Cash Flow Projections for Selected TxDoT Highway Projects

Ra'ed Jarrah[1]; Devdatta Kulkarni[2]; and James T. O'Connor, M.ASCE[3]

Abstract: In planning for contractor payments, an owner with multiple projects needs to estimate the amount of money to be paid to contractors in coming months. For an owner as large as the Texas Department of Transportation (TxDoT), one is faced with the problem of organizing the budget to ensure that there are sufficient funds for all projects. This investigation was requested by TxDoT to provide a tool to forecast future project payments. Recent account summaries of TxDoT projects from 2001 to 2003 were analyzed for creating mathematical models representing monthly payments for various projects. The data were organized into categories of project types and subcategories of project contract amount. A fourth degree polynomial regression analysis was run on the data and the regression curve, when statistically significant, was taken to be the forecast payment curve. Finally, a computer program was developed to implement the results of the investigation for TxDoT's needs. The methodology provided will help other highway agencies to create their own equations to better predict project cash flows and trends. This investigation might also benefit researchers in projecting cash flows and trends, while also allowing for improvements.

DOI: 10.1061/(ASCE)0733-9364(2007)133:3(235)

CE Database subject headings: Resource management; Time factors; Payment; Cost control; Highway construction; Texas; State government.

Introduction

Considering that TxDoT handles hundreds of projects simultaneously, planning capabilities for meeting the budget and ensuring adequate funds are of vital importance. Due to a lack of accurate prediction models for project payment, TxDoT usually estimates payments as uniform distributions over the project duration. In reality, most construction projects do not have such a simple payment curve. With the idea that construction projects tend to have front end loading, TxDoT set out to enhance its budget-predicting formulas by assigning an investigation team to develop methods for better prediction of monthly project payment amounts. The investigation utilized data from past TxDot projects to develop regression models that can be used to predict future payments more accurately.

Objectives and Limitations

The objectives for this investigation were to collect, filter, and analyze recent Dallas district data with the aim of developing regression models to predict project payments per month. The models were to be statistically viable, yet representative. The investigation was not limited to producing just one regression equation, but as many needed to produce a system that would best represent varying project payment profiles.

To help TxDoT make better use of the results, a simple Microsoft EXCEL spreadsheet macro was programmed. The system is a user friendly program that relies on project inputs for multiple projects. The system output is the expected monthly payments for the projects entered in the system.

TxDoT provided the investigation team with the necessary recent project data. The time frame of that data was from September 2001 to December 2003, which encompassed 245 projects. The analysis concentrated on producing payment versus time curves, with individual curves for various project types and project contract amount. The projects which did not pass the elimination criteria set by the investigation team were eliminated from the analysis. These elimination criteria are mentioned later in the paper. For example, long delays in the project which were due to project suspension were not considered in the analysis. Extreme cost variations, such as those that may be indicated by contract change or caused by claims, were also eliminated. The regression analysis was carried out on Microsoft EXCEL with a fourth degree polynomial trend line. A typical S-curve has only one inflection point, and thus the equation would only need a third degree polynomial to represent it, but the investigation team preferred to add an additional degree for accuracy. The models were later refined to obtain the best fit curve by altering the polynomial degree.

Literature Review

It is common consensus that cash flow management and liquidity are key elements in the survival of contractors (Navon 1996). Therefore, many companies forecast and project expenditures to

[1]Doctoral Student, Graduate Research Assistant, Dept. of Civil Engineering, Univ. of Texas, Austin, TX 78712. E-mail: raedjarrah@mail.utexas.edu

[2]Graduate Research Assistant, Dept. of Civil Engineering, Univ. of Texas, Austin, TX 78712. E-mail: kulkarni@alumni.grinnell.edu

[3]C.T. Wells Professor of Project Management and Professor of Civil Engineering, Dept. of Civil Engineering, Univ. of Texas, Austin, TX 78712-1076. E-mail: jtoconnor@mail.utexas.edu

Note. Discussion open until August 1, 2007. Separate discussions must be submitted for individual papers. To extend the closing date by one month, a written request must be filed with the ASCE Managing Editor. The manuscript for this paper was submitted for review and possible publication on January 17, 2006; approved on October 26, 2006. This paper is part of the ***Journal of Construction Engineering and Management***, Vol. 133, No. 3, March 1, 2007. ©ASCE, ISSN 0733-9364/2007/3-235–241/$25.00.

manage their finances. The simplest method to project cash flows is by using an S-curve (Turan et al. 2004). The most popular forms of these S-curves are third, fourth, or fifth degree polynomials (Navon 1996).

Cash flow projection can be as complex as the user requires, with more complex models requiring more data. A variety of cash flow profiles can be applied to cover different situations, but in general the inputs would include project cost, duration, and other characteristics (Turan et al. 2004). By including more details one can produce a cash-flow management tool that can be applied on the company level (Navon 1996). The major obstacle in projecting cash flows would be applying the appropriate cash flow profile. This paper set out to accomplish that task through analysis of the data collected from recent years.

It is very important to use the available resources effectively to improve productivity and reduce the construction costs. Efficient utilization of resources plays an important role in the success of project management (Hegazy and Kassab 2003). In the case of unlimited resources, resources can be allocated intuitively based on an early start schedule (Hinze 2004). When there are not enough resources for the activity, it has to wait till the resources are available. On the other hand, when the project duration is already fixed, one has to add resources based on the schedule. However, one cannot keep hiring and firing the crew based on the schedule, therefore it is important to level the resources so it is within the limits of time and cost. This process is called resource leveling (Hinze 2004). With the use of the better prediction of the cash flow, one can anticipate the resources needed at various projects in coming months (Touran 1991; Touran et al. 2004). This can be achieved by the cash flow equations provided later in the paper for various project types.

Various heuristic solutions are presented for prioritizing the activities. Heuristic methods usually follow a set of hierarchal rules and are used to find where to allocate the resources first. The least total float (LTF) and the earliest late start (ELS) are two most common heuristic approaches used (Davis and Patterson 1975). The value of late start is calculated from the original CPM schedule, whereas total float keeps changing every time the schedule is changed due to lack of resources. Thus ELS can be much easier to apply than LTF approach. Heuristic approaches are easy to apply for large projects and easy to use (Talbot and Patterson 1979). Hagazy and coworkers expanded this approach to allocate multiskilled labor and make the procedure more efficient. They found out that the use of multiskilled labor allocation approach reduced the time and cost of the project compared to when single-skill labor was allocated. The advantage of this approach is its ability to use the underallocated resources to resolve other over-allocated resources (Hegazy et al. 2000). To be able to use the above-mentioned resource allocation methods effectively; one needs to know the need of resources ahead of time. So that enough time is allowed for resource leveling. Number of inspectors required for various projects can be estimated using the results in this paper, and was included in the research application tool. However, the writers will not delve any deeper into the topic of inspectors as it is not part of the scope of this paper.

Methodology

Overview

The TxDoT highway project payment data was provided in the form of monthly account summary reports. The next step was to copy that data to a spreadsheet and organize it in a usable manner. After all the data was transcribed into a spreadsheet, the investigation team developed a plan of approach to deal with the vast amounts of data. In this planning phase the team decided to divide the data into categories of project types (and later into other subcategories if needed), set criteria to eliminate incomplete or abnormal data, and analyze the payment fraction versus the time fraction, called "time charge" by TxDoT, using polynomial regression. After initial organization, preliminary tests revealed the data was too scattered to allow for reasonable regression. This was due to some projects which had incomplete data or had delays due to suspensions or cost overruns due to claims. The team developed elimination criteria for the data to resolve the issue. After the data was filtered, the data were subcategorized by project contract amount. With these final adjustments accomplished, the data were analyzed and the results of the regression were recorded. These regression equations were implemented in a Microsoft EXCEL macro so that it can be used easily by TxDoT.

Data Collection

The information provided by TxDoT was in "html" file format. The investigation team transcribed all that data into a Microsoft EXCEL spreadsheet to enable calculations and analysis. Not all the information was used in the analysis but all the available data from the account summaries was transcribed from the start in order to avoid complications of recollection of information in case it was found lacking in the analysis.

Definitions

- *Bid*: The contract amount for a given project.
- *Amount paid this estimate*: The amount paid in a particular month for a given project.
- *Time charge numerator*: Days passed on the project; time charge is days passed over contract duration.
- *CSJ*: Project unique number in the TxDoT database.
- *Time charge denominator*: Total contract duration in days; time charge is days passed divided by contract duration.
- *Type*: The classification of the type of project. Example: BR=bridge replacement, MSC=miscellaneous, and RER=rehabilitation of existing road. See Table 1 for full list.
- *% of total paid*: The cumulative amount paid for the project up to that date.
- *% time elapsed*: Days passed out of total duration. Also defined as time charge.

The team used 28 monthly reports from September 2001 to December 2003. Each report contained an average of 120 projects. At the end of the data transfer to spreadsheet, 245 different projects were inputted into a Microsoft EXCEL spreadsheet.

Data Organization

Once all the data were transcribed on the spreadsheet, it was organized so that it could be analyzed and would be a better representation for different types of projects. Using the sorting tool in Microsoft EXCEL, the data were sorted by the project type. All the data for each project type were copied into a new file so it could be individually analyzed. Each project type file was sorted by project unique number so that the data for a particular project was together. This file was then sorted by time to arrange the data in chronological order. At the end of this process the

J. Constr. Eng. Manage., 2007, 133(3): 235-241

Table 1. Project Type Definition

Type	Classification
BR	Bridge replacement
BWR	Bridge widening or rehabilitation
CTM	Corridor traffic management
HES	Hazard elimination and safety
INC	Interchange (new or reconstructed)
LSE	Landscape scenic enhancement
MSC	Miscellaneous construction
NNF	New location nonfreeway
OV	Overlay
RER	Rehabilitation of existing road
RES	Restoration
SC	Seal coat
SKP	SKIP (exempt from sealing)
TPD	Traffic protection devices
TS	Traffic signals
UGN	Update to standards nonfreeway
UPG	Update to standards freeway
WF	Widen freeway
WNF	Widen nonfreeway
Other	Collection of all of the above categories with less than seven projects after filtration

different type of projects were in different files. Each file had the data of individual projects in chronological order, so that the entries represented how the project progressed.

Plan of Approach

The first step was to divide the data for easier analysis and more representative results. The data were divided into different types of projects as defined by TxDoT. The next step was to decide on filter criteria to eliminate the incomplete or abnormal data. After filtration, the analysis had to be done on a significant number of data points. The team decided to use a minimum number of seven "complete" projects for analysis. A complete project was defined as one for which 80% or more of its total data were contained over the 28-month period (some projects were started before August 2001, or were finished after December 2003, and thus had missing data). The estimate was derived from the assumption that a complete project had on average seven data points, thus seven complete projects would yield a minimum of 49 data points to analyze. The investigation team decided that was a reasonable number of data points to analyze based on the $n=K^2V^2/D^2$ statistical formula, with assumed values of $K=2$, $V=0.7$, and $D=0.2$, where K=number of standard deviations for a chosen confidence level; D=maximum tolerable error; and V=coefficient of variation or the standard deviation divided by the mean. Therefore, if a project type contained less than seven complete projects, it would be classified with other project types lacking enough data under the category of "other."

The investigation team realized that projects of significantly different contract amounts (even those of the same type) would yield different payment curves due to size of the project. The team assumed that large projects would have more resource reallocation and payment variation than smaller projects. Hence the data was divided into subcategories by contract amount as contract amount was assumed to be an indicator of project size.

Data Division

The data before and after filtration had distributions as illustrated in Tables 2 and 3:

The category miscellaneous (MSC) was subdivided into three

Table 2. Total Number of Projects before Filtration

Type	Classification	Total number	Contract amount				
			<$0.5MM	$0.5–$1MM	$1–V5MM	$5–$20MM	>$20MM
BR	Bridge replacement	**25**	7	10	8		
BWR	Bridge widening or rehabilitation	**4**	3	1			
CTM	Corridor traffic management	**4**			4		
HES	Hazard elimination and safety	**4**	3		1		
INC	Interchange (new or reconstructed)	**4**			1	3	
LSE	Landscape scenic enhancement	**7**	6	1			
MSC	Miscellaneous construction	**55**	24	17	14		
NNF	New location nonfreeway	**6**	2	3			1
OV	Overlay	**29**		1	20	8	
RER	Rehabilitation of existing road	**27**	3	8	12	4	
RES	Restoration	**4**		2	1	1	
SC	Seal coat	**9**		1	8		
SKP	SKIP (exempt from sealing)	**1**				1	
TPD	Traffic protection devices	**1**				1	
TS	Traffic signals	**31**	31				
UGN	Update to standards nonfreeway	**3**			3		
UPG	Update to standards freeway	**4**			1	2	1
WF	Widen freeway	**8**	1	1	4		2
WNF	Widen nonfreeway	**19**			2	15	2
Total		**245**	**80**	**45**	**80**	**34**	**6**

Table 3. Total Number of Projects after Filtration

Type	Classification	Total number	Contract amount				
			<$0.5MM	$0.5–$1MM	$1–$5MM	$5–$20MM	>$20MM
BR	Bridge replacement	**24**	11	7	6		
MSC	Miscellaneous construction	**37**	20	7	10		
Other	Categories with less than seven projects	**27**	6	4	9	8	
OV	Overlay	**14**		1	12	1	
RER	Rehabilitation of existing road	**15**	2	2	7	4	
TS	Traffic signals	**26**	26				
WF	Widen freeway	**8**	1	1	4		2
WNF	Widen nonfreeway	**18**			2	14	2
Total		**169**	**66**	**22**	**50**	**27**	**4**

subcategories for analysis ($0–$0.5MM, $0.5–$1MM, and $1–$5MM) due to the significant number of complete projects available after filtration in each of those brackets.

Data Filtration and Elimination Criteria

To filter the data, the investigation team applied the following elimination criteria prior to analysis:

- Any project with time charge exceeding 120% was removed entirely;
- Any project with percent paid exceeding 120% was removed entirely;
- Any project missing more than 50% of its data was removed entirely; and
- Any project with obvious delays or stalls was removed entirely.

The assumption for the first two conditions was that projects that greatly overran cost or schedule did not behave in a representative manner; their progress shifted. Many of these projects were over budget or delayed because of either claims or suspensions. The reasoning behind this was to derive a model incorporating a limited degree of extension to cost and schedule (20%), and allowing the model users to account for project specific cost or schedule overruns by manipulating the model output [by adding the cost overrun to the period(s) that they were incurred in, and skipping over the delay period(s) when calculating payment dates from time charge]. These overruns would not have occurred yet at the start of the project, and would have to be edited in later as they occur. Excessive prediction of project delays and cost overruns (beyond 20%) were not in the scope of this study, as the heterogeneous nature of construction would not likely yield any conclusive results with the data used for this study (other variables such as location, number of contractors, etc., would have to be factored in).

Projects with 50% or more of the data missing were removed to avoid hetroskedasticity (lack of consistency in data variance). The last filter condition, obvious delays or stalls, was applied later in the analysis when the investigation team discovered some projects with very distinct trends. These projects had several points on the time charge versus percent paid chart that were horizontal. This basically meant that time progressed on the project but the contractor was not paid for a few months. These delays could have been due to seasonal delays, a request by the owner to halt the work, or even unforeseen site conditions. Regardless of the cause, the data was not of regular behavior and therefore were felt to not represent a typical payment curve.

The projects that retained more than 50% but less than 80% of the data were not eliminated, but were not counted as a "complete project." The investigation team labeled them as "partially complete projects." These projects added some statistical value to the analysis, yet violated the rule of homoskedasticity (consistency of data variance) by causing an uneven distribution of data points. However, due to the limited number of projects available, the investigation team decided to accept some violation to that rule to allow for more data points in the regression analysis. These projects were not counted as complete and thus did not affect the initial requirement of at least seven complete projects per regression analysis.

Regression Analysis

After filtration, an *X*-*Y* scatter chart was plotted in Microsoft EXCEL, with time charge on the *X*-axis and percent paid on the *Y*-axis. A regression analysis was then applied by adding a trendline to the data series. Microsoft EXCEL allows the user the option of selecting the type of regression. The investigation team chose a fourth degree polynomial regression. An S-curve has one inflection point, and thus a third degree polynomial is needed to produce the curve. The team decided on a fourth degree polynomial to better grasp the variability in the scatter plot. As mentioned earlier, the analysis was done on at least seven complete projects.

After the analysis, the investigation team looked at the scatter plot results and eliminated any significant irregularities that managed to bypass the filter. With this second filtration, the analysis was complete and the equations (along with their R^2 values) were recorded for later interpretation (see Figs. 1–10). The team inves-

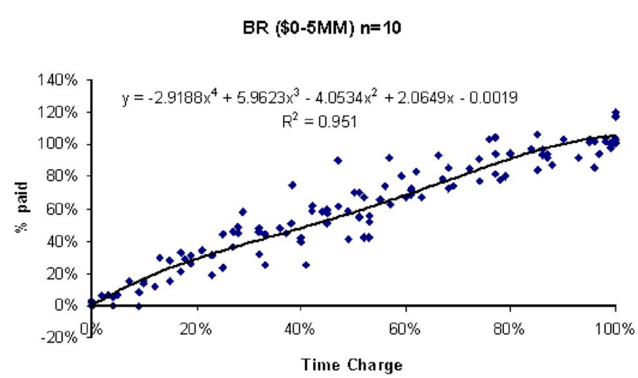

Fig. 1. Cash flow profile for "bridge" projects $0 to $5MM

J. Constr. Eng. Manage., 2007, 133(3): 235-241

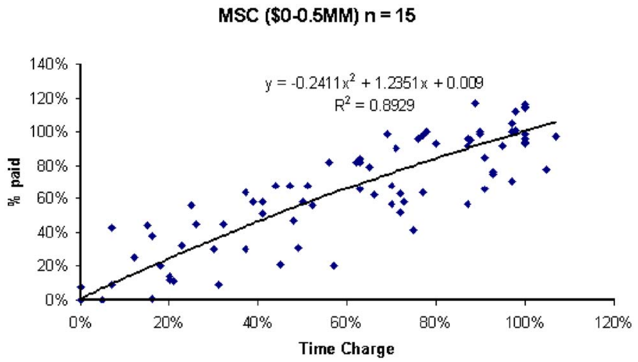

Fig. 2. Cash flow profile for "miscellaneous" projects from $0 to $0.5MM

tigated further into the results by running multiple regression ANOVA tests on the data points.

Interpretation of Results

The scatter plot was a crucial indicator of the trend of a project type and its subcategories. Some projects had typical S-curves, whereas others had very linear profiles. The traffic signal (TS) and MSC (miscellaneous, $0–$0.5MM) plots had the most variation in the data. What is more interesting is that these two categories are small-size projects (project cost being under $0.5MM).

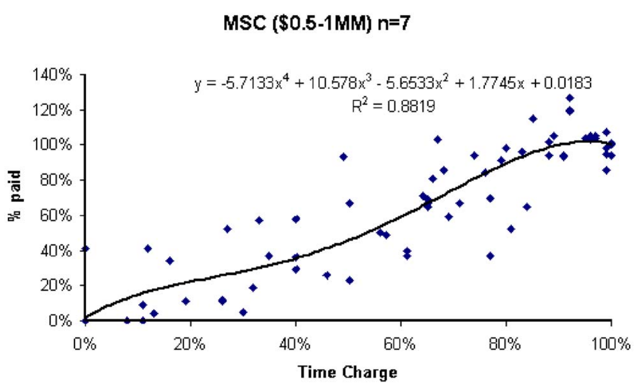

Fig. 3. Cash flow profile for miscellaneous projects from $0.5 to $1MM

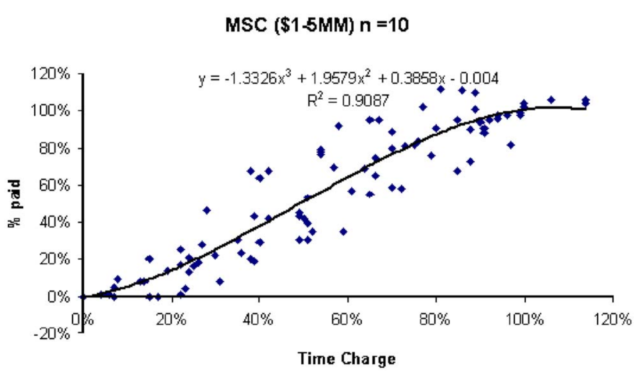

Fig. 4. Cash flow profile for miscellaneous projects from $1 to $5MM

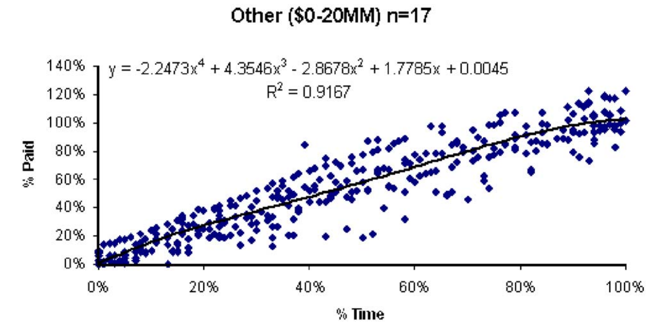

Fig. 5. Cash flow profile for "other" projects from $0 to $20MM

Another issue of interest is that while all the other categies tended to have curved or S-shaped curves, the overlay (OV) category plotted out to be a linear equation.

The results produced acceptable R^2 values for all the projects, and the equations were of reasonable shape. The fourth degree polynomial regression equations sometimes produced a negative slope at the end. This was determined to be due to a collection of points beyond the 100% time charge. However, as the negative slope only occurred outside the scope of application of the equations in the computer program (0–100% time charge), the anomalies were left in the equations. Further, a comparison between the derived regression equations and actual recent data was made on a monthly scale to compare the individual outputs of the computer program versus raw records.

The threshold for statistical significance was set at meeting the number of data points from the $n=K^2V^2/D^2$ formula (with $K=2$ and $D=0.15$) and all regression equation coefficients achieving a

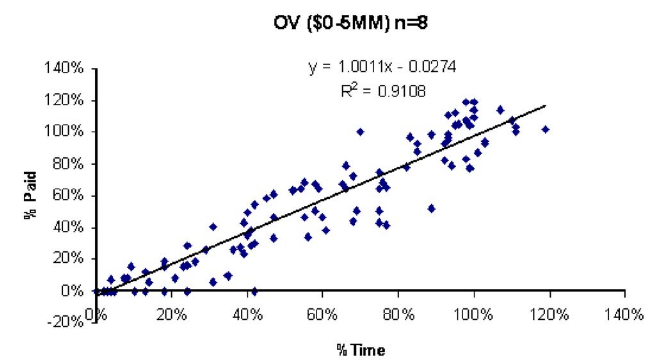

Fig. 6. Cash flow profile for "overlay" projects from $0 to $5MM

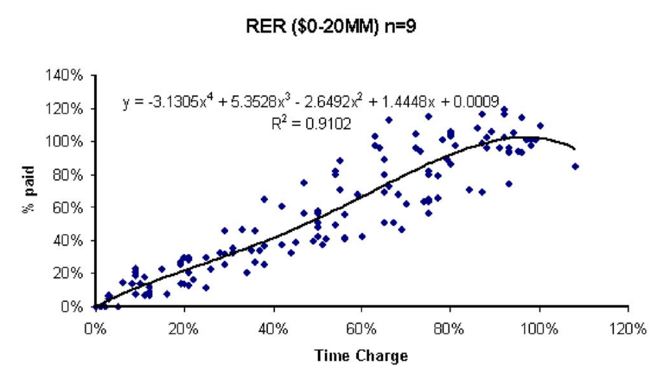

Fig. 7. Cash flow profile for "rehabilitation of existing road" projects from $0 to $20MM

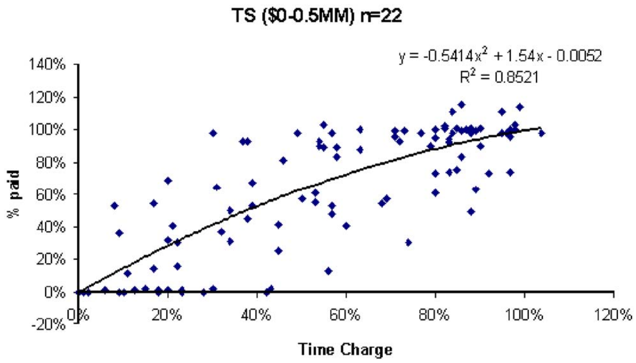

Fig. 8. Cash flow profile for "traffic signal" projects from $0 to $0.55MM

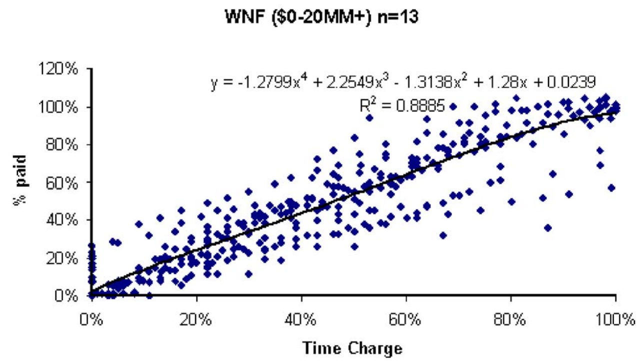

Fig. 10. Cash flow profile for "widen nonfreeway" projects from $0 to $20MM+

P value no more than 0.20 (the models were refined by decreasing the polynomial degree until this criterion is met). The statistical significance of the regression equation was set at an R^2 of 0.85 or higher.

Analysis of Results

The resulting regression formulas underwent an ANOVA analysis to verify they meet the significance criteria mentioned before. The categories bridge replacement (BR), MSC ($0.5MM–$1MM bracket), other, rehabilitation of existing road (RER), and widen nonfreeway (WNF) met all of the statistical significance criteria. The categories MSC ($0–$0.5MM bracket), MSC ($1MM–$5MM bracket), OV, TS, and widen freeway (WF) fell shy of the $n=K^2V^2/D^2$ criterion. This criterion would be met as more data points are provided and analyzed in these categories. It should be noted that all the categories passed the $R^2=0.85$ criteria.

Although some categories did not meet the number of data points for statistical significance ($n=K^2V^2/D^2$), the investigation team decided to develop the computer program. The idea was to provide a working and user-friendly implementation tool that can be edited with more significant regression results as they become available in the future.

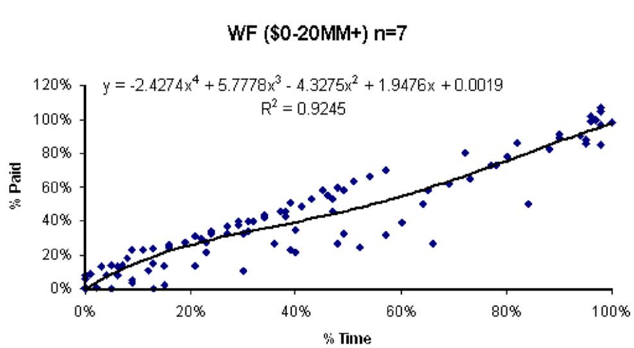

Fig. 9. Cash flow profile for "widen freeway" projects from $0 to $20MM+

System Development

Overview

The investigation team used the projection formulas acquired from the analysis to create a user-friendly tool that would calculate the estimated monthly payments for several projects simultaneously. The team decided to use Microsoft EXCEL to create the tool because of its popularity, flexibility, compatibility, and ease of use. The formulas were programmed in the macro to simplify programming procedures. The system is fully discussed in a later segment of this paper.

The D-CIPS program was developed after the data analysis phase to help the user quickly and easily apply the cash flow profiles developed. The program was written in Microsoft EXCEL using macros and simple functions, and was named D-CIPS (Dallas cash-flow and inspector projection system). The cash flow generator uses "project type," "duration," "contract amount," and "start date" for input and calculates an estimate of payments and inspector assignment at the start of each calendar month for every district manager (the inspector assignment portion of this program is not relevant to this paper, and was implemented as an add-on to the program in later phases of development). The estimates are calculated when costs are incurred and not when they are actually paid. The program is capable of calculating as many as 65,000 different projects, but for the sake of simplicity and practicality, the program was initially designed to calculate up to 300 different projects. The team included an option to increase capacity.

System Details

The program works by utilizing a custom function (created in Microsoft EXCEL macros) that takes project type, duration, contract amount, start date, "working days per week," and "month on the calendar" as inputs and determines the expected monthly payment. A custom function processes the inputs using the regression formulas from the analysis and outputs the projected payment amount. The first month on the calendar is user defined to allow flexibility in viewing the information.

As an added feature, the program is also capable of calculating the cash-flow projections based on working day duration and a specified number of working days per week. This allows the user the freedom of inputting projects in either calendar or working day formats.

Due to the size of the Microsoft EXCEL spreadsheet, the last

J. Constr. Eng. Manage., 2007, 133(3): 235-241

month on the calendar is almost 20 years past the earliest start date. It is not expected that this will cause any problems for the user, as highway construction projects seldom extend beyond 8 years. Due to computer processing limitations, the cash flow generator might need more time to save as more projects are added beyond the initial 300.

Conclusions

- Most of the payment curves turned out to be in the S-shape expected, and those that had a few anomalies in them could be logically explained (such as the negative slope at the end).
- The R^2 values of the regression equations met the statistical significance criterion (above 0.85). This was due in part to the choice in setting up the data filter, which helped keep noncategory data out of the analysis. TS and MSC subcategory $0.5–$1MM had the lowest R^2 values mostly due to the scatter of the data and smaller number of data points, whereas BR projects had the highest R^2 values due an even distribution.
- Analysis results of the MSC ($0–$5MM), MSC ($1–$5MM), OV, TS, and WF data sets revealed that these categories fell short of having the significant number of data points set out in this study.
- The cash flow equations for the various types of projects with their corresponding plots are given below. In the equations shown in the following, x represents the percent of time passed and y represents the corresponding percent of contract amount paid.

Recommendations

The investigation team presents the following recommendations:

- For future system refinement, more data is needed, especially for MSC, OV, TS, and WF type projects.
- The ideal system would the capability to update projection models based on recent project data.
- A future system should make the filter more flexible to allow the user a chance to better deal with the data if need be. This could include features such as choice of subcategory or combination of category types.
- A future analysis (between 5 and 10 years from the date of this investigation) would produce more representative equations for their times.
- The Dallas TxDoT district can monitor the accuracy performance of the prediction models and periodically identify opportunities for enhancement.
- Future system enhancements would be facilitated if the following project data fields were more readily (electronically) accessible.

Acknowledgments

The writers would like to thank Tracey Friggle and TxDoT for funding the investigation and providing the data.

References

Davis, E. W., and Patterson, J. H. (1975). "Resource-based project scheduling: Which rules perform best?" *Project Management Q.*, 6(4), 25–31.

Hegazy, T., et al. (2000). "Algorithm for scheduling with multiskilled constrained resources." *J. Constr. Eng. Manage.*, 126(6), 414–421.

Hegazy, T., and Kassab, M. (2003). "Resource optimization using combined simulation and genetic algorithms." *J. Constr. Eng. Manage.*, 129(6), 698–705.

Hinze, J. W. (2004). "Construction planning and scheduling." *Resource allocation and resource leveling*, Pearson Prentice-Hall, Columbus, Ohio, 110–156.

Navon, R. (1996). "Company-level cash-flow management." *J. Constr. Eng. Manage.*, 122(1), 22–29.

Talbot, F., and Patterson, J. H. (1979). "Optimal methods for scheduling projects under resource constraints." *Project Management Q.*, 10(4), 26–33.

Touran, A. (1991). "Discussion of 'Current float techniques for resources scheduling' by V. Shanmuganayagam." *J. Constr. Eng. Manage.*, 117(3), 574–575.

Touran, A., Atgun, M., and Bhurisith, I. (2004). "Analysis of the United States Department of Transportation prompt pay provisions." *J. Constr. Eng. Manage.*, 130(5), 719–725.

J. Constr. Eng. Manage., 2007, 133(3): 235-241