

Delay in Pipeline Construction Projects in the Oil and Gas Industry: Part 2 (Prediction Models)

Reyadh M. M. Mohammed*, Saad M. A. Suliman

Department of Mechanical Engineering, University of Bahrain, Kingdom of Bahrain

Abstract In part 1 of the paper, factors causing delay in pipeline construction projects in the oil and gas industry in Bahrain were identified and ranked based on their importance and frequency. Risk assessment of 47 critical delay factors was conducted based on a survey and XYZ Company's KPI. In this part of the paper, delay prediction models were developed using multiple linear regression modelling. The multiple linear regression analysis was performed through the Statistical Package for the Social Sciences (SPSS) version 23 using historical data for 80 delayed oil and gas pipeline construction projects. Of the 47 critical delay factors, nineteen were considered high risk according to the risk assessment analysis. Prior to undertaking the analysis, the data was divided into two sets. The first set was comprised of 80 delayed projects for the model development whereas the second set was comprised of 20 delayed projects that were used for the model validation. The delay in days and the cost of delays were estimated using the developed models. Validation of the models showed that the model used for predicting the project costs of delays had a greater accuracy compared to the model used for estimating the project duration delays.

Keywords Prediction models, Project planning, Pipeline construction project, Construction delay, Cost of delay

1. Introduction

Part 1 of this paper focused on the delay factors and their risk assessment. However, development and validation of prediction models for delay was considered the most critical part of the study. Hence the study adopted the multi linear regression methodology to develop models that can be used by different stakeholders in the pipeline construction in oil and gas industry to predict the delay in days and also the cost overrun caused by the delays. Delay in the pipeline construction projects will automatically escalate the construction cost. The various stakeholders in the industry have in past strategized measures to enable completion of the project within the schedule. However, despite the positive interventions, most of the projects experience significant delays which can be otherwise avoided or reduced significantly by deploying prediction models. With the regression models, it is possible to visually examine the relationship existing between the dependent variables which are the cost of delay and the duration of delay in days, in one hand and the independent variables, which are the delay factors, in the other hand.

The main objectives of this paper is to develop and validate models that can predict duration of delay in days and the cost of delay in USD.

2. Literature Review

El-Kholy (2015) carried out a study on two models for predicting cost overruns in construction projects; Case Based Reasoning (CBR) and regression model. The study conducted a comparison of the two models and concluded that the regression model was more accurate than CBR. This study finding corroborated with an earlier study carried out by Dogan, Arditi and Gunaydin (2006). Dogan, Arditi and Gunaydin (2006) did a study to identify attribute weights in Case Based Reasoning (CBR) prediction model to estimate the costs in construction projects. The objective of the study was to investigate the application of the Excel sheet based CBR model of prediction through determining the effects of factor weights that were generated using three different optimization techniques. The three techniques were: counting features, gradient descent, and genetic algorithm. First of all, the spreadsheet was simulated for the study. This simulation was carried out in six steps: organizing and formatting data, calculating attribute similarities, establishing attribute weights, calculating weighted case similarities, identifying similarities of weight attributes and affiliated outputs, and calculating errors. A test was performed by populating data from a residential building construction project into a CBR Excel template. This data

* Corresponding author:

reyadh.mahmood80@gmail.com (Reyadh M. M. Mohammed)

Published online at <http://journal.sapub.org/ijcem>

Copyright © 2019 The Author(s). Published by Scientific & Academic Publishing

This work is licensed under the Creative Commons Attribution International

License (CC BY). <http://creativecommons.org/licenses/by/4.0/>

was from study reports investigating the cost of construction projects in residential buildings in Turkey. The data was then run and the results evaluated. The procedure was done 10 times for the test set and the best results were recorded. The findings from the study indicated that the GA-augmented CBR model was the best, with an estimated 16.23% error, followed by CBR + feature counting with 17.63% error and CBR plus gradient descent with 21.20% error. The study demonstrated the importance of CBR prediction model, especially in the performance of the attribute weights.

A research carried out by Bayram (2017) investigated and validated the “Bromilow’s Time-cost model (BTC)” and “Love et al.’s Time-Floor model (LTF)” on the Turkish public building projects. This study had two objectives, namely: to check the two models (BTC and LTF) appropriateness, and also to compare the two models prediction performances in relation to project durations. The BTC model formulated project duration in terms of cost while the LTF model used both gross floor area and number. Best-fit models were used as bench mark for both BTC and LTF models during the research. The research findings asserted that BTC model is more superior to LTF model, thus concluding that cost is a more significant to both gross floor area and number in predicting construction duration. On the other hand, the proposed best-fit models in the research proposed cost, and the area and height to be vital predictors of duration in construction. This research finding corroborated to earlier research findings by Elhang and Boussabaine (1999) where they conducted statistical analysis in creating priority rating of the prompting factors. Additionally, this research provided ways on how to develop price models by use of both the cost and time factors.

Maria and Georgios (2017) conducted a research using regression analysis in evaluating cost and time prediction for plethora of public works focused on urban renewal in the Thessaloniki district. Data on twelve public projects were provided by the Municipality of Thessaloniki. Moreover, Fast Artificial Neural Network (FANN) tool was used in predicting both duration and final cost of projects by using volume of earthwork as the input variable. The two approaches (regression analysis and FANN tool) facilitated project stakeholders on forecasting the project’s duration and costs hence enabling them to note any deviations from the set deadlines and budgetary allocations. Generally, the research indicated that the artificial fast neural network (FANN) had higher accuracy and efficiency in performance compared to the regression model in the case study. Additional past studies by Hezagy and Ayed (1998) on use of neural network in evaluating cost data and creation of prediction model for highway construction projects emphasized this finding. Here, developers created a model for testing sensitivity of changes in cost related parameters to determine their relative significance. Adaptive modules were incorporated in this study which allowed users to easily input data for re-optimization of the model to the future changing both internal and external project environments.

Gunduz et al. (2017) used different standard statistical methods to evaluate their research outcomes on reasons why projects lagged in Qatar. The main objective of the study was to determine main causes of delays in construction projects. Literature review was used to identify eighty-three causes of delays whereas expert opinions and recommendations reduced the causes by forty-one causes leaving forty-two causes whose outcomes were analysis using standard statistical methods. Cronbach’s Alpha coefficient was used to measure internal inconsistencies of the outcomes whereas the Relative Importance Index (RII) and the Frequency Index (FI) applied on the assessment of different project delay factors. This is corroborated to the Gebrehiwet and Luo (2017) statement that, “timely project completion ranks highly as a project success metric alongside adherence to budget, satisfaction of quality standards, and conformance with benchmarks for safety.” Even though these factors appear standalone, numerous studies across diverse disciplines and industrial fields have reported relationships between these variables (Karami and Olatunji, 2017).

Additionally, Michal, Agnieszka and Krzysztof (2018) also used artificial neural networks for cost estimation in their study. Noting the importance of cost estimation for success of any construction project, they estimated costs of constructing a sports field by use of artificial neural networks which they described as a multi-layer perceptron networks. Moreover, one network in the study was tailored for mapping the relationship between the total construction costs of the works and a selected cost predictor of the sports fields. Results from the study indicated a satisfactory performance of the different selected neural networks in terms of real costs and predicted costs. Therefore, this technique has two key advantages. These are: allows for cost estimation of construction works and support client decisions as they are aware of cost estimation accuracy.

The importance of this paper lies in the fact that most of the previous studies did not cover the modeling of the duration of delay along with the cost of delay. The previous studies were mainly focused on the delay factors and their ranking without developing prediction models for delays. In addition to that, there was no research found to study this specific topic in Kingdom of Bahrain and precisely in pipeline construction projects. Therefore, this study is aimed to identify the causes and effects of the delay in the pipeline construction projects in Bahrain and to model the delay in terms of duration and cost based on real case studies in this field.

3. Development and Validation Delay Models

Regression modelling

Multiple linear regression analysis was performed in SPSS using historical data for 80 oil and gas pipeline construction projects and the KPI data. The data included the

19 high risk delay factors identified in part 1. Prior to undertaking the analysis, the data was divided into two sets comprising 80 and 20 projects used for the development of the model and its validation, respectively.

Table (1)

19 High Risk Delay Factors Resulted from XYZ Company Risk Mapping Matrix			
Factor No.	Factor	Factor No.	Factor
F1	Inadequate project planning, budgeting and scheduling	F10	Poor managerial skills
F2	Scope variation	F11	Delay or long way leave approval process by other parties
F3	Late materials delivery	F12	Weather effect (heat, rain, wind, sand storm, etc.)
F4	Insufficient contractor's manpower (or resources)	F13	Rework due to errors, mistakes and defective works during construction
F5	Lack of communication between client, consultant & contractor	F14	Inadequate contractor experience
F6	Poor site management and supervision	F15	Incompetence of project team
F7	Inefficient and poor performance of contractors/subcontractors	F16	Slow decision making within all project teams
F8	Materials shortage	F17	Delay in drawings preparation of the projects
F9	Contractor's lack of planning at preconstruction	F18	Late materials issuance
		F19	Unforeseen site conditions

Table (2). Binary Analysis for 80 Delayed Projects in XYZ Company using Risk Mapping Matrix Method

Project #	Actual Delay (Days)	Cost of Delay (USD)	F1	F2	F3	F4	F5	F6	F7	F8	F9	F10	F11	F12	F13	F14	F15	F16	F17	F18	F19
1	2.00	12,340.71	0	0	0	0	0	1	1	0	0	0	0	0	1	1	1	0	0	0	1
2	0.67	25,783.95	0	0	0	0	0	1	1	0	0	0	0	0	1	1	1	0	0	0	1
3	3.00	15,435.55	0	0	0	0	0	1	1	0	0	0	0	0	0	1	1	0	0	0	1
4	1.96	43,498.46	0	1	0	0	1	1	1	0	0	0	0	0	1	0	0	0	0	0	1
5	0.23	10,687.79	0	0	0	0	1	1	0	0	0	0	0	0	0	0	1	0	0	0	0
6	0.88	4,227.23	1	0	0	0	1	1	0	0	1	0	0	0	0	0	1	0	0	0	1
7	0.42	59,878.27	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	1
8	1.46	35,539.78	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	1
9	0.09	159.45	0	1	0	0	0	1	1	0	0	1	0	0	1	1	1	0	0	0	1
10	1.67	31,990.41	1	0	0	0	1	1	0	0	0	0	0	0	0	0	1	0	0	0	1
11	3.42	66,186.15	0	1	0	0	0	0	0	0	0	1	0	0	1	0	1	0	0	0	1
12	1.81	22,884.90	1	1	0	0	1	1	0	0	1	0	0	0	0	1	1	0	0	0	0
13	3.67	48,198.13	0	1	0	0	0	0	0	0	0	1	0	0	1	0	1	0	0	0	1
14	2.63	253,404.11	1	0	0	0	1	1	0	0	1	0	1	0	0	0	1	0	0	0	0
15	2.13	5,102.83	0	1	0	0	0	0	0	0	0	1	0	0	1	0	1	0	0	0	1
16	1.67	0.00	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	1
17	0.08	212.27	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	1
18	7.15	34,853.31	1	0	0	0	1	1	0	0	1	0	1	0	0	0	1	0	0	0	0
19	0.65	0.00	1	0	0	0	0	0	0	0	0	0	1	0	0	0	1	0	0	0	1
20	23.10	410,445.20	1	0	0	0	1	0	0	0	0	0	0	0	0	0	1	0	1	0	0
21	13.71	12,394.11	1	1	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0
22	1.30	2,132.16	1	0	0	0	0	0	0	0	0	1	0	0	0	0	1	0	0	0	1
23	5.03	0.00	1	0	0	0	0	0	0	0	0	1	0	0	0	0	1	0	0	0	1
24	0.42	2,927.68	1	0	0	0	0	0	0	0	0	1	0	0	0	0	1	0	0	0	1
25	4.65	37,527.00	1	0	0	0	1	0	0	0	0	0	1	0	0	0	1	0	0	0	1
26	0.60	500.77	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	1
27	0.10	5,312.36	1	0	0	0	0	0	0	0	0	0	1	0	0	0	1	0	0	0	0
28	1.17	14,950.26	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	1	0	0	0
29	0.90	0.00	1	0	0	0	0	0	0	0	0	0	1	0	0	0	1	0	0	0	0
30	9.13	0.00	1	0	0	0	0	0	0	0	0	0	1	0	0	0	1	0	0	0	1
31	1.05	56,535.55	1	0	0	0	0	0	0	0	0	0	1	1	0	0	1	0	0	0	1
32	2.28	5,843.00	1	0	0	0	1	1	1	0	1	0	0	0	0	1	1	0	0	0	1
33	0.46	10,583.20	1	0	0	0	0	0	0	0	0	1	0	0	0	0	1	0	0	0	1
34	0.14	0.00	1	0	0	0	1	1	0	0	0	0	1	0	0	0	1	0	0	0	0
35	0.21	209.23	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0
36	2.03	41,304.18	1	1	0	0	0	0	0	0	0	1	0	0	1	1	1	0	0	0	1
37	15.38	13,013.82	1	0	0	0	1	0	0	0	0	0	1	0	0	0	1	1	0	0	0
38	0.16	8,530.84	1	0	0	0	0	1	1	0	1	0	0	0	0	1	1	0	0	0	0
39	2.31	1,454.60	1	0	0	0	0	0	0	0	0	1	0	0	0	0	1	0	0	0	0
40	1.06	718.81	1	0	1	0	1	0	0	1	0	1	0	0	0	0	1	1	0	1	0
41	30.33	227,010.96	1	1	0	0	0	1	0	0	0	0	0	0	1	0	1	0	0	0	1
42	1.08	0.00	1	0	0	0	0	1	1	0	0	0	0	0	1	0	1	0	0	0	1
43	1.00	854.69	1	0	0	0	1	0	0	0	0	1	0	0	0	0	1	1	0	0	0
44	0.25	4,034.18	1	0	0	0	0	1	1	0	1	0	0	0	0	1	1	0	0	0	0
45	0.27	520.86	1	0	0	0	1	1	0	0	0	1	0	0	0	0	1	0	0	0	0
46	0.25	338.25	1	0	0	1	0	1	1	0	1	1	0	0	0	1	1	0	0	0	0
47	0.40	129.14	1	0	0	0	1	1	0	0	0	1	0	0	0	0	1	0	0	0	0
48	2.42	15,012.92	1	0	0	0	0	0	0	0	0	1	0	0	0	0	1	1	0	0	0
49	1.13	14,644.80	1	0	0	0	0	1	1	0	1	1	0	0	0	1	1	0	0	0	0
50	1.08	2,806.82	1	0	0	0	0	0	0	0	0	1	1	0	0	0	1	0	0	0	0
51	0.35	22.86	1	0	0	0	0	1	1	0	1	1	0	0	0	1	1	0	0	0	0
52	0.79	5,666.20	0	0	0	0	0	1	0	0	0	1	0	0	0	0	1	0	0	0	1
53	0.06	237.41	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1
54	0.13	57.26	0	0	0	0	0	1	1	0	0	0	0	0	1	1	1	0	0	0	1
55	0.10	717.53	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	1
56	2.38	100,137.19	1	0	1	0	1	1	1	1	1	0	0	1	1	1	1	0	0	1	1
57	0.08	316.30	1	0	0	0	1	1	0	0	0	0	0	0	0	0	1	1	0	0	1
58	0.38	822.43	1	1	0	0	1	1	1	0	0	1	0	0	1	1	1	0	0	0	1
59	0.19	4,337.95	0	0	0	0	0	1	1	0	0	0	0	0	1	1	0	0	0	0	1
60	0.44	864.84	0	0	0	0	0	1	1	0	0	0	0	0	1	1	1	0	0	0	1
61	0.31	3,261.07	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	1
62	0.04	375.70	1	0	0	0	0	1	1	0	1	1	0	0	0	1	1	0	0	0	0
63	0.31	749.91	1	0	0	0	1	0	0	0	0	1	0	0	0	0	1	0	0	0	1
64	0.31	903.70	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	1
65	3.00	18,336.33	0	0	0	0	0	1	1	0	0	0	0	0	1	1	1	0	0	0	1
66	1.33	10,599.28	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	1
67	0.81	5,757.83	0	0	0	0	0	1	1	0	0	1	0	0	1	1	1	0	0	0	1
68	0.15	901.97	1	0	0	1	0	1	1	0	1	1	0	0	0	1	1	0	0	0	0
69	0.15	788.10	0	0	0	0	0	1	1	0	0	1	0	0	1	1	1	0	0	0	1
70	0.13	197.14	1	0	0	0	1	0	0	0	0	1	0	0	0	0	1	0	0	0	1
71	0.08	316.34	1	0	0	0	1	1	1	0	0	1	0	0	1	1	1	0	0	0	1
72	1.56	401.82	1	0	0	0	1	0	0	0	0	1	0	0	0	0	1	0	0	0	0
73	0.94	7,682.58	1	0	0	1	1	1	1	0	1	1	0	0	0	1	1	0	0	0	0
74	0.08	639.73	1	0	0	0	1	1	1	0	1	1	0	0	0	1	1	0	0	0	1
75	2.77	8,696.76	1	0	0	1	0	1	1	0	1	1	0	0	0	1	1	0	0	0	1
76	1.13	1,952.57	1	0	1	0	0	1	1	1	1	1	0	0	0	1	1	0	0	1	0
77	0.75	7,235.93	1	0	0</																

A detailed breakdown of the 80 pipeline construction projects that suffered delays during the period 2011 through 2017, and the mapping out of the 19 high risk delay factors was performed (Tables 1 and 2). From the 80 late projects in the XYZ Company KPIs, the researcher had studied the actual reasons of delays to identify the definite causes and correlate the 19 high risk delay factors, resulted from the company's risk assessment matrix, to the actual one by using binary analysis system as shown in Table (2), where a 1 implies presence of the factor and a 0 implies its absence. After filling the binary Excel sheet for the 80 delayed projects, the list was entered in the SPSS software for the delay predication models using the multi linear regression methodology.

Seventeen high risk delay factors were entered into the multiple linear regression model as the predictor (independent) variables as late material delivery and material shortage were excluded since they featured only in 3 out of the 80 projects. The dependent variables were the cost of delay (USD) and the duration of delay (days).

Based on multiple regression model test, the following equation was developed for the prediction of duration of delay:

$$\text{Duration of delay in days} = -1.471 + 2.343 \times F_1 + 3.708 \times F_2 + 0.566 \times F_3 - 2.058 \times F_4 + 2.366 \times F_5 - 1.737 \times F_6 + 1.754 \times F_7 - 0.858 \times F_8 + 0.134 \times F_9 + 1.687 \times F_{10} + 2.892 \times F_{11} - 2.830 \times F_{12} + 1.387 \times F_{13} + 3.388 \times F_{14} + 22.900 \times F_{15} - 0.892 \times F_{16} + 0.089 \times F_{17} \quad (1)$$

Table (3) shows the model summary results. The model summary results for the prediction of duration of delay in days showed that the prediction quality of the 17 high risk delay factors was good based on the multiple correlation coefficient ($R = 0.696$). The Adjusted R^2 coefficient showed that the 17 high risk delay factors could account for up to 34.3% variance in the duration of delay in days ($\text{Adj. } R^2 = 0.343$).

Table (3). Multiple Linear Regression to Predict Duration of Delay in Days Based on 17 High Risk Delay Factors

R	R ²	Adjusted R ²	Std. Error of the Estimate
0.696	0.484	0.343	3.858

The ANOVA results (Table (4)) revealed that the overall multi-linear regression model was a good fit for the data, as a statistically significant F-statistic was achieved ($F(17, 62) = 3.422$, $p < 0.05$). This implied that the 17 high risk delay factors were statistically significant in predicting duration of delay in days.

Table (4). ANOVA Analysis for Multiple Linear Regression to Predict Duration of Delay

Model	Sum of Squares	df	Mean Square	F	p-value
Regression	865.983	17	50.940	3.422	0.000
Residual	922.878	62	14.885		
Total	1788.861	79			

The coefficients of each of the 17 critical delay factors were established as illustrated in the Table (5). The results indicated that only two critical delay factors, namely, scope variation and delay in drawings preparation of the projects were statistically significant predictors of duration of delay ($p < 0.05$).

Table (5). Coefficients' for Multiple Linear Regression to Predict Duration of Delay in Days

Delay Factors		β	p-value
Symbol	Description		
C	Constant	-1.471	0.483
F1	Inadequate project planning, budgeting and scheduling	2.343	0.138
F2	Scope variation	3.708	0.035
F3	Insufficient contractor's manpower (or resources)	0.566	0.808
F4	Lack of communication between client, consultant and contractor	-2.058	0.097
F5	Poor site management and supervision	2.366	0.193
F6	Inefficient and poor performance of contractors/subcontractors	-1.737	0.491
F7	Contractor's lack of planning at preconstruction	1.754	0.370
F8	Poor managerial skills	-0.858	0.476
F9	Delay or long way leave approval process by other parties	0.134	0.938
F10	Weather effect (heat, rain, wind, sand storm, etc.)	1.687	0.440
F11	Rework due to errors, mistakes and defective works during construction	2.892	0.174
F12	Inadequate contractor experience	-2.830	0.206
F13	Incompetence of project team	1.387	0.528
F14	Slow decision making within all project teams	3.388	0.074
F15	Delay in drawings preparation of the projects	22.900	0.000
F16	Late materials issuance	-0.892	0.745
F17	Unforeseen site conditions	0.089	0.944

Similarly, multiple linear regression model was developed to test for the extent to which the 17 high risk delay variables (factors) could predict the cost of delay in USD.

$$\text{Cost of delay in dollars} = -26184.500 + 18040.552 \times F_1 + 11158.961 \times F_2 + 50.141 \times F_3 - 7687.950 \times F_4 + 32754.106 \times F_5 - 46878.255 \times F_6 + 56317.855 \times F_7 - 2580.991 \times F_8 + 20701.724 \times F_9 + 41050.981 \times F_{10} + 70034.984 \times F_{11} - 30061.376 \times F_{12} + 5277.391 \times F_{13} + 12236.998 \times F_{14} + 420999.707 \times F_{15} - 3809.442 \times F_{16} - 1632.185 \times F_{17} \quad (2)$$

Multiple linear regression model was developed to test for the extent to which the 17 high risk delay variables (factors) could predict the cost of delay in USD. The model summary results are summarized in Table (6).

The ANOVA results shown in Table (7) confirmed that the overall regression model was a good fit for the data as a

statistically significant F-statistic was achieved ($F(17, 62) = 11.140$, $p < 0.05$). Thus, it can be said that the 17 high risk delay factors were statistically significant in predicting cost of delay in USD.

Table (6). Model Summary Results for Multiple Linear Regression to Predict Cost of Delay in USD Based on 17 High Risk Delay Factors

R	R ²	Adjusted R ²	Std. Error of the Estimate
0.868	0.753	0.686	33,482.306

Table (7). ANOVA Analysis for Multiple Linear Regression to Predict Cost of Delay in USD

Model	Sum of Squares	df*	Mean Square	F	p-value
Regression	212.314 x 10 ⁹	17	12.489 x 10 ⁹	11.140	0.000
Residual	69.506 x 10 ⁹	62	1.121 x 10 ⁹		
Total	281.820 x 10 ⁹	79			

df* represents degrees of freedom

Table (8). Coefficients' Table for Multiple Linear Regression to Predict Cost of Delay in USD

Delay Factors		β	P-value
Symbol	Description		
C	Constant	-26184.500	0.153
F1	Inadequate project planning, budgeting and scheduling	18040.552	0.187
F2	Scope variation	11158.961	0.459
F3	Insufficient contractor's manpower (or resources)	50.141	0.998
F4	Lack of communication between client, consultant and contractor	-7687.950	0.471
F5	Poor site management and supervision	32754.106	0.040
F6	Inefficient and poor performance of contractors/subcontractors	-46878.255	0.035
F7	Contractor's lack of planning at preconstruction	56317.855	0.001
F8	Poor managerial skills	-2580.991	0.804
F9	Delay or long way leave approval process by other parties	20701.724	0.171
F10	Weather effect (heat, rain, wind, sand storm, etc.)	41050.981	0.033
F11	Rework due to errors, mistakes and defective works during construction	70034.984	0.000
F12	Inadequate contractor experience	-30061.376	0.123
F13	Incompetence of project team	5277.391	0.782
F14	Slow decision making within all project teams	12236.998	0.452
F15	Delay in drawings preparation of the projects	420999.707	0.000
F16	Late materials issuance	-3809.442	0.873
F17	Unforeseen site conditions	-1632.185	0.881

The coefficients' table of the regression model (Table 8) indicated that 6 of the 17 high risk delay factors were significant predictors of the cost of delay poor site management and supervision; inefficient and poor performance of contractors; contractor's lack of planning at preconstruction; weather effect; rework due to errors, mistakes and defective works during construction; and delay in drawings preparation of the projects.

To increase the stated multiple linear regression model's predictive ability, the project delay factors that are statistically significant (p -value < 0.05) and almost significant were only considered for model's development. For the prediction of duration of delay, significant factors were scope variation (p -value = 0.035), and delay in drawing preparation of the projects (p -value = 0.000). The almost significant factors were lack of communication between the client, consultant, and contractor (p -value = 0.097) and slow decision making within all project teams (p -value = 0.074). The following equation was developed for the prediction of duration of delay in pipeline construction:

$$\text{Duration of delay} = 1.295 + 4.512 \times F_2 - 0.466 \times F_4 + 2.860 \times F_{14} + 22.272 \times F_{15} \quad (3)$$

The model summary results for the prediction of duration of delay in days showed that the prediction quality of the four significant and almost significant factors were good based on the multiple correlation coefficient ($R = 0.614$). The Adjusted R^2 coefficient showed that the four significant and almost significant factors could account for up to 34.4% variance in the duration of delay in days ($\text{Adj. } R^2 = 0.344$) as shown in Table (9). The ANOVA results (Table (10)) revealed that the overall multi linear regression model was a good fit for the data, as a statistically significant F-statistic was achieved ($F(4, 75) = 11.45$, $p < 0.05$) from SPSS software.

Table (9). Model Summary Results for Multiple Linear Regression to Predict Duration of Delay in Days Based on Significant and Almost Significant Factors

R	R ²	Adjusted R ²	Std. Error of the Estimate
0.614	0.377	0.344	3.85444

Table (10). ANOVA Analysis for Multiple Linear Regression to Predict Duration of Delay in Days

Model	Sum of Squares	df	Mean Square	F	p-value
Regression	674.608	4	168.652	11.35	0.000
Residual	1114.253	75	14.857		
Total	1788.861	79			

Table (11) revealed that only two critical delay factors were statistically significant predictors of duration of delay in days. These factors were scope variation ($\beta = 4.512$, $p < 0.05$) and delay in drawings preparation of the projects ($\beta = 22.272$, $p < 0.05$).

Although there was no difference between the two prediction models of delay duration in terms of R^2 and the R

coefficient for the first model ($R = 0.696$ for Equation (1)) was a bit higher than the second model ($R = 0.614$ for Equation (3)). However, in order to make the analysis and the computation easier rather than considering all the 17 high risk delay factors, the second model with 4 significant and almost significant factors will be recommended.

Table (11). Coefficients' for Multiple Linear Regression to Predict Duration of Delay in Days Based on Significant and Almost Significant Factors

Delay Factors		β	P-value
Symbol	Description		
C	Constant	1.295	0.024
F2	Scope variation	4.512	0.001
F4	Lack of communication between client, consultant and contractor	-0.466	0.623
F14	Slow decision making within all project teams	2.860	0.056
F15	Delay in drawings preparation of the projects	22.272	0.000

Similarly, the same approach was adopted to develop a more reliable model to predict the cost of delay. The formulation incorporated 6 delay factors which were identified as statically significant in Table (8). The following equation was developed for predicting the cost of delay in pipeline construction:

$$\text{Cost of delay in dollars} = -1326.639 + 26571.714 \times F_5 - 77326.411 \times F_6 + 56808.183 \times F_7 + 21346.484 \times F_{10} + 61517.573 \times F_{11} + 411771.839 \times F_{15} \quad (4)$$

Multiple linear regression model was developed to test for the extent to which the six significant delay factors could predict the cost of delay in USD. The model summary results revealed a good prediction quality of the multiple correlation coefficient ($R = 0.843$). The Adjusted R^2 coefficient showed that the six significant delay factors could account for up to 68.7% variance of the cost of delay in USD (Adj. $R^2 = 0.687$) as shown in Table (12). The ANOVA results shown in Table (13) confirmed that the overall regression model was a good fit for the data as a statistically significant F-statistic was achieved ($F(6, 73) = 29.880$, $p < 0.05$). Thus, the six significant delay factors were statistically significant in predicting duration of delay in days.

Table (12). Model Summary Results for Multiple Linear Regression to Predict Cost of Delay in USD Based on Significant Factors

R	R^2	Adjusted R^2	Std. Error of the Estimate
0.843	0.711	0.687	33423.04477

Table (13). ANOVA Analysis for Multiple Linear Regression to Predict Cost of Delay in USD

Model	Sum of Squares	df*	Mean Square	F	p-value
Regression	200.271 x 10 ⁹	6	33.379 x 10 ⁹	29.880	0.000
Residual	81.548 x 10 ⁹	73	1.117 x 10 ⁹		
Total	281.820 x 10 ⁹	79			

The coefficients' Table (14) of the last regression model showed that 5 of the 6 high risk delay factors were significant predictors of the cost of delay in USD, as indicated in Table (14).

Table (14). Coefficients' Table for Multiple Linear Regression to Predict Cost of Delay in USD Based on Significant Factors

Delay Factors		β	P-value
Symbol	Description		
C	Constant	-1326.639	0.838
F5	Poor site management and supervision	26571.714	0.033
F6	Inefficient and poor performance of contractors/subcontractors	-77326.411	0.000
F7	Contractor's lack of planning at preconstruction	56808.183	0.000
F10	Weather effect (heat, rain, wind, sand storm, etc.)	21346.484	0.080
F11	Rework due to errors, mistakes and defective works during construction	61517.573	0.000
F15	Delay in drawings preparation of the projects	411771.839	0.000

Validation of the developed models

To evaluate the performance of the developed models, twenty delayed projects were randomly chosen to predict both the duration in days and the cost overrun in USD. To ensure robustness in the validation exercise, the study adopted all 17 variables despite some of the critical delay factors ruled out to have no statistical significant predictor power. Actual data obtained from company records concerning duration of delay in days and cost in USD were compared against the estimations obtained from the regression models.

The actual delays in days were calculated by dividing the actual completion hours after deducting 24 hours, which is the XYZ Company target duration of connecting the wells from the wellhead to the well manifold, by 24. On the other hand, the actual cost of delays was calculated by multiplying production loss due to delay by the average oil price at that time. For instance, Project 10 actual delay (1.33 Day) was calculated by dividing the actual duration taken to connect the well to well manifold, which was 56 hours, minus the target company's duration, which is 24 hours, over 24 hours. For the same project, the actual cost of delay (\$10,599.28) was calculated by multiplying the production loss due to delay which was 205.25 barrels per day (bpd) by the oil price at that time, which was \$51.64.

The results are indicated in the Table (15).

In addition, a paired samples t-test was performed in SPSS to determine whether the model estimates for duration of delay and cost of delay were different from the actual duration of delay in days and cost of delay in USD. Table (16) showed that the paired samples correlations revealed strong, near-perfect correlations for actual and model estimated cost

in USD ($R = 0.999$, $p < 0.05$). Similarly, model estimated and actual duration of delay in days registered near-perfect correlation ($R = 0.961$, $p < 0.05$). The paired samples test results (Table 17) showed that the mean differences of the actual and the model estimated costs were not statistically significant ($t(19) = 0.087$, $p > 0.05$). Similarly, the mean

differences for the actual and the model estimated delays in days were not statistically significant ($t(19) = -1.957$, $p > 0.05$). These results were a confirmation that the regression models developed in this study were strong enough to predict duration of delay in days and cost of delay in USD for construction projects in the oil and gas industry in Bahrain.

Table (15). Comparison of Estimated Delay and Cost Using Regression Model and Actual Project Data

Project #	Delay (Days)			Delay Cost (USD)		
	Actual	Estimated	Error	Actual	Estimated	Error
1	23.10	23.10	0.00	410445.20	410445.20	0.00
2	0.15	-0.16	-0.31	788.10	729.17	-58.93
3	0.13	-0.16	-0.29	200.88	729.17	528.29
4	1.05	4.17	3.12	56535.55	57253.96	718.41
5	0.16	1.81	1.65	8530.84	9265.77	734.93
6	1.13	0.06	-1.07	1952.57	2875.34	922.77
7	1.00	2.73	1.73	854.69	-898.50	-1753.00
8	0.79	1.51	0.72	5666.20	7633.82	1967.60
9	0.44	0.70	0.26	864.84	3310.16	2445.30
10	1.33	0.30	-1.03	10599.28	13234.30	2635.00
11	2.03	5.26	3.23	41304.18	44052.84	2748.70
12	0.21	2.26	2.05	209.23	-2866.56	-3076.00
13	0.13	0.70	0.57	57.26	3310.16	3252.90
14	0.08	-1.02	-1.10	639.73	-2635.35	-3275.00
15	2.77	1.61	-1.16	8696.76	5102.74	-3594.00
16	2.38	2.67	0.29	100137.19	104641.17	4504.00
17	0.06	2.35	2.29	237.41	-4498.74	-4736.00
18	0.81	-0.16	-0.97	5757.83	729.17	-5029.00
19	0.25	1.81	1.56	4034.18	9265.77	5231.60
20	1.06	1.84	0.78	718.81	-4707.94	-5427.00

Table (16). Paired Samples Correlations to Compare Model Estimates and Actual Data for Duration of Delay in Days and Cost of Delay in USD

	N	Correlation	Sig.
Pair 1 Actual delay in days & Estimated delay in days	20	0.961	0.000
Pair 2 Actual cost in dollars & Estimated cost in dollars	20	0.999	0.000

Table (17). Paired Samples T-Test to Compare Model Estimates and Actual Data for Difference of Duration of Delay in Days and Cost of Delay in USD

	Paired Differences		t	df	Sig. (2-tailed)
	Mean	Std. Deviation			
Pair 1 Actual duration of delay in days - Estimated duration of delay in days	-0.616	1.408	-1.957	19	0.065
Pair 2 Actual cost in USD - Estimated cost in USD	62.954	3228.514	0.087	19	0.931

4. Discussion

Scope variation and delay in drawing preparations for the project emerged as important factors in predicting duration of delay. The implication of this finding was that scoping and conducting drawing preparations accurately in the project most likely would reduce the length of a project delay. On the other hand, the contractor related delay factors that affecting the cost model were: poor site management and supervision, inefficient and poor performance of

contractors/subcontractors, contractor's lack of planning at preconstruction, and rework due to errors, mistakes and defective works during construction. However, it should be noted that delay factors related to design were also significant predictors of delay costs and delay duration.

The ascertainment received from the paired samples t-test and the regression test was adequate to render the models reliable. The prediction model for cost of delay revealed a higher prediction power with more significant delay variables than the model for duration of delay. The model of

duration of delay had a prediction power of 34.3% while the model of cost of delay had prediction power of 68.6%. The predictive power of the project cost of delays linear regression model was emphasized by the statistical results from t-test paired two samples for mean, which indicated that the means of actual project cost of delays and the means of estimated projected cost of delays seemed to be equal. However, the results from t-test paired two samples for means indicated that the means of actual project duration delays in days seemed to be statistically different and low compared to the means of estimated project duration delays in days.

Based on the two regression models for duration of delay in days and the cost of delays, Figure (1) shows the significant predictor variables for each model.

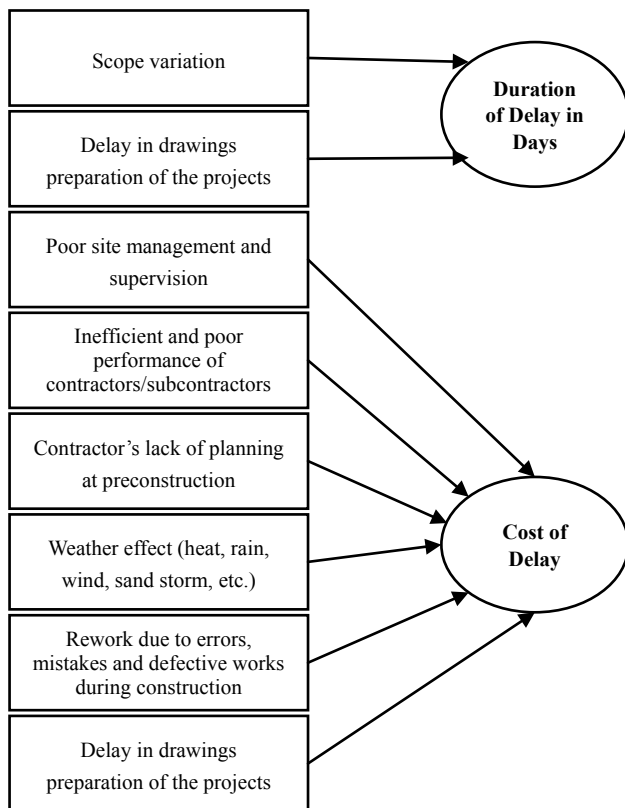


Figure (1). Prediction Models for Delay in Pipeline Construction Projects in Oil and Gas Industry as Developed in This Paper

5. Conclusions and Recommendations

The study acknowledged that based on the outcome from linear regression analysis, project scope variation and delays in drawings preparation for the project have a significant effect in predicting project duration delays. However, the factors explained only 34.3% of the variations in project duration delays. On the other hand, the analysis showed that five delay factors, namely: poor site management and supervision; inefficient and poor performance of contractors/subcontractors; contractor's lack of planning at preconstruction; rework due to errors, mistakes and defective

works during construction; and delay in drawing preparation of the projects have a significant effect in predicting cost of delay. The developed model can explain 68.6% of the variations in project cost of delays. On the basis of results from the multiple linear regression analysis, the study concluded that the project cost of delays prediction model had a greater predictive and forecasting ability compared to the project duration delays prediction model. The findings are also emphasized by the statistical results from t-test paired comparison of means in both prediction of duration delays and cost of delay. Based on the findings, the research recommends the following:

- The significant model predictor variables of project duration delay can be mitigated by having additional measures at the preconstruction phase. This can be achieved by making a precise scope tender documents to avoid variations during the actual construction. In addition, the designer should be asked for clear time line for the design drawing to avoid submitting the drawing late, and
- The significant model predictor variables of cost of delay can be mitigated by providing instant performance feedback to the contractor in a frequent basis during all project stages (planning, execution and closing out the project). Furthermore, spending more time in planning stage from client and contractor side is recommended in order to achieve the target goal with the allocated budget for the project. Some other factors effect can only be minimized and cannot be eliminated such as heat, rain, sand storm... etc. in these cases some extra precautions can be taken in order to speed up the project such as providing shades, special clothes to withstand the heat, face masks for the sand storm and regular rests and breaks for the labors. In some of the cases, the delay happens from rework in the project during construction, to avoid these kind of delays, extra and constant supervision needed on site from both parties the client and the contractor these delay have direct impact on the project duration therefore on the project cost as well.

6. Further Research

Further research is needed on:

- Enhancing the developed model in this study through introduction and testing of additional factors to enhance the prediction power of delay duration and cost in pipeline construction project in the oil and gas industry, and
- Further research is needed to test the model using data from different companies in the same field in Bahrain and other countries pipeline industry sectors to investigate the applicability of the prediction models for other companies.

REFERENCES

- [1] Bayram, S. (2017). Duration prediction models for construction projects: In terms of cost or physical characteristics? *KCSE Journal of Civil Engineering*, 21(6), pp.2049-2060. Retrieved from <https://doi.org/10.1007/s12205-016-0691-2>.
- [2] David, J. L., Margaret, W. E., & Anthony, H. (2006). Predicting Construction Cost Using Multiple Regression Techniques. *Journal of Construction Engineering and Management*, 132(7), ISSN 0733-9364/2006/7-750-758.
- [3] Dogan, S. Z., Arditi, D., & Gunaydin, H. M. (2006). Determining attribute weights on a CBR model for early cost prediction of structural systems. *Journal of Construction Engineering and Management*, 132(10), pp.1092-1098. Du, Y. K., Seung, H. H., Hyoungkwan, K., & Heedae, P. (2009). *Structuring the prediction model of project performance for international construction projects: A comparative analysis*. Retrieved from <https://doi.org/10.1016/j.eswa.2007.12.048>.
- [4] Elhag, T. M. S., & Boussabaine, A. H. (1999). Evaluation of construction costs and time attributes. In *Proceedings of the 15th ARCOM Conference*, Vol. 2, 15-17 September, Liverpool John Moores University, pp. 473-80.
- [5] El-Kholy, M. (2015). Predicting cost overrun in construction projects. *International Journal of Construction Engineering and Management*, 4(4), pp.95-105.
- [6] Hegazy, T. and Ayed, A. (1998). Neural Network Model for Parametric Cost Estimation of Highway Projects. *Journal of Construction Engineering and Management*, 124(3), ASCE, ISSN 0733-9364/98/0003-0210-0218.
- [7] Gebrehiwet, T. & Luo, H. (2017). Analysis of delay impact on construction project based on RII and correlation coefficient: Empirical study. In *proceedings of the Creative Construction Conference 2017, CCC 2017*, 19-22 June 2017, Primosten, Croatia, pp.366-374. doi: 10.1016/j.proeng.2017.07.212.
- [8] Gunduz, M., Nielsen, Y., & Ozdemir, M. (2014). Fuzzy assessment model to estimate the probability of delay in Turkish construction projects. *Journal of Management in Engineering*, 31(4), 04014055. doi: 10.1061/(ASCE)ME.1943-5479.0000261.
- [9] Imad Aleithawe, R. Ralph Sinno and William H. McAnally (2012). Right-of-Way Acquisition Duration Prediction Model for Highway Construction Projects. *Journal of Construction Engineering and Management*, 138(4), ASCE, ISSN 0733-9364/2012/4-540-544.
- [10] Karami, H. and Olatunji, O.A. (2017). Delay causations in marine infrastructures, a review. In *proceedings of International Conference on Durability of Building and Infrastructures*, Miri, Sarawak, Malaysia, pp.167-170.
- [11] Maria, G., & Georgios, N. A. (2017). Neural networks and statistical analysis for time and cost prediction models of urban redevelopments projects. *International Journal of Information Systems and Social Change*, 8(4) pp.37-52. Doi:10.4018/IJSSC.2017100103.
- [12] Michal, J., Agnieszka, L., & Krzysztof, Z. (2018). ANN based approach for estimation of construction costs of sports fields. *Complexity*, Vol. 2018, Article ID 7952434, 11 pages. Retrieved from <https://doi.org/10.1155/2018/7952434>.
- [13] RunZhi Jin, K. C., ChangTaek, H., & Myung Jin, S. (2011). *MRA-based revised CBR model for cost prediction in the early stage of construction projects*. Retrieved from <https://doi.org/10.1016/j.eswa.2011.11.018>.
- [14] Sweis, G., Sweis, R., Abu Hammad, A., & Shboul, A. (2008). Delays in construction projects: The case of Jordan. *International Journal of Project Management*, 26(6), pp.665-674. doi: 10.1016/j.ijproman.2007.09.009.
- [15] Trost, S. M., & Oberlender, G. D. (2003). Predicting accuracy of early cost estimates using factor analysis and multivariate regression. *Journal of Construction Engineering and Management*, 129(2), pp.198-204.
- [16] Yahia, H., Hosny, H., & Abdel Razik, M. (2011). Time Contingency Assessment in Construction Projects in Egypt using Artificial Neural Networks Model. *IJCSI International Journal of Computer Science Issues*, Vol. 8, (Issue 4, No 2), pp.523-530.