Transportation Research Part B 101 (2017) 295–305

Contents lists available at ScienceDirect

Transportation Research Part B

journal homepage: www.elsevier.com/locate/trb

Construction cost estimation: A parametric approach for better estimates of expected cost and variation

Omar Swei[a,*], Jeremy Gregory[b], Randolph Kirchain[c]

[a] U.S. Fulbright Scholar to Jordan, Department of Civil & Environmental Engineering, Material Systems Laboratory, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Building E19-695, Cambridge, MA 02139, USA
[b] Research Scientist, Department of Civil & Environmental Engineering, Material Systems Laboratory, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Building E19-695, Cambridge, MA 02139, USA
[c] Principal Research Scientist, Materials Processing Center, Material Systems Laboratory, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Building E19-695, Cambridge, MA 02139, USA

ARTICLE INFO

Article history:
Received 2 June 2016
Revised 24 April 2017
Accepted 25 April 2017

Keywords:
Cost estimation
Cost overruns
Heteroscedasticity
Least angle regression
Parametric estimation
Roadways

ABSTRACT

As project planners continue to move towards frameworks such as probabilistic life-cycle cost analysis to evaluate competing transportation investments, there is a need to enhance the current cost-estimation approaches that underlie these models to enable improved project selection. This paper presents an approach for cost estimation that combines a maximum likelihood estimator for data transformations with least angle regression for dimensionality reduction. The authors apply the proposed method for 15 different pavement bid items across five states in the United States. The results from the study demonstrate that the proposed approach frequently leads to consistent parametric estimates that address the structural bias and heteroscedasticity that plague the current cost-estimation procedures. Both of these aspects are particularly important for large-scale construction projects, where traditional methods tend to systematically underestimate expected construction costs and overestimate the associated variance.

1. Introduction

Over the last few decades, transportation planners have used several quantitatively-based methods to evaluate the cost-effectiveness of alternative infrastructure investments. For roadways, the focus of this research, decision-makers frequently use life-cycle cost analysis (LCCA), a framework that estimates the total cost of a project over its lifetime, to evaluate the merits of alternative designs (Walls and Smith, 1998). Implicitly, the outputs and accompanying conclusions of any LCCA depend upon assumed values for relevant input parameters. As a result, as practitioners continue to move towards probabilistic-based LCCA, there is an increasing need for techniques that provide more representative estimates of both the expected value and associated variation for input values (Walls and Smith, 1998; Tighe, 2001). This need is particularly true for inputs and life-cycle stages that contribute significantly to total expected life-cycle costs (LCC) and variation.

A previous study conducted by the authors of this paper demonstrated that not only are initial-cost estimates dominant contributors to total agency LCC, but they are also the major driver of variation across a range of contextual (e.g., traffic, climate) conditions (Swei et al., 2015). As a result, the goal of this research is to benchmark the fidelity of current approaches

* Corresponding author.
E-mail addresses: oaswei@mit.edu (O. Swei), jgregory@mit.edu (J. Gregory), kirchain@mit.edu (R. Kirchain).

http://dx.doi.org/10.1016/j.trb.2017.04.013
0191-2615/

Table 1
Statistical methods to estimate construction costs for buildings and bridges.

Study	Approach	Performance metrics
McCaffer et al. (1984)	Linear regression (OLS and alternative approaches)	COV of ratio of predicted over actual
Emsley et al. (2002)	Neural networks and linear regression	Coefficient of determination (R^2) and mean absolute percent error (MAPE)
Trost and Oberlender (2003)	Factor analysis and multiple linear regression (MLR)	Not applicable
Chan and Park (2005)	Principal component analysis and MLR	R^2, MAPE, tests for heteroscedasticity and serial correlation of the residuals
Lowe et al. (2006)	Stepwise MLR	R^2, MAPE, qualitatively assess bias and heteroscedasticity of the residuals
Ji et al. (2010)	Stepwise MLR	Adjusted R^2, MAPE
Kim and Hong (2012)	Case-based reasoning and MLR	MAPE and standard deviation of MAPE

to estimate initial-costs (both expectation and variation) and to identify a modeling approach that leads to better estimates of expected initial-costs and associated variation. The authors compare the performance of the proposed modeling approach and current approaches by applying them to pay items that contribute significantly to the total cost of roadway construction.

Because this study is limited in scope to roadway construction, the findings of this paper should be viewed as one contribution to the broader topic of transportation cost estimation, a domain in which several authors have documented the prevalence of inaccurate and imprecise early cost estimates (Eliasson and Fosgerau, 2013; Flyvbjerg et al., 2002; Molenaar, 2005). Such studies are primarily concerned with mega-infrastructure projects, tremendous financial investments that cost the public billions of dollars to construct (Flyvbjerg, 2003). Flyvbjerg et al. (2002), perhaps the most widely cited paper on this topic, demonstrate that early cost estimates across a range of transportation types (rail, fixed-link, and roadways) are both highly inaccurate, with average cost overruns approaching 30%, and imprecise, with a coefficient of variation (COV) for cost escalation greater than one. Flyvbjerg et al. (2002) subsequently focus their attention on the cause for inaccurate (i.e., biased) estimates and postulate that it is likely attributed to “strategic misrepresentation”, the deliberate misrepresentation of project costs, and possibly “appraisal optimism”, being overly optimistic about the outcomes of a project, on the part of project planners and promoters.

In contrast, this research is concerned with roadway construction, in which the cost to build and maintain a single facility is considerably less than the aforementioned mega-projects. For multiple reasons, the authors believe that this case study is of keen interest to the infrastructure community. Planning agencies charged with maintaining large roadway networks must typically operate within fixed budgets, suggesting that their primary concern is to be efficient with available resources (Markow, 1995). Furthermore, the individual contractors that are responsible for procuring roadway projects typically have extensive experience. As a result, cost escalation stemming from strategic misrepresentation and/or appraisal optimism should be mitigated within this context, which the results of previous studies indirectly suggest. Flyvbjerg et al. (2002), for example, find that roadway projects have on average the lowest cost escalation rate (20%) of the three infrastructure types analyzed. Other studies, such as that of Shrestha and Pradhananga (2010), show an even smaller average cost escalation for pavement construction of 3.5%, while the pavement cost dataset used by Williams (2003) has a cost escalation frequency of 60%. These results are not to say that cost escalation is not present in roadway construction, but rather the systemic bias in cost estimation that stems from economic (i.e., strategic misrepresentation) and psychological (i.e., appraisal optimism) factors for mega-infrastructure projects is limited within this context.

Therefore, it is of particular interest to develop novel methods for cost estimation in a domain where forecast inaccuracy and imprecision are attributed largely to the analytical tools used by planners. The idea that model misspecification can tremendously influence the fidelity of cost estimation tools is not new. As an example, Lowe et al. (2006) developed a series of regression models for large-scale building projects that, as the authors noted, systematically underestimated construction costs due to model specification. However, few studies (as will be further discussed) have searched for techniques to proactively deal with this issue. Consequently, the approaches discussed in this paper for cost estimation are potentially of great significance given (a) the increasing use of frameworks such as LCCA for pavement design and maintenance decisions and (b) the significance of highway expenditures in their aggregate, representing $165 billion in public spending in the United States in 2014 (Congressional Budget Office, 2015). An analytical approach that might offer a small improvement in model fidelity could represent large savings to the public sector.

2. Cost estimation approaches in the literature

Table 1 presents a set of cost estimation research for buildings and bridges, the two most common forms of infrastructure that researchers have studied in-depth. Existing research uses both cross-sectional and panel data to develop parametric and non-parametric models to predict project-level costs. Predictor variables across the studies listed include general site conditions, location, and the geometric layout of a structure. Parametric approaches make use of multiple linear