

# Application of Decision Support Technology for Conceptual Cost Estimation

Gary P. Moynihan<sup>1</sup> & Anh D. Chau<sup>1</sup>

<sup>1</sup> Department of Civil, Construction and Environmental Engineering, The University of Alabama, Tuscaloosa, AL 35487, USA

Correspondence: Gary P. Moynihan, Department of Civil, Construction and Environmental Engineering, The University of Alabama, Tuscaloosa, AL 35487 USA. E-mail: GMoynihan@eng.ua.edu

Received: July 29, 2019      Accepted: September 11, 2019      Online Published: September 13, 2019

doi:10.5539/emr.v8n2p30      URL: <https://doi.org/10.5539/emr.v8n2p30>

## Abstract

Conceptual cost estimates are often made at the beginning of the project when project scope is not yet well defined. Hence, predicting the conceptual costs on time, with high accuracy, presents a considerable challenge. One potential solution is to more effectively utilize historical data via integration with predictive analytical models. In this project, a decision support system was developed which predicts conceptual costs of construction projects and supports decision-making for long-term capital planning in public universities. The prototype system was developed based on historical data for roofing projects at the University of Alabama. We collected this historical data via a web-based data entry form subsystem. The developed system uses ridge regression models to train historical data. This system has a user-friendly interface and supports what-if analysis, allowing the user to see multiple scenarios of the estimation. The system also encompasses capabilities to forecast the effects of inflation on multi-year projects. Subsequent validation has demonstrated improvement in the resulting accuracy of the conceptual estimates.

**Keywords:** construction project, cost estimating, decision support systems, university

## 1. Introduction

Cost estimation is needed to support the decisions regarding funding, developing budget requests and evaluating resource requirements (GAO, 2009). Construction cost estimates are typically classified by their functions. As suggested by Hendrickson and Au (2008), as well as Schmid (2012), a typical project often encompasses three major estimation types: design estimates, bid estimates and control estimates. The design estimate is typically initiated by the owner and developed by designers. It is used in the scope definition, selection of design alternatives, and to keep the owners informed of forecast costs (Peurifoy & Oberlender, 2014). There are three sub-phases under the design estimate, including the order-of-magnitude estimate, conceptual estimate and detailed estimate (Hendrickson & Au, 2008). Once alternatives are approved, contractors conduct the bid estimate based on the subcontractor quotations, quantity takeoffs, and contractors' experience. A bid estimate is submitted to the owner, either for competitive bidding or negotiation, and consists of direct construction costs including field supervision, plus a markup to cover general overhead and profits. After selecting the successful contractor, the bid becomes the contractor's budget and the baseline for managing their resources through the control phase.

To initiate a project, public organizations generally conduct an owner's study including a technical feasibility study and an economic feasibility study of the proposed construction project (Peurifoy & Oberlender, 2014). A conceptual cost estimation (CCE) is a part of this economic feasibility study. It is used to forecast the approximate costs to evaluate the economic feasibility of proceeding with the project. CCE, also referred to as "pre-design cost estimates", "preliminary cost estimating" or "early cost estimating", are approximate estimates early in the project. Since CCE is generally prepared from the project scope, when there is little or no design yet completed, the accuracy will inevitably be lower than later detailed estimates. Some quantifiable definitions have been discussed in recent papers. For example, Asmar, Hanna and Whitehead (2011) identified CCE as an estimate prepared at 30% of the completed design. Rast and Peterson (1999) suggested the accuracy at CCE level should be around +30% to -15%. AACE International (2016) suggested that CCE is conducted at about 1% to 15% of project definition, with typical variations from -15% to -30% with low accuracy, and from +20% to

+50% with high accuracy.

CCE, however, is essential for the owner to decide whether to proceed with the project. Underestimation at this phase will result in overoptimistic projects being approved. On the other hand, overestimation will push the stakeholders to reject the project at the inception. Andersen, Samset and Welde (2008) indicate the roots of underestimation including underestimating risk, overestimating opportunities, inadequate estimation methods and skills, reliance on weak information, and strategic/deliberate scope creep and division of projects. This underestimation can lead to decreasing quality and cancellation of projects (Peurifoy & Oberlender, 2014). Thus, improving the accuracy of CCE is highly essential to the ultimate success of any project.

### *1.1 University Construction Environment*

A report on construction spending by the United State Census Bureau (2016) indicates that public educational construction accounted for more than 70 billion dollars in April 2016, which encompasses 24.1% of the total public construction value. Another report indicates that university construction totaled more than \$11.6 billion in 2014 (Abramson, 2015). University construction has many unique aspects compared to other construction applications. Public postsecondary construction has been decentralized since 1995 when most state's universities and colleges have administrated their own construction program with the administration provided by individual boards of trustees (BOT). Campus projects are funded from a variety of state and non-state sources, which dictate certain project aspects. Generally, all of the construction projects in a university need to conform to a predetermined campus master plan. Such campus master plans are based on assumptions about basic campus characteristics drawn from projections of a broader academic plan (Curuthers & Lazell, 1999). They outline building design and location, campus traffic patterns, utilities need and needed land improvements or acquisitions.

Further, the life of buildings on campus may span hundreds of years and require high maintenance and renovations throughout their lifecycle (Duke, 2013). In many cases, new construction is required to blend with older historic structures. Universities, especially in urban settings, are under pressure to find space for growing programs and changing university needs. Sometimes, this search leads off campus to historic structures, warehouses, and other buildings with a different original use than what is needed (BOM, 1991). Finally, a report by the Office of Program Policy Analysis and Government Accountability (2006) indicates that educational facilities are costlier to build than many other types of construction. The reasons regard the type of facilities built, higher land costs, and the stricter building code, regulations, and standards that educational facilities must meet.

### *1.2 Requirements From the University of Alabama Construction Administration*

The University of Alabama Construction Administration department (UACA) is a function under the Financial Affairs division of the University of Alabama. UACA provides management and support for construction projects on campus. UACA is responsible for implementation of the capital improvement projects and works closely with other departments in reviewing the project requirements based on needs and functions. When necessary, staff retains design professionals for the preparation of drawings and specifications based on the scope of work to be executed. One of the most critical missions of UACA is estimating and controlling construction costs. These often account for an extensive portion of the University's budget. To accomplish that mission, they focus on planning and strategizing the long-term budget and mitigating the inflation risk. For that reason, UACA requires a robust estimation process to provide accurate cost information for making choices among alternatives promptly.

In the university environment, CCE is required by the Board of Trustees (BOT). This CCE is the foundation for establishing project scope, detailed planning, and bidding of campus construction. In late 2016, UACA decided to support a research project focused on developing a cost estimating DSS prototype utilizing big data analytics to support decision-making.

## **2. Literature Review**

The construction industry roughly categorizes conceptual cost estimation methods into five generations. Raftery (1987) suggests three of these generations. The first generation which began in the late 1950's, and continued up to the late 1960's, used the square foot prices method. The second generation was developed during the middle of the 1970's, and was characterized by intensive use of regression analysis making use of increasing availability of fast computing facilities at lower prices. The third-generation appeared to have begun in the early 1980s. This generation was characterized by probabilistic estimates based on Monte Carlo techniques, as well as artificial intelligence and knowledge-based computer systems. Shin (2015) suggests that the fourth generation is an advanced area of the third generation, which focuses on machine learning techniques such as neural networks

and case-based reasoning. Finally, Shin (2015) also postulated that the booting approach, developed by Freund and Schapire (1997), will be the next generation of conceptual cost estimation.

Using data from advanced search facilities in Scopus (2017), the three most common predictive analytical models related to CCE in the construction industry are case-based reasoning (CBR), neural networks (NN), and regression analysis (RA). Among these three models, regression is frequently cited as the most applicable (Kutner, Nachtman & Neter, 2008).

### *2.1 Application of Decision Support Systems*

The conceptual estimation process at the University of Alabama is a time-consuming one involving manually analyzing data without a robust platform. This causes difficulties in collaborating among the analyst, estimator, and managers. Numerous changes in project scope, time and schedule are often involved to meet the project owner's requirements. The conceptual cost estimation process comprises numerous iterations, where the estimators must determine the effects of various scope revisions. In this case, what-if analysis can be used to explore significant changes in parameters. This what-if analysis capability is critical to supporting management, as it struggles to determine the most effective solution matching the owner's needs (such as Moynihan, Saxeena, & Fonseca, 2006).

A major concern of this strategy is balancing the adequate provision for project funding with the anticipated needs. The availability of large quantities of historical data can provide estimators with a better chance to improve accuracy and enable enhanced decision-making. Estimators also have to deal with a large amount of noisy and biased data. Poor analysis can provide imprecise and irrelevant information to managers, leading to confusion. Graham (2009) discussed how incomplete historical data can cause inefficiencies in the estimating process.

These difficulties infer a need for a decision support system to be developed for conceptual cost estimation. The components of a DSS can generally be classified into the following distinct parts: (1) the data management system; (2) the model management system; (3) the user interface; and (4) the user(s) (Moynihan et al., 2006). The model management system is particularly important for the success of a decision making system. The model is mathematical in nature, usually consisting of a management science/operations research (MS/OR) algorithm. These algorithms may also be referred to as predictive analytics. Larose and Larose (2015) define predictive analytics as "the process of extracting information from large data sets in order to make predictions and estimates about future outcomes." They conclude that predictive analytical applications have considerable benefits to a construction project. There are a number of predictive analytical approaches that have been applied to CCE. According to the literature review, most previous research addressed either predicting costs or predicting the accuracy of costs (i.e. quality assessment) in the conceptual phase. Predicting cost seems to be the preferred topic, with a large number of studies conducted (such as Kim, An, & Kang, 2004). Regression was the earliest predictive analytic applied, with examples dating to the 1970s. It is considered to be a well-defined mathematical approach with its rapidity of execution (Kim et al., 2004).

This DSS is especially valuable in the early phases of a project, when more than one alternative is evaluated. A DSS provides an automated platform with fast computation, which can reduce the time spent on estimation and integrating changes (Turban, Aronson, Liang, & Sharda, 2006). With the availability of predictive analytics models, the DSS is capable of dealing with a large volume of historical data to improve the accuracy, and to control noisy and biased data (Sauter, 2010).

### *2.2 Decision Support System Use in Conceptual Cost Estimation*

It appears that most of the previous research in CCE focuses on developing analytical models for CCE while ignoring the actual DSS design. Hegazy and Ayed (1998), as well as Cheng, Tsai and Hsieh (2009), are among the few studies which partially discuss the design of a DSS for CCE. Hegazy and Ayed (1998) developed a decision support system for the conceptual cost estimation of highway projects. The system was built using an MS Excel spreadsheet, which allows simple data management functions. The system uses GeneHunter, which is an add-in for MS Excel to build a NN model. The user interface is built through an MS Excel Macro. The database component in MS Excel was used to store data and allow users to conduct simple data manipulations, such as view, add, and delete data records. The model management subsystem can optimize the NN model, and estimate the costs of new projects based on historical data (Hegazy & Ayed, 1998). The user interface provides visual screens and buttons allowing users to conduct what-if analysis.

Cheng et al. (2009) built a decision support system which estimates project cost using fuzzy logic and neural network models. The system is suggested to be web-based and communicates with the user via the internet. The

system is built in a Windows environment. The database subsystem includes historical data, estimating results, model parameters, net price value, and users' input (Cheng et al., 2009). The model management subsystem supports the project management process, transfer projects, calculates price index and estimates project costs. The model can communicate directly to users for each function of the model via internet servers.

### 3. Objective, Scope and Methodology

The objective of this research is to develop a prototype decision support system which estimates costs and performs what if-analysis among alternatives at the conceptual phase of university construction projects. The requirements of the Construction Administration Department of the University of Alabama provide a framework for this research. The research developed a cost model which takes into consideration the effects of attendant variables regarding historical data of construction projects on campus. This model helps predict construction costs of the University in long-term capital planning, such as preparing the Annual Consolidated Capital Plan Report and Campus Master Plan.

#### 3.1 Initial Investigation

This project was started in August 2016, with the original plan to build a series of modules to support conceptual cost estimation. Meetings with the executive manager of UACA were held to identify project requirements. Data investigation was conducted on the Estus system (Division of Financial Affairs, 2009), Alabama Building Commission (2016) website, and Alabama Commission on Higher Education (2018) database to access potential available data for the DSS. After conducting data feasibility study, four project types were selected for building system modules including roofing, lighting, plumbing and electrical projects. A web-based data entry form was created. On September 1, 2017, the project team was redirected to focus on only roofing projects as the basis for our database and subsequent DSS.

Data from Estus was determined to be the primary source for this analysis. Due to the large volume and high variety of data on Estus, further clarification was needed. It was decided to use Qualtrics software to build a web-based data entry form. Qualtrics is a web-based platform, which provides a convenient way to create an online research survey and report results (Qualtrics, 2018). It allows us to send the entry form via email or share the form link to the participants. The original plan was to contact architects and contractors who worked on UA projects and have them complete the form. However, UACA found difficulties in contacting these past architects and general contractors. Therefore, data was collected from the Estus system with clarification from UACA project managers via Qualtrics.

#### 3.2 Methodology for Developing the DSS

As suggested by Turban et al. (2006), the framework used to develop this system has four main phases: (1) planning, (2) analysis, (3) design, and (4) implementation. Planning maps out the project's overall structure. Analysis determines and understands details of what the system must do to address the problem or needs. The design phase focuses on configuring and structuring the new components. Implementation includes programming, testing and putting the system into operation.

A hybrid method of system development methodology was used, which combines the system development lifecycle (SDLC) approach (Satzinger, Jackson & Burd, 2015) and prototyping approach (Turban et al., 2006). As explained by Satzinger et al. (2015), the system development life cycle is a structured framework that identifies all of the activities required to research, build, deploy, and often maintain an information system. The SDLC approach requires that most of the tasks in one phase are completed before the work proceeds to the next phase. Due to difficulties in identifying and analyzing the problem at the beginning of the project, the SDLC approach was applied during the planning and analysis phases to establish the solid foundation for later phases – design and implementation. Project scope and schedule were verified with UACA management at the end of the analysis, before starting the design phase, to avoid scope creep.

Turban et al. (2006) note that the prototyping approach involves performing the analysis, design and implementation phases concurrently and repeatedly. In this project, prototyping is applied during the design and implementation phases of the project life cycle. This approach builds system in a series of short steps with immediate feedback from users to ensure that the development is proceeding correctly (Turban et al., 2006). An advantage of the prototyping approach in this project is to ensure the satisfaction of the managers (users) with the system interface and system functions. This feedback helped us understand the manager's requirements and reduce the developing time by focusing on important needs.

#### 3.3 Selection of the Predictive Analytical Model

Kutner et al. (2008) discuss regression analysis as a common methodology utilizing the relation between two or

multiple quantitative variables to predict the response variables. Regression discovers which predictors are important, estimates the impact of changing a predictor on the value of response and estimates the future value of response (Weisberg, 2005). Regression analysis using ordinary least squares (OLS) is among the simplest and most widely used methods in prediction (such as Welc & Esquerdo, 2018). Weisberg (2005) postulates that most other modern regression methods are elaborations and modifications of the OLS regression.

To apply OLS, predictor variables are required to be linearly independent (Kutner et al., 2008). When the predictors are correlated among themselves, this phenomenon is referred to as intercorrelation or multicollinearity (Kutner et al., 2008). During the investigation of this research, it was observed that data of predictor variables are potentially correlated by its nature. For example, the data regarding the thickness of roof material covering appeared correlated with the thickness of the insulation layer. Subsequent analysis confirmed this hypothesis. Multicollinearity among variables can impact the result of the prediction when using traditional OLS regression because the method typically has low bias but large prediction variance (Friendly, 2013).

Ridge regression was originally developed by Hoerl (1962), as one of several methods that have been introduced to remedy multicollinearity problems by modifying the method of least squares, allowing biased estimators of the regression coefficients (Kutner et al., 2008). By allowing a small bias, the estimator can improve the precision and probability of being close to the true parameter  $\beta$ . The bias error is taken as the difference between the expected prediction of our model and the correct value which we are trying to predict. Variance error is taken as the variability of a model prediction for a given data point  $x_0$ . Irreducible error comes from noise in data set and cannot fundamentally be reduced as a generalization error. Hastie, Tibshirani and Friedman (2009) discussed the more complex the model  $f(x)$ , the lower the bias will be; however, coefficient estimates will suffer from high variance as more terms are included in the model. Theoretically, if the true model is known and there is infinite data to fit it, both bias and variance can be reduced to 0. With an imperfect model and finite data in the real world, two ideas must be balanced: minimizing the bias (overfitting) and minimizing the variance (underfitting). Thus, at the cost of bias, ridge regression will reduce the variance and the MSE.

#### 4. DSS Functionality, Design and Use

As depicted in Figure 1, the Conceptual Cost Estimation Decision Support System (CCE-DSS) has two primary components: (1) the Data Collection Component and (2) the Data Analysis Component. The Data Analysis Component is comprised of three subsystems: (1) the Access Control Subsystem, (2) the Historical Data Subsystem, and (3) the Estimation subsystem. The Historical Data Subsystem helps control and manage the system database. Users can add, delete, and modify system records on the database. The Control subsystem is responsible for providing different security levels for project participants (i.e. administrators, project managers, architects, contractors). Further, it helps users update subordinate system libraries (i.e. the categorical library and inflation cost index library). It also provides a summary dashboard with buttons to redirect the user to the other screens. The Estimation subsystem is the main part of the decision support system and is responsible for processing estimation calculations. The Estimation subsystem builds fitted models via ridge regression analysis. Users input specific data into the Estimation subsystem. This input data is incorporated into the fitted models to calculate the results and display to users to review.

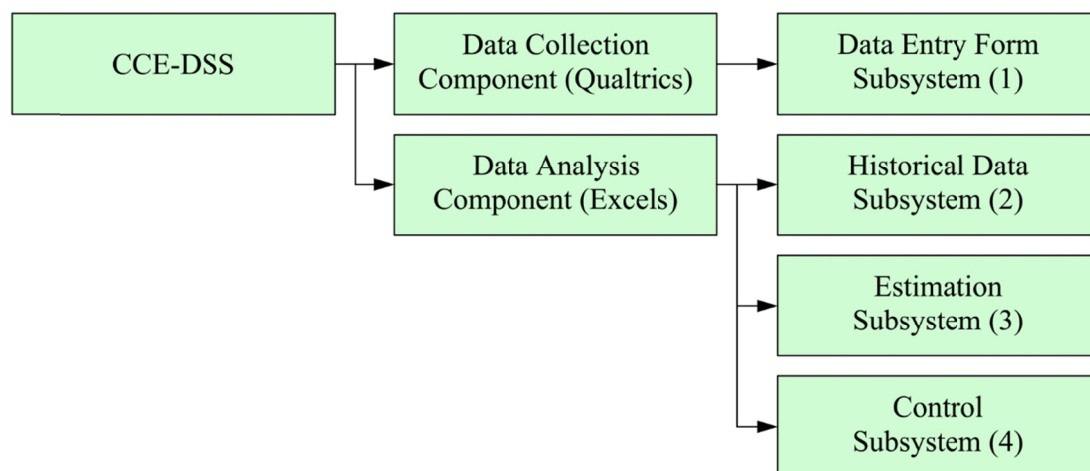


Figure 1. System architecture

#### 4.1 Software and Hardware Considerations

UACA utilizes a series of networked MS-Windows based personal computers. Consistent with the available client resources, the recommended specifications for using CCE-DSS include:

- Operating system: Window 10
- Processor: Dual-core x 86 CPU running at 2 GHz.
- RAM: 4 GB.
- MS Excel 2017

#### 4.2 Input Specifications

This system uses input from two main sources: (1) historical data, and (2) scoring data. In this project, scoring data refers to new data which is applied to the fitted model and generates a fitted value output (SAS Institute, 2014). Figure 2 presents the Control subsystem main dashboard. As shown in the figure, all functional buttons are located on the right side of the dashboard. Users can open the “Settings” dashboard to designate default settings (such as change dashboard size to full screen). Users can save what they have done in the system and close it by clicking on the “Save” and “Close” buttons, respectively. When the system users (such as project participants) opens the decision support system file, the Control subsystem main dashboard is displayed. The associated user ID and password are provided by system administrator. On the Control dashboard, users can also select the system “Help” tab and “About” tab. The “Help” tab provides some quick references which guide users how to utilize the system. It also contains a link to an online user manual which allows the user to view the full system references on a web browser. The “About” tab provides users with some basic information about the system (such as system version, system owner). Only the administrator can open the “Library” tab which allows the administrator to update the categorical library.

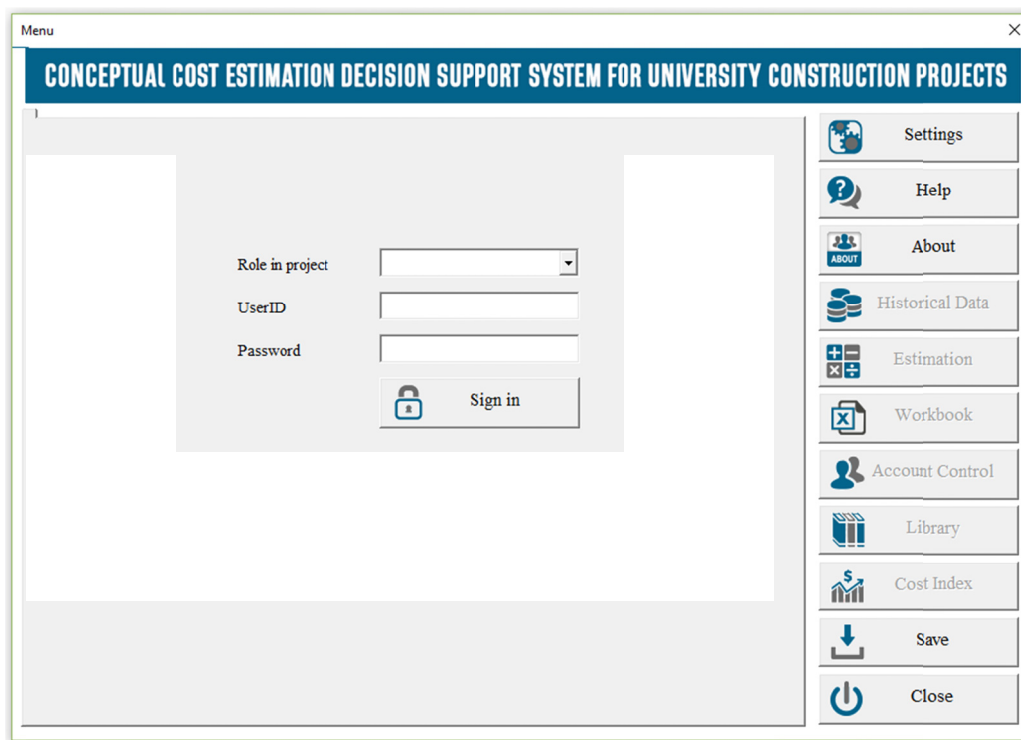


Figure 2. Control system main dashboard

Figure 3 presents inputs to the Architect tab of the Historical Data subsystem. This tab permits entry and editing regarding data on the general design of the building. The tabs for Contractor and Project Manager data have analogous configurations. Data from the Project Management tab, Architect tab, and Contractor tab are referred to as “General” data in this project. Specific roof data may be entered and managed via the Roof tab as depicted in Figure 4.

Users open the Historical Data subsystem from the “Historical Data” button on the Control dashboard (Figure 2).

The system retrieves data from the main system database and displays them in the dashboard. To edit the database, users must change from the view mode to edit mode. The view mode is set as the default, which prevents inadvertent changes by the users.

The system creates temporary data from the main system database. Categorical variables are coded as dummy variables with binary values (0 and 1). Roof data are determined to have some interactions, so interaction values are calculated. Subsequently, roof data are concatenated by roof area. Inflation cost index data is used to build a fitted model by OLS regression. This fitted model is used to adjust prepared data value with inflation consideration to the current time when users run the system. All data are saved in another worksheet in MS Excel and are updated regularly each time users call the “Update” function

Figure 3. Architect tab of the historical data subsystem

Figure 4. Area sub-tab within the roof tab of the historical data subsystem

Similar to the Historical Data subsystem, users open the Estimation subsystem via the “Estimation” button on the Control dashboard (Figure 2). Users input scoring data and execute estimations based on the values selected. The “Re-Estimation” function then allows users to modify the scoring values and redo the estimation until they are satisfied. Based on the variable selections on the “Setting” dashboard of the Historical Data subsystem, the Estimation subsystem will enable users to input scoring data in specific fields on the dashboard.

#### *4.3 Processing Specifications*

After the historical data are well prepared (such as converting categorical variables, or calculating inflation), the system is ready to conduct ridge regression. Scoring data, input via the Estimation subsystem, are stored temporarily. This may cause difficulties in reading the data; however, it helps MS Excel save memory during the calculation. Scoring data are used to build the scoring matrices for each roof area and one matrix for general data. When the “Estimation” button is selected, the VBA sub “GeneralResponse” and VBA sub “RoofResponse” are called to calculate the responses.

The sub “RoofResponse” uses historical data to create a predictor matrix for each roof area based on the scoring data that users select for the roof. One response matrix (containing data of roof area cost (\$) input) is used for the entire roof area. “RoofResponse” also calculates the coefficient for each roof datum. Matrices are standardized via the VBA function “MatrixStandardize”. Since VBA does not have matrix functions such as adding or subtracting matrices, some additional matrix functions are built to support calculations. The sub “RoofResponse” iterates through each roof area and creates ridge regression models for these areas. These coefficients are multiplied with scoring matrices of each roof area to obtain the predicted values. Predicted values are displayed to the “Result” tab of the Estimation subsystem dashboard. Similarly, VBA sub “GeneralResponse” is used to calculate the response values for general data.

Roof areas may use different models depending on the quantity of predictors that users have provided in these areas. For ridge regression, the system standardizes all X and Y matrices, and fits them with ridge regression models to get standardized coefficients. Default lambdas are first used in this fitting process. These coefficients are then unstandardized and formulized default fitted models are used for each roof area. Data are partitioned to conduct cross-validation. This cross-validation uses default fitted models to calculate overall errors. An optimal lambda value is obtained by using the automatic trial and error (Goal Seek in MS Excel) to find the minimum value of the overall cross-validation error. This lambda is used to recalculate the final fitted models. Roof scoring data are plugged into the final fitted model to obtain the material costs (response Ys) of each roof area. These material costs are added up to the total material cost and merged with general scoring data that users input. This merged data becomes the new general scoring data. The system redoes the process, similar to what it has done with the roof data, to obtain final fitted models for the general data.

To evaluate the result of the estimation, mean absolute percentage of errors of regression are calculated, as suggested by Myttenaerea, Goldena, Grandb and Rossic (2017). The VBA function “MeanAbsPercentError” is used in sub “GeneralResponse” and sub “RoofResponse” to calculate the percentage mean of error. In the event that there is a 0 value in the response (such as demolition days, the quantity of subcontractors, and the quantity of change orders), the denominator in Equation 5 would be 0 which would cause VBA function errors. Therefore, another equation is used to calculate the percentage of error for these cases. The symmetric mean absolute percentage error provided is used, as suggested by Kim and Kim (2006).

#### *4.4 Output Specifications*

The Cost Results tab (see Figure 5) displays the results of the estimation regarding material costs, construction cost, and total cost. Percentage of error values of the regression analysis are calculated based on comparing the fitted values and actual values of the response. These guide the users to accept or reject this analysis. The cost result is used to plot summary charts which are displayed to the users. If users reject the results, they can click on the “Re-Estimation” button, return to the scoring tabs, change scoring inputs, and then conduct another estimation. Users can redo this process until they are satisfied with the result. This what-if analysis is intended to provide users with different scenarios of the estimation. By reviewing multiple scenarios, users compare cost values to select the project alternative to proceed with.

Once users accept the estimation result, they can subsequently open the Inflation dashboards to adjust the result with consideration of inflation. Users can select several future points via the Time combo box, then select the “Advance” button to view the prediction chart, change the sources of inflation cost index, and display the new results.



## 5. Verification and Validation

Verification ensures that the product is built correctly and meets the designed specifications (Bourque & Fairly, 2014). As suggested by the International Software Testing Qualifications Board (2018), the developed system was verified via three testing methods: component testing, integration testing and system testing. Component testing is often called unit testing which tests individual hardware or software components (ISTQB, 2018). Bourque and Fairley (2014) discuss that component testing verifies the individual testable system functions which are often done by system programmers. To improve the quality of the system, component testing was performed during the system programming phase. Testing of each VBA function and VBA form component (such as buttons, text boxes, combo boxes) was executed immediately after each of them were created. Any discovered error was completely resolved before proceeding to the next components.

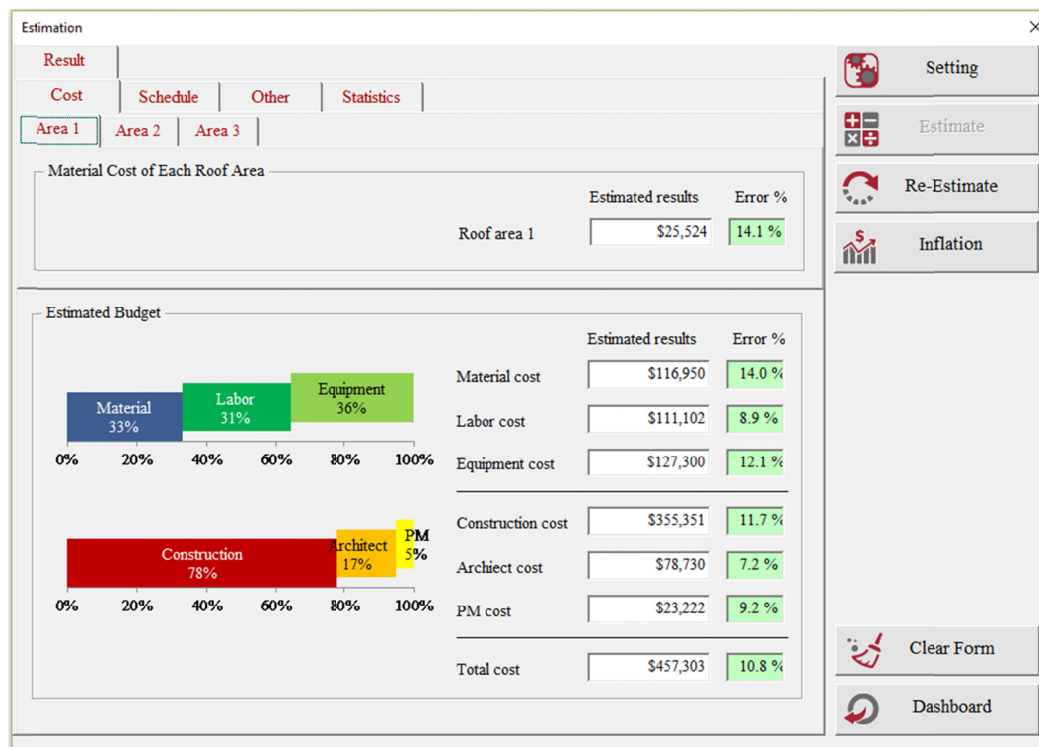


Figure 6. Cost result tab of the estimation subsystem

Integration testing was used to verify complex VBA subs and groups of VBA subs. The integration testing combines individual units and inspects the interactions between integrated components (ISTQB, 2018). A VBA sub performs actions rather than returns a final value. The results of the integration testing of the VBA form components (such as buttons, textboxes, and combo boxes). These components are integrated with multiple VBA subs and related to various VBA forms. Each VBA sub is tested by three processes: (1) visually inspecting the results on the MS Excel spreadsheet, (2) debugging each line of code and comparing local values with manual calculations, and (3) inspecting data on the temporary MS Excel spreadsheet.

System testing evaluates the behavior of the entire system with regards to non-functional system requirements such as security, speed, and reliability (Bourque & Fairley, 2014). Development of the system focused on improving the speed of processing. As discussed previously, the "Update" button was created to help improve the system speed via creating temporary data. This is one of the important improvements that was suggested during the initial system validation with UACA. The speed of the developed system was tested on the standard computer hardware configuration identified previously. The focus was on testing system functions which often take considerable time to execute such as opening the system file, updating historical data, and estimation. The reliability was tested by executing the MS Excel system file on three different MS Windows-based computers since the MS Windows operating system is being used at UACA. The overall result indicates that the system can run correctly on various MS Windows computer configurations.

Validation ensures that the developed system fulfills its intended purpose (Bourque & Fairley, 2014). The prototype system was validated via usability tests (Rubin, Chisnell & Spool, 2008) and test cases ((Chen, Qin,

Koo & Mishra, 2013). The usability testing evaluates the degree of effectiveness, efficiency, and satisfaction when the client uses the developed system in their specified business activities (ISTQB, 2018). Two usability tests were conducted on Apr 25, 2018, and Sep 6, 2018, with our client. The first test focused on improving the system functionality to better match the client's requirements. The second test focused on finalizing the system and developing a transition plan for the UACA to apply this system. The client accepted the developed system with some minor suggestions.

ISTQB (2018) defines a test case as a set of preconditions, inputs, actions expected results and postconditions, developed based on test conditions. The prediction accuracy of the system was tested with actual project values. A new construction project, the Kappa Kappa Gamma Chapter House, was used as the first case. A renovation project, for the Hardaway Hall roof replacement, was used as the second test case. The DSS reduced the prediction errors of the total project cost from 40.5% to 16.4% in these test cases. This is 24.1% average improvement on the accuracy of estimation. This system can also provide UACA detailed construction component costs such as labor, material, equipment costs. This indicates that the system can help improve the current conceptual cost estimation process at UACA.

## 6. Conclusions and Further Research

Using historical data to make actionable decisions has become an accepted approach across multiple industry sectors. Recently, the development of data storage technology has allowed construction organizations and contractor firms to collect more historical data than ever. When data are stored in a single system, such as MS SharePoint, it enables the system's user to extract and analyze for essential information. Accurate information, such as future project costs, can significantly assist managers in important decisions such as long-term capital planning.

This research provides the Construction Administration department, of the University of Alabama, with an advanced decision support system to improve accuracy in conceptual cost estimation. With the help of project managers at UACA, historical data was collected for on-campus roofing projects. The resulting prototype system was programmed in MS Excel and Visual Basic Application. The system architecture encompasses three subsystems: (1) the Historical Data subsystem, (2) the Estimation subsystem, and (3) the Control subsystem. In this research, the ridge regression technique was used. It allows analysis of multiple regression data that suffer from multicollinearity. The Engineering News-Record (2018) index values were utilized to adjust the output of the system to future cost values. The system was validated with two test cases. The test results demonstrate improved accuracy for conceptual cost estimating. This will help the University improve the decision-making process in long-term capital planning.

Future research directions envisioned for the prototype system include migrating the system to other universities that want to improve their conceptual cost estimation and long-term capital planning. This will also help the system acquire more historical data, and thus the predictive accuracy. Modification of the security levels would allow multiple universities to use and share the system. Second, the scope of this research can be expanded to other building components, such as lighting. Third, deployment of this MS Excel system into a web application would enable the users to access the system from multiple locations, further supporting the multi-university initiative. Finally, application of the system beyond the university environment, to the broader construction industry, should be investigated.

## Acknowledgments

The authors would like to thank Mr. Tim Leopard, and the staff of the University of Alabama's Construction Administration Department for their support throughout this project.

## References

- AACE International. (2016). *Cost estimate classification system - as applied in engineering, procurement, and construction for the process industries*. Morgantown, WV: AACE International. Retrieved from <http://www.austintexas.gov/edims/document.cfm?id=280770>
- Abramson, P. (2015). Renewed confidence, continued growth. Retrieved from <https://webcpm.com/Research/2015/02/College-Construction-Report.aspx>
- Alabama Building Commission. (2016). Introduction. Retrieved from <http://www.bc.state.al.us/>
- Alabama Commission on Higher Education. (2018). Home page. Retrieved from <http://ache.state.al.us/>
- Andersen, B., Samset, K., & Welde, M. (2008). Low estimates - high stakes: underestimation of costs at the front-end of projects. *International Journal of Managing Projects in Business*, 9(1), 171-193.

- <https://doi.org/10.1108/IJMPB-01-2015-0008>
- Asmar, M. E., Hanna, A. S., & Whited, G. C. (2011). New approach to developing conceptual cost estimates for highway projects. *Journal of Construction Engineering and Management*, 137(11), 942-949. [https://doi.org/10.1061/\(ASCE\)CO.1943-7862.0000355](https://doi.org/10.1061/(ASCE)CO.1943-7862.0000355)
- Bourque, P., & Fairley, R. E. (2014). *Guide to the software engineering body of knowledge*. (3rd ed.). Piscataway, N.J: IEEE. Retrieved from <https://www.computer.org/education/bodies-of-knowledge/software-engineering>
- Building Operating Management Staff Report. (1991). Tufts University research center show pro's and con's of adaptive reuse, *Building Operation Management*, 50(9), 60-64.
- Caruthers, J. K., & Layzell, D. T. (1999). Campus Master Planning and Capital Budgeting. *New Directions for Higher Education*, 107(7), 73-81. <https://doi.org/10.1002/he.10707>
- Chen, M., Qin, X., Koo, H. M., & Mishra, P. (2013). *System-level validation: High-level modeling and directed test generation techniques*. New York: Springer Science+Business Media. <https://doi.org/10.1007/978-1-4614-1359-2>
- Cheng, M. Y., Tsai, H. C., & Hsieh, W. S. (2009). Web-based conceptual cost estimates for construction projects using evolutionary fuzzy neural inference model. *Automation in Construction*, 18(2), 164-172. <https://doi.org/10.1016/j.autcon.2008.07.001>
- Division of Financial Affairs. (2009). University of Alabama. Estus system. Retrieved from <https://estus.fa.ua.edu/Pages/default.aspx> Accessed 30.01.18
- Duke, W. M. (2013). *Building information modeling: How it can benefit a modern construction project in a university setting*. (Master's Thesis. The University of Alabama, Tuscaloosa, AL).
- Engineering News-Record. (2018). Using ENR indexes. Retrieved from <https://www.enr.com/economics/faq>
- Freund, Y., & Schapire, R. E. (1997). A decision-theoretic generalization of on-line learning and an application to boosting. *Journal of Computer and System Sciences*, 55(4), 119-139. <https://doi.org/10.1006/jcss.1997.1504>
- Friendly, M. (2013). The generalized ridge trace plot: Visualizing bias and precision. *Journal of Computational and Graphical Statistics*, 22(7), 50-68. <https://doi.org/10.1080/10618600.2012.681237>
- Government Accountability Office (2009). *GAO cost estimating and assessment guide: Best practices for developing and managing capital program costs, applied research and methods*. United States Government Accountability Office, Washington, D.C. Retrieved from <https://www.gao.gov/new.items/d093sp.pdf>
- Graham, J. W. (2009). Missing data analysis: making it work in the real world." *Annual Review of Psychology*, 60(1), 549-576. <https://doi.org/10.1146/annurev.psych.58.110405.085530>
- Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The elements of statistical learning: Data mining inference, and prediction*. (2nd ed.). New York: Springer. Retrieved from <https://web.stanford.edu/~hastie/Papers/ESLII.pdf>
- Hegazy, T., & Ayed, A. (1998). Neural network model for parametric cost estimation of highway projects. *Journal of Construction Engineering & Management*, 124(3), 210-218. [https://doi.org/10.1061/\(ASCE\)0733-9364\(1998\)124:3\(210\)](https://doi.org/10.1061/(ASCE)0733-9364(1998)124:3(210))
- Hendrickson, C., & Au, T. (2008). *Project management for construction: Fundamental concepts for owners, engineers, architects and builders*, Pittsburgh: Carnegie Mellon University. Retrieved from <https://www.cmu.edu/cee/projects/PMbook/>
- Hoerl, A. E. (1962). Application of ridge analysis to regression problems. *Chemical Engineering Progress*, 58, 54-59.
- International Software Testing Qualifications Board. (2018) ISTQB glossary. Retrieved from <http://glossary.istqb.org/>
- Kim, G. H., An, S. H., & Kang, K. I. (2004). Comparison of construction cost estimating models based on regression analysis, neural networks, and case-based reasoning. *Building and Environment*, 39(10), 1235-1242. <https://doi.org/10.1016/j.buildenv.2004.02.013>
- Kim, S., & Kim, H. (2016). A new metric of absolute percentage error for intermittent demand forecasts. *International Journal of Forecasting*, 32(3), 669-679. <https://doi.org/10.1016/j.ijforecast.2015.12.003>
- Kutner, M. H., Nachtsheim, C. J., & Neter, J. (2008). *Applied linear regression models*. (4th ed.). New York:

- McGraw-Hill Education. Retrieved from <https://www.mheducation.com/highered/product/0073014664.html>
- Larose, D. T. & Larose, C. D. (2015). *Data mining and predictive analytics*. (2nd ed.). Hoboken, N.J.: John Wiley & Sons. Retrieved from <https://www.oreilly.com/library/view/data-mining-and/9781118868706/>
- Moynihan, G. P., Saxeena, P., & Fonseca, D. J. (2006). Development of a decision support system for procurement operations, *International Journal of Logistics Systems and Management*, 2(1), 1-18. <https://doi.org/10.1504/IJLSM.2006.008214>
- Myttenaerea, A. D., Golden, B., Grandb, B. L., & Rossic, F. (2017). Mean absolute percentage error for regression models. *Neurocomputing*, 192(5), 38-48. <https://doi.org/10.1016/j.neucom.2015.12.114>
- Office of Program Policy Analysis and Government Accountability. (2006). *Higher education facility construction costs are reasonable: Some improvements could maximize use of campus classroom space*. Tallahassee: The Florida Legislature. Retrieved from <http://www.oppaga.state.fl.us/MonitorDocs/Reports/pdf/0631rpt.pdf>
- Peurifoy, R., & Oberlender, G. (2014). *Estimating construction costs*. (6th ed.). New York: McGraw-Hill Education. Retrieved from <https://www.mheducation.com/highered/product/0073398012.html>
- Qualtrics. (2018). Home page. Retrieved from <https://www.qualtrics.com/>
- Raftery, J. (1987). The state of cost/modelling in the UK construction industry: A multi criteria approach. In P. S. Brandon (Ed.), *Building Cost Modeling and Computers: Transactions of the Building Cost Research Conference on Building Cost Modelling and Computers* (pp. 49-71), Abingdon, UK: E & F B Spon.
- Rast, J., & Peterson, M. K. (1999). Parametric cost estimating for environmental remediation projects. In *Proceedings of the American Association of Cost Engineers*, (pp. ENV .09.1-ENV.09.11), Denver: AACE.
- Rubin, J., Chisnell, D., & Spool, J. (2008). *Handbook of usability testing: How to plan, design, and conduct effective tests*. (2nd ed.). Indianapolis: Wiley Publishing. Retrieved from <http://ccftp.scu.edu.cn:8090/Download/efa2417b-08ba-438a-b814-92db3dde0eb6.pdf>
- SAS. (2014). *SAS/STAT® 13.2 user's guide: High-performance procedures*. Cary, NC: SAS Institute. Retrieved from <https://support.sas.com/documentation/cdl/en/stathpug/67524/PDF/default/stathpug.pdf>
- Satzinger, J. W., Jackson, R. B., & Burd, S. D. (2015). *Systems analysis and design in a changing world* (7th ed.). Boston: Cengage Learning. Retrieved from <https://www.cengage.com/c/systems-analysis-and-design-in-a-changing-world-7e-satzinger/>
- Sauter, V. L. (2010). *Decision support systems for business intelligence*. (2nd ed.). Hoboken, NJ: John Wiley & Sons. Retrieved from <https://www.wiley.com/en-am/Decision+Support+Systems+for+Business+Intelligence%2C+2nd+Edition-p-9781118627235>
- Schmid, K. F. (2012). *Construction estimating: A step-by-step guide to a successful estimate*. New York: Momentum Press. Retrieved from <http://ebooks.momentumpress.net/Books/9781606502938>
- Scopus. (2017). Advanced search. Retrieved from <https://www.scopus.com>
- Shin, Y. (2015). Application of boosting regression trees to preliminary cost estimation in building construction projects. *Computational Intelligence and Neuroscience*, 15(2), 1-9. <https://doi.org/10.1155/2015/149702>
- Turban, E., Aronson, J. E., Liang, T. P., & Sharda, R. (2006). *Decision support and business intelligence systems*. (8th ed.). Upper Saddle River, NJ: Prentice Hall.
- United States Census Bureau. (2016). Value of construction put in place at a glance, construction spending. Retrieved from <http://www.census.gov/constructionspending>
- Weisberg, S. (2005). *Applied linear regression*. (3rd ed.). Hoboken, NJ: Wiley.
- Welc, J., & Esquerdo, P. (2018). *Applied regression analysis for business: tools, traps and applications*, Aarhus, Denmark: Springer International Publishing. Retrieved from <https://link.springer.com/book/10.1007/978-3-319-71156-0>

## Copyrights

Copyright for this article is retained by the author(s), with first publication rights granted to the journal.

This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/4.0/>).