, K. (2007). Improving cost estimates of construction projects using phased cost factors. *Journal of Construction Engineering and Management*, *133*(1), 91-95.

Lo, W., Lin, C. L., & Yan, M. R. (2007). Contractor's opportunistic bidding behavior and equilibrium price level in the construction market. *Journal of construction engineering and management*, *133*(6), 409-416.

Marks, H. (2009). Smoothness matters: the influence of pavement on fuel consumption. *HMAT: Hot Mix Asphalt Technology*.

Molenaar, K. R. (2005). Programmatic cost risk analysis for highway megaprojects. *Journal of Construction Engineering and Management*, *131*(3), 343-353.

Motulsky, H. J., & Brown, R. E. (2006). Detecting outliers when fitting data with nonlinear regression–a new method based on robust nonlinear regression and the false discovery rate. *BMC bioinformatics*, *7*(1), 123.

National Asphalt Pavement Association (NAPA). (2018). Engineering Overview. Retrieved from NAPA: http://www.asphaltpavement.org/index.php?option=com_content&view=article&id=14&Itemid=33

Nicholls, J. C., McHale, M. J., & Griffiths, R. D. (2008). *Best practice guide for durability of asphalt pavements*. Wokingham, UK: Transport Research Laboratory.

Odeck, J. (2004). Cost overruns in road construction—what are their sizes and determinants?. *Transport policy*, *11*(1), 43-53.

Pakalapati, K. (2018). *Data Usage Optimization for Cost Estimating In Asphalt Paving Projects Using a Cost Indexing System.* (Master's Thesis, Auburn University).

Project Management Institute (PMI). (2013). *A Guide to the Project Management Body of Knowledge (PMBOK® Guide), Fifth Edition, ANSI/PMI 99-001-2013.* Project Management Institute, Inc.

Phaobunjong, K. (2002). *Parametric cost estimating model for conceptual cost estimating of building construction projects* (Doctoral dissertation).

RSMeans (2018). *Building Construction Costs with RSMeans Data: 76th Annual Edition*. Gordian Group, LLC.

Rueda-Benavides, J. A. (2016). *Indefinite Delivery/Indefinite Quantity project selection framework using stochastic techniques* (Doctoral dissertation, Iowa State University).

Schexnayder, C. J., Weber, S. L., & Fiori, C. (2003). Project cost estimating: A synthesis of highway practice. *Transportation Research Board.*

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

Shane, J. S., Strong, K. C., & Gad, G. M. (2015). *Risk-Based Engineers Estimate* (No. MN/RC 2015-10). Minnesota Department of Transportation, Research Services & Library.

Smith, K. L., Smith, K. D., Evans, L. D., Hoerner, T. E., Darter, M. I., & Woodstrom, J. H. (1997). *Smoothness specifications for pavements* (No. NCHRP Project 1-31).

Sonmez, R. (2004). Conceptual cost estimation of building projects with regression analysis and neural networks. *Canadian Journal of Civil Engineering*, *31*(4), 677-683.

Trost, S. M., & Oberlender, G. D. (2003). Predicting accuracy of early cost estimates using factor analysis and multivariate regression. *Journal of Construction Engineering and Management*, *129*(2), 198-204.

Utah Construction & Design. (2013). UDOT - Cost-Based Estimates. *Utah Construction & Design*, 41-42.

Wickham, H. (2014). Tidy data. *Journal of Statistical Software, 59(10)*, 1-23.

William Manfredonia, J. P. (2016, 11 02). *Cost Estimating*. Retrieved from WBDG: https://www.wbdg.org/design-disciplines/cost-estimating

Xu, J. W., & Moon, S. (2011). Stochastic forecast of construction cost index using a cointegrated vector autoregression model. *Journal of Management in Engineering*, *29*(1), 10-18.

Ye, K., Shen, L., & Lu, W. (2014). A discriminant model for measuring competition intensity of construction market. *Engineering, Construction and Architectural Management*, *21*(2), 152-169.

Zhang, Z., & Sun, Y. (2007). Economic potential of modular reactor nuclear power plants based on the Chinese HTR-PM project. *Nuclear Engineering and Design*, *237*(23), 2265-2274.

Zou, P. X., Zhang, G., & Wang, J. (2007). Understanding the key risks in construction projects in China. *International Journal of Project Management*, *25*(6), 601-614.