Planning-Phase Estimation of Construction Time for a Large Portfolio of Highway Projects

Junghye Son, Ph.D., P.E., M.ASCE[1]; Nabeel Khwaja, P.E.[2]; and Duane S. Milligan, P.E.[3]

Abstract: Estimation of highway project construction time prior to design development has received comparatively little focus from researchers. State transportation agencies (STAs) and their management teams need a method or tool to estimate construction time during the project planning and development phases, where STAs typically manage hundreds of projects at the program level. These programs of projects are updated periodically, which requires an estimation methodology that is efficient, less time consuming, and reliable during the phases where limited project information is available. This paper presents an approach to developing a planning-phase construction time estimation model using linear regression modeling on actual construction time and cost data from 623 highway projects completed by the Dallas District Office, Texas Department of Transportation (TxDOT). The study developed a multiple linear regression model with three project parameters that were identified as key predictors to construction time. The accuracy and reliability of the developed model were higher than those of the well-known Bromilow's time-cost (BTC) model. The model was also validated by the Mann-Whitney test of its applicability to a large group of projects. The proposed model requires less effort to develop, update, and revise with new data, which allows STAs to conduct planning-phase estimation of the construction time of a large portfolio of diverse projects effectively and efficiently. STAs would benefit from using this model in various project planning areas such as financial planning, staff planning, transportation impact mitigation, budget allocation, and project prioritization. This study also contributes to the body of knowledge with the proposed construction time estimation methodology during the planning and development phase that has been less focused. **DOI: 10.1061/(ASCE)CO.1943-7862.0001637.** *© 2019 American Society of Civil Engineers.*

Introduction

Construction duration of transportation projects affects public mobility, access, economy, safety, and environment (Mallela and Sadavisam 2011). Consequently, construction of multiple transportation projects in an area or region can adversely affect local community and environment. By Title 23, US Code Section 135, STAs are mandated to develop and update a statewide transportation improvement program (STIP) covering at least 4 years of projects (FHWA 2017; McCoy et al. 2016; TxDOT 2016). The number of projects that STAs need to plan and manage in accordance with the STIP can reach several thousand projects. For example, the Texas Department of Transportation (TxDOT) will start construction of more than 6,000 projects within 4 years, and approximately 2,000 projects are planned to start in 5–10 years as of early May 2018 (TxDOT 2018). Large districts, such as Dallas, Houston, and San Antonio, routinely plan and develop 200–400 projects in a 4-year time horizon. STIP, a fiscally constrained planning document, includes key project information such as project budget, project limits, county, roadway, and a general description of work. However, it does not contain information on the construction phase duration (TxDOT 2016). Construction time estimation during the planning and development phases could mitigate negative impacts on the local community and environment and enhance various management activities, including general project planning, funding decisions and financial planning, public safety, manpower planning, traffic impact assessment, and public relations (Irfan et al. 2011; Williams 2008). This calls for an efficient and reliable method to estimate construction time as early as the planning and development phases.

Nonetheless, construction time estimation during the planning and development phases has received less attention (Czarnigowska and Sobotka 2014; Williams 2008). Most effort on the time estimation of highway construction has been focused on contract time determination (CTD). For the low-bid design-bid-build projects procured by state transportation agencies (STAs), the contract time is prepared by the engineering department as part of the plans, specifications, and estimate (PS&E) package at the end of design development for project letting (TxDOT 2017b). The Federal Highway Administration (FHWA) requires all STAs to have a formal procedure to determine contract time for federally funded highway projects (FHWA 2002). In contrast to cost estimate preparation being required from the early stages of projects by FHWA and STAs (FHWA 2017; TxDOT 2017b), the requirements on time estimation are mostly focused on CTD. In other words, it is not until the PS&E package preparation that STA engineers prepare the construction time estimate of a project (Williams 2008). As a consequence, many STAs do not have formal procedures for planning-phase estimation of construction time. Williams (2008) conducted interviews with 25 STAs to assess the current practices of early construction time estimation. Whereas 20 respondents (80% of 25 total respondents) perceived the need for construction time estimation during the conceptual design, most respondents answered

[1]Research Associate, Center for Transportation Research, Univ. of Texas at Austin, 1616 Guadalupe, Suite 4.202, Austin, TX 78701 (corresponding author). ORCID: https://orcid.org/0000-0001-9460-0303. Email: junghye.son@austin.utexas.edu

[2]Assistant Director, Center for Transportation Research, Univ. of Texas at Austin, 1616 Guadalupe, Suite 4.202, Austin, TX 78701. Email: khwaja@mail.utexas.edu

[3]Director of Construction, Dallas District, Texas Dept. of Transportation, 4777 East Highway 80, Mesquite, TX 75150. Email: Duane .Milligan@txdot.gov

Note. This manuscript was submitted on March 1, 2018; approved on October 1, 2018; published online on February 14, 2019. Discussion period open until July 14, 2019; separate discussions must be submitted for individual papers. This paper is part of the ***Journal of Construction Engineering and Management***, © ASCE, ISSN 0733-9364.

that there were no procedures or tools, or they prepared the time estimates based on personal experience or historical data from past projects.

The primary challenge in preparing a reliable time estimate during the planning and development phases is the lack of project information. Key project parameters, including location, lane-miles, and project type, are available, but information such as a detailed traffic control plan and bid item quantities are not. The traffic control plan defines construction staging and work sequences, and bid item quantities determine activity durations. Without such detailed project information, it is neither effective nor efficient to apply the current CTD process, which is based on activity durations and work sequences, to estimating construction time during the planning and development phases. This necessitates developing a separate construction time estimation method different from the current CTD process.

A limited number of studies have developed a construction time estimation model for highway projects during earlier project phases (Czarnigowska and Sobotka 2014; Irfan et al. 2011; Williams 2008). Those models presented a methodology to develop a statistical time prediction model and demonstrated that the cost estimate and a few key project parameters were valid predictors of construction time. Besides the insufficient number of studies, the previous studies that used regression models have not fully examined a possible difference between contract and actual cost in developing time estimate models. In developing regression models to estimate construction time from historical data, either the actual or contract cost (or cost estimate) can be used. Considering the fact that only the cost estimate is known before construction finishes, however, there is a need to closely examine which cost is more appropriate for modeling, as Irfan et al. (2011) noted. Another major gap of extant research that needs to be filled is to test the applicability of a time estimation model to a group of projects. From the STA perspective, it is more important to have an effective and efficient tool to estimate construction time for a large number of projects. However, previous studies have not paid attention to this aspect in developing a time estimation model for the planning phases.

The purpose of this study was to develop an efficient and reliable method for construction time estimation of a large portfolio of highway projects for the planning and development phases based on cost and key project parameters. The study subject was limited to highway construction projects. Nonconstruction projects, such as maintenance, traffic signal, and landscaping projects, were excluded. All data were collected from the low-bid, design-bid-build projects procured by TxDOT's Dallas District between 2003 and June 30, 2017. This study did not intend to estimate construction time of highway projects using alternative project delivery or contracting strategies (e.g., A + B bidding method, design-build). This study focused on cost and construction time. Any other cost or time components prior to construction, for example, engineering cost, were not included in the research scope.

Construction Time Estimation

Time Estimation Types

Time estimation can be categorized in various ways, for example, by technical aspects (e.g., simulation, parametric, and expert system), variability (i.e., deterministic or probabilistic), or calculation logic (e.g., critical path method and line of balance). By estimating purposes, time estimation methods can be categorized into three types—conceptual, predictive, and scheduling estimation. Conceptual estimation is a ballpark estimate based on past projects with similar size, location, or type. It is usually performed at project inception. The primary purpose of this estimation is to understand the project scope and timeline of the whole project phases. Because little information is available, conceptual estimation depends mainly on experts' experience and knowledge on similar projects (Williams 2008).

On the other hand, scheduling estimation is generally used for contract time determination, construction work planning, and construction progress management. Popular scheduling techniques include bar chart, critical path method, program evaluation and review technique, and line of balance. These techniques calculate construction time based on the relationships and sequences of activities. Therefore, scheduling estimation typically requires the design information for a project, such as individual activities, work sequences, item quantities, and other relevant details that are not necessarily available prior to detailed design development. The CTD process implemented by STAs (FHWA 2002; Herbsman and Ellis 1995) is a type of scheduling estimation. The CTD process uses bid item quantities, production rates, and productivity factors with a bar chart or critical path method schedule to calculate contract time.

Predictive time estimation is used when project design is still being developed. Thus, it involves key project information and parameters in more detail than conceptual estimation, but it does not necessarily require activity relationships and work sequences, as scheduling estimation does. Therefore, predictive time estimation has higher reliability and accuracy than conceptual estimation and lower reliability and accuracy than scheduling estimation. The regression analysis is the most popular technique of the predictive estimation method, and neural network modeling is one of the emerging techniques (Jin et al. 2016; Odabaşı 2009). A neural network estimation model incorporates project parameters into its neural network program, and the program learns from past cases and refines the learning process by reducing errors and increasing accuracy. Bhokha and Ogunlana (1999) developed an artificial neural network model to predict construction durations of buildings at the predesign phase. The artificial neural network model developed by Mensah et al. (2016) used key work item quantities to determine the construction durations of rural roads. Researchers have found that neural network models would predict construction time more or equally reliably compared with regression models (Bhokha and Ogunlana 1999; Mensah et al. 2016; Nani et al. 2017; Pewdum et al. 2009). Neural networks are a promising technique with a potential for resolving complex problems encountered in the construction management area. The major disadvantages addressed by researchers for its practical application are the "black-box" approach and complexity (Bhokha and Ogunlana 1999; De la Garza and Rouhana 1995; Kim et al. 2004). Although studies have been probing hidden layers of neural network models and starting to uncover its black box (Lu et al. 2001), the usability of neural network modeling for daily practices is yet to be improved. Neural network models require a certain level of expert knowledge to develop and train algorithms, which could be a challenge for practitioners to develop, train, and update time estimation models based on neural networks.

Regression analysis uses a statistical approach that defines a relationship between a response (dependent) variable and a predictor (independent) variable(s). In construction time estimation, the response variable is construction time, and a predictor variable is a project parameter or parameters that have an important impact on construction time. Bromilow's time-cost (BTC) model is one of the first well-known regression models (Hoffman et al. 2007; Jarkas 2016; Love et al. 2005; Odabaşı 2009; Williams 2008). Regression analysis is generally considered cost effective and simple compared with other estimation techniques because it is easier to modify and update the model than neural networks or other artificial intelligence–based models. (De la Garza and Rouhana 1995; Jin et al. 2016; Kim et al. 2004). The major disadvantage of regression analysis is its accuracy. Some researchers have claimed that the regression approach is less accurate in predicting cost or time than neural network modeling (De la Garza and Rouhana 1995; Kim et al. 2004). Nani et al. (2017), on the other hand, demonstrated that there were no significant differences in accuracy between the two approaches in time estimation. Nevertheless, researchers argue that regression analysis is a preferred option for planning-phase time estimation considering its simplicity, cost effectiveness, and practical applicability compared to neural networks or other estimation methods such as simulation or fuzzy theory (Czarnigowska and Sobotka 2014; Sousa et al. 2014).

Time Estimation Model Using Regression

Regression modeling incorporates project parameters to obtain an estimate. For construction time estimation, a regression model set construction time as a response variable and other project parameters as predictor variables. Bromilow conducted regression on the construction time and cost data of 309 building projects and found a linear relationship between the natural logarithm of construction cost and the natural logarithm of construction duration (Bromilow 1969). The BTC model can be expressed by Eq. (1)

$$T = K \times C^B \quad \text{or} \quad \ln T = \ln K + B \ln C \tag{1}$$

where T = actual construction time in working days; C = final construction cost; and K and B = constant. Since Bromilow's study, researchers in building construction have utilized the BTC model. Some studies used a simple linear regression model with only a cost component as a predictor variable, and others developed a multiple linear regression model that included additional predictor variables such as number of floors, project type, and location (Hoffman et al. 2007; Jarkas 2016; Love et al. 2005; Skitmore and Ng 2003). Compared with building construction, not many studies have applied regression modeling to construction time estimation of highway projects (Odabaşı 2009; Williams 2008).

The study by Czarnigowska and Sobotka (2014) is one of those few studies. Their BTC model with multiple predictor variables predicted construction durations of road projects in the preplanning stage. They applied the BTC model to the construction cost and time data of 100 projects. For comparison, they also developed a multiple linear regression model between the natural logarithm of cost and the square root of time. They found that the multiple linear regression model between the natural logarithm of cost and the square root of time performed better than the BTC model. However, only seven projects were used to validate the model, which leaves the need for additional validation of the model.

Williams (2008) developed parametric models to estimate construction durations of highway projects at early project stages. A full model and condensed model were developed for each of two project types—full-depth projects and improvement projects. The condensed model used fewer predictor variables for easier use. The full models did not use a cost estimate as a predictor variable. Rather, the models used key project design parameters, which were used as input to calculate the construction estimate, including traffic volume, curb and gutter (yes, no), median (yes, no), district, and geometric design standard (e.g., urban principal arterials). On the other hand, the primary predictor variable of the condensed model was the cost estimate, and the additional parameter was highway system (for full-depth project type) or area (for improvement project type). It is an interesting approach to develop full and condensed models for each project type, which provided an option for practitioners to use what best suits their use. One weakness of this study is the data size. Less than 100 projects were analyzed for each project type, and, as noted in the study, more data would be needed to strengthen the results. Another shortcoming of the study is that bridge projects were not included in the analysis due to data availability. As discussed in the study, bridge projects should be included in the modeling because they are an imperative part of highway construction.

Zhai et al. (2016) developed five multivariate linear regression models by project type to determine the contract time of highway construction projects with a cost estimate greater than $1 million. The predictor variables of their models were cost estimates and major item quantities such as excavation (cubic yards), stone base (cubic yards), PVC pipe (linear feet), storm sewer (linear feet), culvert pipe (linear feet), steel reinforcement (pounds), granular embankment (cubic yards), and Class AA concrete (cubic yards). Their model was proposed as an alternative to the CTD process for contract time estimation. Their regression models did not transform the time and cost data, which could be an advantage over the BTC model because it would reduce the data treatment steps. The major challenge of using their models for planning phase construction time estimate is that bid item quantities are not available during the planning and development phases. In addition, there is a possibility of collinearity between cost estimates and other predictors (i.e., major item quantities) because a cost estimate is generated based on bid item quantities.

Research Approach

Research Steps

The purpose of this study was to develop a planning phase time estimation method for a large portfolio of highway construction projects. To achieve the purpose, this study first developed the BTC and alternative (ALT) models from the actual construction cost and time data using simple and multiple regression. Second, the developed models were compared and validated to determine the best-fit model for the data. Third, it was examined if there were any significant differences in modeling results between contract and actual cost as a predictor variable. Last, the Mann-Whitney test was conducted to evaluate if the developed model would be effective when applied to a large portfolio of projects for high-level management purposes. To help with understanding, Table 1 lists time-cost models and the denotation used in this study.

Data Collection and Treatment

TxDOT's SiteManager information system retains TxDOT's project contract records such as work type, location, original contract bid price, change order amount, and duration. From the SiteManager system, this study retrieved the records of 623 construction projects, completed between 2003 and June 30, 2017, in TxDOT's Dallas District. Nonconstruction projects, including maintenance and traffic signal projects, were removed from the data set. All projects were low-bid design-bid-build projects. The construction cost of every project was converted to January 2017 dollars using TxDOT's Highway Cost Index (TxDOT 2017a) to eliminate the time effect on cost.

Table 2 presents the descriptive statistics of the construction cost and time data of the 623 projects. The *contract cost* refers to the original contract bid price (i.e., the winning bid), and it is the amount before any change order amounts or price adjustments. The *actual cost* is the revised contract price that includes change

Table 1. Time-cost model denotation

Model	Model type	Regression method	Cost used
$BTCS_{AC}$	BTC model: ln-ln model	Simple regression	Actual cost
$ALTS_{AC}$	ALT model: square root-ln model	Simple regression	Actual cost
$BTCM_{AC}$	BTC model: ln-ln model	Multiple regression	Actual cost
$ALTM_{AC}$	ALT model: square root-ln model	Multiple regression	Actual cost
$BTCS_{CC}$	BTC model: ln-ln model	Simple regression	Contract cost
$ALTS_{CC}$	ALT model: square root-ln model	Simple regression	Contract cost
$BTCM_{CC}$	BTC model: ln-ln model	Multiple regression	Contract cost
$ALTM_{CC}$	ALT model: square root-ln model	Multiple regression	Contract cost

Table 2. Descriptive statistics of data ($N = 623$ projects)

Statistics	Contract cost	Actual cost	Contract days	Actual days
Mean	\$9,969,061	\$10,562,985	232	277
Median	\$2,617,483	\$2,662,209	166	204
Standard deviation	\$24,870,188	\$26,585,323	201	240
Minimum	\$115,573	\$125,776	20	11
Maximum	\$244,854,657	\$264,787,755	1,436	1,352
Skewness	5.66	5.77	1.98	1.74

order amounts, liquidated damages, and other adjustments, if any. The *contract days* indicates the construction duration in working days specified in the contract. The *actual days* refers to the actual construction duration in working days recorded after construction completion. The project size of the data ranged from \$0.1 million to \$265 million, with actual durations between 11 and 1,352 days.

Data Splitting

When it is not possible or practical to collect new data for validation, "reservation of the portion of the available data" is an alternative method for validation (Snee 1977). This study used this approach because collecting data from new projects was not practical. The 623 projects collected from TxDOT's SiteManager system were split into two subsets—80% for modeling and 20% for validation. A total of 125 projects (Subset P) were randomly sampled for validation using the Random Sample function of StatTools, a statistical add-in program for Microsoft Excel. The remaining 498 projects were used as the subset for model development (Subset M). Subset P was reserved for validation and not used for model development.

Development of Time-Cost Regression Model

BTC and Alternative Models

If a linear relationship exists between construction time and cost of a highway project, a time and cost can be expressed by Eq. (2)

$$E(y) = \beta_0 + \beta_1 x \quad (2)$$

where $E(y)$ = expected value of a response variable y (time); x = predictor variable (cost); β_0 = constant (or intercept); and β_1 = regression coefficient associated with x.

As shown in Fig. 1, when the actual time and cost data of Subset M ($N = 498$) were plotted, the relationship was nonlinear, which means that Eq. (2) could not be directly applied. When two variables have a nonlinear relationship, data transformation is conducted to convert the relationship to a linear one. Bromilow used the natural logarithm data transformation for the BTC model

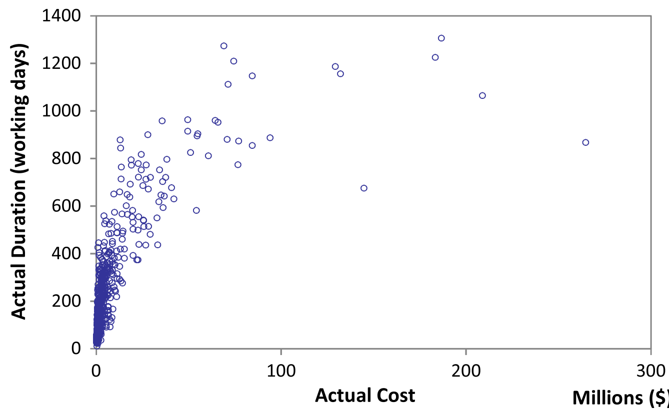

Fig. 1. Scatterplot of actual durations versus actual cost ($N = 498$).

(Bromilow 1969), and Czarnigowska and Sobotka (2014) tested both natural logarithm transformation and square root transformation in developing their models.

To find the data transformation method that would best fit for a linear relationship, this study tested four methods: natural logarithm, square root, square, and reciprocal. These are typical methods used for data transformation in linear regression modeling (Mendenhall et al. 2003). Different combinations of the transformed time and cost data were plotted and visually inspected for a potential linear relationship. Of 16 combinations tested, the square root of time values and the natural logarithm of cost values were finally selected as the alternative (ALT) model because they illustrated a clearer linear relationship between the two variables than the other transformed data. For comparison, the scatterplots of the BTC model and the proposed ALT model are presented in Figs. 2 and 3, respectively. The ALT model (Fig. 3) using the square root of time values and the natural logarithm of cost values shows a better linear fit than Bromilow's model (Fig. 2). The next sections discuss simple and multiple linear regression modeling of these two models and compare the models to identify the best-fit model for the data.

Simple and Multiple Linear Regression Modeling

Simple linear regression relationship has only one predictor variable and one response variable. The BTC model with simple regression (denoted by $BTCS_{AC}$) and the ALT model with simple regression (denoted by $ALTS_{AC}$) on Subset M are presented by Eqs. (3) and (4), respectively

$$BTCS_{AC}\text{: } \ln(\text{Actual Days}) = -2.6604 + 0.5293 \ln(\text{Actual Cost}) \quad (3)$$

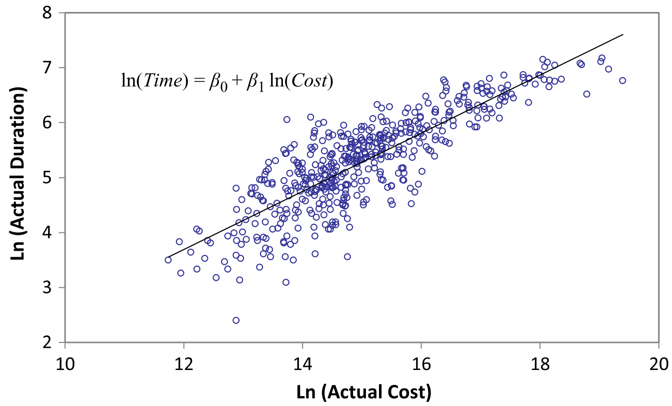

Fig. 2. Scatterplot of BTC model ($N = 498$).

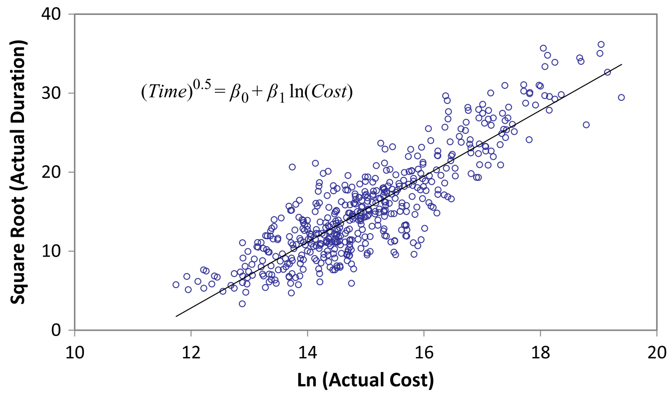

Fig. 3. Scatterplot of ALT model ($N = 498$).

$$\text{ALTS}_{\text{AC}}: \sqrt{(\text{Actual Days})} = -47.2043 + 4.1683 \ln(\text{Actual Cost}) \tag{4}$$

Multiple regression is an approach to improve the model predictability by adding more predictor variables to a regression equation, as expressed by Eq. (5)

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k + \varepsilon \tag{5}$$

where y = response variable; $x_1 \sim x_k$ are predictor variables; β_0 = constant (intercept); $\beta_1 \sim \beta_k$ are regression coefficients $x_1 \sim x_k$; and ε = error term. The primary predictor x_1 was Actual Cost, and variables $x_2 \sim x_k$ were additional predictors with impact on construction time in this study.

Determinants of Construction Time

The BTC model uses cost to predict time. As Sousa et al. (2014) pointed out, the model based on the time-cost relationship is a viable option for time estimation when project scope is not defined and design information is not available and therefore it is difficult to develop estimates with a higher accuracy. The cost estimate of a project is a fundamental deliverable from project initiation and is continuously refined as the project progresses and scope is defined. Using cost as a primary predictor of time estimate is not only because cost and time have a close relationship but also because a cost estimate is almost always available throughout the project life, even in the very early stages. Another key element is that a cost estimate is prepared by design parameters during the planning and development phases. This means that using both cost and design parameters as predictor variables in regression modeling could cause a multicollinearity issue (Williams 2008).

With cost being the primary predictor variable of time, a number of studies have also demonstrated that the regression model with additional predictor variables has higher accuracy than the simple time-cost regression model. Gross floor area, project type, contract type, and location are additional predictor variables commonly found in the time-cost model studies in building research (Hoffman et al. 2007; Jarkas 2016; Love et al. 2005; Skitmore and Ng 2003). In highway construction, the CTD guides developed by STAs listed factors to construction time. Weather conditions, location, traffic, project type, project complexity, permits, and resource availability are the most common factors affecting project construction time (MassDOT 2014; NCDOT 2012; VDOT 2007; WSDOT 2013). TxDOT's Contract Time Determination System introduces productivity factors, including location (rural, small city, and big city), traffic conditions (light, moderate, and high), complexity (low, medium, and high), and soil conditions (good, fair, and poor) (Hancher et al. 1992). Researchers have also incorporated key project factors into time estimation modeling. Project type is one of the most common predictors used in regression modeling besides cost, and researchers have typically developed separate models by project type, for example, new road construction, rehabilitation, and bridge construction (Irfan et al. 2011; Williams 2008; Zhai et al. 2016). Williams (2008) added area (location) and state highway system to the models with a cost estimate as a primary predictor. Williams's models were developed with the data collected from the Virginia Department of Transportation (VDOT), and the additional predictor variables were taken from VDOT's project classification. The area referred to VDOT's area classification—Rural, Small Urban, Medium Urban, and Large Urban per population. The state highway system included five highway types by capacity classified by VDOT (Williams 2008). Czarnigowska and Sobotka (2014) included the length of civil structures, number of winters, and number of civil structures in their models as additional predictors. Contract type and project delivery method have also been used as predictors of construction time (Irfan et al. 2011).

All the previously mentioned factors would have an impact on construction time and improve regression models. However, it is not realistic to model every available predictor. It is necessary to screen key factors that are significant enough to be included in the multiple regression models (Sousa et al. 2014). The primary rule in selecting predictor variables in regression modeling is "parsimony—explaining the most with the least." (Albright et al. 2006) This parsimony rule was applied to two steps. The first step was the identification and selection of predictor variables from previous studies. The second was the final selection of predictor variables through statistical validation, including stepwise regression.

For the first step, this study considered three selection rules: (1) variables should be simple and information about them should be easily obtained, (2) variables should be meaningful and related to project characteristics or scope, and (3) variables should not create excessive complexity for the regression equation. Based on these rules, this study examined the project parameters commonly used as predictors for construction time estimation in previous studies. The candidate parameters identified from the literature review were location, traffic conditions and volume, project type, weather conditions, soil conditions, work item quantity, and contract type. The parameter project type was further divided into two parameters—type of work and highway system—in accordance with the TxDOT's SiteManager system. The type of work here refers to a distinct and dominant characteristic of the

Table 3. Dummy variables for additional predictors

Classification	Category	Dummy variables	
		Variable 1	Variable 2
Type of work			
Rehabilitation of existing road, overlay, restoration	Rehabilitation	Rehab = 1	New = 0
Interchange (new or reconstructed), new location freeway, widening freeway, widening nonfreeway, converting nonfreeway to freeway, upgrade to standard freeway, upgrade to standard nonfreeway	New/upgrade/widening	Rehab = 0	New = 1
Bridge replacement, bridge widening or rehabilitation	Bridge	Rehab = 0	New = 0
County size (population)			
Population with over 1 million	Large county	LargeCo = 1	MediumCo = 0
Population between 0.5 and 1 million	Medium county	LargeCo = 0	MediumCo = 1
Population less than and 0.5 million	Small county	LargeCo = 0	MediumCo = 0
Highway system			
Interstate highways, business interstate highways, United States highways, Business United States highways, state highways, Business state highways, state highway loops, state highway spurs	State highways	SHW = 1	—
Farm to market roads, ranch-to-market roads, county roads	Others	SHW = 0	—

project scope, for example, rehabilitation, new road construction, or bridge construction. Highway system refers to the highway designations of TxDOT. Both parameters are critical to project construction time, and they are defined during the planning and development phases. Project location is also identified as a key predictor. The location of a project is known from the project need assessment phase. It also affects accessibility, traffic control plan, and complexity of project construction. The traffic-related parameters (e.g., traffic conditions, traffic volume) are closely related to the population of project geological location. For the planning phase time estimation, this study assumed the traffic-related parameters were part of the project location parameter rather than an independent parameter.

Weather conditions could have a huge impact on construction time depending on geological location. In some states, all construction work is shut down during winter seasons, and weather is a critical factor of construction time estimation. TxDOT's projects, however, do not have winter shutdowns, and both contact time and actual time in working days recorded in the TxDOT's SiteManager system are inclusive of weather days (i.e., rainy days). This means that the weather impact is equally applied to all highway construction projects in Texas, and it could be redundant to incorporate the weather impact as a predictor variable in regression modeling for the data collected for this study. Work item quantity is a popular component in parametric estimation. Nonetheless, it increases the model complexity, not to mention data availability in planning and development. Including work item quantity in a regression model for time estimation of a single project might not consume a high level of time and effort. However, for a large portfolio consisting of tens or hundreds of projects, it would be extremely time consuming and expensive to collect quantities of key work items for each project.

Contract type is not applicable because all the data of this study were from low-bid design-bid-build projects. Also, the schedule impact of construction methods, including construction techniques and crew or equipment configuration, is not considered in this study. It is difficult to assume specific construction methods other than standard ones during the planning and development phases due to limited project information. Particularly when estimating a large portfolio of projects during those phases, it is likely to be subjective and inaccurate to determine the construction methods of each project without the detailed design and traffic control plan. This is why STAs assume "generic resources, production rates, sequences of construction" (TxDOT 2014) when determining contract time.

Through the screening process, three factors—type of work, county size, and highway system—were selected as additional predictor variables for the modeling. The three factors, however, could not be directly modeled because they were qualitative, nonnumerical variables. Therefore, dummy variables were created as listed in Table 3 to incorporate categorical variables into regression modeling. The classification of type of work and highway system is based on the classification of the TxDOT's SiteManager system. For county size, population data were obtained from the 2010 census (US Census Bureau 2010).

The multiple regression equation with the six predictor variables is expressed by Eq. (6). The response variable y is ln (Actual Days) for the $\mathrm{BTCM_{AC}}$ model and $\sqrt{(\text{Actual Days})}$ for the $\mathrm{ALTM_{AC}}$ model. There would be $(n-1)$ dummy variables when there are n variable categories. For example, the Small County variable is not needed in the County Size part of the regression equation because a response variable y for a project located in county categorized as a Small County would be the equation with the other two dummy variables, LargeCo and MediumCo, being zero. The Bridge, Small County, and Others categories were used as the baseline dummy variables that did not need to be included in Eq. (6)

$$y = \beta_0 + \overbrace{\beta_1 \ln(\text{Actual Cost})}^{\text{Cost}} + \overbrace{\beta_2(\text{Rehab}) + \beta_3(New)}^{\text{Work Type}} + \overbrace{\beta_4(\text{LargeCo}) + \beta_5(\text{MediumCo})}^{\text{County Size}} + \overbrace{\beta_6(\text{SHW})}^{\text{Highway System}} \quad (6)$$

The second step of variable selection of this study was a statistical validation of the six predictor variables. When a regression model includes multiple predictor variables, it is always necessary to check if each predictor variable improves the equation significantly enough to be included in the equation (Albright et al. 2006; Mendenhall et al. 2003). The t-test statistics and variance inflation factors (VIFs) were examined to ensure that the predictor variables were statistically significant. The multiple regression results of the $\mathrm{BTCM_{AC}}$ and $\mathrm{ALTM_{AC}}$ showed that the p-values of the variables, except the LargeCo variable, were less than 0.05, indicating that those variables were statistically significant. The p-values of the LargeCo variable were 0.9331 for the $\mathrm{BTCM_{AC}}$ and 0.7527 for $\mathrm{ALTM_{AC}}$, indicating that the LargeCo variable was not statistically significant. The VIFs of all the variables were less than 2, suggesting that multicollinearity between the variables s negligible. Stepwise regression was also conducted to determine the statistical relevance of the predictor variables. A stepwise approach is one of the most well-known methods to select important predictor variables based on the significance of their t-values at the certain α level (Mendenhall et al. 2003). The stepwise regression results of the model revealed that adding the LargeCo variable did not significantly improve the model, which means that whether including or excluding the LargeCo variable does not make statistically significant differences in the response variable (construction time). Therefore, the LargeCo variable was finally removed from both the $\mathrm{BTCM}_{\mathrm{AC}}$ and $\mathrm{ALTM}_{\mathrm{AC}}$ models. After the LargeCo variable was removed, the County Size part of the equation has only one variable, MediumCo. Therefore, if a project is located in a county with a medium population, the MediumCo variable is 1 and the County Size part becomes β_4. If a project is located in a county with either a large or small population, the MediumCo variable is 0, and the County Size part becomes 0 as well.

The revised regression equation with the five predictor variables is given by Eq. (7), where the response variable y is ln (Actual Days) for the $\mathrm{BTCM}_{\mathrm{AC}}$ model and $\sqrt{(\text{Actual Days})}$ for the $\mathrm{ALTM}_{\mathrm{AC}}$ model. Regression on the revised equation showed that all coefficients were statistically significant, with small p-values less than 0.05. Their VIFs were less than 2, suggesting that multicollinearity between variables could be ignored. All four models satisfied the general assumptions of linear regression

$$y = \beta_0 + \overbrace{\beta_1 \ln(\text{Actual Cost})}^{\text{Cost}} + \overbrace{\beta_2(\text{Rehab}) + \beta_3(\text{New})}^{\text{Work Type}} + \overbrace{\beta_4(\text{MediumCo})}^{\text{County Size}} + \overbrace{\beta_5(\text{SHW})}^{\text{Highway System}} \tag{7}$$

The final regression equations of $\mathrm{BTCM}_{\mathrm{AC}}$ and $\mathrm{ALTM}_{\mathrm{AC}}$ are given by Eqs. (8) and (9), respectively. The regression models indicate that Rehab and New projects tend to have shorter durations than Bridge projects if the other variables remain the same. A project in a medium-sized county would take less time than projects in large or small counties. Projects on state highways finish earlier than non-state highway projects

$$\begin{aligned}\mathrm{BTCM}_{\mathrm{AC}}\colon \ln(\text{Acutal Days}) = & -2.1438 + 0.5257 \ln(\text{Actual Cost}) \\ & -0.4823(\text{Rehab}) - 0.1902(\text{New}) \\ & -0.1865(\text{MediumCo}) - 0.747(\text{SHW})\end{aligned} \tag{8}$$

$$\begin{aligned}\mathrm{ALTM}_{\mathrm{AC}}\colon \sqrt{(\text{Actual Days})} = & -43.3495 + 4.0944 \ln(\text{Actual Cost}) \\ & -3.5131(\text{Rehab}) - 1.0189(\text{New}) \\ & -0.9560(\text{MediumCo}) - 0.6277(\text{SHW})\end{aligned} \tag{9}$$

The outliers were determined by testing the standardized residuals. If the standardized residual of a data point was either larger than 3 or smaller than −3, the data was considered an outlier (Mendenhall et al. 2003). The numbers of outliers of $\mathrm{BTCS}_{\mathrm{AC}}$, $\mathrm{ALTS}_{\mathrm{AC}}$, $\mathrm{BTCM}_{\mathrm{AC}}$, and $\mathrm{ALTM}_{\mathrm{AC}}$ were 4, 1, 3, and 1, respectively. The data identified as outliers were examined, and they were not due to recording errors. However, because the outliers were different across the four models, the regression models would be developed on different data sets if the outliers were removed. Thus, to compare the four models on the same data set, it was decided not to exclude the outliers.

Model Comparisons

The next step was the comparison of the models to identify the best-fit model. Three statistics, global F-test, the coefficient of estimation, and standard error of the estimate, were used to compare the four models ($\mathrm{BTCS}_{\mathrm{AC}}$, $\mathrm{ALTS}_{\mathrm{AC}}$, $\mathrm{BTCM}_{\mathrm{AC}}$, and $\mathrm{ALTM}_{\mathrm{AC}}$). The null hypothesis of the global F-test is that all the coefficients of predictor variables are zero in a multiple regression equation. If the null hypothesis is rejected, at least one of the coefficients is not zero, which indicates that the model is usable (Mendenhall et al. 2003). All four models generated large F-statistics and extremely small p-values (less than 0.0001), indicating that the models were statistically useful.

The coefficients of determination (R^2) and standard errors of the estimates (s_e) of the four models were not directly comparable because they used different units for the response variable [i.e., $\ln(\text{Actual Days})$ and $\sqrt{(\text{Actual Days})}$]. It was necessary to back-transform the data to original values to make the models have the same units so that the R^2 and s_e of each model could be calculated against the original values, Actual Days. The adjusted R^2 and s_e were calculated on the original values using Eqs. (10) and (11)

$$\text{Adjusted } R^2 = 1 - (n-1)\frac{s_e^2}{\sum (y_{\mathrm{i}} - \bar{y})^2} \tag{10}$$

$$s_e = \sqrt{\frac{\sum e_i^2}{n-k-1}} \tag{11}$$

where e_i = residual (the difference between individual values and the fitted values); y_i = individual observed value; $\bar{y}$ = mean value of y; n = sample size; and k = number of predictor variables in the regression equation. Table 4 shows the adjusted R^2 and s_e of the four models on the transformed and back-transformed data for comparison. Three key findings were established from the comparison. First, the adjusted R^2 values on the back-transformed data were close to those of the transformed data. It can be inferred that the statistical significance of the models did not change when the data were back-transformed to the original time and cost values. Second, the ALT models have a higher adjusted R^2 and smaller s_e than the BTC models for both simple and multiple regression modeling. The $\mathrm{ALTM}_{\mathrm{AC}}$ model had the highest adjusted R^2, indicating that 83% of the Actual Days is explained by the Actual Cost for the given data set. This is approximately 13% higher than the $\mathrm{BTCM}_{\mathrm{AC}}$ model. The s_e of the $\mathrm{ALTM}_{\mathrm{AC}}$ model is 100 days, which is less than the s_e of the $\mathrm{BTCM}_{\mathrm{AC}}$ by 33 days. Third, multiple regression performed better than simple regression for the ALT models, but not for the BTC models for the given data set. The $\mathrm{ALTS}_{\mathrm{AC}}$ model has a slightly lower adjusted R^2 and a greater s_e than the $\mathrm{ALTM}_{\mathrm{AC}}$ model, which implies the $\mathrm{ALTM}_{\mathrm{AC}}$ model is a better fit than the $\mathrm{ALTS}_{\mathrm{AC}}$ model. Interestingly, the $\mathrm{BTCM}_{\mathrm{AC}}$ model showed a higher adjusted R^2 and smaller s_e than the $\mathrm{BTCS}_{\mathrm{AC}}$ on the transformed data (i.e., natural logarithms of time and cost

Table 4. Model comparisons (Subset M)

Regression results	$\mathrm{BTCS}_{\mathrm{AC}}$	$\mathrm{ALTS}_{\mathrm{AC}}$	$\mathrm{BTCM}_{\mathrm{AC}}$	$\mathrm{ALTM}_{\mathrm{AC}}$
Data transformation				
Adjusted R^2	0.7081	0.7600	0.7747	0.8148
s_e	0.4671	3.2198	0.4103	2.8284
Inverse data transformation				
Adjusted R^2	0.7171	0.7917	0.7032	0.8311
s_e	129.8637	111.4391	133.0069	100.3274

Table 5. Validation results by model

Model	Sample	Adjusted R^2	s_e
$BTCS_{AC}$	Subset M	0.7171	129.8637
	Subset P	0.4673	161.2458
	Change	(−)0.2498	(+)31.3821
$ALTS_{AC}$	Subset M	0.7917	111.4391
	Subset P	0.7669	106.6632
	Change	(−)0.0248	(−)4.7759
$BTCM_{AC}$	Subset M	0.7032	133.0069
	Subset P	0.4854	158.4942
	Change	(−)0.2178	(+)25.4873
$ALTM_{AC}$	Subset M	0.8311	100.3274
	Subset P	0.8128	95.5851
	Change	(−)0.0183	(−)4.7423

Note: Subset M = modeling sample; Subset P = validation (prediction) sample; (−) = decrease; and (+) = increase.

data), but it had a higher adjusted R^2 and a larger s_e than the $BTCS_{AC}$ model for the back-transformed data, although the differences were very small. The R^2 and adjusted R^2 of the multiple regression models were very close because the models had a relatively small number of predictor variables. In conclusion, the $ALTM_{AC}$ model turned out to be the best fit for the four models tested for the given data set, Subset M.

Model Validation

The purpose of validation in regression modeling is to assess the prediction quality of a model in predicting future values (Albright et al. 2006). This study validated the prediction power and prediction accuracy of the developed models. The prediction power was validated by comparing the R^2 and s_e of the modeling and validation data sets. The prediction accuracy was assessed by the mean absolute error (MAE) and the mean absolute percent error (MAPE).

If the R^2 and s_e of the model on the validation data set are significantly weakened from the modeling data set, the model is not practically usable (i.e., low prediction power). The adjusted R^2 and s_e of the four models for Subset M ($N = 498$) and Subset P ($N = 125$) are presented in Table 5. A noticeable finding was that the adjusted R^2 and s_e of the BTC models significantly deteriorated in Subset P, whereas the changes of the ALT models were negligible. The coefficient of determination R^2 is the portion of the variance of the response variable explained by the predictor variable(s) to the total variance (Washington et al. 2010). The proportion of *explained* variance in the duration of Subset P was notably reduced compared with Subset M. The increased value of the s_e of the BTC models in Subset P also implies that the prediction power is lessened from modeling. On the other hand, the changes of the adjusted R^2 and s_e of the ALT models were slight. These results indicate that the ALT models have more stable and reliable prediction power than the BTC models.

The mean square error, the MAE, and the MAPE are generally used to evaluate the prediction accuracy of transportation modeling (Washington et al. 2010). In highway construction research, MAPE has been used for time-cost models to evaluate prediction accuracy (Czarnigowska and Sobotka 2014; Irfan et al. 2011; Zhai et al. 2016). This study calculated both MAE and MAPE to assess the prediction accuracy in working days as well as in the percentage of the observed values. The equations for calculating the MAE and MAPE are given by Eqs. (12) and (13), respectively

Table 6. Residual analysis (Subset P)

Residual statistics	$BTCS_{AC}$	$ALTS_{AC}$	$BTCM_{AC}$	$ALTM_{AC}$
MAE (days)	92.0266	80.5933	78.2977	67.1807
MAPE (%)	46.08	48.86	35.89	37.61

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \tag{12}$$

$$\mathrm{MAPE} = \frac{100}{n}\sum_{i=1}^{n}\left|\frac{y_i - \hat{y}_i}{y_i}\right| \tag{13}$$

where y_i = observed duration; $\hat{y}_i$ = predicted duration; and n = size of Subset P. Because MAE and MAPE are different expressions of the distance between observed values and predicted values, the larger they are, the lower the accuracy of the model. The MAE and MAPE of the four models on Subset P are presented in Table 6. The MAE of $ALTM_{AC}$ is 67 days, 25 days less than $BTCS_{AC}$. The MAPE of $ALTM_{AC}$ was improved from $BTCS_{AC}$, reducing the percent error by 9%. The four models have relatively large MAPEs, meaning that the prediction accuracy may not be acceptable for circumstances requiring higher accuracy, such as contract time determination. However, the planning-phase estimation may not require a high level of accuracy (Sousa et al. 2014), mainly because not much information is available during those phases. The more important requirements for the planning phase time estimation are usability and simplicity. As projects under planning are further defined and new projects enter the planning phase, the time estimates, along with the cost estimates, need to be continuously updated. Therefore, the accuracy level of the developed models would be acceptable for construction time estimation during planning phases.

In summary, the validation results indicated that the ALT models were a better fit than the BTC models, and multiple regression models outperformed simple regression models. The best-performing model was $ALTM_{AC}$, which was consistent with what was found in modeling.

Contract Cost Models

The relationship between actual cost and actual time would represent the true relationship between the time and cost of a construction project. The challenge is that the actual cost is known only at project completion. For that reason, as Irfan et al. (2011) pointed out, it is arguable if contract cost should be used instead of the actual cost to develop a time prediction model. This study investigated this issue through two approaches. The first was developing a separate time-cost model based on contract cost and actual time. The second was applying contract cost data to the $ALTM_{AC}$ model that was identified as a best-fit model in the previous analyses. The development of the time-cost models based on contract cost followed the same steps used in the actual cost model development. Simple and multiple regression analyses were also conducted on Subset M with contract cost as the primary predictor variable. The other predictor variables included in modeling were the same as in the actual cost models. The regression equations of the contract cost models are given in Eqs. (14)–(17), denoted by $BTCS_{CC}$, $ALTS_{CC}$, $BTCM_{CC}$, and $ALTM_{CC}$, respectively. The contract cost models produced regression results very similar to those of the actual cost models. The alternative multiple regression model appeared the best fit of the four contract cost models tested, the same as observed in the actual cost models

J. Constr. Eng. Manage., 2019, 145(4): 04019018

$$\mathrm{BTCS_{CC}}: \ln(\text{Actual Days}) = -2.5158 + 0.5216 \ln(\text{Contract Cost}) \quad (14)$$

$$\mathrm{ALTS_{CC}}: \sqrt{(\text{Actual Days})} = -46.2913 + 4.1231 \ln(\text{Contract Cost}) \quad (15)$$

$$\begin{aligned}\mathrm{BTCM_{CC}}: \ln(\text{Actual Days}) = &-1.9966 + 0.5182 \ln(\text{Contract Cost}) \\ &- 0.4859(\text{Rehab}) - 0.1685(\text{New}) \\ &- 0.2077(\text{MediumCo}) - 0.1788(\text{SHW})\end{aligned} \quad (16)$$

$$\begin{aligned}\mathrm{ALTM_{CC}}: \sqrt{(\text{Actual Days})} = &-42.4035 + 4.0498 \ln(\text{Contract Cost}) \\ &- 3.5421(\text{Rehab}) - 0.8599(\text{New}) \\ &- 1.1223(\text{MediumCo}) - 0.6692(\text{SHW})\end{aligned} \quad (17)$$

The second approach to use contract cost in time-cost modeling was replacing actual cost with contract cost in the $\mathrm{ALTM_{AC}}$ model, as expressed in Eq. (18)

$\mathrm{ALTM_{AC}}$ with contract cost:

$$\begin{aligned}\sqrt{(\text{Actual Days})} = &-43.3495 + 4.0944 \ln(\text{Contract Cost}) \\ &- 3.5131(\text{Rehab}) - 1.0189(\text{New}) \\ &- 0.9560(\text{MediumCo}) - 0.6277(\text{SHW})\end{aligned} \quad (18)$$

The adjusted R^2 and s_e of the three ALT models—$\mathrm{ALTM_{AC}}$, $\mathrm{ALTM_{CC}}$, and $\mathrm{ALTM_{AC}}$ with contract cost—were compared as shown in Table 7. The three models are very similar in terms of the adjusted R^2, s_e, and residuals, which implies that the time estimates from these three models on the same project would not significantly differ. One explanation could be that the contract and actual cost of the data collected were very close, with a correlation of 99.8%, which would produce similar regression models for predicting actual time.

Although all three models could statistically be used, their utility should be interpreted within a practical context. Actual time and actual cost data include any changes made to the original contract of a project. Whether changes come from differing site conditions or scope changes, the actual time-actual cost relationship represents the actual duration for a given project size (cost) more accurately.

Table 7. Comparisons of three ALT models

Inverse transformation	$\mathrm{ALTM_{AC}}$	$\mathrm{ALTM_{CC}}$	$\mathrm{ALTM_{AC}}$ (contract cost)
Regression results (Subset M)			
Adjusted R^2	0.8311	0.8197	0.8169
s_e	100.3274	103.6645	104.4612
Regression results (Subset P)			
Adjusted R^2	0.8128	0.8126	0.8125
s_e	95.5851	95.6514	95.6754
Prediction quality (Subset P)			
MAE	67.1801	66.3347	65.2840
MAPE (%)	37.61	37.78	36.05

On the other hand, the actual time-contract cost relationship would omit the cost portion associated with duration changes whether it is positive or negative. The actual time-actual cost model would more closely represent the true relationship between time and cost from the practical point. Therefore, $\mathrm{ALTM_{AC}}$ is recommended for estimating construction time. The $\mathrm{ALTM_{AC}}$ model in a general form can be expressed by Eq. (19)

$$\begin{aligned}\sqrt{(\text{Construction Duration})} = &-43.3495 + 4.0944 \ln(\text{Cost Estimate}) \\ &- 3.5131(\text{Rehab}) - 1.0189(\text{New}) \\ &- 0.9560(\text{MediumCo}) - 0.6277(\text{SHW})\end{aligned} \quad (19)$$

where construction time: in working days; cost estimate: in dollars (January 2017 dollars); Rehab = rehabilitation of existing road, overlay, restoration; New = interchange (new or reconstructed), new location freeway, widening freeway, widening nonfreeway, converting nonfreeway to freeway, upgrade to standard freeway, upgrade to standard nonfreeway; MediumCo = population between 0.5 million and 1 million; and SHW = state highways.

When using the model for time estimation, any construction cost estimate data need to be converted to January 2017 dollars.

Model Applicability to a Portfolio of Projects

The purpose of this study was to develop a model as an estimation tool for a large portfolio of projects during the planning and development phases. To achieve the purpose, the developed model was tested using the Mann-Whitney test. The Mann-Whitney test is a statistical two-sample test for samples with nonnormal distributions. It tests whether one sample is significantly different from the other. The test pools two groups, ranks the values, and then compares the rankings of the values of each group separately to determine whether one group is smaller than the other (Washington et al. 2010). In the context of this study, it was assumed that if the distributions of the actual and predicted durations of the models for a group of projects were the same, it could be concluded that the model provides reliable time estimates for a group of projects. The null (H_0) and alternative hypotheses (H_a) for the Mann-Whitney test were established as follows.

H_0: The distributions of actual durations and the predicted durations are the same (i.e., neither distribution is smaller).

H_a: The distributions of actual durations and the predicted durations are not the same (i.e., either distribution is smaller).

The Mann-Whitney test was conducted on 25 sample groups consisting of 30 projects from Subset P. Using the random sample utility function of StatTools, 30 projects were randomly selected from Subset P, and this random sampling was repeated 25 times, resulting in 25 sample groups of 30 projects. This number was selected to include various types and sizes of projects and to be large enough to form a distribution. The test statistics of all 25 sample groups concluded that the null hypothesis could not be rejected at the significance level of 0.01, which indicated that neither the distribution of the actual durations nor that of predicted durations was smaller than the other. The Mann-Whitney test on the entire Subset P ($N = 125$) was conducted as well and resulted in the same conclusions. The p-value was 0.7729, and the null hypothesis could not be rejected.

The model might predict individual project durations with less accuracy than desired, particularly for contract time determination. On the other hand, the Mann-Whitney test demonstrated that the model would perform effectively for a large group of projects from the practical point of view. TxDOT and its districts review hundreds of projects in a 4-year time span to develop STIP. For high-level management purposes, a bigger picture is more useful than a focus on individual projects. In that regard, the developed model ($ALTM_{AC}$) has a great potential to bring value to practitioners in developing and updating transportation plans and programs by facilitating construction time estimation of a large portfolio of projects in their earlier planning phases.

Conclusions

There is an increasing need for construction time estimates for STAs for project-planning purposes. Particularly, there has been a lack of efficient and reliable methods that could help STA practitioners in performing various planning and management activities required during those phases such as forecasting inspection staffing requirements. This study proposed a time-cost model based on actual duration and cost data to provide a construction time estimation method for early project phases when only limited project information is known.

This study employed time-cost regression modeling to develop a planning phase time estimation method. The data were collected from 623 completed highway construction projects in TxDOT's Dallas District; 498 projects were used to develop the model, and 125 projects were used for validation. The time and cost data were transformed using the square root and natural logarithm, respectively, and a linear regression was conducted on the transformed data for modeling.

To improve the quality of the developed model, work type, county size, and highway system were selected as predictor variables in addition to the construction cost. The developed model was compared with the well-known BTC model that is a regression model between the natural logarithms of time and cost. The comparisons of the two models showed that the developed model was a better fit than the BTC model for the data collected. The developed model was validated on the remaining 125 projects to evaluate its prediction power and accuracy. The validation also concluded that the developed model performed better than the BTC model. The adjusted R^2 of the developed model in the prediction data set remained almost identical, whereas the R^2 of the BTC model considerably decreased. This study additionally found that, for the given data set, using actual cost or contract cost did not show differences in estimating construction time in terms of statistical significance and prediction accuracy.

This study sought the applicability of the developed model to a large portfolio of projects. The prediction accuracy of the best-fit model, $ALTM_{AC}$, showed a relatively large variance that may not be acceptable for contract time determination of individual projects where higher estimate accuracy is required. However, the Mann-Whitney test on the model proved that the distributions of actual construction time and predicted construction time of a group of projects were identical. This indicates that the model could be an effective tool to estimate construction time of a large group of projects during the planning and development phases where a higher estimating accuracy is hardly achievable due to limited project information and where the estimating purpose is for planning and management rather than determining contract time.

One of the strengths of the developed model is its simplicity and ease of use in practice. The required input data are the construction cost estimate and basic project parameters known prior to engineering design development. This strength allows time estimation of a large portfolio of hundreds of projects with considerably less time and effort. The model does not require complicated mathematical calculations and can be updated easily as needed by adding more project data with minimal maintenance effort. A major limitation of this model is that data were collected from one geographical region, the Dallas area of Texas. Although this helped reduce data variability for regression modeling, the model might not be directly applicable to projects in different areas or states. However, the modeling methodology is still applicable, and the agencies of other areas or states can develop their own model with their data using the methodology. Another limitation might be cost conversion. The model was developed using the cost data converted to January 2017 dollars, which means that any future project cost data also needs to be converted to January 2017 dollars to apply the developed model. Nonetheless, this should not be a major limitation because the FHWA and TxDOT maintain cost indices for construction projects. Moreover, all the cost data used in the model were January 2017 values, which means that the conversion of the model to another time base would not affect the prediction power and accuracy because the same cost index would be applied to all the cost data.

The main benefit of this study is that it provides STA practitioners with a quick and simple method to estimate the construction time of a diverse portfolio of highway projects during the planning and development phases. For example, STAs can use the model to forecast inspection manpower needs for the construction phase of a large portfolio of road, bridge, and other surface transportation projects. From the academic perspective, this study suggested a possible use of a time-cost regression model in long-term planning of projects and programs. Unlike cost estimation, planning-phase estimation of construction time has not been an area of focus for researchers. Moreover, there has been a lack of research on portfolio and program management for highway construction projects. This study proposed a time-cost regression model that has potential use for such purposes.

A future research opportunity will be the sensitivity analysis between the time and cost in the model. The cost estimate is the primary predictor variable of the model. The change of the cost estimate drives the change of the time estimate. During the planning and development phases, however, the cost is still an estimate and subject to change. To understand the time estimation risks associated with the accuracy of the cost estimate, it needs to assess how sensitive the time estimate is as the cost estimate changes during the planning and development phases. Another avenue for future research can be identifying and exploring areas of implementation of the model in the planning phases. Pilot tests of the model with potential planning activities would increase the strength of the model as well as revealing additional research needs in those areas. In addition, developing planning-phase estimation of construction time methods for alternative project delivery methods or contracting types other than low-bid design-bid-build could be promising research.

Data Availability Statement

Data generated or analyzed during the study are available from the corresponding author by request. Information about the *Journal*'s data-sharing policy can be found here: http://ascelibrary.org/doi/10.1061/(ASCE)CO.1943-7862.0001263.

Acknowledgments

The authors would like to thank the Dallas District of the Texas Department of Transportation for providing support for data collection and analysis.

References

Albright, S. C., W. Winston, and C. Zappe. 2006. *Data analysis and decision making*. 3rd ed. Boston: Cengage Learning.

Bhokha, S., and S. O. Ogunlana. 1999. "Application of artificial neural network to forecast construction duration of buildings at the predesign stage." *Eng. Constr. Archit. Manage.* 6 (2): 133–144. https://doi.org/10.1108/eb021106.

Bromilow, F. J. 1969. "Contract time performance expectations and the reality." *Build. Forum* 1 (3): 70–80.

Czarnigowska, A., and A. Sobotka. 2014. "Estimating construction duration for public roads during the preplanning phase." *J. Eng. Prof. Prod. Manage.* 4 (1): 26–35. https://doi.org/10.32738/JEPPM.201401.0004.

De la Garza, J. M., and K. G. Rouhana. 1995. "Neural networks versus parameter-based applications in cost." *Cost Eng.* 37 (2): 14.

FHWA (Federal Highway Administration). 2002. "FHWA guide for construction contract time determination procedures." Accessed June 18, 2017. https://www.fhwa.dot.gov/construction/contracts/t508015.cfm.

FHWA (Federal Highway Administration). 2017. *The transportation planning process briefing book*. FHWA-HEP-18-005. Washington, DC: FHWA.

Hancher, D. E., W. F. McFarland, and R. T. Alabay. 1992. *Construction contract time determination*. No. FHWA/TX-92/1262-1F. Washington, DC: FHWA.

Herbsman, Z. J., and R. Ellis. 1995. *Determination of contract time for highway construction projects*. NCHRP Synthesis, 215. Washington, DC: National Cooperative Highway Research Program.

Hoffman, G. J., A. E. Thal Jr., T. S. Webb, and J. D. Weir. 2007. "Estimating performance time for construction projects." *J. Manage. Eng.* 23 (4): 193–199. https://doi.org/10.1061/(ASCE)0742-597X(2007)23:4(193).

Irfan, M., M. B. Khurshid, P. Anastaspoulous, S. Labi, and F. Moavenzadeh. 2011. "Planning-stage estimation of highway project duration on the basis of anticipated project cost, project type, and contract type." *Int. J. Project Manage.* 29 (1): 78–92. https://doi.org/10.1016/j.ijproman.2010.01.001.

Jarkas, A. M. 2016. "Predicting contract duration for building construction: Is Bromilow's time-cost model a panacea?" *J. Manage. Eng.* 32 (1): 05015004. https://doi.org/10.1061/(ASCE)ME.1943-5479.0000394.

Jin, R., S. Han, C. Hyun, and Y. Cha. 2016. "Application of case-based reasoning for estimating preliminary duration of building projects." *J. Constr. Eng. Manage.* 142 (2): 04015082. https://doi.org/10.1061/(ASCE)CO.1943-7862.0001072.

Kim, G. H., S. H. An, and K. I. Kang. 2004. "Comparison of construction cost estimating models based on regression analysis, neural networks, and case-based reasoning." *Build. Environ.* 39 (10): 1235–1242. https://doi.org/10.1016/j.buildenv.2004.02.013.

Love, P. E., R. Y. Tse, and D. J. Edwards. 2005. "Time-cost relationships in Australian building construction projects." *J. Constr. Eng. Manage.* 131 (2): 187–194. https://doi.org/10.1061/(ASCE)0733-9364(2005)131:2(187).

Lu, M., S. M. AbouRizk, and U. H. Hermann. 2001. "Sensitivity analysis of neural networks in spool fabrication productivity studies." *J. Comput. Civ. Eng.* 15 (4): 299–308. https://doi.org/10.1061/(ASCE)0887-3801(2001)15:4(299).

Mallela, J., and S. Sadavisam. 2011. *Work zone road user costs: Concepts and applications*. FHWA-HOP-12-005. Washington, DC: Federal Highway Administration.

MassDOT (Massachusetts Department of Transportation). 2014. "Construction contract time determination (CTD) guidelines for designers/planners." Accessed June 18, 2017. https://www.massdot.state.ma.us/Portals/8/docs/projectManagement/controlsGuidelines/CTDguidelines.pdf.

McCoy, K., A. Ingles, and W. Lyons. 2016. *STIP State of the practice review: Development and use of statewide transportation improvement programs*. FHWA-HEP-16-061. Washington, DC: Federal Highway Administration.

Mendenhall, W., T. Sincich, and N. S. Boudreau. 2003. Vol. 6 of *A second course in statistics: Regression analysis*. Upper Saddle River, NJ: Prentice Hall.

Mensah, I., T. Adjei-Kumi, and G. Nani. 2016. "Duration determination for rural roads using the principal component analysis and artificial neural network." *Eng. Constr. Archit. Manage.* 23 (5): 638–656. https://doi.org/10.1108/ECAM-09-2015-0148.

Nani, G., I. Mensah, and T. Adjei-Kumi. 2017. "Duration estimation model for bridge construction projects in Ghana." *J. Eng. Des. Technol.* 15 (6): 754–777. https://doi.org/10.1108/JEDT-04-2017-0029.

NCDOT (North Carolina Department of Transportation). 2012. "Guidance for determining contract times." Accessed June 18, 2017. https://connect.ncdot.gov/municipalities/Pages/Bid-Proposals-for-LGA.aspx.

Odabaşı, E. 2009. "Models for estimating construction duration: An application for selected buildings on the Metu campus." M.S. thesis, Dept. of Architecture, Middle East Technical Univ.

Pewdum, W., T. Rujirayanyong, and V. Sooksatra. 2009. "Forecasting final budget and duration of highway construction projects." *Eng. Constr. Archit. Manage.* 16 (6): 544–557. https://doi.org/10.1108/09699980911002566.

Skitmore, R. M., and S. T. Ng. 2003. "Forecast models for actual construction time and cost." *Build. Environ.* 38 (8): 1075–1083. https://doi.org/10.1016/S0360-1323(03)00067-2.

Snee, R. D. 1977. "Validation of regression models: Methods and examples." *Technometrics* 19 (4): 415–428. https://doi.org/10.1080/00401706.1977.10489581.

Sousa, V., N. M. Almeida, and L. A. Dias. 2014. "Role of statistics and engineering judgment in developing optimized time-cost relationship models." *J. Constr. Eng. Manage.* 140 (8): 04014034. https://doi.org/10.1061/(ASCE)CO.1943-7862.0000874.

TxDOT (Texas Department of Transportation). 2014. "Standard specifications for construction and maintenance of highways, streets and bridges." Accessed June 18, 2017. ftp://ftp.dot.state.tx.us/pub/txdot-info/des/spec-book-1114.pdf.

TxDOT (Texas Department of Transportation). 2016. "2017–2020 Statewide transportation improvement program." Accessed June 18, 2017. http://ftp.dot.state.tx.us/pub/txdot-info/tpp/stip/2017-2020/introduction.pdf.

TxDOT (Texas Department of Transportation). 2017a. "Highway cost index (1997 base) index report for May 2017." Accessed June 18, 2017. http://ftp.dot.state.tx.us/pub/txdot-info/cst/hci-binder.pdf.

TxDOT (Texas Department of Transportation). 2017b. "Project development process manual." Accessed June 18, 2017. http://onlinemanuals.txdot.gov/txdotmanuals/pdp/manual_notice.htm.

TxDOT (Texas Department of Transportation). 2018. "Project tracker." Accessed April 30, 2018. http://apps.dot.state.tx.us/apps-cq/project_tracker/#.

US Census Bureau. 2010. "Centers of population by county: 2010." Data Files. Accessed June 18, 2017. https://www2.census.gov/geo/docs/reference/cenpop2010/county/CenPop2010_Mean_CO48.txt.

VDOT (Virginia Department of Transportation). 2007. "Contract time determination guidelines." Accessed June 18, 2017. http://www.vdot.virginia.gov/business/resources/const/CTDR_Guidelines_VDOT.pdf.

Washington, S. P., M. G. Karlaftis, and F. Mannering. 2010. *Statistical and econometric methods for transportation data analysis*. Boca Raton, FL: CRC Press.

Williams, R. C. 2008. "The development of mathematical models for preliminary prediction of highway construction duration." Ph.D. dissertation, Dept. of Civil Engineering, Virginia Tech.

WSDOT (Washington State Department of Transportation). 2013. "Plans preparation manual." Accessed June 18, 2017. http://www.wsdot.wa.gov/publications/manuals/fulltext/M22-31/M22-31.05Revision.pdf.

Zhai, D., Y. Shan, R. E. Sturgill, T. R. Taylor, and P. M. Goodrum. 2016. "Using parametric modeling to estimate highway construction contract time." *Transp. Res. Rec.* 2573 (1): 1–9. https://doi.org/10.3141/2573-01.