Wiley StatsRef:
Statistics Reference Online

Project Management in the Oil & Gas Industry – A Bayesian Approach

By *Franco Caron*[1] *and Fabrizio Ruggeri*[2]

Keywords: *project control, estimate to complete, Bayesian approach, earned value management*

Abstract: A reliable "estimate at completion" from the early stage of project execution is essential in order to enable efficient and proactive project management. The *nonrepetitive* and *uncertain* nature of projects and the involvement of *multiple stakeholders* require the use and integration of multiple informative sources in order to provide accurate forecasts. Moreover, in the Oil & Gas industry, projects are characterized by a high level of complexity and financial impact.

The article aims at multiple objectives such as introducing the need for the identification and utilization of all the available knowledge in order to improve the forecasting process; developing a Bayesian approach in order to integrate the diverse knowledge sources; exploring the integration of data records and experts' judgment related to the ongoing project; exploring the integration of data records related to projects completed in the past and to the ongoing project; and finally developing a Bayesian model capable of using three different knowledge sources, data records and experts' judgments related to the ongoing project and data records related to similar projects completed in the past.

The model has been tested in a set of large and complex projects in the Oil & Gas industry, in order to forecast the final duration and the final cost. The results show a higher forecasting accuracy of the Bayesian model compared to the traditional earned value management (EVM) methodology.

1 Introduction

From a survey analyzing more than 300 megaprojects, it appeared that 65% of the industrial projects with a minimum budget of 1 billion US dollars did not succeed in meeting the objectives of cost, duration, and quality[1]. Even though advanced project management systems have been extensively implemented in the recent years, failure to meet project objectives is very common, in particular in large engineering projects such as in the Oil & Gas industry. However, it remains an open question whether these failures are due to

[1]Politecnico di Milano, Milan, Italy
[2]Istituto di Matematica Applicata e Tecnologie Informatiche, Milan, Italy

Wiley StatsRef: Statistics Reference Online, © 2014–2016 John Wiley & Sons, Ltd.

DOI: 10.1002/9781118445112.stat07893

a lack of project performance during the execution stage or to a lack of forecasting accuracy during the planning and control process. In the former case, both positive and negative deviations from the baseline should be expected, depending on the specific performance of each project. On the contrary, a systematic overrun in terms of cost and/or time may be easily explained as a weakness in the forecasting process since the beginning of the project.

In both cases, that is, initial planning (before execution) and project control (during execution), the forecasting process plays a critical role, as, relying upon sound estimates about the future, the project team can steer the ongoing project in order to meet specific time and cost objectives[2]. Without a good forecast, there can be no rationale in making a decision and we will have to rely just on project capability to adapt to emerging circumstances.

Planning and forecasting are strictly intertwined, both at the project outset when the project baseline must be determined and then across the entire project life cycle when corrective measures must be taken[3]. Forecasting feeds (re)planning and (re)planning feeds forecasting. This approach to project control corresponds to a *feed-forward* control loop[4], as analysis of the future informs present-day decisions.

At a given time in the project life cycle, that is, the time now (TN), a certain amount of the work will already be completed (work completed, WC), while the rest of the work will have to be completed in the future, corresponding to the work remaining (WR). The cost and time performance related to the WC will be known in detail, while an estimate as accurate as possible has to be developed for the WR (Figure 1), in particular about possible budget overrun and completion delay.

It should be noted that both the *accuracy of the forecast about WR* and the *impact of the corrective actions* that may be implemented based on the above forecast will depend on the progress of the project at the TN. The effectiveness of the corrective actions is greater in the early stages of the project execution and progressively diminishes as progress increases: in fact, as progress increases, the degrees of freedom available to steer the project tend to reduce progressively. On the other hand, the capability to forecast the project final duration and cost follows an opposite trend. In fact, at an early stage in the execution phase, the knowledge available to the decision maker is scant and rapidly evolving; therefore, the capability to provide a *reliable forecast* is reduced, particularly if the forecast is only based on the analysis of the performance of the ongoing project until the TN, without considering any other knowledge source. Moreover, the planning process is affected by different sources of bias, which influence the accuracy of final cost and duration estimates, including cognitive, psychological, and political sources[5]. The influence of such bias sources may depend on an exclusively *internal* approach to forecasting, that is, based on knowledge related only to the ongoing project[6].

From a motivational point of view, a typical example of bias is given by excessive optimism aiming at emphasizing the positive and downplaying the negative in order to get the project approved or funded. An outside view, also known as *reference class forecasting*, may mitigate both cognitive and motivational biases

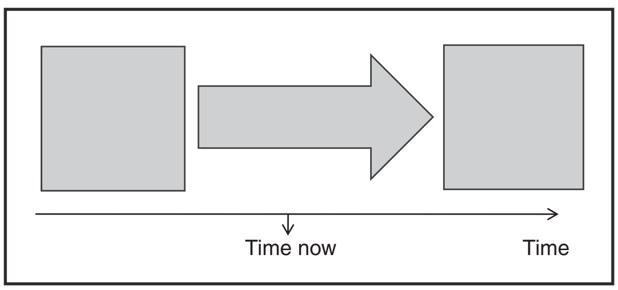

Figure 1. Estimation at completion at time now *(internal view)*.

by considering data records related to a class of similar projects completed in the past. From this situation, the need emerges to exploit other knowledge sources outside the project, in order to minimize any bias. In fact, integration between the knowledge stemming from the *internal* view and the knowledge stemming from the *external* view is needed, the latter related to the projects completed in the past (Figure 2)[7].

In fact, the ongoing project can be seen as belonging to a cluster of similar projects developed by the company. Note that the selection of the cluster is basically subjective as it depends on the similarity criteria adopted[8]. In fact, in some cases, similarity is affected by a strong ambiguity. For example, if a company has to estimate the cost of an investment in a new technology and, moreover, in an unfamiliar technological domain, should it take into account the set of highly innovative projects developed in different technological domains or the set of barely innovative projects but belonging to the same technological domain? Neither the former nor the latter option may be a satisfactory solution but both should be taken into consideration. Moreover, a similarity assessment should also take into account the trade-off between using a large number of past projects, leading to the risk of including projects substantially different from the current one, and using a small number of projects strictly similar to the current one, leading to the risk of losing statistical significance. Unfortunately, owing to differences in technical innovation, geographical area, type of customer, execution time, size, and so on, in a typical engineering and construction company, it is difficult to collect a large number of cases, for example, larger than five, similar to the ongoing project, in particular in a small/medium-size company.

In summary, all the available knowledge should be used in order to improve the planning and control process for a complex project[9−11].

In general, the knowledge available to the project team may be classified in two ways: explicit/ tacit and internal/external. Explicit internal knowledge corresponds to the data records related to the WC, allowing for an evaluation of the performance trend at the TN. Tacit internal knowledge is related to the experts' judgment about possible events/situations affecting the project's WR. Explicit external knowledge corresponds to the data records about similar projects completed in the past. Tacit external knowledge concerns the identification of similarities between the current project and past projects in order to transfer past data to the current project.

In general, the specific contribution given by tacit knowledge, that is, by experts about the future development of the project, may concern the following:

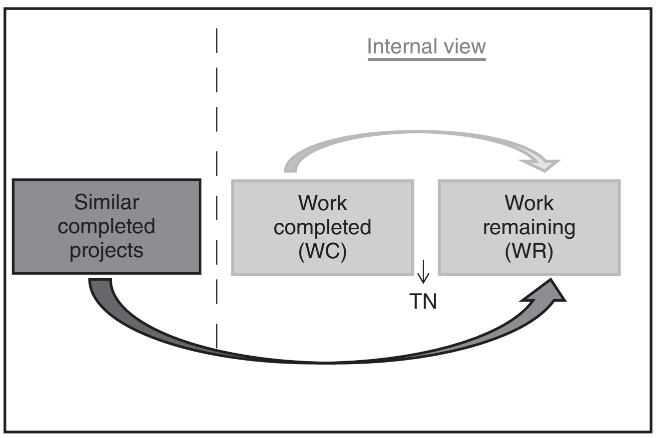

Figure 2. Internal and external views.

- Hindsight, that is, monitoring drivers explaining the project development during WC and presumably affecting also WR, that is, answering the question: what kind of plausible drivers may have generated the actual development of the project till the TN and how they will also influence the future? (e.g., drivers such as schedule aggressiveness, degree of engineering completeness, owner involvement, turnover in project leadership, anomalous low bid from subcontractors, unsatisfied stakeholders, new technology, project team integration, and project team staffing);
- Insight, that is, making sense of the weak signals that anticipate the emerging situations possibly affecting project performance (e.g., weak signals such as frequent scope changes, permits delay, engineering sequence not aligned with construction, high rate of rework in construction, and missing data in design process)[1,12,13].
- Foresight, that is, anticipating certain/uncertain events or conditions affecting project performance during WR, which may originate both internally and externally to the project. Certain events may include planned corrective actions or contractual constraints, while uncertain events, that is, risks, may arise both in terms of threats (e.g., adverse weather conditions) and opportunities (e.g., more efficient solutions deriving from suppliers' collaboration); in particular, anticipating possible behaviors of the stakeholders involved in the project, for example, opportunistic behavior.

It should be noted that in the model, the contribution deriving from the data records to the estimate at completion may be considered accurate just for the near future and losing accurateness if a more extended horizon is considered. As the reliability of the estimation based just on data records rapidly decreases, thus the contribution deriving from the subjective information expressed through experts' opinions becomes predominant.

2 Knowledge Integration – Bayesian Approach

The **Bayes' Theorem** represents a formal technique for the integration of explicit and tacit knowledge. In a Bayesian framework, the experts' preliminary opinions are based on subjective probability, which is the only probability concept applicable to nonrepetitive processes such as projects. The concept of subjective probability is strictly related to the level of available information.

Subjective probability is defined as the degree of belief in the occurrence of an event, by a given person at a given time and with a given set of information. It should be noted that increasing the level of knowledge available may modify the value of the subjective probability assigned to a future event or the degree of belief about it[14]. In general, we can assume that in terms of impact on project performance, any proactive action involving the future may be considered a gamble, such as a corrective action implemented during the project control process. Several Bayesian models have been proposed in literature to formulate reliable forecasts in the project control process[15,16]. In particular, the Bayesian adaptive model introduced by Gardoni *et al.*[17] and the model by Kim and Reinschmidt[18] aim at forecasting the actual duration of a project, by deriving the S curve describing its progress along the life cycle. In the first paper, the results show the importance of integrating the information extracted from a cluster of similar projects to obtain reliable predictions, when the project is in its early stage. In the second paper, the proposed method provides reliable time estimates as it is able to extract more information from the common monthly reports. Both methods provide a higher accuracy than the traditional earned value management (EVM) approach; however, both approaches focus uniquely on time performance and require quite complex computations, if compared to the extremely simple EVM formulas. Another paper[16] presents a probabilistic project cost forecasting model that systematically integrates project cost risk assessment and actual performance data into a computationally efficient probabilistic project cost forecast.

The essence of Bayesian inference is in the use of probability to describe our state of knowledge about some event or parameter of interest (e.g., the cost of an item). A prior distribution (based on expert's tacit knowledge) is updated by means of experimental observations (data records) collected during the project execution process, in order to obtain a posterior distribution, integrating both knowledge sources. For instance, the project team may assume a prior distribution of the final budget overrun, based on subjective expectations about the development of the ongoing project, and this prior distribution may be updated based on the actual performance of the ongoing project until TN[19,20].

The above statement is easier to understand if the formulation of the Bayes theorem is considered:

$$f(\mu|y_1, y_2, \ldots, y_n) \propto L(\mu; y_1, y_2, \ldots, y_n) \cdot f(\mu) \tag{1}$$

where we omit the normalizing constant (i.e., probability integrates to one) in order to enlighten the three main components:

- $f(\mu)$ is the prior distribution of the parameter of interest for inference, μ: it summarizes the initial subjective opinion detained by the decision maker about the probability density function of the parameter μ;
- $L(\mu; y_1, y_2, \ldots, y_n)$ is the likelihood function, obtained after the collection of the n experimental data $y_1, y_2, \ldots, y_n$;
- $f(\mu|y_1, y_2, \ldots, y_n)$ is the posterior distribution, that is, the distribution that expresses the knowledge acquired about the parameter μ after updating the initial subjective judgments with the experimental data.

Equation (1) represents a formal method to update the prior information $f(\mu)$, which reflects experts' tacit actual knowledge, taking into account a series of n past observations x_is, through the likelihood $L(\mu; y_1, y_2, \ldots, y_n)$.

3 Earned Value Management System

The basis for any forecasting process is that the future has its seeds in the present (and the past). In project planning and control, trend analysis is a very popular approach. The most popular version of the trend analysis applied to project planning and control is the EVM[21]. It is based on linear extrapolation, as the performance related to WC is linearly extrapolated to WR[22]. This linear extrapolation normally concerns the parameters' cost performance index (CPI_f) and schedule performance index (SPI_f), where f means future. These indexes can take into account not only data records related to WC but also subjective expectations made by the decision makers about the WR. EVM is an efficient and popular performance measurement and reporting technique for estimating cost and time at completion (TAC)[20]. In an EVM framework, the role of the estimate to complete (ETC), that is, how much money and time is needed in order to complete the project, is critical, as the information drawn from the ETC is essential in order to identify suitable corrective actions aiming to achieve the project objectives.

The following basic parameters are used in EVM, where TN indicates the TN, that is, the time along the project life cycle at which the control process is implemented:

- Planned Value (PV), the budget cost of work scheduled at TN;
- Earned Value (EV), the budget cost of work completed at TN;
- Actual Cost (AC), the actual cost of WC at TN.

EVM was improved by Lipke[23], who introduced the concept of earned schedule (ES) for obtaining a measure of the SPI based on time units and overcoming the flaws associated with an SPI, defined as the ratio between EV and PV, both converging toward BAC (budget at completion) and consequently determining a decreasing reliability of the ETC with the project progress. ES is the time at which the EV achieved at TN should have been obtained according to the project baseline. The new $SPI_{(t)}$ at TN, defined as the ratio between ES and TN, allows for a more accurate linear extrapolation. The above three parameters and the ES, all of them evaluated at TN, allow for the calculation of a set of indices and variances at TN. Firstly, the basic parameters indicating the performance of the project at TN:

- Cost Variance $CV = EV - AC$
- Schedule Variance $SV_{(t)} = ES - TN$

Secondly, the basic parameters allowing for the linear extrapolation of the index to the future:

- Cost Performance Index $CPI = EV/AC$
- Schedule Performance Index $SPI_{(t)} = ES/TN$;

Variances CV and $SV_{(t)}$ summarize the project's past performance during WC, while indexes CPI and $SPI_{(t)}$ may be used in order to extrapolate the current trend and estimate the future performance during WR[24].

Many formulas for estimate at completion have been proposed during almost 50 years of applying EVM but none of them has proved to be always more accurate than the others. In the basic approach, the estimate of final cost (i.e., EAC) and final duration (i.e., TAC) is based on the following equations:

$$EAC = AC + (BAC - EV)/CPI_f \tag{2}$$

where

BAC = budget at completion.
CPI_f = cost performance index estimated for the WR.

$$TAC = TN + (SAC - ES)/SPI_f \tag{3}$$

where

SAC = scheduled at completion, that is, the planned duration of the project.
SPI_f = schedule performance index estimated for the WR.

The traditional EVM's approach was based on the use of CPI and SPI values related exclusively to WC in Equations (2) and (3). Nevertheless, it should be noted that future performance values may significantly differ from past performance. The new performance indices CPI_f and $SPI_{(t)f}$ have been introduced in Equations (2) and (3) with reference to the WR and may consider a different evolution of the project from the expected, leading to a different performance. While the generic indices CPI and $SPI_{(t)}$ are related to the overall WC, CPI_f and $SPI_{(t)f}$ are related to the overall WR. In fact, relying only on past performance while developing a forecast could be misleading, as considering only past values of CPI and $SPI_{(t)}$ is similar to driving a car while looking just in the rear-view mirror, thus making it impossible to dodge the obstacles that may lie on the road ahead. Therefore, the contribution given by experts' judgment, the only contribution looking forward, reveals very important. Both Equations (2) and (3) indicate that the values assigned to the performance indexes CPI_f and SPI_f play a critical role in obtaining an accurate estimate of the final cost and duration. As a consequence, forecasting capability can be improved by utilizing all the available knowledge about the performance indexes CPIf and SPIf. A Bayesian approach allows for the integration of the knowledge content stemming from diverse knowledge sources in a rigorous way.

Using a set of case studies from the Oil & Gas industry, this article summarizes the results deriving from the application of a Bayesian approach to determine the ETC for a project. Three types of projects in the Oil & Gas industry will be considered: offshore, onshore, and subsea projects, respectively.

Offshore projects concern offshore facilities for drilling and extraction of hydrocarbons. Fixed or floating platforms can be used, depending on water depth. The extracted hydrocarbons are then transported onshore through a sea-line system.

Onshore projects are characterized by the construction and installation of onshore facilities. The liquid and/or gas hydrocarbons extracted from the wells are firstly stored and then the liquids are sent through flow lines to a preliminary treatment unit and then to a refinery. The gas follows a similar process until the preliminary treatment; afterward, it is sent for further treatment through a gas pipeline.

Subsea projects are characterized by undersea facilities for extraction and production. The need to install undersea facilities is due to the technical or economical unfeasibility to utilize offshore platforms. When there are more undersea wells, the wells are interconnected through flow lines, in addition to the sea lines that link the wells of extraction and production with onshore facilities.

In the Oil & Gas industry, the project life cycle can be divided into a sequence of phases. At the end of each single phase, a review of the project must be completed and a new decision to proceed must be made:

- *Evaluation*. Carrying out the feasibility study of the project concerning a previously identified opportunity and the evaluation of its alignment to the business strategy;
- *Concept Selection*. Developing alternative concepts in terms of technical and economical solutions and choosing the alternative that maximizes the project value;
- *Concept Definition*. Developing a detailed design and planning of the selected concept;
- *Execution*. Executing the project while aiming at meeting the project baseline;
- *Commissioning, Start-Up, and Performance Test*. Preparing and completing the final test representing the prerequisite for the start-up of the operation phase (i.e., first oil).

Three models will be presented, each using a different set of knowledge sources:

- The first model utilizes data records related to WC and experts' judgment, by adjusting through experts' judgment the trend resulting from data records about the expected performance during WR[19];
- The second model integrates the "internal" view, that is, data records related to the ongoing project, and the "external" view, that is, data records related to past similar projects[25]
- The third model summarizes all the knowledge sources, utilizing data records related to WC and experts' judgment related to WR, with data records deriving from similar projects completed in the past[26].

In the last approach, besides the use of internal knowledge, both explicit and tacit, external knowledge related to similar projects completed in the past has been used.

4 Model 1 – Integration of Data Records and Experts' Judgment

The first model explores the improvement for the forecasting process deriving from the integration of two knowledge sources: data records (related to WC) and experts' judgment (related to WR), both related to the ongoing project. The proposed model is based on the use of a Bayesian approach that takes into account available information in terms of experts' opinions and data records, in order to establish the value of the future indices CPI_f and $SPI(t)_f$.

The proposed Bayesian model is made up of three phases: data records analysis, elicitation of experts' judgment, and calculation of posterior distribution of CPI_f and $SPI_{(t)f}$.

Firstly, in order to implement the forecasting process, at each TN along the project life cycle, available data records corresponding to the monthly values CPI_m and $SPI_{(t)m}$, are collected. The prior distribution of the indices CPI_f and $SPI_{(t)f}$ reflects experts' knowledge about project trends, future threats and opportunities, expected stakeholders' behavior, impact of corrective actions, etc., corresponding to the main contribution given by the experts to the forecasting process. The model aims at combining prior distribution based on experts' judgment with the evidence provided by the data records collected during WC, in order to update the forecast values of performance indices concerning WR. The posterior distribution of CPI_f and $SPI_{(t)f}$ concerning WR enables to identify a confidence interval for the estimate at completion both in terms of cost and time, that is, EAC and TAC. In fact, the Bayesian model expresses the estimates in terms of a probability density function for the forecasted final cost and duration. Hence, the project manager has an indication of the degree of confidence about the expected value forecasted, which results in better quality information available for the decision-making process.

The Bayesian model has been applied to the construction of two pipelines within a development project for a sweetening and stabilization plant in the Oil & Gas industry, in order to compare its accuracy with the traditional EVM formulas.

Regardless of the actual values of $CPI = 1.292$ and $SPI_{(t)} = 0.678$ at TN (April 2010), the experts do not agree that the past trend that emerges from WC data should be extended to WR and think that the estimate for WR will be more similar to the PVs (experts' mean estimate is 1.086 for CPI_f and 0.971 for $SPI_{(t)f}$). This is a typical case in which subjective knowledge exercises a significant influence in the forecasting process by changing the trend deriving from data records.

The posterior distribution for CPI_f resulted wider than that of $SPI_{(t)f}$. In general, in the Oil & Gas sector, the schedule performance is the most critical, as delays in the first oil date (the day in which the plant starts the production) can cause a huge financial loss for the owner. For this reason, the project team focuses their efforts on completing the project as soon as possible, in spite of the cost performance trend that is expected to worsen due to the incentives for the subcontractors in order to accelerate project progress.

The results of the Bayesian model are compared (Table 1) with traditional EVM formulas.

In Table 1, the estimated values of CPI_{AC} and $SPI_{(t)AC}$, that is, the values of the indexes at completion, are reported for different EVM cases and the Bayesian model (see column "case"). The first column indicates the underlying assumptions for each case. As for the first case, in the last column of Table 1, *m* indicates the monthly value and 3 and 6 indicate the 3 and 6 months moving average values, respectively. In comparison with the predicted values of the indices listed in Table 1, the actual performance of the project, related to 96% cumulative physical progress, shows a value of 1.46 for CPI_{AC} and a value of 0.92 for $SPI_{(t)AC}$.

Table 1. Comparison of different estimate at completion formulas.

Assumptions	Case	CPI_{AC}	$SPI_{(t)AC}$	Index type/time interval
Traditional EVM	EVM 1	1.995	0.581	*m*
	EVM 1	1.649	0.591	3
	EVM 1	1.369	0.603	6
	EVM 1	1.435	0.996	SCI_m
	EVM 1	1.136	0.854	SCI_3
	EVM 1	0.922	0.738	SCI_6
Bayesian approach	1 – Bayesian model ETC	1.806	0.993	Lower bound
	1 – Bayesian model ETC	1.465	0.813	Mean
	1 – Bayesian model ETC	1.128	0.598	Upper bound
Actual performance computing at 96% of physical progress		1.46	0.92	Progress at 96%

Actual project performances are included in the range provided by the Bayesian model and demonstrate its accuracy.

Among the different indices reported in Table 1, the most accurate seems to be SCI_m, hardly ever used in estimate at completion because of its pessimistic bias. SCI_3 appears to be a good estimator of schedule performance but doesn't perform well in cost performance. In general, SCI_m, SCI_3, and SCI_6 seem to work quite well in this case because there is a clear trade-off between time and cost performances. This trade-off has been recognized also by subjective judgments, which provide accurate estimates at completion for both cost and schedule as actual performances are within the model range and, moreover, close to its mean.

5 Model 2 – Integration Between Internal View and External View

The second model explores the improvement deriving for the forecasting process from the integration of the internal and external knowledge sources. It should be noted that the previous model was based exclusively on an internal view of the project.

This model aims at using the information related to the actual performance of the current project, corresponding to the "internal view," and the information related to the cluster of similar projects completed in the past, corresponding to the "external view." In order to integrate both types of information, a Bayesian model has been developed, allowing for the updating of a *prior* estimate based on the external view by means of the data records collected during the progress of the ongoing project, in order to obtain a *posterior* estimate of the final cost and duration of the project. This approach allows for the mitigation of possible optimistic biases, which can affect the project control process, particularly at the project early stage.

The process leading to identify the cluster of similar past projects comprises the following steps:

1. recording the data related to past projects;
2. classifying projects (by similarity criteria, by transferability of past experience to the current projects, etc.);
3. selecting clusters of similar projects;
4. analyzing each cluster in terms of relevant performance parameters such as the distribution of the final budget overrun.

The Bayesian model is described by the following two equations.

The *likelihood* function is given by

$$\frac{x(t)}{K(t)} \sim N\ \ (Z;\ \ \sigma_x) \tag{4}$$

where

$K(t)$: physical progress percentage at TN;
$x(t)$: overrun percentage cumulated at TN;
$\frac{x(t)}{K(t)}$: overrun percentage extrapolated to the end of the project;
Z: true value, to be estimated, of the final overrun percentage;
σ_x: standard deviation of the observation $\frac{x(t)}{K(t)}$.

Note that in the model, the likelihood function used to update the *prior* distribution is based on a single observation $\frac{x(t)}{K(t)}$ obtained at TN, where $\frac{x(t)}{K(t)}$ represents the observed value affected by error of the true value Z of the final overrun percentage. The project's physical progress percentage $K(t)$ at TN may be defined as the ratio between the amount of work already done, that is, WC, and the overall amount of work to be done, that is, WC+WR. The quantity $x(t)$ may indicate both cost overrun and time overrun percentages.

The prior distribution is given by:

$$Z \sim N\ (\theta; \sigma_{rc}) \tag{5}$$

where

Z: parameter to be estimated, overrun percentage at the end of the project;
θ: expected value of the parameter to be estimated obtained by the *prior* distribution;
σ_{rc}: standard deviation of the parameter to be estimated obtained by the *prior* distribution.

As shown in Equation (6), the *posterior* probability density function of the parameter Z to be estimated is obtained by a combination between *prior* probability density function and likelihood function:

$$f\ \ (Z|(x(t))/K(t)) \propto f\ \ ((x(t))/K(t)|Z) \cdot f\ \ (Z) \tag{6}$$

The Bayesian model has been tested on three industrial projects in the Oil & Gas industry, each associated to a cluster of similar projects completed in the past, in order to test its forecasting accuracy compared to the traditional formulas applied in the *Earned Value Management System* (*EVMS*).

The clusters of similar projects considered in the case study are the following:

- *Subsea* projects (wellheads and pipeline installation under the sea);
- *Offshore* projects (facilities construction and oil extraction in the sea);
- *Onshore* projects (facilities and pipelines construction on land).

The corresponding *prior* distributions (Equation 5) were obtained for each reference cluster.

After having identified the reference clusters of projects, a single project currently in progress was chosen for each of the reference clusters in order to test the accuracy of the forecasting model.

Data analysis pointed out that the three clusters present a different behavior in terms of cost performance and a similar behavior in terms of schedule performance. In particular, the *subsea* cluster is characterized by the lowest values of cost overrun percentage and related dispersion, while the onshore cluster presents the highest values.

The following characteristics explain the different behaviors of the *subsea* cluster:

- More standard technologies;
- Larger contingency used to cover unforeseen events;
- Partnership with a small number of specialized suppliers;
- Highly skilled workforce.

On the other side, the *onshore* cluster is characterized by greater values of cost overrun percentage and related dispersion, mainly due to the different geographic areas where the company operates, often requiring different technologies and numerous interfaces with local stakeholders that adversely influence the project performance.

Two EVM cases have been considered for $\mathrm{CPI_f}$ (*cost performance index future*) and $\mathrm{SPI_f}$ (*schedule performance index future*), respectively:

1. $\mathrm{CPI_f} = \mathrm{CPI_p}$, ($\mathrm{SPI_f} = \mathrm{SPI_p}$), that is, the future performance of the project, related to WR, will correspond to the past performance, related to WC;
2. $\mathrm{CPI_f} = 1$, ($\mathrm{SPI_f} = 1$), that is, the future performance of the project will correspond to the initially planned performance.

As shown in Table 2, the Bayesian model has achieved a high accuracy in estimating cost and duration at completion, in particular at the project outset, thanks to the integration between *internal view* and *external*

Table 2. Output of the Bayesian model versus EVMS formulas versus actual results (cost).

	$K(t)$	BAC	EAC Bayesian	EVMS ($CPI_f = CPI_p$)	EVMS ($CPI_f = 1$)	Final actual cost
Case A (subsea)	12.06	525	509	218	456	523
	58.8		523	380	435	
	86.7		530	516	517	
Case B (offshore)	6.83	81	103	95	82	89
	62.2		90	71	75	
	80.6		91	90	82	
Case C (onshore)	20.08	320	416	314	319	491
	68.1		385	338	332	
	89.57		473	470	455	

view. For example, focusing on the forecast in case A (subsea), at 12.06% of physical progress in February 2010, the exclusive use of the *internal view,* based on the observation at TN of $\frac{x(t)}{K(t)} = -1.09$, would have given a very optimistic estimate. Thanks to the integration with the *external view* information, at this early stage of the project, the final cost estimation becomes 509 million, which is close to the 523 million AC at completion. At 50% progress the estimate becomes even more accurate.

The results achieved by the Bayesian model do not only show a greater accuracy compared to the traditional EVMS formulas at the outset of the project, when it is necessary to make significant decisions and the available information is scant, but also indicate a greater stability of the forecast along the life cycle of the project. Indeed, the results achieved by the traditional EVMS formulas feature a greater volatility along the project life cycle, as shown in Table 2. In the Bayesian model, the stability has been obtained by means of the contribution of the *external view*, which, in general, represents a stabilization factor for the internal view, particularly in the early phase of the project. In the early stage of the project life cycle, the *prior* probability density function, that is, the *external view*, has a greater weight than the *internal view*. Across the project life cycle, the weight of the *internal view* overtakes progressively the *external view*.

6 Model 2 – Integration of Experts' Judgment and Data Records Related to the Ongoing Project and to Past Similar Projects

The third model uses three different sources of knowledge: data records and experts' judgment about the ongoing project and data records about similar projects developed in the past. Internal data records highlight the current performance trend of the project, experts' judgments express a forward looking view, and data records related to past similar projects increase realism and stability of the estimate.

In order to collect current performance values, in terms of EVM metrics, two performance indexes can be introduced as follows:

$$\mathrm{ND}_{(\mathrm{c})\mathrm{TN}} = \frac{\mathrm{AC}_{\mathrm{TN}} - \mathrm{EV}_{\mathrm{TN}}}{\mathrm{EV}_{\mathrm{TN}}} \tag{7}$$

$$\mathrm{ND}_{(\mathrm{t})\mathrm{TN}} = \frac{\mathrm{TN} - \mathrm{ES}_{\mathrm{TN}}}{\mathrm{ES}_{\mathrm{TN}}} \tag{8}$$

where $\mathrm{ND}_{(\mathrm{c})\mathrm{TN}}$ stands for normalized deviation (cost) at the TN and similarly $\mathrm{ND}_{(\mathrm{t})\mathrm{TN}}$; when referring to the performance indicator in general, without distinction between time and cost, the acronym ND, normalized deviation, will be used.

In the Bayesian model, the parameter of inference is the normalized deviation of cost and TF, identified by the Equations (9) and (10):

$$\mathrm{ND}_{(c)} = \frac{\mathrm{AC}_{\mathrm{TF}} - \mathrm{BAC}}{\mathrm{BAC}} \tag{9}$$

$$\mathrm{ND}_{(t)} = \frac{\mathrm{TF} - \mathrm{SAC}}{\mathrm{SAC}} \tag{10}$$

where $\mathrm{AC}_{\mathrm{TF}}$ and TF indicate the final actual cost and duration of the project, while BAC and SAC indicate the planned cost and duration of the project, respectively. When a budget overrun is observed, ND is greater than zero, whereas if an underrun is recorded, ND is less than zero.

The sequence of steps required for the development of the model is as follows:

- Distribution based on experts' opinions
- Distribution of similar past projects
- Definition of the weight of each knowledge source
- Prior distribution based on a mixture of the two previous distributions
- Likelihood function
- Posterior distribution.

The prior distribution is a summary of the initial knowledge available to the decision maker for the forecasting process. For this purpose, the project team can rely on two main information sources: experts' opinions and data records related to similar projects completed in the past. By integrating these two components, a starting assumption about the distribution of $\mathrm{ND}_{\mathrm{TF}}$ can be formulated. In order to translate these concepts in Bayesian terms, firstly two distributions have to be defined: one representing the experts' opinions and one the cluster of similar past projects. This approach should safeguard the subjectivity of the similarity analysis in order to confirm the subjective nature of the overall prior distribution.

Secondly, a unique distribution has to be derived by the integration of these two. In mathematical terms, the integration can be obtained by the concept of *mixture:* assigning a value ranging from 0 to 1 to a weight ε, the prior distribution is obtained as the weighted sum of the distribution of experts' opinions and the distribution of the cluster of similar projects:

$$\prod(\cdot) = \varepsilon \prod_{\mathrm{E}}(\cdot) + (1-\varepsilon)\prod_{\mathrm{S}}(\cdot) \tag{11}$$

In Equation (11):

- is a generic point of the prior distribution;
- $\prod_{\mathrm{E}}(\cdot)$ is the evaluation of the distribution of experts' opinions in a generic point $\cdot$;
- $\prod_{\mathrm{S}}(\cdot)$ is the evaluation of the distribution of the cluster of similar projects in a generic point $\cdot$;
- ε is the weight.

Adjusting the value assigned to ε, it is possible to give more or less weight to one of the two informative sources. When ε increases, the weight of the experts' distribution increases, while the weight of the distribution of similar projects decreases. Consequently, the prior distribution is characterized by a high level of flexibility, as it can be formulated according to the different weight of the sources of information available to the project team.

As for the likelihood function, the experimental data collection will be represented by a single observation of the performance index, that is, $\mathrm{ND}_{(c)\mathrm{TN}}$ and $\mathrm{ND}_{(t)\mathrm{TN}}$ (see Equations 7 and 8). The posterior function is a mixture of the posterior distribution of similar projects $\prod_{\mathrm{S}}(\mu|x)$ and the posterior distribution of experts' opinions $\prod_{\mathrm{E}}(\mu|x)$ through the posterior weight ε^*.

Table 3. Comparison of the mean squared error.

Mean squared error	Bayesian model	EVM
MSE cost	1.23×10^{-2}	0.1683
MSE time	9.3×10^{-3}	3.55×10^{-2}

After the three projects reached completion, it was possible to check the error made by the Bayesian model and to compare it with the one that would have been made if traditional EVM forecasting formulas had been applied. Table 3 shows the mean squared error made by the Bayesian Model and by EVM in the three Oil & Gas industry cases.

It is clear that the Bayesian model is more accurate: its mean squared error is 10 times lower, both for cost and time performance, owing to the larger amount of information used, even though the model was tested on very complex and risky projects.

7 Conclusions

Using a set of case studies from the Oil & Gas industry, this article describes the results obtained in project planning and control by three forecasting models based on a Bayesian approach. The article achieved multiple objectives: a general objective related to the identification and utilization of all the available knowledge sources in order to improve the forecasting process; the development of a Bayesian approach in order to integrate diverse knowledge sources in a formal and rigorous way when addressing the project planning and control process; integration of data records and experts' judgment related to an ongoing project; integration of data records related to projects completed in the past and to the ongoing project; development of a Bayesian model capable of using three different knowledge sources: data records related to WC, experts' judgments related to WC, and data records related to similar projects completed in the past; and test of the models in a set of case studies related to large and complex projects in the Oil & Gas industry.

Related Articles

Statistics of Risk Management; **Risk Measurement, Foundations of**; **Statistics in Industry**; **Management Science, Statistics in**; **Rare-Event Risk Analysis**; **Analytical Methods for Risk Management: An Engineering Systems Perspective**; **Methods of Risks Estimation and Analysis of Business Processes**; **Methods of Risks Estimation and Analysis of Business Processes**; **Bayesian Inference**; **Bayesian Robustness**; **Bayesian Forecasting**; **Elicitation**; **Subjective Probabilities: Theory**.

References

[1] Merrow, E.W. (2011) *Industrial Megaprojects: Concepts, Strategies and Practices for Success*, John Wiley & Sons, Inc., Hoboken, NJ.

[2] Dvir, D. and Lechler, T. (2004) Plans are nothing, changing plans is everything: the impact of changes on project success. *Res. Policy*, **33**, 1–15.

[3] Hogarth, R.M. and Makridakis, S. (1981) Forecasting and planning: an evaluation. *Manag. Sci.*, **27** (2), 115–138.

[4] Christensen, D. (1996) Project advocacy and the estimate at completion problem. *J. Cost Anal. Manag.*, **13**, 35–60.

[5] Lovallo, D. and Kahneman, D. (2003) Delusion of success: how optimism undermines executives' decisions. *Harv. Bus. Rev.*, **81**, 56–63.
[6] Kahneman, D. and Tversky, A. (1979) Intuitive prediction: biases and corrective procedures. *TIMS Stud. Manag. Sci.*, **12**, 313–327.
[7] Flyvbjerg, B. (2006) From Nobel Prize to project management: getting risk right. *Project Manag. J.*, **37**, 5–15.
[8] Green, K.C. and Armstrong, J.S. (2007) Structured analogies for forecasting. *Int. J. Forecast.*, **23**, 365–367.
[9] Caron, F. (2014) *Managing the Continuum: Certainty, Uncertainty, Unpredictability in Large Engineering Projects*, Springer, Milan, Heidelberg, New York, Dordrecht, London.
[10] Reich, B.H., Gemino, A., and Sauer, C. (2014) How knowledge management impacts performance in projects: an empirical study. *Int. J. Project Manag.*, **32** (4), 590–602.
[11] Schindler, M. and Eppler, M.J. (2003) Harvesting project knowledge: a review of project learning methods and success factors. *Int. J. Project Manag.*, **21**, 219–228.
[12] Williams, T., Klakegg, O.J., Walker, D.H.T., *et al.* (2012) Identifying and acting on early warning signs in complex projects. *Project Manag. J.*, **43** (2), 37–53.
[13] Haji-Kazemi, S., Andersen, B., and Krane, H.P. (2013) A review on possible approaches for detecting early warning signs in projects. *Project Manag. J.*, **44** (5), 55–69.
[14] De Finetti, B. (1937) La prévison: ses lois logiques, ses sources subjectives. *Annales de l'Institut Henri Poincaré*, **7** (1), 1–68.
[15] Palomo, J., Ruggeri, F., Rios Insua, D., *et al.* (2006) On Bayesian forecasting of procurement delays: a case study. *Appl. Stoch. Model. Bus. Ind.*, **22**, 181–192.
[16] Kim, B. (2015) Integrating risk assessment and actual performance for probabilistic project cost forecasting: a second moment Bayesian model. *IEEE Trans. Eng. Manag.*, **62** (2), 158–170.
[17] Gardoni, P., Reinschmidt, K.F., and Kumar, R. (2007) A probabilistic framework for Bayesian adaptive forecast. *Comput. Aided Civ. Infrastruct. Eng.*, **22**, 182–196.
[18] Kim, B. and Reinschmidt, K.F. (2009) Probabilistic forecasting of project duration using Bayesian inference and the beta distribution. *J. Constr. Eng. Manag.*, **135** (3), 178–186.
[19] Caron, F., Ruggeri, F., and Merli, A. (2013) A Bayesian approach to improve estimate at completion in earned value management. *Project Manag. J.*, **44**, 3–16.
[20] Project Management Institute (2013) *A Guide to the Project Management Body of Knowledge (PMBOK)*, *5th edn* edn, Project Management Institute, Newton Square, PA.
[21] Marshall, R.A., Ruiz, P., and Bredillet, C.N. (2008) Earned value management insights using inferential statistics. *Int. J. Manag. Project. Bus.*, **1** (2), 288–294.
[22] Fleming, Q.W. and Koppelman, J. (2006) *Earned Value Project Management*, *3rd edn* edn, Project Management Institute, Newton Square, PA.
[23] Goodwin, P. (2005) How to integrate management judgment with statistical forecasts. *Foresight*, **1**, 8–12.
[24] Anbari, F. (2003) Earned value: project management method and extension. *Project Manag. J.*, **34**, 12–23.
[25] Caron, F., Ruggeri, F., and Borgarucci, C. (2013) Bayesian integration of internal and external views in forecasting project performance. *J. Mod. Project Manag.*, **1** (2), 112–121.
[26] Caron, F., Ruggeri F. and Pierini B., A Bayesian approach to improving estimate to complete, *Int. J. Project Manag.*, , to be published.
[27] Lipke, W. (2003) Schedule is Different. The Measurable News, Summer 2003, 31–34.

Further Reading

Lipke, W. (2003) Schedule is Different. The Measurable News, Summer 2003, 31–34.