Statistical Model for Schedule Prediction: Validation in a Housing-Cooperative Construction Database

N. Rudeli[1]; A. Santilli, Ph.D.[2]; I. Puente, Ph.D.[3]; and E. Viles, Ph.D.[4]

Abstract: There are often considerable differences between the planned schedule for a construction project and what later develops during actual construction. This paper introduces an innovative approach that uses Markov Chain models to support predictions during earned value analyses. A statistical model was developed to predict possible deviations in a project schedule and the future progress of a project. This model, based on Markov chains, uses data from the past to adjust future predictions. A case study was built from a database of 90 housing cooperative construction projects and was validated in 12 more projects. A cross validation of three interactions was also carried out, obtaining an error of 2.38% in the prediction of future progress and an error of 4.29% in the prediction of construction timing. The innovative prediction model presented in this paper contributes to the management body of knowledge by introducing a new tool for the management and control of construction timing. The method presented improves construction management because it predicts future deviations in schedules with reduced errors and determines total deviation from a construction schedule with great precision. This allows better control over work timing and represents important input in determining strategies and future actions. **DOI: [10.1061/(ASCE)CO.1943-7862.0001396](https://doi.org/10.1061/(ASCE)CO.1943-7862.0001396).** © 2017 American Society of Civil Engineers.

Author keywords: Prediction; Schedule; Earned schedule; Earned value management; Cost and schedule.

Introduction

The problem of project schedule overruns in the construction industry is a global one (Sambasivan and Soon 2007). Alkhathami (2004) defined a schedule overrun as extra time required to finish a given construction project beyond its original planned duration. Assaf and Al-Hejji (2006) defined a schedule overrun as time either beyond the completion date specified in a contract or beyond the date that the parties agreed on for delivery of a project. It is basically a project slipping over its planned schedule and is considered a common problem in construction projects worldwide. Delivery of a project within the contract-stipulated time is one of the yard sticks of a successful project. Despite its proven importance, however, construction projects that fail to achieve their objectives are not uncommon (Memon et al. 2012).

The construction process can be divided into three important phases: conception, design, and construction. Usually, the vast majority of project delays occur during the construction phase, where many unforeseen factors are always involved (Chan and Kumaraswamy 1997). This is why study of the construction phase and development of a delay prediction tool are of relevance.

One of the most popular and accepted tools for controlling and monitoring construction timing is schedules (Aristondo 2003). This paper presents a statistical model for predicting schedule deviation and validates it in a Uruguayan case study using a database of 90 housing cooperative construction projects plus 12 more projects for cross validation. The model allows managers to take into account historic behaviors for the prediction of schedule deviations in order to make corrections and take management decisions during the construction phase.

Literature Review

Earned value analysis (EVA) is an accepted theoretical technique advocated for the control of projects. This section describes the literature relevant to EVA and the state of the art relating to Markov chains and their use in construction schedule prediction.

Earned Value Management

Earned value management (EVM) is a methodology used to measure and communicate the actual physical progress of a project that integrates the three critical elements of project management: scope, time, and cost. It takes into account work done, time taken, and costs incurred to complete the project, and it helps to evaluate and control project risks (Fleming and Koppelman 2010). According to Lipke (2003), EVM is a technique for controlling the performance of a project by comparing the amount of work up to a certain moment with estimates made before the project was started. In this way, there is a measure of the amount of work finished and the amount of work remaining. Moreover, EVM allows the efficiency of the original schedule to be determined.

The concept of earned schedule (ES) is an extension of the concept of EVM, and it refers to the amount of additional time needed to reach the established progress goals when construction is behind schedule. If construction is ahead of schedule, it can be defined as the range of time in which construction can make no progress

[1]Ph.D. Candidate, Dept. of Housing Programs, Agencia Nacional de Vivienda, Cerrito 400, 11000 Montevideo, Uruguay.

[2]Assistant Professor, Dept. of Civil Engineering, Universidad de Montevideo, Luis P. Ponce 1307, 11300 Montevideo, Uruguay (corresponding author). E-mail: asantilli@um.edu.uy

[3]Professor, Dept. of Mechanical Engineering and Materials, TECNUN Escuela de Ingenieros, Universidad de Navarra, Manuel de Lardizábal 13, 20018 San Sebastián, España.

[4]Associate Professor, Dept. of Industrial Organization, TECNUN Escuela de Ingenieros, Universidad de Navarra, Manuel de Lardizábal 13, 20018 San Sebastián, España.

© ASCE 04017083-1 J. Constr. Eng. Manage.

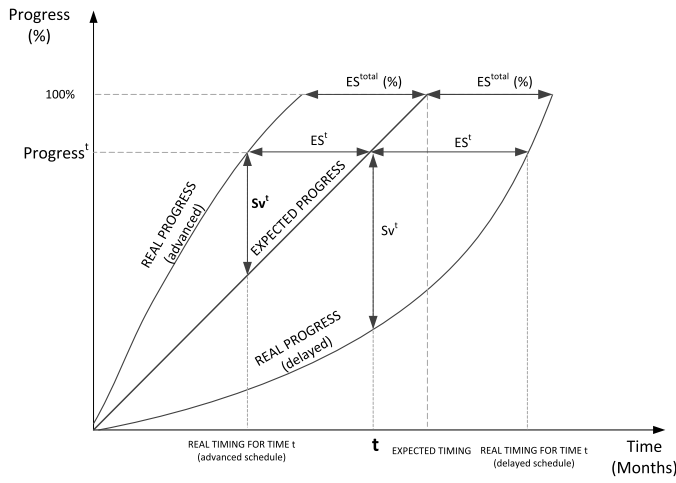J. Constr. Eng. Manage., 2017, 143(11): 04017083

Fig. 1. Definition of the ES indicator

without producing a delay in the schedule. This concept is shown in Fig. 1. Therefore, the indicator used for this work, denoted ES^{total} and measured as the percentage of expected progress, indicates the amount of additional time needed to reach 100% progress (i.e., project completion). If construction is ahead of the original schedule, ES^{total} is measured as the percentage of time (in terms of expected progress) that construction may take without falling behind schedule.

Earned schedule is used to determine the efficiency of schedules. For example, Lipke (2003) carried out research with real data from 16 building projects in the United States, comparing the efficiency of some EVM methods and ES in estimating and predicting schedules. Anbari (2003), Fleming and Koppelman (2010), and Vanhoucke (2010, 2014), studied EVM and ES, showing their applicability in measuring scheduling efficiency. Liberatore et al. (2001) and De Marco and Narbaev (2013) used EVM indicators to evaluate construction project performance. Forbes and Ahmed (2010) validated the use of EVM indicators for several studies in civil construction.

Given the importance of some EVM indicators in the control and management of construction, this paper presents a statistical model that predicts and adjusts construction schedules, making it possible to predict total deviation from the schedule's ES^{total} and its deviation percentage at a specific time of construction, Sv^t (Fig. 1) for future construction. The model is based on Markov chains, which not only make it possible to adjust predictions while construction is taking place but also use data collected during construction to adjust future predictions.

Stochastic Process and Markov Chains

A series of observations $X_1, X_2, \ldots, X_n$ is called a stochastic process if the values of these observations cannot be exactly predicted but the probabilities for different possible values can be specified at any time (Manjia et al. 2013). For example, X_1 defines the initial state and X_n defines the state at time n. For each possible value of the initial state s_1, and for each succeeding value of the states X_n ($n = 2, 3, \ldots, k$), it can be said that in a stochastic process the probability of an event at moment $n + 1$ can be defined as in Eq. (1) if what happened moments before is known

$$P(X_{n+1} = s_{n+1}|X_1 = s_1, X_2 = s_2, \ldots, X_n = s_n) \quad (1)$$

A Markov chain corresponds to a specific class of stochastic process in the field of probabilistic models (Bertsekas and Tsitsiklis 2000; Ross 2003) in which the current state X_n and

the previous states $X_1, \ldots, X_{n-1}$ are known. Therefore, for $n = 1, 2, \ldots, k$ and for any succession of states $s_1, \ldots, s_{n+1}$, Eq. (1) can be redefined as

$$P(X_{n+1} = s_{n+1}|X_1 = s_1, X_2 = s_2, \ldots, X_n = s_n)$$
$$= P(X_{n+1} = s_{n+1}|X_n = s_n) \quad (2)$$

Markov chains have the probabilities of stationary transitions if for each pair of states s_i and s_j there is a probability of transition p_{ij}, such that what is established in Eq. (3) holds

$$P(X_{n+1} = s_j|X_n = s_i) = p_{ij} \quad \text{for } n = 1, 2, \ldots, k \quad (3)$$

For a stochastic process with k possible states $s_1, \ldots, s_k$, denoted $p_{ij}^{(m)} = P(X_{n+1} = s_j|X_n = s_i)$, where $p_{ij}^{(m)}$ is the element in row i and column j of the transition matrix $P^{(m)}$ with ($m = 2, 3, \ldots, k - 1$), the matrix m of transition $P^{(m)}$ of the Markov chain is defined as

$$P^{(m-1)} = \begin{array}{c|cccccc} \text{State} & 0 & 1 & 2 & \ldots & k \\ \hline 0 & p_{00}^{(m)} & p_{01}^{(m)} & p_{02}^{(m)} & \ldots & p_{0k}^{(m)} \\ 1 & p_{10}^{(m)} & p_{11}^{(m)} & p_{12}^{(m)} & \ldots & p_{1k}^{(m)} \\ 2 & p_{20}^{(m)} & p_{21}^{(m)} & p_{22}^{(m)} & \ldots & p_{2k}^{(m)} \\ \ldots & \ldots & \ldots & \ldots & \ldots & \ldots \\ k & p_{k0}^{(m)} & p_{k1}^{(m)} & p_{k2}^{(m)} & \ldots & p_{kk}^{(m)} \end{array} \quad (4)$$

There are several cases in which Markov chains have been used for predictions. Yuan (1999) used Markov chains for predicting protein subcellular locations; Logofet and Lesnaya (2000) used them for an ecological prediction model; and Zhu et al. (2002) used them for prediction in adaptive websites. However, in the literature no evidence of Markov chains being used in civil construction management or earned value analysis has been found, and for this reason the present research can be considered valuable and innovative.

Statistical Model for Forecasting Schedule Deviations Based on Markov Chains

For the construction of the prediction model, some formal aspects had to be taken into account. Because projects have different construction times, to standardize and compare data the values of estimated completion time needed to be discretized. Thus the expected construction time was divided into 10 equal parts over ten tenths (Fig. 2) so that two projects with different construction times could be compared based on the percentage of progress up to a certain time.

In the model, x_t indicates the state of a construction project at moment t. This defines a stochastic process that corresponds to the sequence $x_1, x_2, x_3, \ldots, x_{10}$, which represents the gradual progress percentage of a construction project from one state to another, where the value of x_t usually relies on the previous values in the sequence x_{t-1}. As time goes on, the state changes in probabilistic terms, represented by state transition probabilities. The hypotheses contemplated in the Markovian model for this stochastic process are as follows:
- There are a finite number of states describing the behavior of construction projects; for the prediction model, there are 9 transition states among the 10 tenths of a construction project and an additional transition state to predict ES^{total} when the gradual

© ASCE   04017083-2   J. Constr. Eng. Manage.

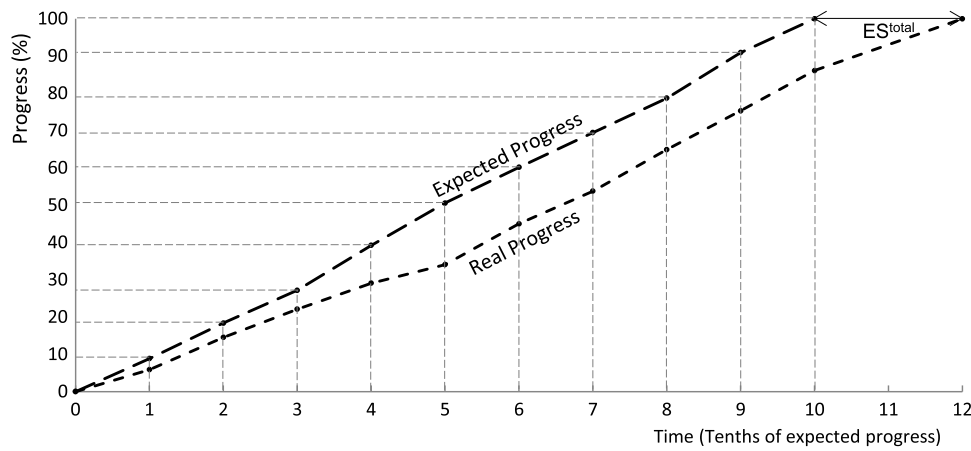J. Constr. Eng. Manage., 2017, 143(11): 04017083

Fig. 2. Example of real progress and expected progress for a construction project

progress percentage of the last tenth is known, totaling 10 transition matrices;
- There is a known distribution of probabilities at the beginning of the study's projection ($t = 1$); and
- The transition from a current state to a future one depends on the current state (Markovian property), which means that progress in tenth x_{n+1} is influenced by the progress made in tenth x_n.

The model for predicting the gradual percentage of construction progress using Markov chains consists of building 10 transition matrices between states t and $t + 1$ according to the selected database. In the matrix's rows and columns, incremental progress of 1% is considered from one period of study to the next. Thus there is a probability that in tenth m, construction will progress between 2 and 3% if it progressed between 0 and 1% in tenth $m - 1$, and that probability is in place $p_{2\text{–}3\%/0\text{–}1\%}^{(m)}$ in the $P^{(m-1)}$ transition matrix, as shown in Eq. (5)

$$
P^{(m-1)} =
\begin{array}{c|ccccc}
\text{State} & 0\text{–}1\% & 1\text{–}2\% & 2\text{–}3\% & \ldots & 15\text{–}16\% \\
\hline
0\text{–}1\% & p_{0\text{–}1\%/0\text{–}1\%}^{(m)} & p_{1\text{–}2\%/0\text{–}1\%}^{(m)} & p_{1\text{–}2\%/15\text{–}16\%}^{(m)} & \ldots & p_{15\text{–}16\%/0\text{–}1\%}^{(m)} \\
1\text{–}2\% & p_{0\text{–}1\%/1\text{–}2\%}^{(m)} & p_{1\text{–}2\%/1\text{–}2\%}^{(m)} & p_{2\text{–}3\%/1\text{–}2\%}^{(m)} & \ldots & p_{15\text{–}16\%/1\text{–}2\%}^{(m)} \\
2\text{–}3\% & p_{0\text{–}1\%/2\text{–}3\%}^{(m)} & p_{1\text{–}2\%/2\text{–}3\%}^{(m)} & p_{2\text{–}3\%/2\text{–}3\%}^{(m)} & \ldots & p_{15\text{–}16\%/2\text{–}3\%}^{(m)} \\
\ldots & \ldots & \ldots & \ldots & \ldots & \ldots \\
15\text{–}16\% & p_{0\text{–}1\%/15\text{–}16\%}^{(m)} & p_{1\text{–}2\%/15\text{–}16\%}^{(m)} & p_{2\text{–}3\%/15\text{–}16\%}^{(m)} & \ldots & p_{15\text{–}16\%/15\text{–}16\%}^{(m)}
\end{array}
\tag{5}
$$

In this way, the transition matrices $P^{(m)}$ (16×16) take into account a gradual progress of 1% from one period to another (varying between 0 and 16%). These values are calculated taking into consideration previous experiences, where at no time did the progress of a project exceed 15% between periods of study.

Subsequently, in order to predict the behavior of a new construction project, knowing the gradual progress percentage in one of the tenths is enough to enable the prediction of the following tenth with the corresponding transition matrix. The progress prediction is obtained as the weighted mean between the row containing the advance of the previous month of the corresponding transition matrix and Vector $(0.5, 1.5, 2.5, 3.5, \ldots, 15.5\%)$, which contains the gradual progress middle point for each of the k intervals considered in the method. If it is necessary to determine ES^{total} from the $P^{(10)}$ transition matrix, a multiplication operation has to be carried out between the row related to the monthly progress of Tenth 10 and Vector: $(4, 12, 20, 28, \ldots, 124\%)$.

Statistical Model Validation

Sample Characterization

The indicator ES^{total} refers to the additional amount of time needed to complete construction as a proportion of the expected progress. However, if progress is greater than expected, it is measured as the percentage of time construction may advance without any delay (and compared with expected progress) For example, the expected construction time for the construction project in Fig. 2 is 24 months, which represents progress up to Tenth 10. The actual construction duration was 29 months—5 months more than expected—which is 20.83% more than the initially expected time. Consequently, the actual 100% progress value on the x-axis shown in tenths of expected progress is an additional 20.83% from the last expected tenth—that is, 12.08.

The database of 90 housing construction projects includes both planned and executed schedules. The Markovian matrices

© ASCE 04017083-3 J. Constr. Eng. Manage.

J. Constr. Eng. Manage., 2017, 143(11): 04017083

Table 1. Summary of Real Progress and ES^{total} Data Measured in Tenths of Expected Progress for the 90 Construction Projects Studied

Index (%)	ES^{total}	1t/10	2t/10	3t/10	4t/10	5t/10	6t/10	7t/10	8t/10	9t/10	10t/10
Mean	34.68	5.50	13.20	21.30	29.31	38.06	47.85	57.35	66.36	75.34	83.47
Deviation	27.74	4.17	6.25	8.52	10.36	12.31	14.00	15.17	15.50	14.75	13.64
Maximum	120.00	17.62	37.62	54.31	65.31	73.52	83.19	96.67	99.99	100.00	100.00
Minimum	−6.25	0.00	1.39	3.88	5.99	11.31	17.48	22.05	25.97	30.85	37.96

Table 2. Summary of Results for Management Indicators EVM and ES Applied to the Database

Index	SV (%)	SPI	ES^{total} (%)[a]
Mean	−9.73	0.83	34.68
Deviation	9.92	0.26	27.74
Upper limit 99%	−7.68	0.89	40.41
Lower limit 99%	−11.78	0.78	28.95

[a]Actual timing average = 33 months; expected timing average = 24.62 months.

presented in the Supplemental Data are constructed with the executed schedules, taking into account the procedure described in the section "Statistical Model for Forecasting Schedule Deviations Based on Markov Chains." In this way, data were collected for ES^{total} and the actual progress for each tenth of the database. Table 1 summarizes mean progress for each tenth in the database, mean ES^{total}, and calculated standard deviation for the data.

The database was also characterized by measuring certain management indicators that are provided by EVM and comparing them against the ES^{total} indicator. In EVM the planned value (PV) is defined as the amount of time expected to be spent on the completion of construction at any point in the original schedule's timing (in this case, the percentage of construction progress). Earned value (EV), then, is the sum of construction progress up to a certain point, also measured as a percentage of construction progress.

All schedule performance indicators used in this research refer to the percentage of construction progress and not to monetary amounts. The database is made up of turnkey projects, where it makes more sense to evaluate the construction completion percentage (Alsakini et al. 2004; Flatscher 2015; Assaf and Al-Hejji 2006). Regardless of the methodology, the results are still valid for use with the monetary measurement of EVM indicators depending on the database selected. They are valid when the database refers to schedules expressed as a percentage of work progress and schedules expressed as the cost of construction.

The two indicators used to characterize progress, schedule variance (SV) and the schedule performance index (SPI), are defined by Eqs. (6) and (7), respectively

$$SV = EV - PV \qquad (6)$$

$$SPI = \frac{EV}{PV} \qquad (7)$$

The SV indicator (the difference in the percentage of progress against the expected plan) and the ES indicator are shown in Fig. 1. An SV indicator below zero indicates that construction is behind the original schedule, whereas an SV indicator above zero indicates that construction has advanced beyond the original schedule. As for ES^{total}, a value above zero represents the delay percentage according to the expected timing (Fig. 1). Finally, the SPI indicator indicates the progress percentage according to the expected plan. For construction to proceed efficiently, this indicator must be higher than 1; a value lower than 1 indicates a delay relative to the planned value.

The results of analysis of the 90 construction projects in the database are provided in Table 2: mean values for the three previously defined indicators (SV, SPI, and ES^{total}) and both actual and initially expected timing. For the three management indicators, the standard deviation and upper and lower limits for a confidence interval (CI) of 99% are also provided.

From the case study, it can be deduced that the SV indicator is −9.73%; in other words, the construction of the housing cooperatives under study fell 10% behind the initially expected schedule. The SPI indicator is 0.83, which reaffirms the hypothesis that cooperative construction tends to fall behind expected timing. Regarding ES^{total}, the case study shows that in general this project needs 34.68% more time (measured as a percentage of expected construction time) to reach 100% progress.

Research Method

Ten transition matrices were constructed for the database in order to predict future schedule deviations and validate the forecasting method presented. The proposed method allows prediction of the gradual progress percentage in one tenth if the advance of the previous tenth and the corresponding transition matrix are known. For instance, knowing the gradual progress percentage of construction in Tenth 4 allows the gradual progress percentage for Tenth 5 to be predicted using the $P^{(4)}$ transition matrix and the known progress. If the gradual progress percentage of a construction project at Tenth 4 is between 2 and 3%, the progress prediction for Tenth 5 is obtained as the weighted mean between Row 3 of the $P^{(4)}$ transition matrix and Vector (1×16). Then, if the $P^{(4)}$ matrix shown in Table 3 is considered, predicting progress for the fifth tenth of the project only requires calculating the weighted mean of the matrix's third row [Eq. (8)] and Vector (1×16), which contains the considered averages for gradual progress [Eq. (9)]

$$[0.0 \quad 0.0 \quad 0.5 \quad 0.0 \quad 0.0 \quad 0.0 \quad 0.5 \quad 0.0 \quad 0.0 \quad 0.0 \quad 0.0 \quad 0.0 \quad 0.0 \quad 0.0 \quad 0.0 \quad 0.0] \qquad (8)$$

$$[0.5 \quad 1.5 \quad 2.5 \quad 3.5 \quad 4.5 \quad 5.5 \quad 6.5 \quad 7.5 \quad 8.5 \quad 9.5 \quad 10.5 \quad 11.5 \quad 12.5 \quad 13.5 \quad 14.5 \quad 15.5] \qquad (9)$$

© ASCE — 04017083-4 — J. Constr. Eng. Manage.

J. Constr. Eng. Manage., 2017, 143(11): 04017083

Table 3. Transition Matrix between States X_4 and X_5: Gradual Progress Percentage

$P^{(4)}$	0–1%	1–2%	2–3%	3–4%	4–5%	5–6%	6–7%	7–8%	8–9%	9–10%	10–11%	11–12%	12–13%	13–14%	14–15%	15–16%
0–1%	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
1–2%	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
2–3%	0.00	0.00	0.50	0.00	0.00	0.50	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
3–4%	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
4–5%	0.00	0.00	0.00	0.14	0.14	0.14	0.00	0.29	0.14	0.14	0.00	0.00	0.00	0.00	0.00	0.00
5–6%	0.00	0.00	0.11	0.00	0.11	0.22	0.22	0.11	0.00	0.22	0.00	0.00	0.00	0.00	0.00	0.00
6–7%	0.00	0.00	0.00	0.00	0.08	0.08	0.38	0.15	0.23	0.00	0.00	0.08	0.00	0.00	0.00	0.00
7–8%	0.00	0.00	0.00	0.13	0.00	0.25	0.00	0.13	0.50	0.00	0.00	0.00	0.00	0.00	0.00	0.00
8–9%	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.33	0.07	0.20	0.20	0.13	0.00	0.00	0.07	0.00
9–10%	0.00	0.10	0.00	0.00	0.00	0.00	0.00	0.00	0.10	0.30	0.20	0.20	0.00	0.00	0.10	0.00
10–11%	0.00	0.00	0.00	0.00	0.14	0.00	0.29	0.00	0.00	0.29	0.00	0.14	0.14	0.00	0.00	0.00
11–12%	0.00	0.00	0.00	0.00	0.00	0.00	0.14	0.14	0.00	0.14	0.00	0.00	0.29	0.00	0.14	0.14
12–13%	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.25	0.25	0.00	0.00	0.50
13–14%	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00
14–15%	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00
15–16%	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.25	0.00	0.25	0.00	0.25	0.00	0.25

Table 4. Actual Gradual Progress as a Percentage at Tenth 4 of Construction for the 12 Verifying Construction Projects

Interval	O1	O2	O3	O4	O5	O6	O7	O8	O9	O10	O11	O12
4/10t	10.17	6.33	9.08	9.71	4.15	7.97	9.22	10.99	8.91	5.06	13.21	8.15

For this example, X_5 is calculated in Eq. (10)

$$X_5 = 0.50 \times 2.50\% + 0.50 \times 5.50\% = 4.00\% \qquad (10)$$

As mentioned previously, the methodology to determine future progress by knowing past monthly progress is based on building 10 transition matrices between states. For the database, transition matrices between the states $x_1, x_2, x_3, \ldots, x_{10}$ have been built; these are presented in the Supplemental Data. In each of these matrices, the probability of transition from one state to the next based on the behavior of the 90 construction projects is shown. Following that, another 12 projects were taken into account in order to verify and validate the methodology. There was an expected schedule for these 12 projects and real percentage values for gradual progress (in each tenth).

The expected monthly prediction was made with these facts and using Markov matrices for each state.

Results

For illustrative purposes, the validation is presented for one of the 12 verifying projects using the $P^{(4)}$ transition matrix between the gradual progress percentage values for the fourth and fifth tenths. The $P^{(4)}$ transition matrix is then considered. Real progress values obtained for the 12 verifying projects, O1, O2, O3, ..., O12, in the fourth tenth of each project (Table 4) should also be considered.

To predict progress in the fifth tenth of Cooperative O1, progress at the fourth tenth must be known and the weighted mean of Row 10 of the $P^{(4)}$ matrix [Eq. (11)] and Vector (1×16) [Eq. (9)] must be calculated. State X_5 is obtained through Eq. (12)

$$[0.0 \quad 0.0 \quad 0.0 \quad 0.0 \quad 0.14 \quad 0.0 \quad 0.29 \quad 0.0 \quad 0.0 \quad 0.29 \quad 0.0 \quad 0.14 \quad 0.14 \quad 0.0 \quad 0.0 \quad 0.0] \qquad (11)$$

$$X_5 = 0.14 \times 4.50\% + 0.29 \times 6.50\% + 0.29 \times 9.50\% + 0.14 \times 11.50\% + 0.14 \times 12.50\% = 8.63\% \qquad (12)$$

By knowing the gradual progress percentage values for states $x_1, x_2, x_3, \ldots, x_{10}$ (Table 5) and using the 10 transition matrices included in the Supplemental Data, the expected progress for each of the 12 verifying cooperatives (Table 6) is determined.

To discover the precision of the method in predicting monthly progress and ES^{total}, percentage differences between real progress values for each verifying construction project (for each state studied) and expected progress values were determined—that is, the percentage differences between the values in Table 5 and the values in Table 6. Table 7 lists these percentage differences in the Means row, where the mean between the nine transition states predicting monthly advances for each construction project appear. The average between mean values in the case of

monthly advances is 2.74%, and the average between ES^{total} values is 4.75%.

A cross validation was carried out taking into account the same criteria. The error or deviation estimation could vary depending on the data in the selected database and on the data used for validating the methodology. Cross validation is used to assess the results of the statistical analysis and to ensure that they are independent of the partition between training and testing data. It consists of repeating and calculating the arithmetic mean obtained from evaluation measures over different partitions. It is used when the main aim is the prediction and estimation of a model's precision (Garcia 2005).

According to Schneider (1997), the cross-validation method derives from the holdout method, which consists of dividing the

© ASCE

04017083-5

J. Constr. Eng. Manage.

Table 5. Actual Gradual Progress Percentages for the 12 Verifying Construction Projects for States $x_1, x_2, x_3, \ldots, x_{10}$

Interval	O1	O2	O3	O4	O5	O6	O7	O8	O9	O10	O11	O12
1t/10	2.25	4.52	2.94	11.20	1.06	4.46	0.50	0.00	4.76	5.41	8.90	4.74
2t/10	4.68	9.55	9.90	17.81	6.37	5.01	3.28	2.65	9.25	3.80	8.28	17.24
3t/10	9.47	10.39	7.04	17.96	4.66	4.80	6.51	8.21	9.99	13.34	10.12	10.48
4t/10	10.17	6.33	9.08	9.71	4.15	7.97	9.22	10.99	8.91	5.06	13.21	8.15
5t/10	13.71	11.35	11.13	8.72	8.78	8.42	14.32	11.44	9.94	2.13	12.24	13.69
6t/10	11.03	10.31	12.11	8.79	8.67	13.52	12.50	12.24	11.32	6.12	9.49	11.49
7t/10	6.38	9.34	8.69	6.19	4.12	12.31	10.99	4.97	12.61	2.51	10.88	8.05
8t/10	7.47	7.71	5.18	6.80	4.51	10.56	11.18	4.55	9.77	7.84	12.05	10.68
9t/10	8.50	11.55	10.84	3.36	14.73	7.95	12.57	10.78	8.85	13.67	6.69	11.10
10t/10	10.67	14.60	13.97	5.25	6.43	7.60	6.46	13.12	9.13	10.83	0.00	4.07
ES^{total}	18.63	13.88	16.66	7.69	22.21	22.78	23.08	25.32	4.16	23.45	11.32	9.10

Table 6. Predictions of Gradual Progress Percentages for the 12 Verifying Construction Projects for States $x_2, x_3, x_4, \ldots, x_{10}$

Interval	O1	O2	O3	O4	O5	O6	O7	O8	O9	O10	O11	O12
2t/10	8.87	6.58	8.95	7.50	6.41	6.58	5.17	5.17	6.58	6.92	8.50	6.58
3t/10	6.75	8.50	8.50	11.50	7.60	6.86	7.25	4.83	8.50	7.25	8.97	11.50
4t/10	9.25	9.68	7.36	12.00	7.00	7.00	6.59	9.75	9.25	12.50	9.68	9.68
5t/10	8.64	7.27	9.70	9.70	6.64	7.00	9.70	8.64	9.57	6.39	15.50	9.57
6t/10	9.50	10.50	10.50	9.80	9.80	9.80	11.50	10.50	9.73	7.00	9.00	9.50
7t/10	9.30	9.94	10.33	10.07	10.07	0.00	10.33	10.33	9.30	8.14	9.88	9.30
8t/10	8.30	8.67	7.86	8.30	5.17	10.75	8.80	5.17	10.75	8.50	8.80	7.86
9t/10	8.90	8.90	9.00	7.19	9.50	10.75	8.88	8.88	9.50	8.00	10.00	10.75
10t/10	8.04	10.00	8.79	0.00	15.50	8.33	10.75	8.79	8.04	11.25	6.07	10.00
ES^{total}	15.08	20.00	20.00	19.00	23.33	23.33	23.33	22.00	15.43	15.08	18.00	8.00

Table 7. Summary of Differences between Real and Expected Progress for the 12 Verifying Construction Projects

Interval	O1	O2	O3	O4	O5	O6	O7	O8	O9	O10	O11	O12
2/10t	4.19	2.97	0.95	10.31	0.04	1.57	1.89	2.52	2.67	3.12	0.22	10.66
3/10t	2.72	1.89	1.46	6.46	2.94	2.06	0.74	3.38	1.49	6.09	1.15	1.02
4/10t	0.92	3.35	1.72	2.29	2.85	0.97	2.63	1.24	0.34	7.44	3.53	1.53
5/10t	5.07	4.08	1.43	0.98	2.14	1.42	4.62	2.80	0.37	4.26	3.26	4.12
6/10t	1.53	0.19	1.61	1.01	1.13	3.72	1.00	1.74	1.59	0.88	0.49	1.99
7/10t	2.92	0.60	1.64	3.88	5.95	12.31	0.66	5.36	3.31	5.63	1.00	1.25
8/10t	0.83	0.96	2.68	1.50	0.66	0.19	2.38	0.62	0.98	0.66	3.25	2.82
9/10t	0.40	2.65	1.84	3.83	5.23	2.80	3.69	1.28	0.85	4.77	3.31	0.35
10/10t	2.63	4.60	5.18	5.25	9.07	0.73	4.29	4.33	1.09	0.42	6.07	5.93
Mean	2.36	2.37	2.06	3.95	3.33	2.86	2.43	2.58	1.41	3.70	2.48	3.30
ES^{total}	3.55	6.12	3.34	11.31	1.12	0.55	0.25	3.32	11.27	8.37	6.68	1.10

Note: ES^{total} average = 4.75%; monthly advances average = 2.74%.

sample data into two complementary groups, the analysis of one subgroup (called the training set), and the validation analysis of the other subgroup (called the test set). The approximation function is only adjusted to the training set, and from that the estimate is calculated depending on the division between training and testing data.

A k-fold cross validation was used for this study, where the sample data were divided k times into subgroups analyzing the 12 projects used for verification and the 90 projects from the database, where $k = 3$. Then the randomly obtained subgroup with the greatest amount of data in each division was used as testing data and the other 12 construction projects were used as training data. The cross-validation process was repeated during the three interactions, and finally the arithmetic mean of the results was calculated in order to obtain a single result.

Discussion

After carrying out the validation, it is shown that the proposed methodology predicts progress with a deviation of 2.47%.

Furthermore, this methodology allows for the prediction of ES^{total} with a deviation of 4.75%.

Using Markov chains, it is possible to predict the deviation of 9.73% with an error of 2.74% in the first tenth, but this error is reduced when more is known about the behavior of the work. The estimation error for construction progress (Es^{total}) reaches a value of 4.75%. This method of determining progress in construction and schedule deviations is a helpful tool for controlling the scheduling of a construction project in that it enables the anticipation of possible delays with a considerably lower error rate than the rate initially estimated.

Furthermore, cross-validation analysis shows that the statistical prediction method resulted in a mean difference of 2.38%. Additionally, a mean deviation of 4.29% was obtained in the prediction of ES^{total} from the three cross-validation interactions. This also represents a considerable improvement. The possibility of making more accurate predictions about future progress and the ES^{total} indicator allows a construction project to be managed in a more effective way, providing the opportunity to correct deviations in advance.

© ASCE 04017083-6 J. Constr. Eng. Manage.

J. Constr. Eng. Manage., 2017, 143(11): 04017083

Although the case study was constructed from a data set of 90 construction projects (and validated for 12 more), the prediction method can be used for any similar database. For this study, the Uruguayan database was selected because of the large amount of data it contains, but the applicability of the prediction method is independent of the selected database.

Conclusions

A statistical prediction model has been presented that makes it possible to predict future progress for a construction project by knowing its past behavior and using Markov chains. The model was validated using data from 90 housing construction projects. From this, it could be deduced that the tool makes it possible to anticipate future progress with a mean error of 2.38% and total deviation in construction time with an error of 4.29%. Moreover, the model was validated using 12 other construction projects. More important, a cross validation of three interactions was carried out and similar results were obtained.

The innovative prediction model presented in this paper contributes to the body of project management knowledge by introducing a new tool for managing and controlling construction timing. The method presented allows construction management to be improved because it predicts future deviations in schedules with reduced errors and determines total deviation in construction timing with great accuracy. Moreover, the model is the first of its kind to use Markov chains to predict deviations in construction projects.

Although the case study was constructed from a database of 90 housing projects (and validated for 12 more), the prediction method can be used for any similar database.

Data Availability Statement

Data generated or analyzed during the study are available from the corresponding author by request. Information about the *Journal*'s data sharing policy can be found here: http://ascelibrary.org/doi/10.1061/%28ASCE%29CO.1943-7862.0001263.

Acknowledgments

Supplemental Data

Tables S1–S10 are available online in the ASCE Library (www.ascelibrary.org).

References

Alkhathami, M. (2004). "Examination of the correlation of critical success and delay factors in construction projects in the Kingdom of Saudi Arabia." Ph.D. thesis, Univ. of Pittsburgh, Pittsburgh.

Alsakini, W., Wikström, K., and Kiiras, J. (2004). "Proactive schedule management of turnkey projects in developing countries." *Int. J. Project Manage.*, 22(1), 75–85.

Anbari, F. T. (2003). "Earned value project management method and extensions." *Project Manage. J.*, 34(4), 12–23.

Aristondo, G. (2003). "Cooperativismo, ayuda y autogestión." *Revista electrónica de geografía y ciencias sociales*, Universidad de Barcelona, Barcelona, Spain, 741–798.

Assaf, S., and Al-Hejji, S. (2006). "Causes of delay in large construction projects." *Int. J. Project Manage.*, 24(4), 349–357.

Bertsekas, D., and Tsitsiklis, J. (2000). *Introduction to probability*, American Mathematical Society, Providence, RI.

Chan, D., and Kumaraswamy, M. (1997). "A comparative study of causes of time overruns in Hong Kong construction projects." *Int. J. Project Manage.*, 15(1), 55–63.

De Marco, A., and Narbaev, T. (2013). "Earned value-based performance monitoring of facility construction projects." *J. Facilities Manage.*, 11(1), 69–80.

Flatscher, W. (2015). *7 top mistakes in applying earned value management*, ConCost, Dubai, UAE.

Fleming, Q., and Koppelman, J. (2010). *Earned value project management*, 4th Ed., Project Management Institute, Newtown Square, PA.

Forbes, L., and Ahmed, S. (2010). *Modern construction: Lean project delivery and integrated practices*, CRC Press, Boca Raton, FL.

Garcia, A. (2005). *Métodos avanzados de estadística aplicada: Métodos robustos y de re muestreo*, Universidad Autónoma de Madrid, Madrid, Spain.

Liberatore, M., Pollack-Johnson, B., and Smith, C. (2001). "Project management in construction: Software use and research directions." *J. Constr. Eng. Manage.*, 10.1061/(ASCE)0733-9364(2001)127:2(101), 101–107.

Lipke, W. (2003). *Project duration forecasting: A comparison of earned value management methods to earned schedule*, Project Management Institute, Newtown Square, PA.

Logofet, D., and Lesnaya, E. (2000). "The mathematics of Markov models: What Markov chains can really predict in forest successions." *Ecol. Modell.*, 126(2), 285–298.

Manjia, M., Abanda, B., and Pettang, C. (2013). *Using Markov decision process for construction site management in Cameroon*, Univ. of Yaoundé, Yaoundé, Cameroon.

Memon, A. H., Rahman, I. A., and Azis, A. A. A. (2012). "Time and cost performance in construction projects in Southern and Central Regions of Peninsular Malaysia." *Int. J. Adv. Appl. Sci.*, 1(1), 45–52.

Ross, S. (2003). *Introduction to probability models*, Academic Press, Cambridge, MA.

Sambasivan, M., and Soon, Y. W. (2007). "Causes and effects of delays in Malaysian construction industry." *Int. J. Project Manage.*, 25(5), 517–526.

Schneider, J. (1997). *The holdout method*, Wright State Univ., Dayton, OH.

Vanhoucke, M. (2010). "Measuring time—Improving project performance using earned value management." *International series in operations, research and management science*, Vol. 136, Springer, Berlin.

Vanhoucke, M. (2014). "Integrated project management and control: First comes the theory, then the practice." *Management for professionals*, Springer, New York.

Yuan, Z. (1999). "Prediction of protein subcellular locations using Markov chain models." *FEBS Lett.*, 451(1), 23–26.

Zhu, X., Ghahramani, Z., and Lafferty, J. (2002). "Semi-supervised learning using Gaussian fields and harmonic functions." *Proc., 20th Int. Conf. on Machine Learning (ICML-03)*, Vol. 3, Association for the Advancement of Artificial Intelligence, Menlo Park, CA, 912–919.

© ASCE   04017083-7   J. Constr. Eng. Manage.

J. Constr. Eng. Manage., 2017, 143(11): 04017083