DEVELOPMENT OF AN ESTIMATING TOOL FOR TRANSPORTATION PROJECT

DESIGN EFFORT

Submission Date: 11/8/12

Word Count: 4200

Authors:
Trefor P. Williams, Ph.D., P.E. (Corresponding Author)
Center for Advanced Infrastructure and Transportation
Rutgers University
100 Brett Road
Piscataway, NJ 08854-8014
Telephone number: 848-445-2880
Email: tpw@rci.rutgers.edu
Fax: 732-445-3325

Neville Parker, Ph.D., P.E.
Herbert G. Kayser Professor of Civil Engineering
Director, CUNY Institute for Transportation Systems
The City College, City University of New York
Convent Avenue & 138th Street
New York, NY10031
Telephone number: 212-650-8050/8054(direct); Fax:
Email: parker@utrc2.org
Fax: 212-659-8374

James Klotz, P.E.
Civil Engineer II
New York State Department of Transportation
NYS DOT Consultant Management Bureau POD 22
50 Wolf Rd
Albany, NY 12232
Telephone number: 518-485-5429
Email: jklotz@dot.state.ny.us
Fax: 518-457-5933

TRB 2013 Annual Meeting
Paper revised from original submittal.
ABSTRACT
The development of an estimating tool to assist New York State Department of Transportation (NYSDOT) managers is described. The tool was developed using Microsoft Access. It enables managers to input project characteristics and then search a data base of historical projects. The estimating tool returns the staffing levels for similar completed projects. The system was also developed to allow newly completed projects to be added to the data base and to allow for the monitoring of design hours expended for projects that are in design. Regression analysis was also explored as a means of predicting total project design hours. With the assistance of NYSDOT experts important factors that influence required design hours were identified. Estimating tool users can make a regression prediction for the total project hours using historical projects as input data.
INTRODUCTION

State transportation agencies typically negotiate with designers to determine the cost and staffing levels for the design of highway and bridge infrastructure. In the traditional Design-Bid-Build project format the design consultants are selected based on the merit of their proposal. After the designer is selected, the designer enters a negotiation with the state transportation agency to determine the design fee and the allowable billable hours for the various design activities.

The New York State Department of Transportation (NYSDOT) regularly contracts with engineering consultants for design services for some of their capital improvement projects. The contracts are administered through their Consultant Management Bureau. The Consultant Management Bureau is responsible for establishing the scope of work for a project and preparing an independent staffing estimate for the design effort involved.

The goal of this paper is to describe a database that has been developed by the NYSDOT to provide Consultant Management Bureau managers with a predictive and comparative tool to aide in negotiations with design consultants. The purpose of the tool is to provide comparisons of the hourly staffing level of similar completed projects and to make a prediction of the total required hours for the project using regression analysis. The tool was developed using Microsoft Access. The database is continuously updated as additional projects are completed so it serves the dual purpose of assisting managers in determining appropriate staffing levels and as a repository of historical data.

COLLECTION AND ANALYSIS OF EXISTING DATA

At the initiation of the project a panel of NYSDOT experts identified the major factors that influenced required design-staffing levels. Based on expert opinion, the following project characteristics were established and evaluated:

1. Complexity
2. Project type
3. Number of sub-consultants
4. Construction costs
5. Number of lanes
6. Number of plan sheets
7. State Environmental Quality Review (SEQR) classification
8. National Environmental Policy Act (NEPA) classification
9. Predominant bridge type
10. Number of bridges
11. Highway classification
12. Length of project

The research focused on project characteristics 3-6 and 12 because there was insufficient data to include the other factors in the analysis.

Data Collection

Data were collected from 73 past projects. Data collected included the 12 factors identified by the expert panel. In addition a spreadsheet of consultant design hours for these projects was developed. Upon examination of the data, it was found that 64 projects had enough data to record hours associated with task levels outlined in the department’s base scope of services (BSOS).
The spreadsheet also included the factors that were identified by the NYSDOT experts that influence the number of design hours. Variables included the project complexity, the type of project, the total cost of the project, and the number of plan sheets. Figure 1 shows the median values found for the various design functions from the 64 projects analyzed. The 1000 level tasks are planning, 2000 tasks are design survey, 3000 are right-of-way surveys, 4000 tasks are preliminary design, 5000 tasks are in the area of socio/environmental impacts, bridge design is 7000, and 8000 are management tasks. It was noted that the highest values were in the 6000 categories, which is final design.

![FIGURE 1 Median Values for Project Design Tasks](image)

**Statistical Analysis of Data**

A correlation analysis was conducted to determine the relationship between the total project hours and other project variables including the number of sub-consultants, the number of plan sheets, the total project cost, and the project complexity. The Pearson correlation coefficient was calculated using the SPSS statistical software (1). The coefficient describes the linear relationship between variables. The coefficient can take any value between -1 and +1. A zero coefficient means the variables are not correlated. Coefficients of +1 mean the variables are completely correlated. Values of -1 means the variables are perfectly negatively correlated,
where as one variable increases, the other decreases. Table 1 shows the correlation table generated for the project data. The table shows that the total project hours have statistically significant positive correlations with the total project cost, the total number of plan sheets, and the number of lanes. The number of sub-consultants is also highly correlated; however, the expert opinion of the NYSDOT panel indicated that the number of sub-consultants is not an

**Correlation is significant at the 0.01 level (2-tailed).**

**TABLE 1 Correlations**
independent variable that should be included in further analysis because it is dependent on the number of estimated design hours, and requires Disadvantaged/Minority and Woman Owned Business Enterprise participation and specialized services.

**REGRESSION ANALYSIS**

Regression was studied to determine if simple or multiple linear regression models could be used to provide accurate predictions of total project hours and then be incorporated in the developed estimating support tool. The object of linear regression is to use the linear relationship between a response (dependent) variable and factor (independent) variables to predict or explain the behavior of the response variable. For this project the response variable was the design hours, and the independent variables considered for inclusion in the regression model were the number of plan sheets, project cost and length of the project.

**Development of Regression Models**

To develop the linear regression equations SPSS statistical software was employed (2). The software provided the capability to experiment with the stepwise equation-building technique and the backward elimination equation building technique. The stepwise procedure examines the significance of each input variable. Variables that are statistically significant are added to the regression equation. The order of insertion is determined by using the partial autocorrelation coefficient as a measure of importance of variables not yet in the equation (3). Depending on the significance of the input variables, the technique can yield a simple or a multiple regression equation. This technique is implemented in the SPSS statistical software package.

Using the stepwise procedure, all of the variables identified by the NYSDOT expert panel were considered for inclusion in the model. The stepwise procedure examines the significance of each input variable. Variables that are statistically significant are added to the regression equation. One simple equation was produced that relates the total project hours to the project construction cost.

The backward technique was also used to develop regression models. It is a method for sequentially removing variables. It begins with a model containing all independent variables and removes the variables that change $R^2$ the least. Using this technique multiple regression models were produced using several variables. It yielded a model that predicted total hours based on the Number of Plan Sheets, the number of lanes and the project length in miles.

**Regression Models Using Design Consultant Data**

Historical data from 64 NYSDOT projects designed by consultants were analyzed. Several regression models were constructed from this database. Regression models were constructed using project cost, number of lanes; project length and number of plan sheets as factor (independent) variables.

The database contained cases with data missing so some of the combinations of independent variables had very few cases to use in building a model. It is anticipated that as more data is collected the regression models can be recalculated and that other combinations of the variables could emerge as the “best” regression equation.

Several models have been studied. The two best performing models are shown in Table 2. $R^2$ is a measure of the performance of regression models. An $R^2$ of 1 indicates the model and the data are perfectly correlated. At this time it is recommended that equation one be employed as a predictor of total hours because it is the most parsimonious model and has the highest $R^2$ value.
Table 2. Regression Models for Consultant Data

<table>
<thead>
<tr>
<th>Model Number</th>
<th>No. of Cases</th>
<th>Model Equation</th>
<th>R</th>
<th>R²</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>22</td>
<td>Total Hours = (.001) Construction Cost + 8223.14</td>
<td>.830</td>
<td>.689</td>
</tr>
<tr>
<td>2</td>
<td>12</td>
<td>Total Hours = (28)No. Plan Sheets + (793.14)No. of Lanes - (1108.60)Length + 10593.61</td>
<td>.792</td>
<td>.628</td>
</tr>
</tbody>
</table>

Regression Analysis of State Designed Bridge Projects

Regression was also applied to bridge projects that had been designed internally by the NYSDOT. This data were used to determine if a larger data set would result in higher R² value and to similar factors also governed staffing levels for internal design projects. The stepwise regression method was applied and there 202 cases in the database used for analysis. The regression equation was found to be:

\[
\text{Total Hours} = 519.25 + 93.41(\text{Number of Plan Sheets})
\]

The R-value for this regression equation is 0.792 and the R² value is 0.628. This regression R² was coincidently similar to the R² results reported for regression model 2 using the consultant data.

Regression through the Origin

When using the regression equations derived from the consultant data, it became apparent that the equation gave impractical predictions. In particular, if there were zero plan sheets, because of the constant in the regression equation, the model produced a positive number as the prediction. To make the model output more realistic, it was decided to force the regression equation through the origin to insure that when there were zero plan sheets, no design hours would be predicted. This indicates that the relationship between the number of plan sheets and total hours may be non-linear. When forcing the regression through the origin, there is no constant in the equation produced. The best performing model using the SPSS software when regression is forced through the origin was found to be:

\[
\text{Total Hours} = 95.04(\text{Number of Plan Sheets})
\]

Based on this analysis it was decided to use this simple relationship between the total hours and the number of plan sheets in the Consultant Management decision support tool. However, the database has been programmed to continuously recalculate the regression equation every time there is a search so that new data is automatically included in the calculation of the simple regression model. It is anticipated that as more data is added the regression predictions will become more accurate.
THE CONSULTANT MANAGEMENT ESTIMATING TOOL

A tool that provides guidance and predictions of required design hours for projects has been developed. The tool was developed using the Microsoft Access software. It incorporates the 64 projects that were found during the data collection phase. The estimating tool has the capability to accept new projects and to monitor actual design hours expended for in-progress projects. Microsoft Excel spreadsheets have been developed that can automatically upload data for new projects into the database. The database can be used for two major functions: searching for similar completed projects and the design hours they required; and producing a prediction of project total design hours.

Searching for Similar Projects

A primary function of the estimating tool is to allow NYSDOT managers to review data from past projects that are similar to projects they are currently negotiating. This allows the NYSDOT managers to identify appropriate levels of project staffing for the new project. Figure 2 shows a search form that is used to find projects in the database that match important project criteria. The search allows users to view aggregated estimated and actual hours of the projects that are matched with the search criteria that the user has specified. Figure 3 shows an output table that is produced. It is a table of all of the projects in the database that matches the user’s search criteria. Each column shows the number of hours required for different types of design tasks. This report shows the estimated hours for the project. Estimated hours is the design effort estimated at the initiation of the project. It is also possible to click on the “actual hours” button and see a similar table showing the actual design hours that were expended. The output provides NYSDOT managers with an indication of the appropriate hours required for new projects and provides an aid for negotians with design consultants.
Estimates of Total Design Hours Required
The Estimate Total Hours Form shown in Figure 4 allows a user to estimate hours for a new project by providing an estimate of the number of plan sheets required. Experienced managers typically know an expected range of values for the number of plan sheets.

The Estimate Total Hours tools have two sections. In the first section the user can select the criteria in order to create a subset of projects that have characteristics similar to the new project (the project for which we are estimating the hours) and use that subset for a regression analysis that predicts total project hours. If the user does not select anything in the first section then all available projects will be used in the regression analysis operation. After all selections have been made in the top portion of the form the user clicks on “Filter Data” to create a list of the projects matching the selected criteria. These projects are used as the input data for the regression analysis. To see the list of filtered projects the user can optionally click on the “View Data” button.

In the second section the user must provide an estimate of the number of plan sheets the new project will have. Click on the “Calculated Hours” button to obtain a prediction. The tool will automatically calculate the estimated hours for the new project with the number of plan sheets as the independent variable. Figure 4 shows the prediction output in red. The total hours prediction and the error range of the predictions is shown. The prediction is not to be regarded as an exact number. The prediction should be considered as a guide to be modified by the users experience and the exact details of each particular project.

ESTIMATING TOOL IMPLEMENTATION
The system was extensively tested by the NYSDOT prior to implementation. The estimating tool has been actively used since August of 2011. NYSDOT managers have initially focused on expanding the database of projects. As projects are completed they are added to the database.
The size of the database is expected to increase from the original 64 projects with the establishment of NYSDOT business practices. It is anticipated that as the database grows in size, the output of the estimating tool will become more useful. In particular, the estimating tool has been designed to continuously recalculate the regression model coefficients based upon user inputs. It is expected that regression model prediction will become more accurate as the database increases in size. NYSDOT managers have reported that the information provided by the estimating tool has augmented the information available to managers determining project-staffing levels.

**CONCLUSIONS**

Initial reports of Estimating Tool use have indicated that it provides useful information to NYSDOT managers. The system provides managers with additional information about the appropriate staffing levels for a project design based on the project's type and characteristics. This additional information assists in the development of more accurate initial staffing plans.

The system has also proven to be useful as a systematic way of logging historical project data. It is anticipated that as the database expands in size, the utility of the system will increase. As more data becomes available, future research will explore the use of revised regression models and data mining algorithms to enhance the quality of the predictions made by the Estimating Tool.

For the estimating tool, we forced the regression through the origin to get more realistic results. This suggests that the relationship between the number of plan sheets and the design hours is non-linear. An area of future research is to study non-linear models to predict the required number of design hours.

**ACKNOWLEDGEMENTS**

This project was funded by the New York State Department of Transportation through the University Transportation Research Center, Region 2.
REFERENCES

