LEARNING CURVE PREDICTORS FOR CONSTRUCTION FIELD OPERATIONS

By John G. Everett,[1] Member, ASCE, and
Sherif Farghal,[2] Student Member, ASCE

Downloaded from ascelibrary.org by Mississippi State Univ Lib on 07/07/19. Copyright ASCE. For personal use only; all rights reserved.

ABSTRACT: Many repetitive construction field operations exhibit a learning curve, over which the time or cost per cycle decreases as the cycle number increases. This paper evaluates several mathematical models to determine which best describes the relationship between the activity time or cost and the cycle number. For completed activities, cubic learning curve models are found to provide the most reliable statistical fit, and linear models provide the least reliable fit. The real potential value of learning curves is their ability to predict the time or cost needed to perform future activities. This paper presents a methodology for predicting future activity time or cost based on completed activity data. The best predictors of future performance are found to be linear models. The cubic models that best describe completed activities are poor predictors of future performance.

INTRODUCTION

Anyone who has ever performed the same complex task several times in succession knows that it takes less time to perform the second cycle than the first, less time to perform the third cycle than the second, and so on. Many construction field operations exhibit this phenomenon known as the learning or experience effect. A learning curve is generated when the time or cost required to complete one cycle of an activity is plotted as a function of the cycle number. Learning curve data can be presented in several terms such as man-hours/cycle, dollars/cycle, minutes/cycle, and so on. In this paper, time/cycle is used as the unit of measure, but other units could be used as well.

Several factors contribute to the decreasing activity time/cycle with increasing number of cycles: greater familiarity with the task, better coordination, more effective use of tools and methods, and more attention from management and supervision (Oglesby et al. 1989).

LEARNING CURVE MODELS

Past research in learning curve theory has focused on developing mathematical models that describe the time/cycle as a function of the cycle number and fitting the models to historical data. Wright first described a logarithmic learning curve for the production of airplanes in the 1930s (Wright 1936). This model assumes a constant rate of learning or improvement. Each time the number of cycles doubles, the time/cycle decreases by a constant percentage. This model is often called the straight-line model because it produces a straight line when plotted on log-log coordinates.

---

[1]Asst. Prof., Dept. of Civ. and Envir. Engrg., Univ. of Michigan, Ann Arbor, MI 48109-2125.

[2]Grad. Res. Asst., Dept. of Civ. and Envir. Engrg., Univ. of Michigan, Ann Arbor, MI.

Note. Discussion open until February 1, 1995. To extend the closing date one month, a written request must be filed with the ASCE Manager of Journals. The manuscript for this paper was submitted for review and possible publication on January 25, 1993. This paper is part of the ***Journal of Construction Engineering and Management***, Vol. 120, No. 3, September, 1994. ©ASCE, ISSN 0733-9364/94/0003-0603/$2.00 + $.25 per page. Paper No. 5600.

J. Constr. Eng. Manage., 1994, 120(3): 603-616

The Stanford B model was developed in the late 1940s to account for a crew's previous experience in the first few cycles (*An improved* 1949). After an initial hump in the learning curve, the remaining activities follow the logarithmic curve with a constant rate of learning.

A cubic model was developed in the 1970s to account for a crew's experience in the first few cycles and also a leveling off in improvement as the project approaches completion or as production reaches a steady state. The cubic model assumes that the rate of learning may change over time (Carlson 1973).

The piecewise model approximates the cubic model with three distinct phases, each with a constant rate of learning. On a log-log plot, this model appears as three straight line segments. The first phase, analogous to the hump in the Stanford B model, is called the operation learning phase. A second phase, called the routine acquiring period, resembles the logarithmic curve. The third phase, standard production, occurs when the production rate levels off or ceases to improve (Thomas et al. 1986).

An exponential model was developed in the 1960s by the Norwegian Building Research Institute (*Effect* 1965). The model assumes that part of the time/cycle is fixed and part is subject to improvement through repetition. The model states that the part subject to improvement will be reduced by one-half after a constant number of cycles, and the time/cycle will gradually approach an ultimate or lowest value.

The mathematical formulas describing the logarithmic and cubic learning curves are presented in the following. For a more detailed discussion of the Stanford B, piecewise, and exponential models, the reader is referred to Thomas et al. (1986).

According to Thomas et al. (1986)

> the researcher generally has the choice of developing forecasting models by using unit data or cumulative average data. The cumulative average curve has considerable power to smooth out the unit data. . . . However, the cumulative average curve can be deceptive because it is a smoothing process whose power increases as the cumulative quantity increases; it has a tendency to make the basic data look better (smaller variation) than similar curves using unit data. For controlling current operations, the unit curve or perhaps a moving average curve contains more relevant information. . . . However, the cumulative average and unit curves are complementary rather than competing forms, because both are derivatives of the same data.

The research described in the remainder of this paper is based on unit data.

OBJECTIVE

This paper has two objectives. The objective of the following section is to measure the correlation between several previously untested mathematical models and historical data. The new models will be compared to the standard logarithmic and cubic models just described to determine which models provide the best statistical fit for completed construction activities.

The objective of the section headed "Using Learning Curves to Predict Future Performance" is to determine which mathematical models provide the best prediction of time/cycle for future activities. Previous research in

J. Constr. Eng. Manage., 1994, 120(3): 603-616

Downloaded from ascelibrary.org by Mississippi State Univ Lib on 07/07/19. Copyright ASCE. For personal use only; all rights reserved.

learning curve models has concentrated on finding mathematical models to describe and fit completed activities. These data may be of historical interest, but prediction of the time/cycle for future activities will be much more valuable to construction estimators, schedulers, and managers.

CORRELATION OF MATHEMATICAL MODELS WITH HISTORICAL DATA

Methods

Historical data for 60 construction field operations were gathered from several published sources (*Effect* 1965; Everett and Slocum 1993; McClure et al. 1980a,b; Oglesby and Parker 1989). Least-squares regression was used to find the best-fit learning curve for each of 12 mathematical models. Pearson's coefficient of correlation, R^2, is calculated to measure the correlation of each model to each of the 60 sets of data. To assess the overall correlation of each learning curve model to historical data, the average R^2-value for each model over 60 sets of data was calculated.

The methodology employed in this part of the research is similar to that of Thomas et al. (1986) except that the present paper evaluates unit data exclusively; Thomas et al. used both unit data and cumulative average data. Also, this paper only examines construction examples with eight or more data points, as opposed to other research that uses examples with as few as four data points. The somewhat arbitrary minimum number of eight points was chosen for two reasons. First, "it is always possible to fit a polynomial of degree $n - 1$ to n data points . . . and the researcher should not consider using a model that is "saturated," that is, that has very nearly as many independent variables as observations" (Hines and Montgomery 1990). With only four data points, most of the learning curve models to be evaluated would be saturated and the results suspect. By increasing the minimum number of points, the saturation problem is reduced. Second, the technique used in the section headed "Using Learning Curves to Predict Future Performance" of this paper requires that learning curves based on the first $n/2$ cycles be correlated with data from the second $n/2$ cycles. A reasonably large n is required to perform this part of the analysis.

This paper evaluates two models, the logarithmic and cubic models previously tested by Thomas et al. (1986) and 10 previously unreported models. The three other models described by Thomas et al.—the Stanford B, piecewise, and exponential models—are not evaluated here. The Stanford B model is identical to the logarithmic model except for the first few cycles. The piecewise model and exponential model each requires assumptions about the values of certain parameters, in particular, the ultimate or steady-state time/cycle. This paper assumes that the construction manager does not know the ultimate time/cycle, and is applying learning curve theory to help predict it.

Mathematical learning curve models should account for several characteristics of repetitive work. Except for random scatter, the time/cycle for successive cycles should be monotonically decreasing. The decrease in time/cycle between successive cycles should be greatest near the beginning of a series of similar cycles and should decrease as the number of cycles becomes larger. The time required to complete a cycle can never reach zero or become negative.

Many mathematical functions satisfy these requirements for at least part of their range. Twelve mathematical functions are described in this investigation. Several other models, not described in this paper were also eval-

J. Constr. Eng. Manage., 1994, 120(3): 603-616

Downloaded from ascelibrary.org by Mississippi State Univ Lib on 07/07/19. Copyright ASCE. For personal use only; all rights reserved.

uated, and it would be possible to develop many more. The other models tested but not reported here were found to be unreliable for describing past and future construction work. As is shown in the following, increasingly complex mathematical functions have little practical use in describing construction field operations.

The general form of the 12 mathematical models evaluated in this study are

Linear x, y

$$y = a + bx \tag{1}$$

Linear, x, log y

$$\log y = a + bx \tag{2}$$

Linear log x, y

$$y = a + b \log x \tag{3}$$

Linear log x, log y

$$\log y = a + b \log x \tag{4}$$

Quadratic x, y

$$y = a + bx + cx^2 \tag{5}$$

Quadratic x, log y

$$\log y = a + bx + cx^2 \tag{6}$$

Quadratic log x, y

$$y = a + b(\log x) + c(\log x)^2 \tag{7}$$

Quadratic log x, log y

$$\log y = a + b(\log x) + c(\log x)^2 \tag{8}$$

Cubic x, y

$$y = a + bx + cx^2 + dx^3 \tag{9}$$

Cubic x, log y

$$\log y = a + bx + cx^2 + dx^3 \tag{10}$$

Cubic log x, y

$$y = a + b(\log x) + c(\log x)^2 + d(\log x)^3 \tag{11}$$

Cubic log x, log y

$$\log y = a + b(\log x) + c(\log x)^2 + d(\log x)^3 \tag{12}$$

The linear log x, log y model is the logarithmic or straight-line model described by Wright (1936). The cubic log x, log y model is the cubic model described by Carlson (1973) and Thomas et al. (1986). For all models the independent variable x = cycle number; and the dependent variable y = time required to perform that single cycle.

Pearson's coefficient of determination, R^2, is used as a measure of the adequacy of a regression model or how well a best-fit curve of that model

J. Constr. Eng. Manage., 1994, 120(3): 603-616

Downloaded from ascelibrary.org by Mississippi State Univ Lib on 07/07/19. Copyright ASCE. For personal use only; all rights reserved.

correlates to a given set of data. Let n = number of cycles of a construction activity; $y_1, y_2, \ldots, y_n$ be the measured time/cycle for cycles 1, 2, . . . , n; and $y'_1, y'_2, \ldots, y'_n$ be the time/cycle corresponding to cycle 1, 2, . . . , n, on the least-squares best-fit curve. Let

$$T = \sum_{i=1}^{n} (y_i - \bar{y})^2 \tag{13}$$

where

$$\bar{y} = \frac{\sum_{i=1}^{n} y_i}{n} \tag{14}$$

which is the mean of the measured time/cycle; and $(y_i - y)$ = deviation of each measured time/cycle from the mean. Let

$$E = \sum_{i=1}^{n} (y_i - y'_i)^2 \tag{15}$$

where $(y_i - y'_i)$ shows the deviation of the time/cycle on the best-fit curve from the measured time/cycle for each cycle. Now

$$R^2 = \frac{T - E}{T} \tag{16}$$

If E is small, indicating that the time/cycle values on the best-fit curve are nearly equal to the corresponding measured values, R^2 approaches 1, and the best-fit curve is very well correlated to the measured data. As the deviations of the time/cycle values on the best-fit curve compared to the measured data increase, E increases and R^2 approaches 0. For reasons too detailed to be explained here, E cannot exceed T. Thus, R^2 ranges from 0 to 1, where 0 denotes no correlation and 1 denotes perfect correlation. One interpretation of the meaning of R^2 is that it explains the percentage variation in the time/cycle due to the change in the cycle number.

Results

Table 1 shows the 60 construction examples in column 1 and the corresponding values of R^2 for the linear log x, log y; cubic log x, y; and cubic log x, log y models in columns 2, 3, and 4, respectively. R^2 values were calculated for all of the other models, but their results have been omitted for space and clarity. Complete results are tabulated in Farghal and Everett (1993). The linear log x, log y model is shown because it is discussed in more detail in the following. The cubic log x, y model is shown because this model gives the highest average value for R^2. The cubic log x, log y model is shown because this is the cubic model described by Thomas et al. (1986) and Carlson (1973).

Table 2 shows a summary of the results for all 12 models. Column 2 shows the average R^2-value for each model. The highest average value of R^2 (0.75) is given by the cubic log x, y model.

Conclusion

Cubic models in general provide better correlation to historical data than the quadratic models, which are superior to the linear models. This is not surprising. Higher-order polynomials always produce higher R^2-values. Carlson

Downloaded from ascelibrary.org by Mississippi State Univ Lib on 07/07/19. Copyright ASCE. For personal use only; all rights reserved.

TABLE 1. Correlation of Mathematical Models with Historical Data

Activity (1)	R^2 linear log x, log y (2)	R^2 cubic log x, y (3)	R^2 cubic log x, log y (4)
Residential construction (*Effect* 1965)			
Erection of pairs of houses (site A)	0.67	0.74	0.72
Erection of pairs of houses (site B)	0.67	0.72	0.71
Erection of pairs of houses (site C)	0.47	0.51	0.50
Erection of pairs of houses (site D)	0.08	0.32	0.32
Erection of pairs of houses (site E)	0.25	0.49	0.48
Assembling and dismantling formwork (A)	0.88	0.96	0.95
Assembling and dismantling formwork (B)	0.72	0.99	0.99
Assembling formwork	0.79	0.90	0.88
Formwork and concreting	0.51	0.95	0.94
Assembling of framework for hoistways	0.98	1.00	1.00
Brickwork operations	0.78	0.80	0.81
Building element assembly	0.79	0.92	0.88
Concreting (A)	0.85	0.98	0.98
Concreting (B)	0.33	0.42	0.41
Concreting (C)	0.86	0.99	0.99
Concreting of first-floor walls and roof	0.50	0.78	0.72
Dismantling formwork	0.16	0.50	0.40
Electrical carcassing	0.83	0.92	0.86
Erection of 15-storey buildings	0.81	0.98	0.97
Erection of one-family houses	0.95	1.00	1.00
Erection of non-traditional houses	0.69	0.74	0.73
Erection of pairs of traditional houses	0.83	0.95	0.93
Erection of plaster board partitions	0.84	0.93	0.87
Erection of tunnel formwork	0.76	0.99	0.97
Facing of outside wall	0.37	0.58	0.42
Fitting of mouldings and window-sills	0.14	0.27	0.18
Fixing of wall measurements	0.97	1.00	1.00
Floor-boarding	0.83	0.89	0.83
Handling of bolts (A)	0.63	0.72	0.79
Handling of bolts (B)	0.96	1.00	0.99
Making of roof trusses on site	0.56	0.87	0.65
Placing of light concrete insulation	0.98	1.00	1.00
Placing of panels (A)	0.72	0.99	0.97
Placing of panels (B)	0.85	0.95	0.94
Production of dwellings	0.98	1.00	1.00
Production of identical dwellings (A)	0.99	0.99	1.00
Production of identical dwellings (B)	0.86	1.00	1.00
Production of raw structure	0.87	0.93	0.91
Sawing and erecting structure (A)	0.68	0.77	0.68
Sawing and erecting structure (B)	0.73	0.80	0.74
Setting up kitchen and bathrooms	0.73	0.84	0.75
Support-pole setting	0.00	0.21	0.20
Formwork element assembly	0.93	0.99	0.99
Wooden shuttering (A)	0.57	0.72	0.65
Wooden shuttering (B)	0.85	0.93	0.90

Downloaded from ascelibrary.org by Mississippi State Univ Lib on 07/07/19. Copyright ASCE. For personal use only; all rights reserved.

TABLE 1. (*Continued*)

(1)	(2)	(3)	(4)
Crane operations (Everett and Slocum 1993)			
Conventional crane operations	0.44	0.52	0.51
Using CRANIUM video system	0.36	0.51	0.51
Precast concrete fabrication (McClure et al. 1980a)			
Applying epoxy	0.00	0.18	0.17
Critical part of mixing epoxy	0.17	0.09	0.19
Hookup, mate, and set down	0.34	0.88	0.55
Mating after epoxy	0.03	0.04	0.05
Removing excess epoxy from ducts	0.23	0.56	0.35
Removing excess epoxy from segment joint	0.12	0.24	0.19
Temporary posttensioning cycle	0.36	0.67	0.50
Tensioning a pair of diagonal bars	0.32	0.47	0.37
Tensioning bars	0.01	0.29	0.29
Segmental bridge construction (McClure et al. (1980b)			
Casting and finishing operation	0.86	0.96	0.95
Form preparation	0.92	0.95	0.94
Void form placement and securement	0.64	0.91	0.85
Tunnelling (Oglesby et al. 1989)			
Tunnel excavation	0.80	0.99	0.98
[Average]	0.61	0.75	0.72
[Standard deviation]	0.30	0.27	0.28

TABLE 2. R^2-Values, Average, and Standard Deviation for 60 Examples

Learning curve model (1)	R^2 average (2)	R^2 standard deviation (3)
Linear x, y	0.47	0.25
Linear x, log y	0.49	0.26
Linear log x, y	0.63	0.30
Linear log x, log y	0.61	0.30
Quadratic x, y	0.67	0.29
Quadratic x, log y	0.67	0.29
Quadratic log x, y	0.72	0.30
Quadratic log x, log y	0.68	0.30
Cubic x, y	0.72	0.28
Cubic x, log y	0.71	0.29
Cubic log x, y	0.75	0.27
Cubic log x, log y	0.72	0.28

(1973) and Thomas et al. (1986) concluded that a cubic model, $\log y = a + b(\log x) + c(\log x)^2 + d(\log x)^3$—called cubic log x, log y in this paper—was superior to the other models they tested for cumulative average data.

This research shows that for the unit data of the 60 examples selected, the new cubic log x, y model provides a slightly higher average R^2-value than the old cubic log x, log y model (0.75 versus 0.72). Some of the possible reasons for the difference in results are the selection of examples, especially with regard to the minimum number of cycles required, the difference be-

Downloaded from ascelibrary.org by Mississippi State Univ Lib on 07/07/19. Copyright ASCE. For personal use only; all rights reserved.

Downloaded from ascelibrary.org by Mississippi State Univ Lib on 07/07/19. Copyright ASCE. For personal use only; all rights reserved.

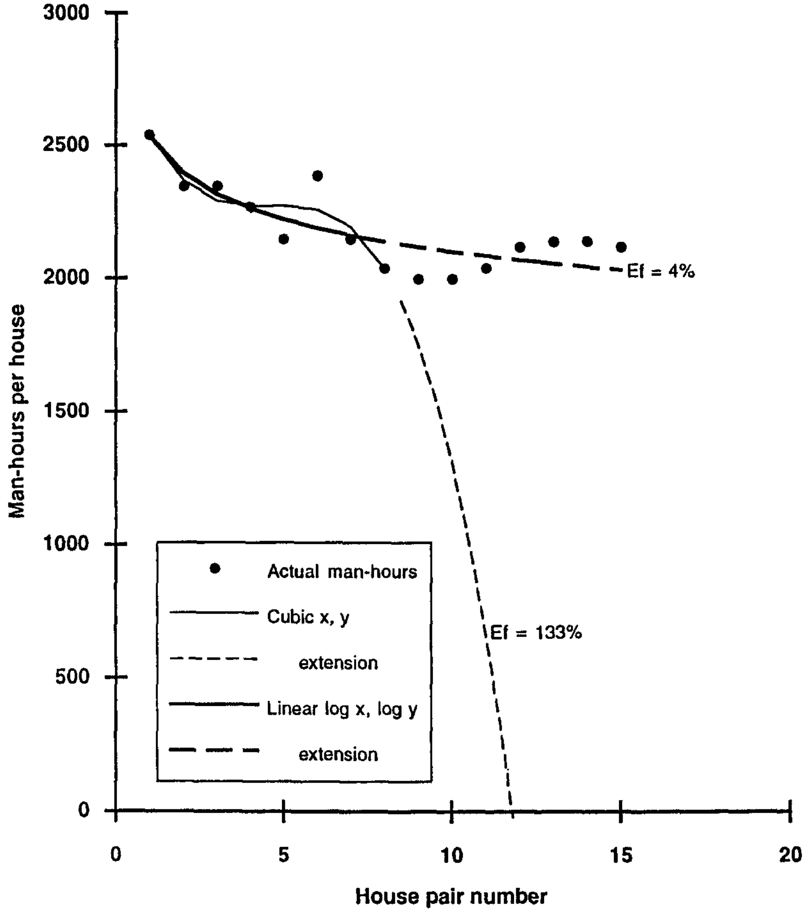

FIG. 1. Best Fit Cubic x, y and Linear log x, log y Models (——) and Extensions (- - - -)

tween unit and cumulative average values, and the fact that all of the various cubic models may not have been evaluated previously. In any case, the differences in the various cubic models' abilities to fit measured data are small and not statistically significant. There is a statistically significant difference ($p < 0.05$) between R^2-values for the cubic curves and R^2-values for linear curves. As is shown subsequently, the issue of which cubic model is best becomes moot when the cubic models are compared to other models as predictors of future performance.

USING LEARNING CURVES TO PREDICT FUTURE PERFORMANCE

The real potential value of learning curves lies in their ability to predict the time/cycle needed to perform future work. Past research, including the first part of this paper, has focused on measuring a learning curve model's ability to correlate to completed work.

No published research investigates which learning curve models are the

J. Constr. Eng. Manage., 1994, 120(3): 603-616

Downloaded from ascelibrary.org by Mississippi State Univ Lib on 07/07/19. Copyright ASCE. For personal use only; all rights reserved.

TABLE 3. R^2_{1-8} **and** $E_{f(9-15)}$ **Values for Example "Erection of pairs of houses (site A)"**

Learning curve model (1)	R^2_{1-8} (cycles 1–8) (2)	$E_{f(9-15)}$ (cycles 9–15) (%) (3)
Linear x, y	0.67	10.1
Linear x, log y	0.67	8.7
Linear log x, y	0.68	4.2
Linear log x, log y	0.67	4.0
Quadratic x, y	0.67	6.9
Quadratic x, log y	0.67	8.0
Quadratic log x, y	0.68	4.7
Quadratic log x, log y	0.67	4.7
Cubic x, y	0.79	133.6
Cubic x, log y	0.79	58.2
Cubic log x, y	0.74	17.0
Cubic log x, log y	0.73	14.7

best predictors for work cycles that have not yet occurred. This paper now presents a methodology for performing such an investigation and gives results based on the 60 construction examples previously cited.

Methods

To assess a learning curve model's ability to predict future performance, the following methodology has been developed. Data from a construction activity with n cycles is collected. For each learning curve model, the equation of the least-squares best-fit curve for the first $n/2$ or m observations is determined. Selection of the midpoint of the data set is somewhat arbitrary, but the sensitivity analysis presented in the following validates this selection. The learning curve is extended according to the best-fit equation beyond m, out to the nth or last cycle. The first m cycles become the historical data, and the remaining cycles become the future data. The correlation between the extended curve and the future data is calculated. For 60 examples, the correlations for all 12 learning curve models are compared to determine which model generally provides the most reliable prediction of future performance.

Mathematically, m = number of cycles to be fitted; k = number of cycles to be predicted; and $n = m + k$ = total number of cycles. Pearson's coefficient of determination, R^2, used in the section headed "Correlation of Mathematical Models with Historical Data," is not valid for correlating points with best-fit curves outside the range of points used to determine the best-fit curve, so the following procedure is used.

A statistic to be called E_f gives the average percentage difference between the values $y'_{m+1}, y'_{m+2}, \ldots, y'_{m+k=n}$ found on the extension of the best-fit curve for cycles $1, 2, \ldots, m$, and the actual measured values $y_{m+1}, y_{m+2}, \ldots, y_{m+k=n}$.

$$E_f = \frac{\sum_{i=1}^{k} \frac{|y'_{m+i} - y_{m+i}|}{y_{m+i}}}{k} \times 100 \qquad (17)$$

J. Constr. Eng. Manage., 1994, 120(3): 603-616

TABLE 4. Average R^2 and E_f-Values for 60 Examples

Learning curve model (1)	Average R^2 (2)	Average R^2_{1-m} (3)	Average $E_{f(m+1-n)}$ (%) (4)
Linear x, y	0.47	0.60	81.6
Linear x, log y	0.49	0.61	32.2
Linear log x, y	0.63	0.67	38.4
Linear log x, log y	0.61	0.65	20.0
Quadratic x, y	0.67	0.70	192.7
Quadratic x, log y	0.67	0.70	9,758.9
Quadratic log x, y	0.72	0.73	52.0
Quadratic log x, log y	0.68	0.71	26.9
Cubic x, y	0.72	0.77	1,591.2
Cubic x, log y	0.71	0.77	5,664.1
Cubic log x, y	0.75	0.80	125.9
Cubic log x, log y	0.72	0.77	506.6

E_f can range from 0% indicating a perfect correlation between the extended best-fit curve and the actual data to large positive values indicating no correlation.

Results

To illustrate the procedure used in this study, the first activity in Table 1, "Erection of pairs of houses (site A)," is presented. The example is of "total time consumption in the serial erection of pairs of one-family houses" in the United Kingdom (*Effect* 1965). Fig. 1 shows a scatter plot of the man-hours per house versus the number in order of construction of pairs of houses. The number of available observations is $n = 15$.

A least-squares best-fit curve of the form of each of the mathematical models was determined for the first $m = 8$ ($\approx n/2$) cycles. The coefficients of correlation for the first eight cycles for each model are shown as R^2_{1-8} (indicating R^2 for observations 1 through 8) in Table 3, column 2. In this example the cubic x, y and cubic x, log y models give the highest R^2_{1-8} value (0.79). The best-fit curves for each model were then extended through the ninth to 15th cycles. The average percentage error, $E_{f(9-15)}$, between the observed data and the extended curves is shown in Table 3, column 3. The lowest $E_{f(9-15)}$ value (4.0%) is generated by the linear log x, log y model. The cubic x, y and cubic x, log y models, which give the highest R^2_{1-8}, are poor predictors for cycles nine through 15, with $E_{f(9-15)}$ values of 133.6% and 58.2%, respectively.

The fundamental flaw with the cubic models can be seen qualitatively in Fig. 1. The heavy and light solid lines show the least-squares best-fit curves for the first eight cycles for the linear log x, log y and cubic x, y models, respectively. For the first eight cycles, the cubic x, y model ($R^2_{1-8} = 0.79$) gives a higher correlation than the linear log x, log y model ($R^2_{1-8} = 0.67$).

The dashed lines in Fig. 1 show the continuations of the best-fit curves beyond the range of the observations used to generate the curves. The extended cubic x, y curve turns sharply downward into the meaningless range of negative man-hours/house. Clearly such a curve does not provide a useful prediction of future performance, even for one or two cycles. The linear log x, log y curve, however, correlates well to the actual data beyond

Downloaded from ascelibrary.org by Mississippi State Univ Lib on 07/07/19. Copyright ASCE. For personal use only; all rights reserved.

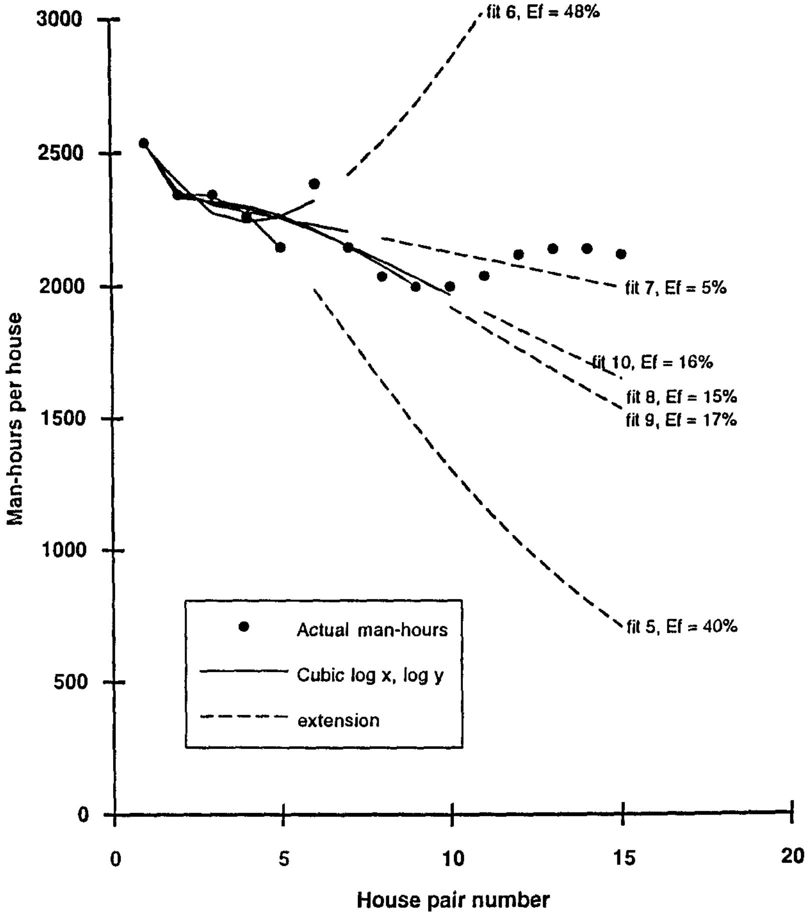

FIG. 2. Sensitivity of Extensions of Cubic log x, log y Model to Number of Points for Best-Fit Curve

the range used to generate the curve. This example clearly demonstrates the unreliability of one cubic learning curve model for predicting future performance based on past experience. In other examples, not shown here, the extended cubic curves eventually turn downward into negative values or upward into very large positive values.

This analysis was performed for all 12 learning curve models on all 60 of the construction activities listed in Table 1. The average value for R^2_{1-m} (R^2 for the first half of the cycles) over 60 examples is shown in Table 4, column 3. The average E_f for the remaining cycles ($m + 1$ to n) is shown in column 4. The R^2_{1-m} values in column 3 show the same trend as the R^2-values for all observations in column 2. (Table 4, column 2 is identical to Table 2, column 2.) The cubic models in general are the most reliable, the quadratic models are slightly less reliable, and the linear models are the least reliable indicators of past performance. The trend for the E_f-values in column 4 is opposite that of the R^2-values. The most reliable predictors of future performance are the linear models, and the least reliable predictors are the

J. Constr. Eng. Manage., 1994, 120(3): 603-616

Downloaded from ascelibrary.org by Mississippi State Univ Lib on 07/07/19. Copyright ASCE. For personal use only; all rights reserved.

Downloaded from ascelibrary.org by Mississippi State Univ Lib on 07/07/19. Copyright ASCE. For personal use only; all rights reserved.

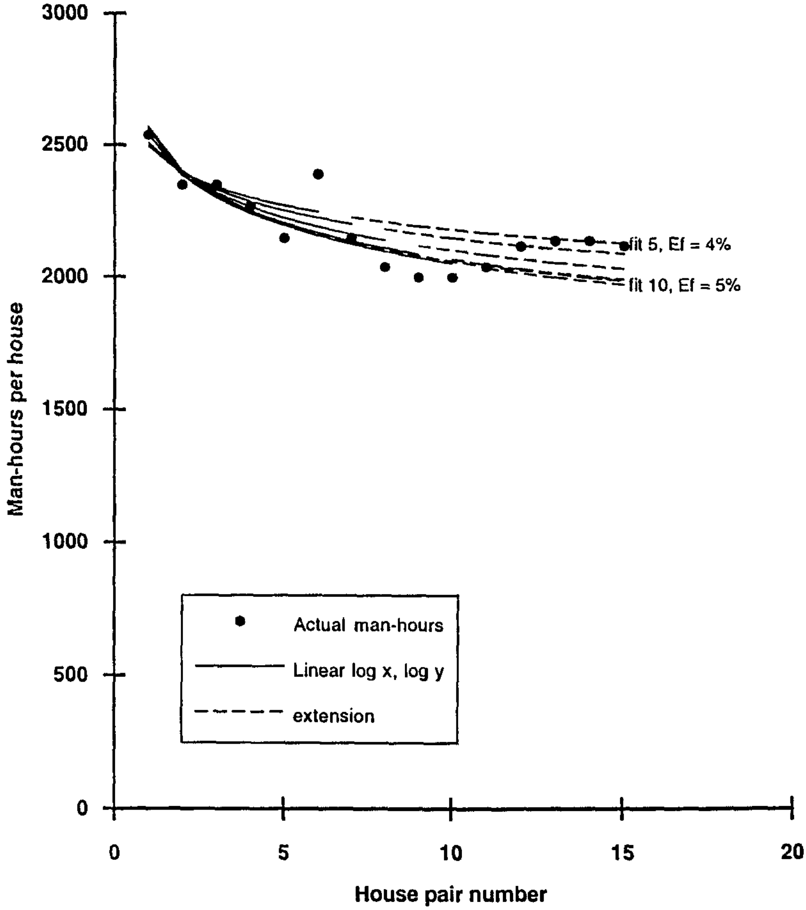

FIG. 3. Sensitivity of Extensions of Linear log x, log y Model to Number of Points for Best-Fit Curve

cubic models. The single most reliable predictor is the linear log x, log y model with $E_f = 20.0\%$. The cubic log x, y model, which has the highest R^2-values, has $E_f = 125.9\%$. The cubic model described by Carlson (1973) and Thomas et al. (1986), the cubic log x, log y model, has $E_f = 506.6\%$. For the cubic log x, log y model, one example gives a very high E_f. If that one example is ignored, $E_f = 39.7\%$, which is still significantly higher ($p < 0.05$) than E_f for the linear log x, log y model.

The quadratic log x, log y model also appears to be a reasonable predictor ($E_f = 26.9\%$), but for long-term prediction the quadratic models suffer the same problem as the cubic models. The extended curves eventually continue upward or downward toward unreasonable values. The quadratic models just take more cycles for the divergence to develop.

As mentioned earlier, selection of the midpoint of the set of observed data as the dividing point between past and future performance was arbitrary. A sensitivity analysis was performed to evaluate the effect of selecting other dividing points in the example of "Erection of pairs of houses (site

J. Constr. Eng. Manage., 1994, 120(3): 603-616

A)" (*Effect* 1965). Fig. 2 shows the same actual man-hours per house pair as Fig. 1. The solid lines show the best-fit curves for the cubic log x, log y model fit to the first 5 through 10 cycles. The dashed lines show the extensions of the best-fit lines through cycle 15. At the end of the dashed lines are labels indicating how many cycles were used to derive the best-fit curve. For example, the lowest dashed line, labeled "fit 5," is the extension of the best-fit curve for cycles one through five. Fig. 2 shows how sensitive the extended cubic curves are to the number of cycles used to derive the extended curves. The figure also shows how poorly the cubic model predicts future performance in this example. Only one of the dashed lines, the extension from cycles one through seven, gives a reasonable prediction of future performance with $E_f = 14.7\%$. The other dashed lines are all over the chart.

Fig. 3 shows the same sensitivity analysis for the linear log x, log y model. For clarity, only two of the individual lines are labeled, but it is clear that this model is not very sensitive to the number of cycles used to derive the extended best-fit curves. All of the curves give good predictions of future performance with E_f-values between 4% and 5%.

Conclusion

This paper has shown that various forms of cubic learning curve models generally give the highest correlation to completed repetitive construction activities. Despite their high correlation to completed activities, the cubic models are poor predictors of future performance and should not be used to estimate performance beyond known historical data.

This paper has shown that for the unit data of the 60 construction examples selected, the linear log x, log y learning curve model ($\log y = a + b \log x$) is the most reliable predictor of future performance. This model is the original learning curve first described in the 1930s (Wright 1936), but it was never tested for its predictive capacity in construction field operations. The cubic models fit the early data best and the linear models fit the later data best. This is in agreement with the Stanford B (*An improved* 1949) and piecewise models.

The bottom line is that if the construction manager knows the time or cost of each cycle in the first half of an evolving process, after the Stanford B hump or operation learning phase, the manager should not be fooled into believing a cubic learning curve model will reliably predict future performance. In the long run, the best predictor will be a linear learning curve model. Early data can be used to predict the constants in the long-term curve.

FUTURE RESEARCH

The methodology developed in this research for measuring future performance is based on using half of the available data and projecting learning curve models out to the final cycle. More research is needed to determine the minimum number of cycles necessary before a reliable projection can be made and how far into the future the extended curve remains reliable.

The examples used in this research cover a wide range of construction work, from macro activities such as building entire houses, to micro activities associated with precast concrete construction. As part of this research, similar activities were grouped and analyzed to see if there were any obvious differences among different types of construction. No trends were apparent.

J. Constr. Eng. Manage., 1994, 120(3): 603-616

Downloaded from ascelibrary.org by Mississippi State Univ Lib on 07/07/19. Copyright ASCE. For personal use only; all rights reserved.

More research is needed into the nature of repetitive work and the managerial actions that make it subject to learning effects. Only then will construction managers be able to estimate, before the field operation commences, learning curves and future performance in construction.

APPENDIX. REFERENCES

Carlson, J. G. H. (1973). "Cubic learning curves: precision tool for labor estimating." *Manufacturing Engrg. and Mgmt.*, 67(11), 22–25.

Effect of repetition on building operations and processes on site. (1965). *Rep. ST/ECE/HOU/14*, United Nations Committee on Housing, Building, and Planning, United Nations, New York, N.Y.

Everett, J. G., and Slocum, A. H. (1992). "CRANIUM: device for improving crane safety and productivity." *J. Constr. Engrg. and Mgmt.*, ASCE, 119(1), 1–17.

Farghal, S. H., and Everett, J. G. (1993). "Learning curves as time and cost predictors for construction field operations." *Tech. Rep. 94-21*, Center for Construction Engineering and Management, Dept. of Civ. and Envir. Engrg., University of Michigan, Ann Arbor, Mich.

Hines, W. W., and Montgomery, D. C. (1990). *Probability and statistics in engineering and management science.* J. Wiley and Sons, New York, N.Y.

An improved rationale and mathematical explanation of the progress curve in airframe production. (1949). Stanford Research Institute, Stanford, Calif.

McClure, R. M., Thomas, H. R., and Hehler, J. E. (1980b). "Construction of experimental segmental bridge." *Rep. to the Commonwealth of Pennsylvania Department of Transportation*, Pennsylvania Transportation Institute, Philadelphia, Pa.

McClure, R. M., Willenbrock, J. H., and Henderson, J. D. (1980a). "Fabrication of segments for an experimental segmental bridge." *Rep. to the Commonwealth of Pennsylvania Department of Transportation*, Pennsylvania Transportation Institute, Philadelphia, Pa.

Oglesby, C. H., Parker, H. W., and Howell, G. A. (1989). *Productivity improvement in construction.* McGraw-Hill, New York, N.Y.

Thomas, H. R., Mathews, C. T., and Ward, J. G. (1986). "Learning curve models of construction productivity." *J. Constr. Engrg. and Mgmt.*, ASCE, 112(2), 245–258.

Wright, T. P. (1936). "Factors affecting the cost of airplanes." *J. Aeronautical Sci.*, (Feb.), 124–125.

J. Constr. Eng. Manage., 1994, 120(3): 603-616

Downloaded from ascelibrary.org by Mississippi State Univ Lib on 07/07/19. Copyright ASCE. For personal use only; all rights reserved.