Project 2148 | May 2023

SJSU SAN JOSÉ STATE UNIVERSITY

MINETA TRANSPORTATION INSTITUTE

Preconstruction Support Cost Hours Estimating on Caltrans Pavement Rehabilitation Projects

Nigel Blampied, PhD   Tariq Shehab, PhD   Elhami Nasr, PhD   Laxmi Sindhu Samudrala

CSUTC California State University Transportation Consortium

SJSU SAN JOSÉ STATE UNIVERSITY

Mineta Transportation Institute

Founded in 1991, the Mineta Transportation Institute (MTI), an organized research and training unit in partnership with the Lucas College and Graduate School of Business at San José State University (SJSU), increases mobility for all by improving the safety, efficiency, accessibility, and convenience of our nation's transportation system. Through research, education, workforce development, and technology transfer, we help create a connected world. MTI leads the Mineta Consortium for Transportation Mobility (MCTM) funded by the U.S. Department of Transportation and the California State University Transportation Consortium (CSUTC) funded by the State of California through Senate Bill 1. MTI focuses on three primary responsibilities:

Research

MTI conducts multi-disciplinary research focused on surface transportation that contributes to effective decision making. Research areas include: active transportation; planning and policy; security and counterterrorism; sustainable transportation and land use; transit and passenger rail; transportation engineering; transportation finance; transportation technology; and workforce and labor. MTI research publications undergo expert peer review to ensure the quality of the research.

Education and Workforce

To ensure the efficient movement of people and products, we must prepare a new cohort of transportation professionals who are ready to lead a more diverse, inclusive, and equitable transportation industry. To help achieve this, MTI sponsors a suite of workforce development and education opportunities. The Institute supports educational programs offered by the Lucas Graduate School of Business: a Master of Science in Transportation Management, plus graduate certificates that include High-Speed and Intercity Rail Management and Transportation Security Management. These flexible programs offer live online classes so that working transportation professionals can pursue an advanced degree regardless of their location.

Information and Technology Transfer

MTI utilizes a diverse array of dissemination methods and media to ensure research results reach those responsible for managing change. These methods include publication, seminars, workshops, websites, social media, webinars, and other technology transfer mechanisms. Additionally, MTI promotes the availability of completed research to professional organizations and works to integrate the research findings into the graduate education program. MTI's extensive collection of transportation-related publications is integrated into San José State University's world-class Martin Luther King, Jr. Library.

Report 23-07

Preconstruction Support Cost Hours Estimating on Caltrans Pavement Rehabilitation Projects

Nigel Blampied, PhD

Tariq Shehab, PhD

Elhami Nasr, PhD

Laxmi Sindhu Samudrala

May 2023

A publication of the
Mineta Transportation Institute
Created by Congress in 1991

College of Business
San José State University
San José, CA 95192-0219

TECHNICAL REPORT DOCUMENTATION PAGE

1. Report No. 23-07	**2. Government Accession No.**	**3. Recipient's Catalog No.**
4. Title and Subtitle Preconstruction Support Cost Hours Estimating on Caltrans Pavement Rehabilitation Projects		**5. Report Date** May 2023 **6. Performing Organization Code**
7. Authors Nigel Blampied, PhD https://orcid.org/0000-0001-8597-6691 Tariq Shehab, PhD Elhami Nasr, PhD Laxmi Sindhu Samudrala		**8. Performing Organization Report** CA-MTI-2148
9. Performing Organization Name and Address Mineta Transportation Institute College of Business San José State University San José, CA 95192-0219		**10. Work Unit No.** **11. Contract or Grant No.** ZSB12017-SJAUX
12. Sponsoring Agency Name and Address State of California SB1 2017/2018 Trustees of the California State University Sponsored Programs Administration 401 Golden Shore, 5th Floor, Long Beach, CA 90802		**13. Type of Report and Period Covered** **14. Sponsoring Agency Code**
15. Supplemental Notes		

16. Abstract

Because the construction phase accounts for the majority of project costs for pavement rehabilitation projects, most research on infrastructure project cost estimating focuses on that phase, rather than on the preconstruction phases. Nevertheless, costs incurred prior to construction, referred to in this report as "preconstruction costs" are significant and worthy of consideration (See Section 2.1 of the report for a more detailed and precise definition of preconstruction). In the 2020–2021 fiscal year, for instance, the California Department of Transportation (Caltrans) spent more than $169 million on preconstruction work for pavement rehabilitation projects. This report presents the results of a study of preconstruction cost estimating for pavement rehabilitation projects undertaken by Caltrans. It uses data on the 139 pavement rehabilitation projects for which Caltrans opened bids in the five-year period from April 26, 2016 to May 11, 2021. A data set was developed that combined the preconstruction hours for each project with the primary bid items for the pavement rehabilitation projects. Two models were developed to estimate preconstruction hours from the bid items, one using an Artificial Neural Network (ANN) and the other a parametric exponential model developed using multiple regression. The models had coefficients of determination of 0.85 and 0.80, respectively. Tools were then developed to assist professional users in validating their preconstruction cost estimates using each of the models. CTC staff or Caltrans can use these tools to evaluate the reasonableness of the preconstruction estimate on an individual project, or on the sum of an entire biennial SHOPP pavement rehabilitation portfolio, in order to assure the most efficient use of infrastructure funding to best serve the community's transportation needs.

17. Key Words Cost estimating, Neural networks, Regression analysis, Planning and design, Pavements	**18. Distribution Statement** No restrictions. This document is available to the public through The National Technical Information Service, Springfield, VA 22161.		
19. Security Classif. (of this report) Unclassified	**20. Security Classif.** (of this page) Unclassified	**21. No. of Pages** 50	**22. Price**

Form DOT F 1700.7 (8-72)

DOI: 10.31979/mti.2023.2148

Mineta Transportation Institute College of Business
San José State University San José, CA 95192-0219

Tel: (408) 924-7560
Fax: (408) 924-7565
Email: mineta-institute@sjsu.edu

transweb.sjsu.edu/research/2148

Acknowledgments

The authors acknowledge the valuable advice and assistance provided by personnel from the California Transportation Commission (CTC) and the California Department of Transportation (Caltrans). CTC personnel who suggested this research, endorsed the proposal, and met the authors to discuss the progress included Timothy Sobelman, Jonathan Pray, Gurtej Bhattal, Jaden Gales, and Sheila Ennes. Caltrans personnel who met with the authors or provided data include Donna Berry, Chad Baker, Jeffrey Wiley, Rich Williams, Kien Le, John Ackerman, and Andriy Baydala.

Dr. Hilary Nixon provided the authors with valuable advice on the procedures to be followed on this California State University Transportation Consortium Grant, and Dr. Lisa Blampied very kindly reviewed and commented on the draft report, as did James Davis, retired Chief Deputy Director of Caltrans, and Robert So, retired Deputy Director for Program and Project Management in Caltrans District 7 (Los Angeles).

This research was funded, partially or entirely, by a grant from the Trustees of the California State University and the prime sponsoring agency State of California AB115 2017/2018 and contract number 08720115-SJAUX. The authors gratefully acknowledge this support. Any opinions, findings, and conclusions or recommendations expressed in this report are those of the authors and do not necessarily reflect the views of the California State University system in whole or part, the Trustees of the system, the CTC, or Caltrans.

CONTENTS

List of Figures

List of Tables

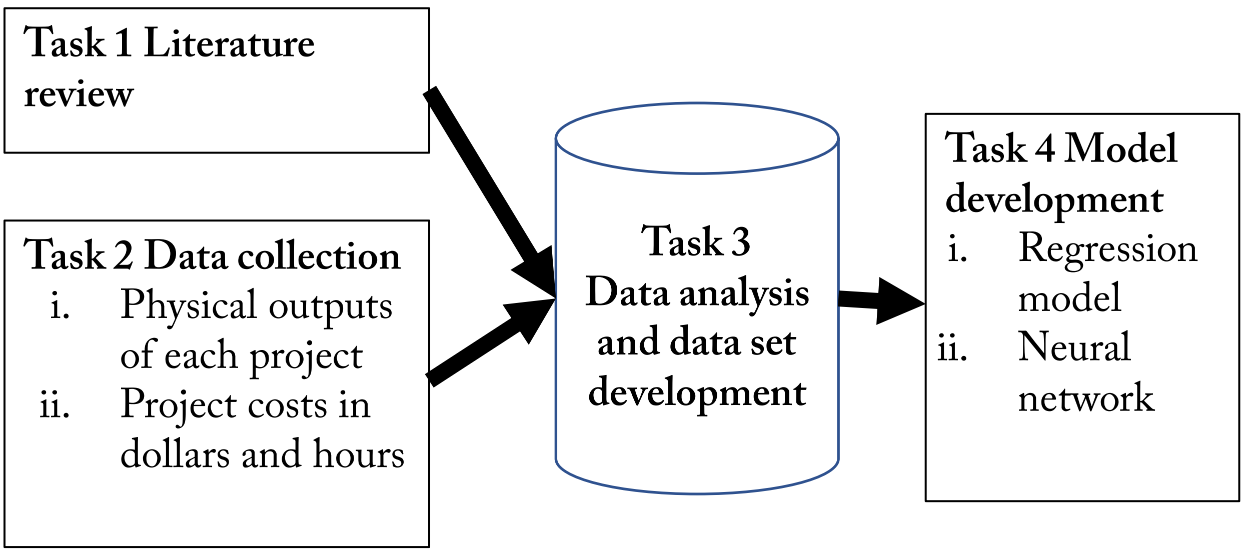
Executive Summary

This report presents the results of a research project on preconstruction cost hour estimating for pavement rehabilitation projects undertaken by the California Department of Transportation (Caltrans). It is laid out according to, and describes, the four-step process followed in the research, namely (1) a literature review; (2) data collection; (3). data analysis; and (4) model development. In the literature review, the report develops the case for using two types of models for estimating preconstruction hours, namely an Artificial Neural Network (ANN) and a Parametric Model. Data for 139 pavement rehabilitation projects was obtained from Caltrans, and a data set was developed that combined the preconstruction hours for each project with the primary bid items for pavement rehabilitation projects. The report describes how the two models were developed and their success in explaining 85% and 80%, respectively, of the variation in the preconstruction hours. Finally, the report concludes with suggestions for further research.

1. Introduction

This report presents the results of a research project on preconstruction cost estimating for pavement rehabilitation projects undertaken by the California Department of Transportation (Caltrans). The research project consisted of the four principal tasks illustrated in Figure 1.

Figure 1. Four Principal Tasks in the Research Plan



The research project responded to a need expressed by personnel in the California Transportation Commission (CTC). The CTC is required to oversee the preconstruction cost estimating process for the State Highway Operation and Protection Program (SHOPP), of which pavement rehabilitation projects are a part. The requirement in law reads as follows:

> Government Code 14526.5. (g) On or after July 1, 2017, to provide sufficient and transparent oversight of the department's capital outlay support resources composed of both state staff and contractors, the commission shall be required to allocate the department's capital outlay support resources by project phase, including preconstruction. Through this action, the commission will provide public transparency for the department's budget estimates, increasing assurance that the annual budget forecast is reasonable. The commission shall develop guidelines, in consultation with the department, to implement this subdivision. Guidelines adopted by the commission to implement this subdivision shall be exempt from the Administrative Procedure Act. (Chapter 3.5 (commencing with Section 11340) of Part 1)

The 2022 four-year SHOPP consists of more than 1,000 projects valued in total at $17.9 billion (Caltrans, 2022). Caltrans, the “department” referred to in the code, indicates that 14% of this amount is dedicated to the funding of preconstruction work. Preconstruction therefore accounts for about $2.5 billion of the four-year SHOPP, or about $600 million per year. The intent of this research, and similar additional future research, is to provide tools to help the CTC and Caltrans evaluate the estimates for this $600 million per year. Each of the four tasks in the research plan is presented as a chapter in the report that follows.

2. Literature Review

This chapter introduces some of the principal guidance documents on cost estimating that are relevant to this study. It begins with a background on Caltrans phases, proceeds to discuss estimating methods, considers the inputs to estimating, and gives an example of a tool formerly used by Caltrans for support cost estimating.

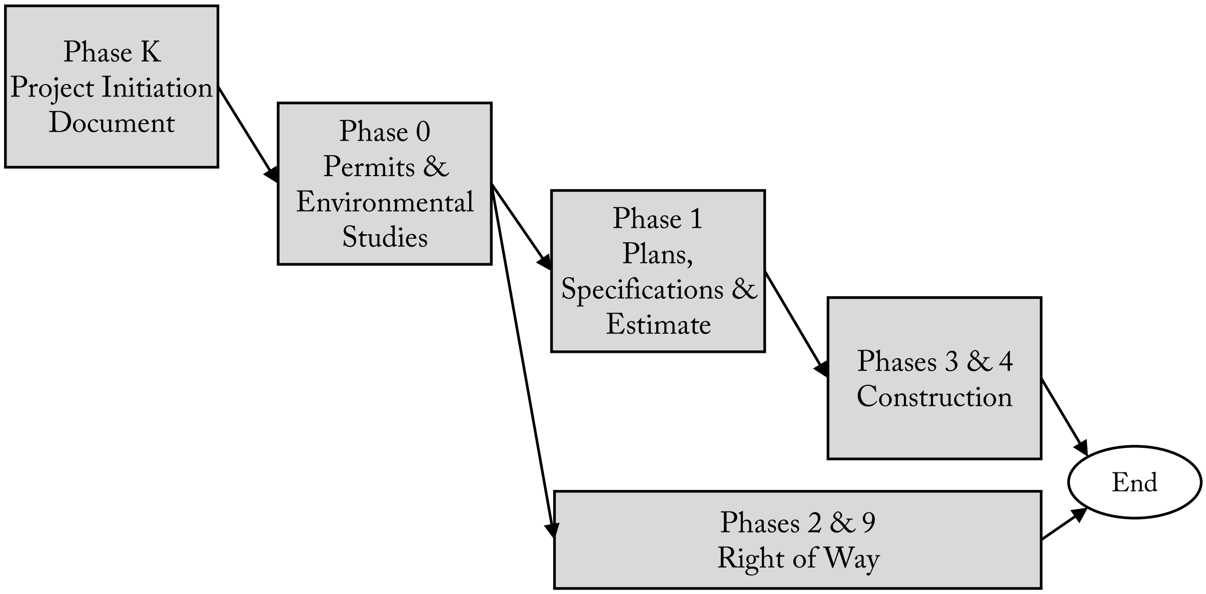
2.1 Caltrans Phases: *Project Management Handbook*

This report is concerned with the preconstruction phases of pavement rehabilitation projects. Caltrans records and reports project in seven phases, as follows:

- Project Initiation Documents, Phase K. Some of these documents are also referred to as Project Study Reports and Feasibility Studies.
- Permits and Environmental Studies, Phase 0, which ends in the Project Approval and Environmental Documents milestone (PA&ED) and is sometimes referenced by it.
- Plans, Specifications, and Estimates, Phase 1.
- Right of Way Operations, Phase 2.
- Construction Engineering, Phase 3.
- Construction Capital, Phase 4.
- Right of Way Capital, Phase 9.

Phases 2 and 9 occur simultaneously and can be collectively described as a single phase, "Right of Way." The same is true for Phases 3 and 4, which can be described as a single phase, "Construction." The separate recording and reporting of Phases 2 from 9 and 3 from 4 are required by Section 3.00 of California's annual budget and its *State Administrative Manual* (DGS, 2017). With these collective phases, then, the seven phases become five. The sequencing and relationship between these five phases are illustrated in Figure 2.

Figure 2. Caltrans Project Phases



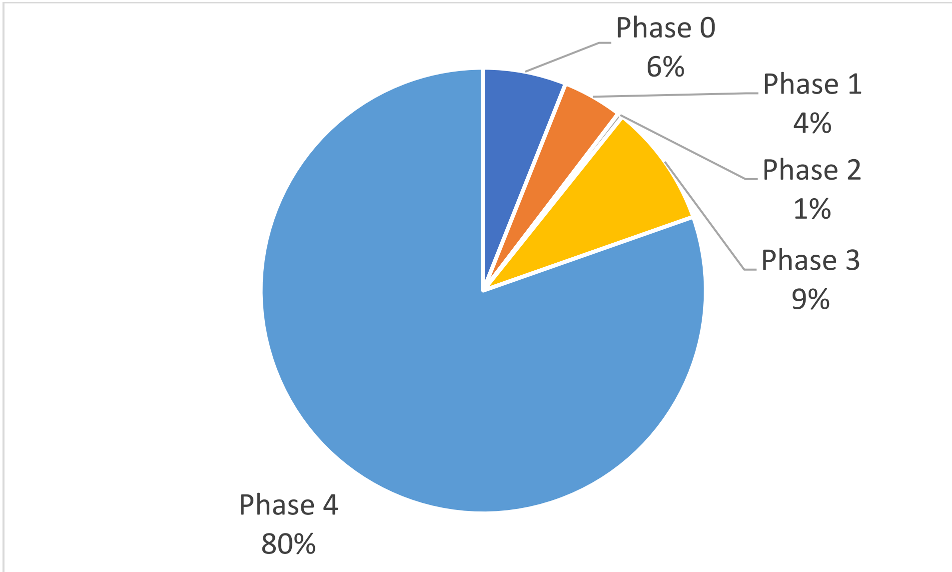
In this report, "preconstruction" is understood to refer to the sum of Phases 0 and 1 ("Permits and Environmental Studies" and "Plans, Specifications, and Estimates"). The cost estimate for these phases is presented to the CTC for approval before the start of Phase 0, that is, at the end of Phase K ("Project Initiation Document").

Figure 3 illustrates how pavement rehabilitation project costs were split across the phases, by percentage, excluding the Phase 9 cost, in the year from July 1, 2020 to July 1, 2021, the last complete year for which data was available when this research was undertaken. The data received from Caltrans did not include Phase 9, Right of Way capital or Phase K, as the Project Initiation Document is considered to be a pre-project cost rather than a project cost. Nevertheless, it is clear that construction capital costs comprise the bulk of the pavement rehabilitation costs. Construction capital refers almost exclusively to payments made to the contractors who perform construction work.

The preconstruction effort that is the topic of this report, Phase 0 and 1, constituted only 10 percent of the costs represented in Figure 3. Although only 10 percent of the total, these preconstruction costs in 2020–2021 amounted to more than $169 million, which is a sum worth examining. This $169 million is the portion of the $600 million of SHOPP preconstruction work, discussed in the Introduction, that is dedicated to pavement rehabilitation projects.

It should be noted that the data in Figure 3 is for all pavement rehabilitation projects that incurred expenses in 2020–2021. This is a different population and time period from the 138 projects that will be discussed later in this report, although some of the 138 projects did incur expenses in 2020–2021 and are therefore partially included in Figure 3.

Figure 3. Percentages of Caltrans Pavement Rehabilitation Costs by Phase in 2020–2021



2.2 Estimating Methods

This section introduces some of the principal guidance documents on cost estimating relevant to this study. The section moves from the general to the specific and discusses the Project Management Institute's (PMI) *Practice Standard for Estimating* (PMI, 2011), the American Association of State Highway and Transportation Officials' (AASHTO) *Practical Guide to Cost Estimating* (AASHTO, 2013), and the Transportation Research Board (TRB) *Guidebook on Estimating Highway Preconstruction Services Costs* (Gransberg et al., 2016). Table 1 indicates some characteristics of these sources that informed the choice to include them here.

Table 1. Characteristics of the Selected Literature Sources

	Specifically about Estimating	Focuses on Highways	Limited to Estimating of Costs	Specific to Preconstruction
PMI, 2011	✓			
AASHTO, 2013	✓	✓	✓	
Gransberg et al., 2016	✓	✓	✓	✓

2.2.1 Project Management Institute Practice Standard for Estimating

The PMI (2011) indicates that estimating methods fall into just three categories: analogous, bottom-up, and parametric. PMI defines analogous estimating as "a technique for estimating the duration or cost of an activity or a project using historical data from a similar activity or project" (PMI, 2017) and adds:

> Analogous techniques, also known as top-down estimating, are used when very little information is available about the project, or the new project is very similar to a previous project or the estimators have great experience with what is going to be estimated. This category of technique results in a total project estimate and is the technique of choice for early estimates where detailed information is not available. (PMI, 2011)

The analogous approach is frequently used in conceptual cost estimating where, as PMI indicates, "detailed information is not available." It both requires that the estimator have knowledge of similar comparable projects and relies upon the estimators' expert judgement.

PMI defines the second method, bottom-up estimating, as "a method of estimating project duration or cost by aggregating the estimates of the lower-level components of the work breakdown structure" (PMI, 2017) and adds:

> Bottom-up techniques are applied as the estimating tool of choice when the detailed project data becomes available. Using this technique, the expenditure of every resource of every component of the project is estimated as a prelude to rolling up these estimates to the intermediate levels and to the total project. This technique will result in a transparent and structured estimate for the project, which can be tracked and managed. (PMI, 2011)

The bottom-up approach is not feasible in conceptual cost estimating since one must have detailed project data, and this is not the case at the conceptual stage.

PMI defines the third method, parametric estimating, as "an estimating technique in which an algorithm is used to calculate cost or duration based on historical data and project parameters" (PMI, 2017) and notes: "Parametric techniques use statistical relationships between historical data and other variables (e.g., square meters in construction) to calculate an estimate for an activity cost, duration, or resource" (PMI, 2011). A discussion of parametric estimating follows in this report.

2.2.2 AASHTO Practical Guide to Cost Estimating

Although PMI indicates that analogous and parametric estimating are used at the early stage of a project and bottom-up estimating later, it does not identify a clear correlation between estimating methods and project phases. AASHTO's *Practical Guide to Cost Estimating* (AASHTO, 2013) provides a clearer correlation, as illustrated in Table 2.

Table 2. Cost Estimating Methods by Phase, According to AASHTO

Project Development Phase	Project Maturity (% project definition completed)	Purpose of the Estimate	Estimating Methodology	Estimate Range
Planning	0 to 2%	Conceptual Estimating—Estimate Potential Funds Needed (20-year plan)	Parametric (Stochastic or Judgment)	–50% to +200%
	1% to 15%	Conceptual Estimating—Prioritize Needs for Long-Range Plans (IRP—10-year plan)	Parametric or Historical Bid-Based (Primarily Stochastic)	–40% to +100%
Scoping	10% to 30%	Design Estimating—Establish a Baseline Cost for Project and Program Projects (IRP and STIP)	Historical Bid-Based or Cost-Based (Mixed, but Primarily Stochastic)	–30% to +50%
Design	30% to 90%	Design Estimating—Manage Project Budgets against Baseline (STIP, Contingency)	Historical Bid-Based or Cost-Based (Primarily Deterministic)	–10% to +25%
Final Design	90% to 100%	PS&E Estimating—Compare with Bid and Obligate Funds for Construction	Cost-Based or Historical Bid-Based Using Cost Estimate System (Deterministic)	–5% to +10%

Source: AASHTO (2013).

It will be noticed that AASHTO uses different terms from PMI. It substitutes as follows:

- PMI's "analogous" is AASHTO's "judgment"
- PMI's "parametric" is AASHTO's "stochastic"
- PMI's "bottom-up" is AASHTO's "deterministic"

As indicated in Table 2, the appropriate estimating methods at the early planning phase of a project are analogous/judgement or parametric/stochastic. From the discussion of Caltrans phases, in Section 2.1, preconstruction cost estimates are needed at such an early phase. This study therefore focuses on parametric methods and an alternative not considered to any great extent by PMI or AASHTO, namely artificial neural networks, as introduced in the next section.

2.2.3 TRB Guidebook on Estimating Highway Preconstruction Services Costs

Gransberg et al. (2016) write specifically about estimating of preconstruction costs. They offer three methods for estimating such preconstruction costs:

- Multiple regression modeling (yet another term for parametric and stochastic estimating)
- Decision tree analysis
- Artificial neural network modeling

This expands upon the methods described by PMI and AASHTO by adding two new methods, decision tree analysis and artificial neural network modelling, but does not include the analogous or judgmental method.

2.2.4 Summary

Table 3 provides side-by-side definitions of the estimating methods, as provided by PMI, AASHTO, and Gransberg et al., using the equivalent terms in these three sources.

Table 3. Summary of Estimating Methods, by Source

PMI, 2017	AASHTO, 2013	Gransberg et al., 2016
<u>Analogous</u>: “a technique for estimating the duration or cost of an activity or a project using historical data from a similar activity or project.”	<u>Judgment</u>: not defined.	
<u>Parametric</u>: “an estimating technique in which an algorithm is used to calculate cost or duration based on historical data and project parameters.”	<u>Stochastic</u>: not defined.	<u>Multiple Regression</u>: “. . . statistical method for studying the relationship between a single dependent variable and one or more independent variables” (Allison, 2009).
		<u>Artificial neural networks</u>: “. . . are capable of learning complex relationships in data. By mimicking the functions of the brain, they can discern patterns in data, and then extrapolate predictions when given new data” (Palisade Corporation, 2010).
		<u>Decision Tree</u>: “A decision tree describes graphically the decisions to be made, the events that may occur, and the outcomes associated with combinations of decisions and events. Probabilities are assigned to the events, and values are determined for each outcome” (TreePlan Software, Inc., 2016).
<u>Bottom-up</u>: “a method of estimating project duration or cost by aggregating the estimates of the lower-level components of the work breakdown structure.”	<u>Deterministic</u>: Historic bid-based or Cost-based	<u>Bottom-up</u>: “Detailed estimates of work packages usually made by those who are most familiar with the task (also called micro estimates)” (Larson and Gray, 2011).

In this report, because we are working with estimates at the Caltrans Project Initiation Document phase, which AASHTO calls the Planning Phase, we consider only estimates that use two methods:

- Parametric estimates (PMI), which AASHTO calls "Stochastic" and Gransberg et al. call "Multiple Regression".
- Artificial Neural Networks (ANN), which are discussed by Gransberg et al., but not by PMI or AASHTO.

At the Project Initiation Document / Planning Phase, it would also be appropriate to use Gransberg's "Decision Tree" method, but that would require a larger collection of input variables than what were available for this study. PYPSCAN, discussed in Section 2.4, did follow a decision tree approach, but it required a multi-year data gathering effort and an ongoing mandate for Caltrans districts to maintain the data. Input variables are discussed in Section 2.5, and an example of their use is provided in Section 2.4.

2.3 Input Variables

To develop a cost estimating model, whether parametric or ANN, one needs both a set of projects for which one knows the values of given input variables and the final costs. The model would then be expressed as "given inputs x, the expected cost is y." Blampied (2018) reviewed 40 papers on cost estimating and found that their input variables fell into five categories:

- Project outputs: Items that the project will produce for the project customers and for which the project is undertaken.
- Bid items: Items for which contractors submit prices before the start of the contract and for which they receive payment during or after the contract. Bid items are often needed to achieve the project outputs and may be a proxy for project outputs.
- Project characteristics: Items, other than project outputs and bid items, that affect a project and that must be taken into consideration during the project's development.
- Economic: Indicators of the economic situation in the area in which the project is to be developed.
- Geographic: Physical characteristics of the area in which the project is to be developed.

Appendix A provides examples of input variables in each of these categories, modified from Blampied (2018). To use any of these input variables, the data needs to be available. In the present study, we used only project outputs and bid items. A future study could add a consideration of project characteristics such as environmental document type, right-of-way characteristics, and urban-versus-rural settings. Economic and geographic input variables are less likely to be useful for a study that is confined to a single state, however, because they refer to input variables that differ from country to country or, at least, region to region.

2.4 PYPSCAN: Prior Use of the Parametric Exponential in Caltrans

Caltrans has a precedent for developing parametric estimates of support effort by phase on projects named the Person-Year Project Scheduling and Cost Analysis (PYPSCAN). This precedent is described in Blampied et al. (2017) and Blampied (2018). It came into service in 1980 after several years of development (Caltrans 2017). Caltrans has not updated PYPSCAN since 1997, and it is no longer used as the official Caltrans system for estimating support effort (Caltrans 2017).

PYPSCAN calculated expected person year (PY) resource needs based on several variables:

Project Type: Each project in a database of over 12,000 projects was assigned to a project type. At first there were 107 project types (McManus 1981) and, over time, additional types were added to reach 119 by 1992. Each project was assigned only one project type (Caltrans 1992).

Function: Expected PYs were found for each of five "functions": (1) highway preliminary engineering (or "project development," PJD), (2) right of way (RWO), (3) structures design (STD), (4) structures construction (STC), and (5) highway construction (CON). In the early stages, PYSCAN also calculated day labor (D/L), but it was later dropped (McManus 1981 compared with Caltrans 1992).

Capital Cost: The user entered three capital costs: (1) the total construction cost, (2) right-of-way capital, and (3) structures construction cost. The system adjusted these costs for inflation using the Caltrans Construction Cost Index. A different capital cost was used for each function: PJD, CON, and D/L used total construction cost; STC and STD used structure construction cost; and RWO used right-of-way capital.

Environmental Type: Each project had one of three environmental types, as specified in State and Federal law (CEQA and NEPA): (1) Categorical Exemption / Exclusion (CE), (2) Environmental Impact Report / Environmental Impact Statement (EIR/EIS), (3) Negative Declaration / Finding of No Significant Impact (ND/FONSI).

Location: California was divided into Urban and Rural areas.

Right-of-Way Information: This included numbers of appraisals, acquisitions, utilities, relocation assistance cases, demolitions, railroad agreements, and condemnations.

Weather Zone: Each project was assigned to one of five weather zones, with Zone 1 being the driest and Zone 5 the wettest.

Considering the input variables in Appendix A, PYPSCAN used a combination of three types of input variables to develop the expected PY workload:

Bid items: The construction capital costs are totals of the bid item costs on each project.

Project characteristics: Environmental document type, urban / rural location, and right-of-way information.

Geographic: Weather zone.

PYSCAN formulas had an exponential form:

$$P = aX^b \qquad \text{(equation 1)}$$

Where *P* is the expected PY need, *a* is a constant, *X* is the inflation-adjusted estimated capital cost for the particular function, and *b* is a constant, with $0 < b < 1$. This produced a possible 12 formulas for each project type (6 for PJD: 3 environmental types x 2 location types; and 2 each for each location type—STD, STC, and CON). A single formula, regardless of project type, was developed for the RWO calculated from the number and complexity of the Right of Way parcels that were affected. There were thus a possible 1,429 formulas (12 x 119 + 1). Each formula was developed by performing regression analysis on the projects from the database that matched the particular combination of input variables. In practice, there were fewer than 1,429 formulas because some project types could not include structures, and most had only one possible environmental type.

PYPSCAN is an example of Gransberg's "Decision Tree" method. One begins with 1,429 possible formulas and then eliminates groups of formulas as one queries the project type, function, environmental type, and location to end up with a single formula that is applicable to the specific situation.

Until 1996, Caltrans used PYPSCAN to develop the annual Capital Outlay Support budget. Through PYPSCAN, an expected number of PY was calculated for each project for each year by function. These numbers were added and submitted to the Legislature as the need for direct project work in the coming year. The number for any given project would almost certainly be higher or lower than the expected number, but the overages and underages would cancel each other out and, when added over thousands of projects, would be an accurate estimate of the department-wide need. In 1996, Caltrans used PYSCAN, but then made an adjustment to the total.

The Legislative Analyst reported:

> Caltrans uses a statistical model to estimate its capital outlay support staff requirements, based upon the number, size, and complexity of scheduled projects. For 1996-97, this workload model calculated a higher staffing requirement than in the current year. However, Caltrans reduced the modeled workload by 19 percent in order to attain the

> staffing level proposed in the budget. Caltrans reports that it made the adjustments in order to account for anticipated efficiencies and shortcomings in the model. (LAO 1996)

Later in the same report, the Legislative Analyst continued, “Caltrans must, therefore, improve its workload forecasting models and practices.”

At that time, Caltrans had recently introduced a new commercial project management system, eXpert Project Manager (XPM). In response to the Legislative Analyst’s criticism, Caltrans introduced a process in 1997 whereby task managers could estimate the hours required for their units’ work on each project, and those estimates were entered into XPM. The basis of the task managers’ estimates was not documented but if it had been, it would probably have been based on the task managers’ experience and the estimates would therefore be analogous estimates.

In 2014, Caltrans introduced a new project management system to replace XPM, CA PPM, now called Broadcom Clarity. Caltrans refers to its installation of Broadcom Clarity as “Project Resourcing and Schedule Management” (PRSM).

3. Data Collection

Sections 2.5 and 2.6 have discussed some of the input variables that could be used in developing an estimate of preconstruction effort. In this study, the researchers chose to use only project outputs and bid items as input variables. This choice was made for two reasons:

(1) The data was known to be available in existing databases and would therefore not require an extensive search for data or the assembly of a new database of original data.

(2) The project outputs are, by definition, the items desired by the customer. The model this study produced would therefore tell the customers what the costs of their desired outputs would be. Bid items, as discussed above, can be proxies for project outputs, even though they themselves might not be the actual outputs desired by the customers. For instance, on the pavement rehabilitation projects considered in this study, lane miles of rehabilitated pavement are the desired project output. Bid items such as cubic yards of concrete pavement or tons of hot-mix asphalt bear a relationship to the lane miles and can thus serve as proxies for the lane miles.

3.1 Project Performance Targets

As previously noted, project outputs are frequently used as predictors of cost. Examples of these outputs are listed in Appendix A. The use of such outputs is facilitated by the 2012 Federal Surface Transportation Reauthorization Act, known as MAP-21 (US Congress, 2012), which requires each state to adopt performance measures and to set targets for its expected achievements against each measure. Caltrans has assigned specific performance targets to each of its projects. For pavement rehabilitation, the performance target is measured as lane miles that are to be repaved.

Caltrans provided the researchers with data on the performance targets for its projects, where these targets are available. As the MAP-21 requirement is fairly new, targets are available only for projects that have been programmed since the targets were established.

3.2 Project Bid Information

Caltrans provided the researchers with the bid data for the 1,055 major projects for which bids were opened over the five-year period from April 26, 2016 to May 11, 2021.

3.3 Expenditure Information

Caltrans provided the researchers with actual project expenditure data for phases 0, 1, 2, 3, and 4 of all Caltrans state highway projects from July 1, 2009 to January 10, 2022. The researchers had previous expenditure files for all phases of all Caltrans state highway projects from July 1, 1982 to June 30, 2009.

4. Data Analysis

4.1 Data Combination

The researchers combined the bid data with the expenditure data to find the Caltrans programs that funded each project. In some cases, projects received funding from more than one program, and in those cases, Caltrans split the funding between programs. The list of programs and the number of projects in each program is provided in Appendix B. From this list of programs, the researchers identified three programs listed in Table 4 that, together, comprise "pavement rehabilitation" for this report's purposes. Under this classification, 139 pavement rehabilitation projects had bid openings in the five-year period from April 26, 2016 to May 11, 2021.

Table 4. Three Pavement Rehabilitation Programs

Program	Number of bid openings
20.201.121 Pavement Rehabilitation	100
20.201.122 Pavement Preservation	26
20.201.120 Roadway Rehabilitation	13

4.2 Pareto Analysis

The researchers considered the bid information on each of the 139 pavement projects and, for each project, found the fewest bid items that together contributed more than 80 percent of the awarded bid amount, in dollars, on that project. These bid items were then analyzed across the 139 projects, and it was found that 20 primary bid items together contributed more than 80 percent of the awarded bid amount, in dollars, on every one of the 139 projects.

4.3 Data Set Creation

After the Pareto Analysis, three data sets were created as described below.

4.3.1 Loaded State Employee Cost

The first data set listed the sums, by fiscal year, for all projects in the three programs in Table 4, of the:

a. State Employee non-overtime preconstruction hours;

b. State Employee non-overtime preconstruction personal services cost dollars; and

c. Overhead assessments in dollars (these are assessed to projects in proportion to non-overtime hours).

All of this data was drawn from the expenditure information discussed in Section 3.3.

Using this data set, a loaded hourly State Employee cost, *d*, was calculated for each fiscal year, being d = (b + c) / a, using the factors above.

4.3.2 A&E Hours

The second data set listed, for each of the 139 projects, the Architectural and Engineering (A&E) consultant preconstruction dollar expenditures, e, by fiscal year, from the expenditure information discussed in Section 3.3. Using the loaded hourly State Employee costs, *d*, by fiscal year from the first data set, A&E dollars were then converted into an hour equivalent, f = e / d.

4.3.3 Preconstruction Hours for the 139 Projects

The third data set listed the bid quantity, state employee preconstruction hours, and sum of equivalent A&E preconstruction hours:

g. Bid quantity for each of the 20 primary bid items (data from the Bid Information described in Section 3.2, selected based upon the Pareto Analysis described in Section 4.2)

h. State Employee preconstruction hours from the expenditure information discussed in Section 3.3.

i. Sum of equivalent A&E preconstruction hours, f, from the second data set.

Total preconstruction hours, being j = h + i.

This data set focused on preconstruction hours rather than dollar costs because hours are not subject to inflation. If dollar costs were used, each annual cost would need to be inflation-adjusted to a baseline date, using an inflation index. Indexes are averages of sample data and have significant margins of error.

Caltrans estimates preconstruction effort in hours, a standard practice in most project cost estimating. The hours are then entered into PRSM where average dollar costs per hour are applied. The priced preconstruction costs are entered into the Caltrans programming system, California Transportation Improvement Program System (CTIPS), where they are inflation-adjusted to the years in which work will be performed.

By using hours rather than dollars, this research avoided several adjustments for inflation, both up and down, and the use of average costs. These adjustments and averages increase the variation and uncertainty in the data and consequently decrease the reliability of the final results.

5. Model Development

Using the third data set described in Section 4.3.3, the researchers developed two models of estimating the preconstruction hours on Caltrans pavement rehabilitation projects: an Artificial Neural Network (ANN) and a Parametric Model.

5.1 Artificial Neural Network (ANN) Model

5.1.1 ANN Concept

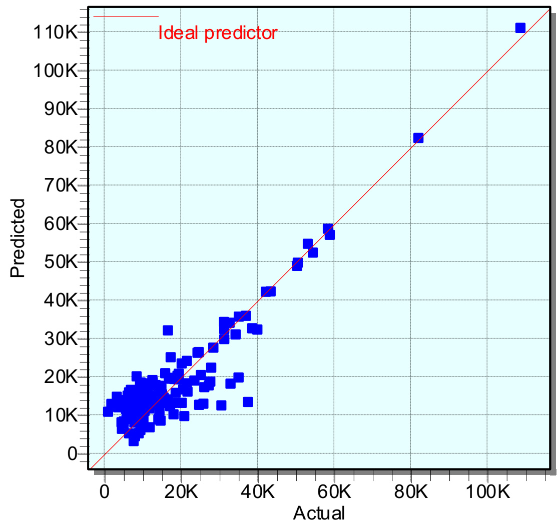
ANNs are designed to mimic human thought and are a form of artificial intelligence. Blampied (2018) found the first suggestion that ANNs might be used in cost estimating in a paper from 1993 (Rao et al., 1993) and the first successful development of an ANN for cost estimating in 1998 (Siqueira and Moselhi, 1998). Since then, ANNs have become more widely used in cost estimating. This development is part of the increasing use of artificial intelligence in many spheres of life. ANNs operate with "layers": an input layer, one or more hidden layers, and an output layer.

The ANN for Caltrans pavement rehabilitation projects, based on the 139 projects in the third data set (see Section 4.3.3), has a coefficient of determination (R^2) of 0.85 and an average error of 4,407.7. This will described in more detail in a forthcoming paper by Shehab et al. "Estimation of the Preconstruction Cost of Pavement Rehabilitation Projects".

5.1.2 Actual vs. Predicted Hours

Figure 4 compares the actual preconstruction hours on the 139 projects in the ANN version of this study to the hours predicted using the ANN model. The figure includes a diagonal line along which the actual hours equal the predicted. All points above the diagonal line have actual hours that exceed the predicted hours, while points below the diagonal have actual hours that are fewer than the predicted hours.

Figure 4. Comparison of Actual to Predicted Preconstruction Hours Using the ANN Model



Although there is a relationship between the predicted and the actual preconstruction hours, as attested by the coefficient of determination (R^2) of 0.85, there remains a significant spread in the data, both above and below the best fit line. This is discussed in Section 6.1.

5.2 Parametric Models

5.2.1 Least Squares Regression Analysis

The researchers wrote MATLAB scripts to test the pavement rehabilitation project data, with the goal of finding the lowest least squares best fit relationship of the pavement rehabilitation project lane miles and primary bid items to the preconstruction effort in hours. Each script found the minimum value of a parametric function when varying a single unknown and keeping all other unknowns constant. The scripts iterated through the unknowns, adjusting each one in turn, until no further adjustment could reduce the total of the squares of the errors in the expression (i.e., providing a least-squares regression). The MATLAB scripts are available upon request and are not repeated here.

Upon examination of the data in the third data set (see Section 4.3.3), it was found that one of the 139 projects had no recorded preconstruction hours. This project was therefore omitted from the least squares regression. As a result, the parametric estimates are based on 138 projects, whereas the ANN was based on 139 projects.

5.2.2 Single-input-Variable Parametric Models

MATLAB code was first written to consider the lane mile project output and each of the primary bid items as single input variables. One of the primary bid items, Time Related Overhead, occurred on every project but took two forms, namely payment by Work Days (WDAYS) or payment as a

Lump Sum (LS). For the parametric analysis, this bid item was considered as two separate items. A total of 21 input variables were therefore considered: the 1 project output, lane miles, and now 20 primary bid items. The single-input-variable parametric model takes the form:

Preconstruction hours (Phases 0 + 1) = $a + b.X^c$ (equation 2)

Where *a*, *b*, and *c* are constants, and *X* is each of the input variables. It should be noted that the form of this equation is similar to that used for PYPSCAN (equation 1) with the addition of a leading constant, *a*. The constant is added because some activities on a project are independent of the project size. Establishing a project in the accounting system, one of the first project activities, for instance, requires the same effort on every project regardless of size.

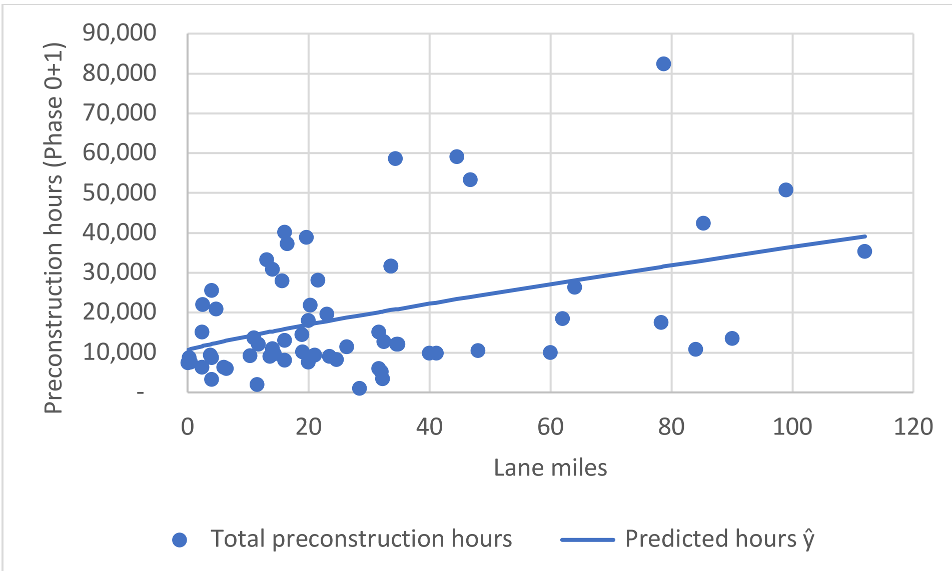
Table 5 lists the resulting determinations of factors a, b, and c along with the number of projects that included each of the input variables X, the number of times the script iterated through the constants, and the coefficient of determination, R2. This coefficient indicated how much of the variation in hours is accounted for by the input variable X. For instance, the first listed input variable, Time-Related Overhead in working days, accounts for 49.4 percent of the variation in hours.

Table 5. Least Squares Best Fit Factors for Equation 2

Input variable X	Unit	Number of projects	Least squares iterations	a	b	c	R^2
X_1 Time-Related Overhead – Days	WDAY	70	48	1,674.651	75.562	1.000	0.494
X_2 Structural Concrete	CY	66	2,525	17,099.381	52.742	0.798	0.474
X_3 Lean Concrete	CY	25	1,034	0.000	2,663.737	0.312	0.425
X_4 Temporary Railing	LF	86	6,751	10,206.483	169.207	0.427	0.338
X_5 Roadway Excavation	CY	93	4,358	15,399.599	10.707	0.640	0.277
X_6 Jointed Concrete Pavement	CY	29	5,086	25,415.463	5.084	0.789	0.222
X_7 Aggregate Base	CY	79	1,458	16,379.988	228.608	0.408	0.206
X_8 Continuous Reinforced Concrete Pavement	CY	18	2,220	0.000	1,163.964	0.288	0.185
X_9 Lane Miles	miles	66	1,721	10,581.393	476.250	0.867	0.174
X_{10} Grind Existing Concrete Pavement	SQYD	36	7,939	16,127.312	22.931	0.526	0.154
X_{11} Hot Mix Asphalt	Ton	131	577	14,410.965	0.146	1.000	0.139
X_{12} Minor Concrete	CY	94	2,609	7,709.038	6,393.685	0.133	0.058
X_{13} Vegetation Control	SQYD	81	11,035	10,814.140	735.777	0.317	0.048
X_{14} Guardrail System	LF	108	16,995	5,722.603	3,440.373	0.168	0.041
X_{15} Mobilization	LS	138	276	10,787.163	8,050.130	0.000	0.013
X_{16} Modifying or Removing Electric System	LS	138	37	16,271.534	3,587.036	0.000	0.013
X_{17} Rubberized Hot Mix Asphalt	Ton	104	366	17,222.739	0.039	1.000	0.008
X_{18} Cold Plane Asphalt Concrete Pavement	SQYD	128	66	17,580.246	0.006	1.000	0.006
X_{19} Slab Replacement	CY	34	32,046	0.000	21,683.010	0.036	0.003
X_{20} Traffic Control System	LS	138	207	17,357.374	1,034.112	1.000	0.0005
X_{21} Time--Related Overhead – LS	LS	68	1	15,223.170	0.000	1.000	0.000

Input variable X_9, Lane Miles, is the project output for pavement rehabilitation projects. It would be desirable to say that a given preconstruction effort would produce a lane mile of pavement, but the data indicates that lane miles is not a good predictor of preconstruction effort. This is illustrated in Figure 5 which charts the preconstruction hours against the lane miles on the 138 projects. There is a wide scatter of data above and below the best-fit "predicted hours" line.

Figure 5. Lane Miles vs. Preconstruction Hours



5.2.3 Multiple Regression Analysis

After considering the single-variable least squares regression parametric models, MATLAB scripts were written to examine situations that considered the full set of 21 input variables. Several different versions were developed, and these will be described in more detail in a forthcoming paper by Blampied et al., "Parametric Estimation of the Preconstruction Cost of Pavement Rehabilitation Projects". The highest correlation was found from an additive exponential parametric model that excluded three input variables and found a further four null input variables. An additive exponential parametric model takes the form:

$$\text{Preconstruction hours (Phases 0 + 1)} = a + b_1.X_1^{c1} + b_2.X_2^{c2} + \ldots + b_n.X_n^{cn} \text{ (equation 3)}$$

The expression with the highest correlation had a coefficient of determination (R^2) of 0.80 and an average error of 6,429.8. It therefore had a lower accuracy than the ANN.

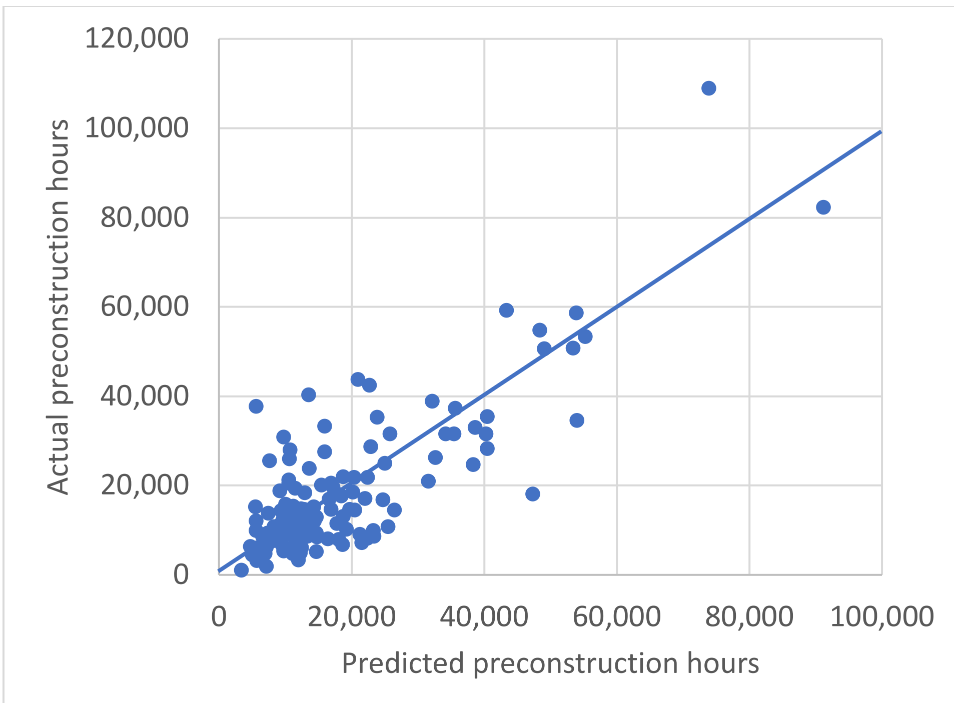
The version that produced the above results excluded three input variables from consideration because they could potentially be manipulated by the user: X_1 Time-Related Overhead – Days, X_{15} Mobilization, and X_{21} Time-Related Overhead – LS. The least squares calculations then found that another four input variables had no impact on the preconstruction hours. These were X_5

Roadway Excavation, X_6 Jointed Concrete Pavement, X_8 Continuous Reinforced Concrete Pavement, and X_9 Lane Miles. As a result, 14 input variables were found to be significant.

5.2.4 Actual vs. Predicted Hours

Figure 6 compares the actual preconstruction hours on the 138 projects in the parametric version of this study to the hours predicted using the multiple regression analysis. The figure includes a diagonal line along which the actual hours equal the predicted. All points above the diagonal line have actual hours that exceed the predicted hours, while points below the diagonal have actual hours that are fewer than the predicted hours.

Figure 6. Comparison of Actual to Predicted Preconstruction Hours Using the Parametric Model



Although there is a relationship between the predicted and the actual preconstruction hours, as attested by the coefficient of determination (R^2) of 0.80, there remains a significant spread in the data, above and below the best fit line. This is discussed in Section 6.1.

6. Summary & Conclusions

6.1 Dealing with the Unexplained Variation

This study has produced two models, an ANN with a coefficient of determination (R^2) of 0.85 and a parametric model with a coefficient of determination (R^2) of 0.80. They thus address 85% and 80% of the variation in the data, respectively, and leave 15% and 20% of unexplained variation. The 85% and 80% is a significant achievement, but the unexplained variations still pose problems. Possible approaches to addressing this variation include searching for additional variables, assuming estimate ranges, developing probability density functions, and adopting a portfolio-level perspective. These approaches are considered below.

6.1.1 Search for Additional Variables

This study has used two types of input variable, namely, project outputs and bid items. Section 2.5 discusses other variables that could be used. Some of them would probably increase the coefficient of determination of the models and therefore explain a larger percentage of the variation; however, this might not be the case. For instance, Gardner et al. (2016) studied the correlation of project costs using 29 project characteristics on State Highway pavement preservation projects in Montana and found that the coefficient of determination did not increase significantly when they considered more than the eight highest-impact project characteristics. They found that a significant effort was required to gather data on additional project characteristics and that the cost of data gathering outweighed the value of increase in the coefficient of determination. Gardner et al. suggested that agencies identify their highest-impact variables and then maintain data only on those input variables. This would avoid the costly effort needed to assemble hard-to-find data on low-impact variables.

It seems reasonable to expect only marginal changes in the coefficient of determination as additional low-impact variables are added to a parametric model, and the finding by Gardner et al. (2016) is not surprising.

6.1.2 Assuming Estimate Ranges

A common method of addressing the unexplained variation is to ascribe it to random factors and establish empirical estimate ranges. AASHTO uses this approach, indicating that conceptual estimates should typically have a range +200 percent to -50 percent (AASHTO, 2013).

6.1.3 Probability Density Function

A third method of addressing the unexplained variation is to develop a probability density chart. Figure 7 provides an illustrative hypothetical example of a simple case in which there is only one input variable, the “units of output.”

Figure 7. Illustration of “best fit” Parametric Cost Estimate

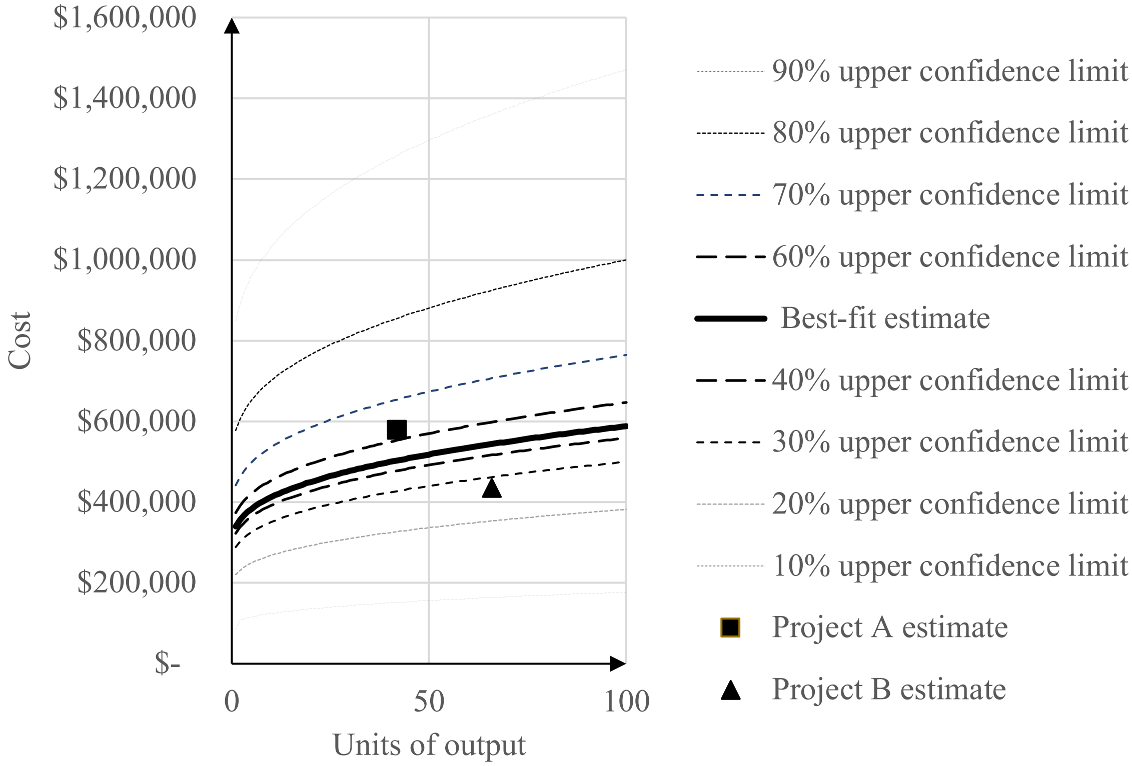


Figure 7 illustrates a best fit parametric cost estimate and then adds probability lines for various confidence levels above and below the best fit expected cost. At the 10 percent upper confidence limit line, for instance, there is 10 percent confidence that the cost will be below that limit line and 90 percent confidence that the cost will be above the limit line. At the 90 percent upper confidence limit line, the reverse is true—there is 90 percent confidence that the cost will be below that limit line and 10 percent confidence that the cost will be above the limit line.

A chart such as Figure 7 could be used by project portfolio owners to evaluate cost estimates that are submitted by project teams. In the example in Figure 7, the team for Project A has submitted an estimate that is higher than average, being slightly above the 60-percent upper confidence limit. The team for Project B has submitted an estimate that is lower than average, being slightly below the 30 percent upper confidence limit.

The estimates for Projects A and B are not necessarily wrong. Almost 40 percent of historic projects with outputs similar to those of Project A have cost more than the Project A estimate, and almost 30 percent of projects with outputs similar to those of Project B have cost less than the Project B estimate. The chart provides a tool that the project team can use to evaluate the team’s estimates. There may good reasons why the estimate for a project is above or below the historic average, and it is incumbent upon the team to both consider and provide those reasons.

6.1.4 Portfolio-level Perspective

A fourth approach to addressing the unexplained variation is to use a portfolio-level perspective. Rather than focusing on individual projects, one could consider the entire project portfolio. As an example, California's State Highway Operation and Protection Program (SHOPP) consists of more than 1,000 projects with a total four-year programmed budget of $17.9 billion, as was noted in the Introduction (Caltrans, 2022). In a portfolio such as the SHOPP, some projects can be expected to have estimates above the "best fit" line and others below (as illustrated in Figure 7), and these surpluses and deficits can be expected to cancel each other out. Across more than 1,000 projects, for instance, the aggregate of the estimates by the project teams should be close to the aggregate of the "best fit" expected conceptual estimates.

Nevertheless, it should be noted that while this portfolio approach could be useful to estimate the total budget at the portfolio level, it does little to correct potential inaccuracies in individual projects.

6.2 Usefulness to the CTC and Caltrans

Separately from this report, the researchers have provided the CTC with tools to use this research. These consist of both ANN and parametric estimating tools. Suggestions for their use are discussed below.

6.2.1 ANN Model

The use of the ANN model has been demonstrated to CTC staff. The application of this method requires that the user obtain the ANN software. As noted above, the ANN provides a more reliable estimate than the estimate provided by the parametric model.

6.2.2 Parametric Model

A spreadsheet has been provided that can be used by CTC or Caltrans staff to obtain the expected number of preconstruction (Phases 0 and 1) hours for a given pavement rehabilitation project when the project is submitted for programming. In addition to the expected hours, this spreadsheet provides upper and lower confidence limits which are based on the AASHTO limits (AASHTO, 2013). This spreadsheet is available for use by other interested parties.

Figure 8 is an image of the input form for the parametric model. To use it, the capital cost estimate in the Project Initiation Document would need to include estimated quantities for each of the 14 expected bid items listed in the input form. After entering those 14 quantities, the spreadsheet would return an expected number of preconstruction hours, along with upper and lower confidence limits based on the AASHTO factors (AASHTO, 2013).

Figure 8. Image of the Input form for the Parametric Model

Program "Task" Other- this model does not apply	District	Project number
Expected Bid Item	**Quantity**	**Unit of measure**
Traffic Control System	-	LS 1 = Yes, included in project, 0 = No, not included in project
Temporary Railing	-	LF
Aggregate Base	-	CY
Lean Concrete	-	CY
Hot Mix Asphalt	-	Ton
Rubberized Hot Mix Asphalt	-	Ton
Cold Plane Asphalt Concrete Pavement	-	SQYD
Slab Replacement	-	CY
Grind Existing Concrete Pavement	-	SQYD
Structural Concrete	-	CY
Minor Concrete	-	CY
Guardrail System	-	LF
Vegetation Control	-	SQYD
Modify or Remove Electrical System	-	LS 1 = Yes, included in project, 0 = No, not included in project
Upper confidence limit	**9,420**	*Using AASHTO estimate ranges, 90% certain that the actual Phase 0 + 1 hours (state employees and consultants combined) will be less that this upper confidence limit.*
Best estimate	**3,140**	*Best estimate of the Phase 0 + 1 hours (state employees and consultants combined)*
Lower confidence limit	**1,570**	*Using AASHTO estimate ranges, 90% certain that the actual Phase 0 + 1 hours (state employees and consultants combined) will be more that this lower confidence limit.*

CTC staff or Caltrans can use these tools to evaluate the reasonableness of the preconstruction estimate on an individual project or on the sum of an entire biennial SHOPP pavement rehabilitation portfolio. The researchers recommend that the CTC consider the results at the aggregate portfolio level and delegate the further consideration of individual projects to Caltrans.

6.3 Further Research

This report and its associated research are of limited scope due to constraints of time and funding. Further opportunities for research include:

A. Research that can be accomplished with the existing data:

1. Estimating for other project types
2. Estimating for other phases
3. Estimating upper and lower confidence limits at the bid stage
4. Developing portfolio-level cost estimates for entire biennial tranches of projects
5. Evaluating of efficiencies
6. Evaluating alternative project development processes
7. Evaluating optimum levels of contracting-out
8. Evaluating multi-year “support to capital”

B. Research that would need some additional data:

9. Estimating upper and lower confidence limits at the conceptual stage
10. Using project characteristics as input variables
11. Estimating project durations

Each of these opportunities is discussed below.

6.3.1 Estimating for other Project Types

This report has focused on pavement rehabilitation projects using a set of 138 projects that had bid opening from April 26, 2016 to May 11, 2021. As noted above, Caltrans opened bids on 1,055 projects in that time period, and the most common type of projects was Safety Improvements, for which 252 projects had bid openings (see Appendix B). The researchers chose to focus first on pavement rehabilitation because they believed that this was a major project type that would have a reasonably high R^2 factor, which was accurate. Safety improvements might be more challenging because a wide variety of improvements might be characterized as "safety." Now that the data is available, similar studies could easily be performed for safety improvements and the other types of projects listed in Appendix B.

6.3.2 Estimating for other Phases

The current research project has focused on the preconstruction phases, Caltrans Phases 0 and 1. On pavement rehabilitation projects, these phases account for 10 percent of the total project cost, as illustrated in Figure 3. On some other types of projects, the preconstruction phases account for larger percentages of the total project cost. Research similar the current research could be performed on the remaining 90 percent of project costs. That would use different data selections and might need some supplemental data because the remaining phases, right of way and construction, continue through construction whereas preconstruction, as its name implies, ends when construction starts. The data selections used for right of way or construction would need to consist of projects that have completed construction and therefore have complete cost data.

6.3.3 & 6.3.9 Estimating upper and Lower Confidence Limits

The discussion above includes consideration of a probability density function, as shown in Figure 7 and discussed in Section 5.1.3, that would allow one to say that there is a 10 percent (or other percent) certainty that a given estimate is too high or too low. There is very little prior research on this type of estimate. The spreadsheet provided to the CTC uses the AASHTO confidence limits, which are based upon the expert opinions of panels of experts, but not upon a statistical analysis. AASHTO gives upper and lower confidence limits of +200 percent and -50 percent respectively. From Figure 7, though, it appears that the upper and lower confidence lines ought not to be percentages of the estimate, but rather some other relationship. The upper and lower confidence lines might be fixed numerical deviations rather than percentages, for instance.

The data provided to the researchers by Caltrans is sufficient to do a statistical analysis of cost estimate ranges at the bid stage, which would be a significant contribution to existing knowledge. To analyze the ranges at the conceptual, or Project Initiation Document, stage would require additional data, but the researchers believe that this data is available in the Caltrans CTIPS system, and the researchers would ask for the needed data if more analysis is requested.

6.3.4 Developing Portfolio-level Cost Estimates for Entire Biennial Tranches of Projects

This report has presented two methods of estimating the preconstruction effort needed on individual projects. The normal outcome is that some projects will require a higher-than-expected effort while others will require a lower-than-expected effort and that the highs will balance the lows. Averaged over an entire portfolio of projects, then, the methods used in this report should be a good predictor of preconstruction effort. As an alternative to using the upper and lower confidence limits, the CTC could consider the portfolio as a whole while leaving Caltrans to manage the highs and lows. It is a widely accepted management best practice for upper-level executives to consider and try to manage overall trends in an organization's data, while leaving lower-level managers to manage individual variations within the data.

The SHOPP is developed in two-year cycles. The CTC could use tools such as those developed in this report to estimate and evaluate the entire effort for the projects in each cycle, while ignoring the deviations of individual projects from their expected values and leaving Caltrans to manage those deviations.

6.3.5 Evaluating Efficiencies

SB1 of 2017 requires Caltrans to submit an annual report on its achievement of efficiencies. The requirement reads:

> Streets and Highways Code 2032.5. (d) The department shall implement efficiency measures with the goal to generate at least one hundred million dollars ($100,000,000) per year in savings to invest in maintenance and rehabilitation of the state highway system. These savings shall be reported to the commission. (California Legislature, 2017)

Caltrans defines efficiencies as "Caltrans will consider efficiencies that result in cost avoidance or a reduction in support or capital costs." (Caltrans, 2021). One of the challenges in evaluating efficiencies is the fact that organizations are complex systems. That is, each part in an organization affects other parts. A cost avoidance or reduction in cost in one part of an organization almost always causes a cost increase in some other part of the organization. In evaluating efficiencies, one needs to determine whether the organization as a whole is operating more efficiently rather than considering or counting only those portions of the organization that have become more efficient.

The tools developed in this report could be used to establish whether Caltrans as a whole is becoming more efficient. It could also evaluate whether individual projects with claimed efficiencies truly are more efficient than the preceding projects. Whether considering the whole organization or individual projects, the approach would be based upon recognized statistical methods.

6.3.6 Evaluating Alternative Project Development Processes

At various times, Caltrans has obtained legislative permission for the experimental use of alternative project development processes, including construction manager/general contractor, design-build, and design sequencing. There has also been consideration of processes that do not require legislation, such as lean scheduling. The legislature often tasks the CTC with evaluating the success of these measures. In the same manner as the evaluation of efficiencies described above, the tools developed in this report could be used to establish whether the alternative processes do, in fact, result in cost savings. As with the evaluation of efficiencies, the evaluation of alternative processes would be based upon recognized statistical methods. An evaluation of this nature has previously been performed on the use of lean scheduling on a drainage project in Los Angeles County (Caltrans District 7).

As noted below, the methods discussed in this report could be used to develop expected project durations. If this were done, those durations could be used to evaluate the impact of alternative processes on project durations.

6.3.7 Evaluating Optimum Levels of Contracting-out

Caltrans began contracting out some of its engineering work in the 1986–1987 Fiscal Year, and since then, private consultants have formed a part of the overall Caltrans project development workforce. Occasionally, there have been debates, both in Caltrans and in other transportation departments, about how to use consultants most efficiently. This report has already described the use of the study's tools to evaluate efficiencies, which could equally be well used to evaluate the efficiency of different levels of consultant involvement in the various phases and types of projects. As with the evaluation of other efficiencies, this evaluation would use recognized statistical methods.

6.3.8 Evaluating Multi-year "support to Capital"

Caltrans and other departments of transportation have, at times, used the "support to capital" ratio as a tool in evaluating their operations. That is the ratio of the costs of Phases 0, 1, 2, and 3 combined with the costs of Phases 4 and 9. This ratio has limited use because a change in the ratio does not necessarily mean that the organization is becoming more or less efficient. In principle, an increased design effort should produce a more competitive PS&E package with a resulting decrease in construction costs. Figure 3 illustrates the fact that, on average, construction costs are much greater than design costs. A small percentage of savings in construction costs can compensate for a relatively large increase in design costs.

The data that the researchers have obtained are sufficient to make a forty-year examination of the support to capital ratio in Caltrans. Such an examination could be useful in evaluating the effectiveness of major changes that Caltrans has made over this time period and might be useful in informing the CTC and Caltrans as they consider future changes.

6.3.10 Using Project Characteristics as Input Variables

This report has introduced the concept of project characteristics, and some examples of project characteristics are provided in Appendix A. Each of the characteristics listed in that appendix has been used by various researchers to predict project costs and could be considered to supplement the current research on Caltrans projects. Such supplemental research would require the gathering of data on the selected project characteristics. The research could be useful in improving the cost estimates. As discussed in the section on PYPSCAN, Caltrans has previously developed cost estimating models using project characteristics that include the weather zone, urban or rural location, environmental document type, and right-of-way information, including numbers of appraisals, acquisitions, utilities, relocation assistance cases, demolitions, railroad agreements, and condemnations.

6.3.11 Estimating Project Durations

The discussion to this point has focused almost entirely of estimating effort and costs. The methods used here could equally well be used for estimating the durations of projects and phases. Such an investigation would be based upon milestones and would require the collection from Caltrans of actual achieved milestone data. The researchers believe that this data does exist and could be obtained.

6.4 Conclusion

As has been discussed in this report, conceptual cost estimates are made at the earliest stage of projects and can be developed through one of three approaches: analogous estimating, parametric estimating, and artificial intelligence. While the analogous approach relies on the estimator's expert judgement, the other two approaches use computations from large sets of historic data on similar projects. This report has developed, and then compared, the three types of parametric estimates: additive exponential, linear, and multiplicative exponential. The immediate purpose of a conceptual cost estimate is to provide decision makers with an order of magnitude to assist them in deciding whether to start a feasibility study. The parametric and artificial intelligence approaches, because they have a statistical basis in historic costs, can also have other uses:

- Assisting executives to evaluate whether cost estimates at later stages in a project are reasonable. This includes determining both where a given estimate on one project is reasonable and whether the collective estimates of the entire portfolio is reasonable.

- Assisting executives to determine whether their organizational efficiency measures are succeeding. To demonstrate that an innovation has achieved an efficiency, one must know both the actual cost after the innovation and what the cost would have been without the innovation. While the cost after the innovation would be readily available from the actual costs incurred, it is more difficult to determine what the cost would have been without the innovation. The parametric and artificial intelligence approaches, based on historic costs, give an indication of what the pre-innovation cost would be. This is particularly useful when combined with an approach that accommodates unexplained variance, such as the chart in Figure 7 or the portfolio-level perspective.

- Incentivizing project teams to improve upon past performance. The cost estimating models provide teams with an expected cost using their historic delivery methods. Given that information, it is possible (and likely) that teams would rise to the challenge and work to complete their projects at lower-than-expected costs.

- Finally, this research is generalizable; although this research used data on pavement rehabilitation projects, the ANN and parametric estimating models could be applied to any set of projects on which input variables and actual costs are known. However, the coefficients of determination of those models might not be as high as the 0.85 and 0.80 observed in this study.

Appendix A: Examples of the Input Variables in Five Categories

Modified from Blampied (2018).

Project outputs	Bid items	Project characteristics	Economic	Geographic
building area floor area heavy manufacturing high or low building Lane miles of pavement multi-use or single-user building number of apartments number of elevators Number of wheelchair ramps number of floors single or double track railroad size of parking area specially engineered building stack-up building	cubic meters of asphalt concrete cubic meters of base cubic meters of concrete cubic meters of earthwork kilograms of steel / tons of reinforcing bars square meters of formwork	average daily traffic earthwork electrical work electro- mechanical work environmental document type geo-technical work structural work height of overburden height of piers interior decoration length length of structures number of piles number of spans plumbing work project duration Right of way needs site area sub-structure super-structure terrain type urban / rural location	fuel consumption Gross National Income number of bidders number of local bidders oil prices bidding and labor climate	climate earthquake impact geology region terrain

Appendix B: Caltrans Project Count by Program

This table shows the count of major projects with construction awards by Caltrans with bid openings from April 26, 2016 to May 11, 2021, by program.

Program	Number of bid openings
20.201.010 Safety Improvements	252
20.201.235 Roadside Safety Improvements	112
20.201.121 Pavement Rehabilitation	100
20.201.015 Collision Severity Reduction	79
20.201.131 Major Damage (Permanent Restoration)	74
20.201.110 Bridge Rehabilitation	55
20.201.315 Transportation Management Systems	50
20.201.335 Storm Water Mitigation	48
20.201.151 Drainage System Restoration	42
20.201.119 Capital Bridge Preventative Maintenance Program	32
20.201.361 ADA Curb Ramps	28
20.201.122 Pavement Preservation	26
20.400.100 Locally Generated Funds	22
20.201.113 Bridge Seismic Restoration	20
20.201.378 Pedestrian Infrastructure	20
20.201.112 Bridge Rail Replacement and Upgrade	18
20.400 Locally Funded State Highway Projects	17
20.201.310 Operational Improvements	17
20.201.120 Roadway Rehabilitation	13
20.201.321 Weigh Stations & Weigh-In-Motion Facilities	12
20.201.111 Bridge Scour Mitigation	11
20.201.150 Roadway Protective Betterments	9
20.201.170 Signs and Lighting Rehabilitation	9
20.075.600 Regional Improvement Program	8
20.201.322 Transportation Permit Requirements for Bridges	8
20.201.999 Other SHOPP	7
20.800.100 Other State Funds	5
20.400.232 Local Surface Transportation Program	4
20.201.352 Maintenance Facilities	4
20.800.200 Generic Non-STIP/SHOPP State Fund – Support	3
20.201.210 Highway Planting Restoration	3
20.722.000 Proposition 1B Funding of State Route 99 Improvements	2
20.201.130 Major Damage Restoration	2
20.201.351 Equipment Facilities	1
20.723.000 Trade Corridors Improvement Fund	1
20.400.200 Federal High Priority Projects/Demonstration Projects	1
20.025.700 New Programming Interregional Improvement Program	1
20.723.100 Trade Corridors Improvement Fund	1

Program	Number of bid openings
20.400.246 Local Highway Safety Improvement Program (Infrastructure)	1
20.400.330 Regional Surface Transportation Program	1
20.201.250 Safety Roadside Rest Area Restoration	1
20.400.210 Congestion Management and Air Quality Improvement	1

Bibliography

AASHTO. (2013). *Practical guide to cost estimating*. American Association of State Highway and Transportation Officials.

Allison, P. D. (1999). *Multiple Regression: A Primer.* Pine Forge Press, Inc.

Blampied, N. B. (2018). Parametric functions for conceptual and feasibility estimating in public highway project portfolios. University of California, Berkeley.

Blampied, N., Tommelein, I. D., Jayamanne, M. E., McKeever, B. (2017). *PRSM Review: Year 1 Report Part A: Review of PRSM Use at Caltrans*, Project Production Systems Laboratory, and Institute for Transportation Studies, University of California, Berkeley.

Caltrans (1992). *Project Management Control System (PMCS) User Manual*, California Department of Transportation.

Caltrans (2007). *Caltrans Project Management Handbook, Fifth Edition*. California Department of Transportation.

Caltrans (2017). *Chronology of Project Management in Caltrans*. http://www.dot.ca.gov/hq/projmgmt/chron_1980.htm visited on October 10, 2017.

Caltrans (2022). State Highway Operation and Protection Program, Fiscal Years 2022–23 through 2025–26. California Department of Transportation.

Caltrans (2021) C*altrans Efficiencies Report, 2020–21*. California Department of Transportation.

DGS (2017). Capital Outlay versus State Operations and Local Assistance – 6806. (Revised: 11/2107) *State Administrative Manual*, California Department of General Services.

Gardner, B. J., Gransberg, D. D., and Jeong, H. D. (2016). Reducing Data-Collection Efforts for Conceptual Cost Estimating at a Highway Agency, *Journal of Construction Engineering and Management*, 142(11), 04016057

Gransberg, D., Jeong, H. D., Craigie, E. K., Rueda-Benavides, J. A., and Shrestha, K. J. (2016). *Estimating Highway Preconstruction Services Costs. Volume 1: Guidebook*. National Academies of Sciences, Engineering, and Medicine.

LAO (1996). *LAO Analysis of the 1996–97 Budget Bill Transportation, Part I.* State of California Legislative Analyst's Office.

Larson, E. W. and Gray, C. F. (2011). *Project Management: The Managerial Process.* Fifth Edition, McGraw Hill.

McManus, J. F. (1981). Automating California's Capital Project Delivery Plan: ACSP / PYPSCAN Unit 1, California Department of Transportation.

Palisade Corporation. (2010). *Neural Tools, Version 5.7*, Palisade Corporation.

PMI. (2011). *Practice Standard for Project Estimating.* Project Management Institute.

PMI. (2017). A Guide to the Project Management Body of Knowledge (PMBOK® Guide), Fifth Edition. Project Management Institute.

Rao, G. N., Grobler, F., and Kim, S. (1993). Conceptual cost estimating: A hybrid neural-expert system approach. *Computing in Civil and Building Engineering* (pp. 423–430).

Siqueira, I., and Moselhi, O. (1998). An integrated cost estimating method for prefabricated construction, *Proceedings, Annual Conference – Canadian Society for Civil Engineering*.

TreePlan Software, Inc. (2016). TreePlan = The Decision Tree Add-in. http://treeplan.com/.

US Congress (2012). Public Law 112-141, *Moving Ahead for Progress in the 21st Century, MAP-21*, 126 Statutes 405, enacted July 6, 2012, amending 23 United States Code.

About the Authors

Nigel Blampied, PhD

Dr. Blampied's research and teaching focuses on project management in public transportation agencies. He teaches in the Master of Transportation Management program at San José State University and is a Research Associate at the Mineta Transportation Institute. His doctoral dissertation at the University of California, Berkeley discussed parametric cost estimating at the early conceptual phase of projects, the subject, in part, of this report, and used data from the Caltrans State Highway Operation and Protection Program (SHOPP), as does this report. While his earlier work considered pedestrian accessibility projects in the SHOPP, this report focuses on pavement rehabilitation projects.

Tariq Shehab, PhD

Dr. Shehab is an expert in developing cost estimating applications and the field of artificial intelligence. He has more than 25 years of experience in developing intelligent cost estimating applications, during which he developed systems for highway projects, utility pipe networks, and school buildings. His contributions included also the determination of key factors that attribute to the maintenance costs of major infrastructure facilities. He is the author and co-author of numerous articles published in many ASCE journals and other equally prestigious periodicals.

Elhami Nasr, PhD

Dr. Nasr is a Subject Matter Expert in Construction and Project Management. He built his career of over three decades with dual careers in academia and industry. He is a Professor of Construction Engineering Management at California State University, Long Beach. Previously, he served as Vice Provost, Interim Provost, and Vice President of Academic Affairs at Florida Polytechnic University, industry (California Department of Transportation (Caltrans)) and various national and international consulting engagements where he initiated, developed, and delivered hundreds of specialized and customized companywide project management training programs for large-scale professional organizations, in both private and government sectors.

Laxmi Sindhu Samudrala, MS

Laxmi Sindhu Samudrala received her Master's degree in Data Analytics from San José State University and is currently working for PayPal Inc. as a software engineer in the Product Security Team. Her work involves data engineering and machine learning, and her interests lie in machine learning and artificial intelligence. She helped assemble the data sets for this report, and she has also worked on projects requiring object detection and behavior classification for intelligent autonomous vehicle safety.