POLITECNICO DI TORINO

Master's Degree Course in Engineering and Management

A.y. 2021/2022

December 2022

Literature Review of project Cost Estimate at Completion forecasting methods

Supervisor:
De Marco Alberto
Co-supervisor:
Ottaviani Filippo Maria

Candidate:
Matarese Rosa Pia

TABLE OF CONTENTS

LIST OF FIGURES

ABSTRACT

The estimation of the time and cost at completion represents one of the most important concepts in Project Management. Its importance increased a lot when the Navy's A-12 project of the US Department of Defense (DoD) failed due to cost overruns and time delay. The identification of the best method for the calculation of the Estimate at Completion, became one of the major challenges in the evaluation of the acquisition costs in the purchasing process of military weapons. Also today, the determination of the performance factors and the estimate at completion of time and cost of a project is fundamental for the monitoring and control of the project progress and for the implementation of corrective actions, if some problems arise, to bring back the project on time and on budget. This thesis has the purpose of conducting a literature review of the major models and comparative studies of the EAC calculation methods. The first chapter of this thesis introduce the basic elements of Project Management, following the definitions of the Project Management Institute (PMI). Moreover, there is a description of the Earned Value Management, a management methodology aimed at measuring the progress of the project through its performance factors such as BCWS, BCWP and ACWP and forecasting new ones such a CEAC and TEAC. The second chapter describes the methodology used to conduct the research, identifying the basic steps for a correct Literature Review. The third chapter lists most of the studies conducted by authors regarding the different methods for the calculation of the cost estimate at completion during the years. The estimate at completion literature can be divided into sections based on the different approaches used. The first one considered is the index-based method, which include formulas based on the so-called Past Performance Factors. Then regression-based methods are described, which uses the regression analysis to model the nonlinear cumulative cost growth curves, to overcome the limitations of the linear models. Finally, other methodologies are considered, including also new methods for the EAC calculation which are finalized to fill eventual gaps in previous models. From these studies, as result, emerged that it does not exist a general model for the EAC's forecast that can be used for all the projects; on the contrary the choice of the most accurate method will depend on the sector, type and characteristics of the project and on its percent at completion.

INTRODUCTION

Many of project management techniques that are used nowadays have origins in the US Department of Defense (DoD) which needed to improve its management control system due to increasingly complex weapon systems development contracts. When the Navy's A-12 project of the DoD failed due to cost overruns and time delay, the identification of the best method for the calculation of the Estimate at Completion, became one of the major challenges in the evaluation of the acquisition costs in the purchasing process of military weapons.

Also today, the determination of the performance factors and the estimate at completion of time and cost of a project represents one of the most important concepts in Project Management and it is fundamental for the monitoring and control of the project progress and for the implementation of corrective actions, if some problems arise, to bring back the project on time and on budget.

This thesis has the purpose of conducting a literature review of the major models and comparative studies of the EAC calculation methods.

The first chapter of this thesis introduce the basic elements of Project Management, following the definitions of the Project Management Institute (PMI). Moreover, there is a description of the Earned Value Management, a management methodology aimed at measuring the progress of the project through its performance factors such as BCWS, BCWP and ACWP and forecasting new ones such a CEAC and TEAC.

The second chapter describes the methodology used to conduct the research, identifying the basic steps for a correct Literature Review.

The third chapter lists the most important studies, for the Estimate at Completion forecasting, conducted by authors in the literature, among years, until today. The first one is the is the index-based methods, which include formulas based on the Past Performance Factors. Authors proposed different methods and conducted comparative studies among existing models, starting from the late seventies in the field of military projects. Then regression-based methods are described, in which the EAC is estimated with the use of linear and/or nonlinear regression equations. The index-based method and the regression-based method are the two major categories in the EAC forecasting. However, during the years, other studies were conducted with the aim of overcoming the limitations found in the traditional models and thus to find a better forecasting model. Few probability-based methods are described. A combined index-regression methodology based on the Gompertz growth model is considered with the aim of overcoming the absence of the time factor in the regression-based models and a derived model, which considers also the risk factor, is described. A method based on the use of the Particle Filter, as a solution for the Bayesian filter execution, is described. Soft Computing and Artificial Intelligence approaches, such as neural networks (NN), fuzzy logic (FL), genetic algorithms (GA) and other derivative ones were applied in the EAC forecasting methods with the aim of improving the accuracy of results. Finally, the last method considered and the newest one, developed in 2022, is a multiple linear regression model with the aim of improving the project cost estimate approaches.

CHAPTER 1.

PROJECT MANAGEMENT

The origins of project management go back to ancient times. Throughout history, it was necessary to organize the planning, the coordination and the building of ancient monuments, palaces, and temples. For this reason, humans applied project management practices (even if unknowingly) to their works, long before the introduction of guides, methods, tools, and techniques. Just to name a few, Pyramid of Giza, the Great Wall of China, the Coliseum, Roman Aqueducts and Panama Canal represent the concrete evidence of impressive projects delivered in the ancient world.
Project management, in modern sense, began in the mid-20th century with Henry Gantt who invented the chart that bears his name, and it is still today considered an essential tool for breaking large projects into smaller manageable tasks.
From construction, engineering and defense activity, the field of project management is still evolving and in more recent times is including software implementation.

1.1 BASIC ELEMENTS

To understand the discipline of project management, it’s important first to comprehend some key concepts.

Project Management
According to Project Management Institute, Project Management is *“the application of knowledge, skills, tools, and techniques to project activities to meet project requirements”*.[1]
PMI has divided the huge field of project management into ten knowledge areas (what to know) and five process groups (what to do), which are combined in a matrix forming forty-nine individual processes.

The Project Management Knowledge Area describes project management knowledge and practice in terms of its component processes associated with a particular topic in project management. The Knowledge Areas are ten and are briefly described as follows.

[1] Project Management Institute. (2017). *A guide to the Project Management Body of Knowledge: (Pmbok Guide)* (6th edition).

Project Integration Management: it includes the processes required to identify and define the several activities and ensure that the various elements of the project are properly coordinated and that the different team members follow one overall plan.

Project Scope Management: it includes the processes required to guarantee that all and only the necessary work is included in the project in order to complete it successfully.

Project Schedule Management: it includes the processes required to ensure the completion of the project in terms of time.

Project Cost Management: it includes the processes involved in anything that has to do with costs in order to ensure that the project is completed within the agreed budget.

Project Quality Management: it includes the processes required to ensure that the work delivered from a project owns the quality requirements in order to meet stakeholders' expectations.

Project Resource Management: it includes the processes required to detect, obtain, and manage the resources needed in order to complete successfully the project.

Project Communications Management: it includes the processes required to ensure the appropriate communication of the project information between all the members involved.

Project Risk Management: it includes the processes involved in the risk for a project.

Project Procurement Management: it includes the processes required to obtain external products or services needed.

Project Stakeholder Management: it includes the processes required to identify all the individuals and organizations who are actively involved in the project, or whose interests may be affected by its completion, to analyze stakeholder expectations and their impact on the project and to engaging stakeholders in project decisions and execution.

Project Management Process Groups is a set of project management processes needed to meet project purpose. The Process Groups are five and are briefly described as follows.

Initiating Process Group: the process(es) implemented to determine a new project or a new project phase having obtained the formal approval to start it.

Planning Process Group: the process(es) related to establish the scope of the project, plan the activities and their execution, in order to achieve the project objectives.

Executing Process Group: the process(es) related to the completion of the activities and tasks established in the project management plan in order to fulfill its requirements.

Monitoring and Controlling Process Group: the process(es) implemented to track the project progress in order to identify the performance of the project and undertake corrective actions to bring the project back on time and on budget.

Closing Process Group: the process(es) implemented to complete or close a project, phase, or contract.

Project

According to Project Management Institute, a project is *"a temporary endeavor undertaken to create a unique product, service, or result"*.[2]

Temporary means that every project has a specific beginning and end.

Unique means that the deliverable produced by the fulfillment of a project (product, service, result, combination of one or more products, services, or results) is different from other similar deliverables, remaining unique in its key characteristics.

A project is successfully completed when it is able to deliver a product with the specific features e functions that were originally set. However, a project is subject to constraints: scope, time, cost and quality. The combination of these elements is represented through a triangle, called Project Management Triangle of Constraints.

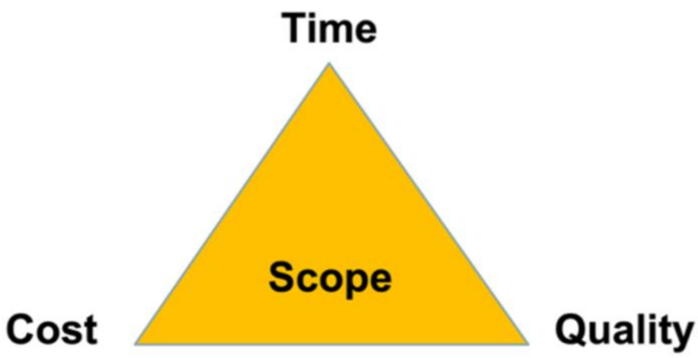


Figure 1: Project management triangle of Constraints

These constraints are defined "competitive", since as one of them changes so do the others. Thus, it is important to reach a trade-off between cost, quality, and time, with the resources available, to achieve the predetermined scope of the project, i.e., the objective of the project and the way the work should be done in order to deliver a product with the specified features and functions. The project manager must therefore make a choice: set one constraint that he or she believes is the most important to the optimal success of the project and vary the others accordingly.

Since projects are unique undertakings, have limited resources and constraints, they involve a degree of uncertainty. For this reason, the monitoring and control of a project is a crucial aspect of project management.

[2] Project Management Institute. (2017). *A guide to the Project Management Body of Knowledge: (Pmbok Guide)* (6th edition).

1.2 MONITORING AND CONTROL

Monitoring and Control (often shortly called Project Control) are two parts of a feedback system, illustrated in Figure 2, aimed at detecting and correcting deviation from desired.[3]

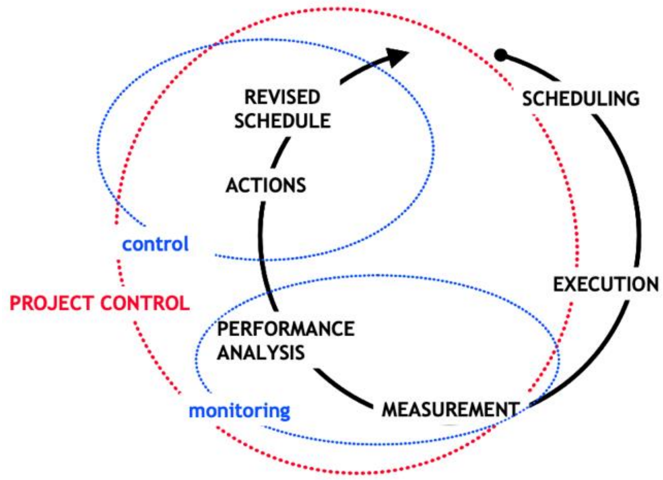


Figure 2: Monitoring and Control as part of a feedback system

Detection is made through monitoring; it follows the planning phase all along the project execution and it is done at regular intervals and/or in exceptional events.

Monitoring purpose is the tracking of the actual evolution of a project through a set of methods and practices to identify, analyze and report real-time project performance data. So, the first step is the measurement of the current cost and time progress, based on performance metrics. Then, after having the complete picture of the actual project status, it is important to compare it with the project management plan and project baseline. Since the project plan is made of numerous activities, a WBS (Work Breakdown Structure) and a CBS (Cost Breakdown Structure) can be used to perform an effective monitoring. The WBS is a deliverable-oriented grouping of project elements (i.e., work packages and elementary activities) that organize and define the total scope of the project. While the CBS, commonly, is the WBS filled with costs data for each work package and detailed tasks, resulting in a hierarchical representation of the budget cost of a project.

This comparison leads the project manager to see if there are discrepancies and so if the project isn't in compliance with the given constraints, measuring the project progress. Generally, it is done by

[3] De Marco, A. (2018). *Project management for Facility Constructions*. (2nd edition). SPRINGER.

looking at the actual resources consumed, that is time and money of each activity of the WBS. For tangible activities, the progress can be calculated in percentage, based on the quantity of the work physical output, through the following equation:

$$Progress = performed\ quality / total\ scheduled$$

However, sometimes simply comparing actual performances against the plan is not sufficient, and moreover, measuring the percentage progress with the quantity is difficult.
So, one method used to estimate the overall project progress is to define the original budget and give a cost weight to each project activity, then measure the activity progress and actual cost and finally calculate the progress as a weighted sum of each activity percent progress.
Finally, with the results of the previous analysis, it is possible to make calculations to estimate the cost and the time at completion of the project. This will be discussed deeply in the following chapters.

The correction is the goal of the control process; it follows the monitoring and is intended to detect the causes and the reason why performances vary from the original plan. Through control activities, corrective actions can be implemented to put in practice appropriate changes needed to bring the project back to what was initially scheduled or to undertake preventive actions to anticipate possible problems.
Monitoring and controlling the ongoing project activities against the planned one is important not only for the project manager who can supervise the project progress, but it also provides the project stakeholders with insight into the status of the project, the possible concerns, and the continuous improvement activities via trend graphs. In this way, the exchange of information with stakeholders makes easier to understand and meet their needs and expectations from the project.

1.3 PERFORMANCE MEASUREMENT: EARNED VALUE MANAGEMENT

Earned value management *"is a management methodology for integrating scope, schedule, and resources; for objectively measuring project performance and progress; and for forecasting project outcome"*[4].
It's a valuable management tool for the monitoring of project progress. It enables the measurement of the actual cost, schedule and technical performances and its comparison with the agreed ones.

[4] Project Management Institute. (2011). *Practice standard for earned value management* (2nd edition).

Based on the progress measurement and trend analysis, it also allows to forecast a reliable future total cost and date of completion of the project.
This analysis permits the project manager to have the capacity to predict if the project will finally end up with a success or a failure and if so to undertake corrective actions in time to improve the performances and bring the project back to the optimal conditions.

1.3.1 EVM Origins

EVM originated in the late 1960s, when the US Department of Defense (DoD) needed to improve its management control system due to increasingly complex weapon systems development contracts. In 1967 the Department of Defense issued Instruction 7000.2, "Performance Measurement for Selected Acquisitions", based on the first EVM approach implemented by Air Force: Cost/Schedule Planning Control System (C/SPCS) specification. These 35 criteria derived from the best management practices used by the American industry and outline the minimal requirement that a contractor's control system must meet. These criteria were then finalized into the American National Standards Institute/Electronic Industries Alliance Standard 748, Earned Value Management Systems (ANSI/EIA-748) (Abba, 2001).

During the 1980s, emerged the real necessity to have a reliable management tool dealing with the Navy's A-12 "Avenger" aircraft development program. It was a stealth bomber that was supposed to replace the A-6 "Intruder" with a scheduled first flight for the 1990, but it ended up being a failure. This was due to the inability to meet all the tree project constraint resulting in a huge cost overrun and a time delay.
Although there were numerous motivations behind this failure, the most critical one, that tipped the Secretary of Defense's decision toward a cancellation of the project, was the fact that no one was able to carry out a reliable estimation of cost and time at completion. In fact, there were more values for the time and cost estimate at completion, some more reliable than others. Unfortunately, the estimations considered the less credible and most optimistic ones, even if the earned value analysis showed that there were some problems long before the contractors, and program managers acknowledged them; this episode strongly exposed the issue about the value of EVM for early warning.
Moreover, it was also figure out that EVM information was not being used correctly for management purposes throughout the DoD, but on the contrary, it was viewed as a financial reporting requirement. Soon EVM was redefined in defense regulations as a project management tool and the industry issued

its own standard, the ANSI/EIA 748-98 "Earned Value Management Systems," which was adopted by the Department of Defense in 1999.
The same year, the Project Management Institute dedicated a convention in Houston, sponsored by the College of Performance Management, to the EVM, recognizing it as a crucial tool for the project management. Finally, PMI included the methodology in its standards.

Consequently, the method got across in many industries including research, energy, oil and gas, construction, manufacturing, software, nuclear, production, becoming one of the best practices and the most reliable among the main techniques for evaluating cost, time and scope performance in the project management.

1.3.2 Fundamentals of EVM

Determining a Performance Measurement Baseline (PMB) and monitoring the project's performance in relation to the PMB constitute the two primary EVM practices.
Over the years, several formulas have been studied from EVM's measurement, forecast and analysis of the project cost and time performance using monetary information.
Regardless of the kind of project they are based on four key values, which are BCWS, BAC, ACWP and BCWP, described in detail as follows.

- Planned Value (PV) also known as Budgeted Cost of Work Scheduled (BCWS)

  It is the planned cost assigned to accomplish the scheduled work in a specific time period. The total of the PV is sometimes referred to as the Performance Measurement Baseline (PMB) and represents the standard against which the performances are compared, highlighting the deviations. It is calculated with the following formula:

$$BCWS = Budgeted\ Cost * Work\ Scheduled$$

- Budget At Completion (BAC)

  It is the sum of all the budgets established for the work to be performed in the whole project, so it is the total planned value for the project.

- Actual cost (AC) also known as Actual Cost of Work Performed (ACWP)

  It is the cost actually incurred to accomplish the work performed within the control time. It is calculated with the following formula:

$$ACWP = Actual\ Cost * Work\ Performed$$

  The direct comparison between ACWP and BCWS, in monetary value, is formally incorrect since, from it, it can be derived only the difference between what was originally planned and what is actually spent, without taking into account what is the work performed at the specific time being. Therefore, estimating costs and time at completion only through the Actual Cost Work Performed would be wrong and would lead to incorrect forecasts due to the absence of work done information in the ACWP.

  For this reason, a data representing the budget of the work performed is crucial in the Earned Value Analysis, that is the Earned Value.

- Earned Value (EV) also known as Budgeted Cost of Work Performed (BCWP)

  It is the budget value of the work performed within the control time. It is calculated with the following formula:

$$BCWP = Budget\ Cost * Work\ Scheduled$$

  It is essential that the AC must be recorded in the same time period and for the same WBS component as EV, in order for EV analysis to be reliable.

- Budget Cost of Work Remaining (BCWR)

  Another value, which is not usually used for the calculation of the other performance factors is the Budget Cost Work Remaining (BCWR), that is, the budget cost of the work remaining at instant t. This is a variable that is not traditionally used for the calculation of the Estimate at Completion, but it appears in some studies conducted for the forecast of the EAC. Thus, it is added in this list for completeness.

All the other data items can be derived from these four data points, such as the variances of actual against planned performance and their related indices, and the forecasted cost and time at completion of a project.

1.3.3 S-Curves

In earned value analysis, the three parameters planned value, earned value, and actual cost can be graphically represented with S-Curves in a chart, to better understand their behavior and to enable a quick understanding of the project status. They are monitored and reported on both period-by-period basis (typically weekly or monthly) and cumulative basis, which represent x-axis and y-axis respectively.

The reason why S-Curves are so named is because the three parameters of the project form a “S” shape. In the early stages, the processes start gradually, and the project’s growth is slow since the team members are either conducting market research or just beginning the initial phase of execution. As the project moves forward, and progress begins to become more rapid, it starts a period of strong growth, reaching the peak at the so called “point of inflation”, giving rise to the middle part of the s-shape. This phase is very important; team members are working heavily on the project since many of the major expenses are incurred and many of the major resource are used.

After the point of inflection, the growth starts to decline, forming the upper part of the “S” known as the upper asymptote and the mature part of the project begins. In this phase, most of the progress has been made and therefore only secondary tasks, such as reviews and final approvals are missing.

Going into the specifics of each performance, the planned value is represented remaining constant, as it is the budget at completion while the earned value and actual costs are updated as the project progresses.

The Earned Value Analysis compares EV to the AC to recognize if there are any cost variances and to the PV to evaluate any schedule alterations.

The three variables would overlap throughout the project life cycle if there were not misalignment and so if they are in compliance with the schedule and the cost planned. However, such a situation is very rare to happen.

Comparing the position of the three S-shaped Planned Value, Actual Cost and Earned Value graphs, six different possible arrangements can be displayed as in the chart presented in Figure 3.

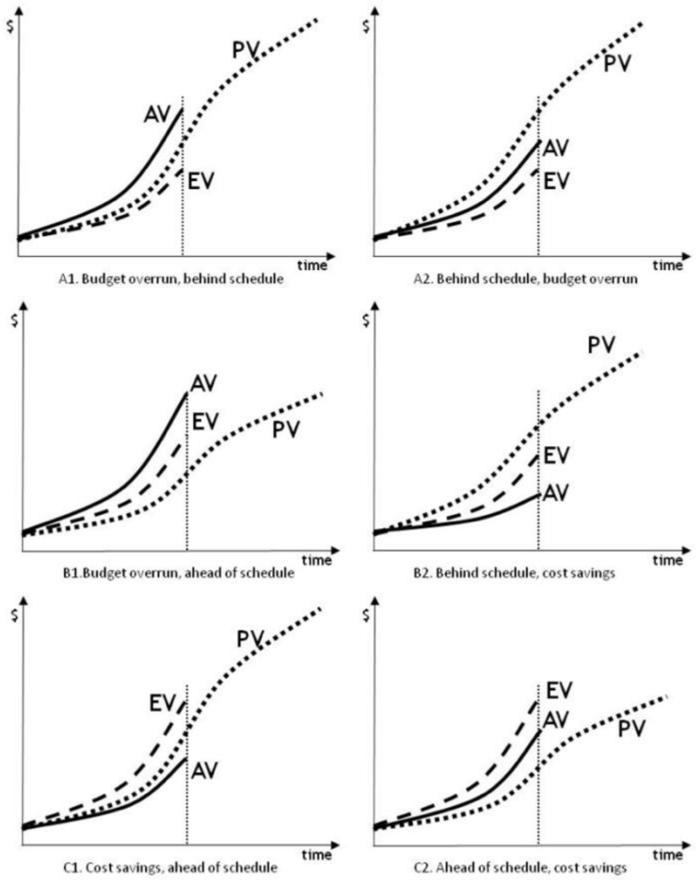


Figure 3: Possible arrangements of S-curves

Looking at these graphs, general comments can be deducted. It can be said that whether AV is above EV means that the project is over budget, meaning that the cost actually spent on the work performed are higher that the budget one, on the contrary if AV is below EV means that the project is cost saving. Comparing the PV with the EV, the same analysis can be done looking instead at the time: if PV is above AV, the project is behind the schedule, meaning that there is a delay in the project scheduling, on the contrary if PV is below AV the project is ahead of schedule.

The worst-case scenario is the one when the project is both in a condition of budget overrun and behind the schedule. This situation is represented in the A1 graph with more significant problems on cost and in A2 graph with more significant problems on schedule delay.

Similarly, the best case-scenario is the one when the project is both in a condition of cost savings and ahead of schedule. This situation is represented in the C1 graph with more importance on cost better than scheduled and in C2 graph with importance on the duration better than cost.

A better analysis can be conducted by looking at the variances, examined in the following subparagraph.

1.3.4 Variance analysis

PMI's PMBOK® Guide defines a variance as *"a quantifiable deviation, departure, or divergence away from a known baseline or expected value"*[5]. Variance analyses can be performed to identify variances between the project baseline and the actual project performance, revising continuously the status of the project throughout its life. There are three basic expressions of variance: resource variance, cost variance and schedule variance, which give precise monetary values of positive or negative status.

The related indices, unlike variances, are used as indicators of actual performances without giving a monetary value and are Resource Flow Index, Cost Performance Index and Schedule Performance Index.

Overall, both the variances and indices are indicators of past behavior and can be used to forecast cost and time at completion and to understand whether there is the necessity to undertake corrective actions.

Variances

- Resource Flow Variance (RV)

  It measures how much is presumed to spend in a time period respect to what is actually spent, regardless the quantity of work done. It is calculated by the difference between the Budget Cost of Work Scheduled (PV) and the Actual Cost of Work Performed (AC).

$$RV = PV - AC$$

  A negative value of Resource Flow Variance means that the project is overrun, a positive value means that the project is underrun, and a zero value means the project is on target.

- Cost Variance (CV)

  It is a measure of the project cost performance since it is the amount of the cost incurred in excess or in deficit respect to the budget cost. It shows whether a project is in a condition of

[5] Project Management Institute. (2017). *A guide to the Project Management Body of Knowledge: (Pmbok Guide)* (6th edition).

over or under budget. Cost Variance is measured as the difference between the Budgeted Cost of Work Performed (EV) and the Actual Cost of Work Performed (AC). Dividing the cost variance by the earned value, the CV can be expressed in percentage terms.

$$CV = EV - AC$$

$$CV\% = CV \ / \ EV$$

At the end of the project, the cost variance is expressed through the difference between the budget at completion (BAC) and the actual amount spent.

A negative value of Cost Variance means that the project is overrun, so over the budget, having a loss due to the costs higher than the budget, which is often a situation difficult to recover. On the contrary, a positive value means that the project is underrun, leading to a gain due to cost savings. A zero value means the project is on budget.

- Schedule Variance (SV)

It is a measure of the project schedule performance. Sometimes, the SV is taken as time-based indicator in defining whether the project is early or late against its baseline schedule. This is an incorrect interpretation of the SV since it measures not the time but the amount by which a project is ahead of or behind schedule in accomplishing work at a given point in time, and so it is a physical status indicator on how much of the project work has been completed. It is measured by subtracting the Budgeted Cost of Work Scheduled (PV) from the Budgeted Cost of Work Performed (EV). Dividing the schedule variance by the earned value, the SV can be expressed in percentage terms.

$$SV = EV - PV$$

$$SV\% = SV/EV$$

A negative value of Schedule Variance means that the project is behind schedule meaning a loss of time since the work that has been accomplished is less than the planned one. On the contrary, a positive value indicates that the work that has been accomplished is more than the planned one, leading to a gain of time and so the project is ahead the schedule. A zero value means the project is on schedule.

Graphical representation

Also, resource, cost and schedule variances can be graphically represented in chart, where the ACWP, BCWS and BCWP S-curves are drawn. In the figure 4 the worst-case scenario is represented: both the CV and SV are negative, highlighting the criticality of the performances.

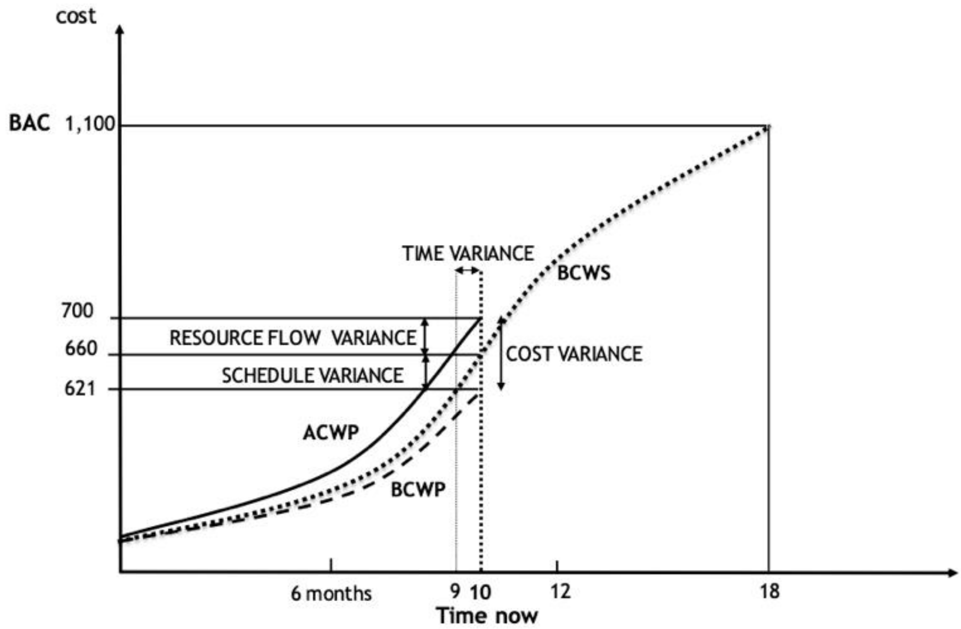


Figure 4: Representation of variances on a S-curve chart

Indexes

- Resource Flow Index (RI)

It is evaluated as the ratio between the Budget Cost of Work Scheduled (PV) and the Actual Cost of Work Performed (AC).

$$RI = PV \; / \; AC$$

Analogously to the analysis of the RV, a RI with a value less than one means that project is overrun; while a RI more than one means that project is underrun. A RI exactly equal to one means that the project is on target.

However, it’s important to point out that these metrics do not necessary indicate a negative or positive scenario since for instance a project can proceed more quickly and with a lower cost or more slowly and with higher costs than expected.

- Cost Performance Index (CI)

It is an indicator that evaluates the performance of the project in terms of cost since it measures the cost efficiency of the project team in using the resources. It is considered the most critical metric in the earned value analysis, and it is calculated dividing the Budgeted Cost of Work Performed (EV) and the Actual Cost of Work Performed (AC).

$$CI = EV/AC$$

Analogously to the analysis of the CV, a CI with a value less than one means that project is overrun, indicating that the money spent to accomplish the work is higher than planned; while a CI more than one indicates a budget underrun of performance to date. A CI exactly equal to one means that the project is on budget.

- Schedule performance index (SI)

It is an indicator that evaluates the performance of the project in terms scheduling since it measures the efficiency of the project team in accomplishing the project work. It is calculated dividing the Budgeted Cost of Work Performed (EV) and the Budgeted Cost of Work Scheduled (PV).

$$SI = EV/PV$$

Analogously to the analysis of the SV, a SI with a value less than one means that project is behind schedule; while a SI more than one indicates that the project is ahead the schedule. A SI exactly equal to one means that the project is on schedule.
Despite this consideration, to determine correctly if the project is ahead or behind the schedule, SI is not sufficient and the performance on the critical path also needs to be considered.

Starting from CI and SI it is also possible to identify an additional performance indicator, the Critical Ratio.

- Critical ratio (CR), also known as Schedule-Cost Index (SCI)

  It is an indicator of the overall project performance, and it is calculated by the product between the Cost Performance Index and the Schedule Performance Index.

$$CR = SI * CI$$

  Its value is of fundamental importance because, based on this, different corrective actions can be taken to align the project with the project baseline.
  In fact, a value of the CR equal to one means that the overall project is on target. However, this not necessarily means that both CI and SI are on target, but it can also happen that one of them denote a poor performance, while the other one a good performance. In this case, it's important for the project manager to have a specific understanding on both the indexes and find a trade-off to reach the required goals.
  A value of the CR greater than one indicates that there is a good overall project performance. This might happen if the CI and SI are both over target, or if one of them denote a poor performance, while the other one an excellent performance.
  On the contrary, a value of the CR less than one indicates that there is a good overall project performance. This might happen if the CI and SI are both worse than the target, or if one of them denote a good performance, while the other one a really poor performance.
  After having done an accurate analysis, control actions can be undertaken, based on the so-called Control Limits of the critical ratio, displayed in the following Figure 5.

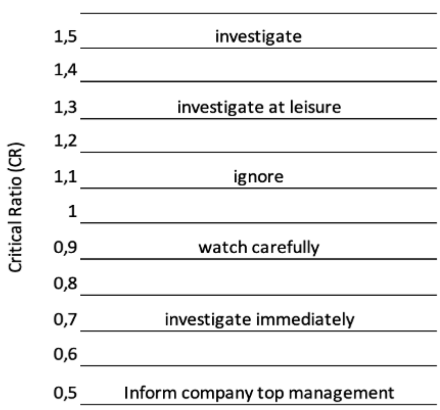


Figure 5: Critical ratio and control limits

The considerations made so far on the cost performance index and the scheduled performance index can be summarized in a chart having on the axis the values of the CI and SI, in Figure 6.
This allows to have a general picture on the progress of the project considering the values of the CI and the SI evaluating them whether are more or less than 1. The values are not taken singularly, concentrating the attention just to the cost or time performance, but they are analyzed in conjunction, leading to 4 combinations: project on budget but not on schedule, project on schedule but not on budget, project both not on budget and on schedule and project both on budget and on schedule. It can be easily deducted that the bast-case scenario is this last case in the upper-right part of the graph, with a value of both the CI and SI higher than one. This situation of a project whose time and cost are lower than expected can be given by, on one side, some positive factors such a very good organization of the project, the presence of clever team members, favorable external factors, but on the other side by an overestimation of the project performances and less importance given to the quality of the work done, the third element of the Project management Triangle of Constraints.

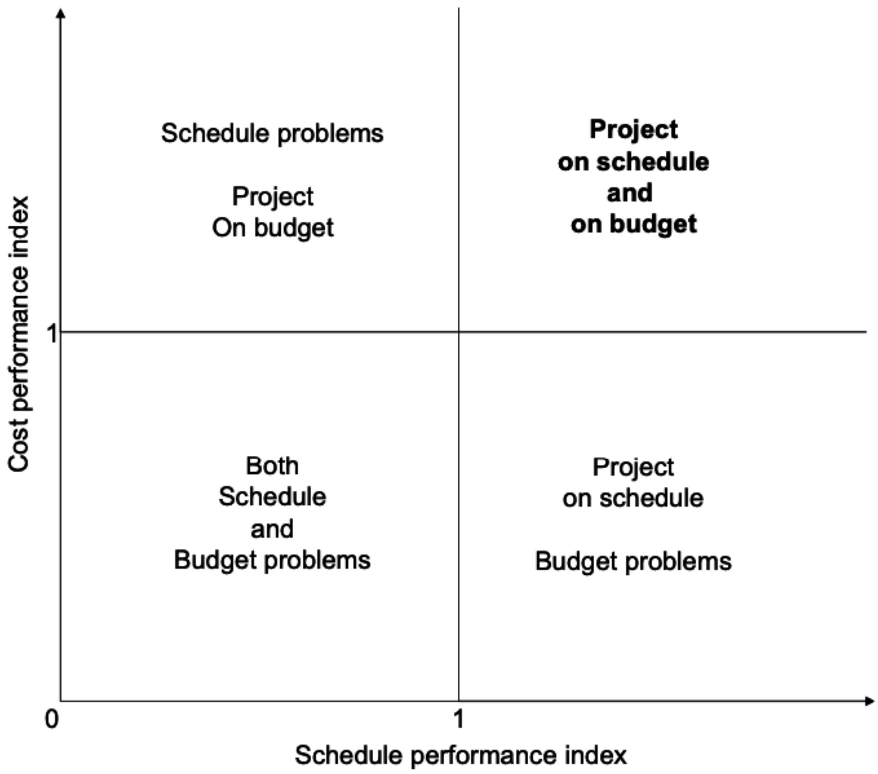


Figure 6: Areas of cost and time performance using Cost and Schedule Performance Indexes

This chart is therefore a tool used to support the decision-making process in terms of possible corrective actions, identifying which is the area and the task that has the most critical condition and so has the priority of intervention.

1.3.4 Earned Schedule

The Earned Value method presents some limitations regarding the schedule performance since it is measured in currency units and not in time units. Another problem is that, when the project is about to reach the completion the Schedule Variance (SV) tends to zero, and the Schedule Performance Index (SI) tends to unity regardless the progress of the project. This can cause problems since even if the project is behind the schedule, the schedule trend indicates that instead the project is in time. For this reason, the schedule indicators are not considered as reliable as cost indicators as far as the project progress goes to its end, which in most projects is after 2/3 of the project is completed. To overcome this problem, the Earned Schedule (ES) methodology was ideated by Lipke, in 2003, and today it represents an emerging practice in the EVM.

The Earned Schedule is analogous to Earned Value, but it measures the schedule performances using time units instead of currency units. The ES is obtained by comparing the cumulative
EV (BCWP) earned to the performance baseline PV (BCWS), projecting to actual time AT the EV S-curve onto the PV S-curve. The cumulative value of ES is calculated as follow:

$$ES_t = c + \frac{EV - PV_c}{PV_{c+1} - PV_c}$$

Where c is the time units for which EV exceed PV.
Given the value of ES it's possible to compute appropriate values of Schedule Variance and Schedule Performance Index expressed in time units, which behave analogously to the cost indicators.

- Schedule Variance (time) (SV_t)

  It is a measure of the project schedule performance in time units, calculated using ES.

$$SV_t = ES_t - AT$$

  Where AT is the Actual time, which replace the Earned Value Management concept of PV.
  A positive value of SV_t is given when the ES exceeds AT and it indicates a good condition, and a negative value means that the project lags its expected performance, indicating a critical condition. Differently from SV expressed in cost terms (indicated with $SV_\$$), the SV_t is equal to zero at project completion only if the project is really on schedule.

- Schedule Performance Index (time) (SI_t)

  It is a measure the schedule efficiency in time units of a project, calculated using ES.

$$SI_t = ES/AT$$

  SI_t is greater than one when the ES exceeds AT, and it indicates a good condition while SI_t is less than one when the ES is less then AT and it indicates a negative status of the project. Differently from SI expressed in cost terms (indicated with $SI_\$$), the SI_t is equal to one at project completion only if the project is really on schedule.

The use of the Earned Schedule concept is helpful and significant in having schedule performance indicators (SV_t and SI_t) which behave correctly throughout the entire period of project.

1.3.5 Cost and time forecasting

As the project proceeds, cost and schedule performance can be forecasted using the trends of past performance behaviors. Thus, cost and schedule forecasts are predictions of the status of the project in the future, based on available performances at the time of the forecast and are used to determine if the project is in time and on budget, identifying whether corrective actions are needed and design the project control procedure in advance.
Using the variables and performance indicators presented so far in the earned value analysis, it is possible to calculate cost and time at project completion.

Cost Estimate at Completion (CEAC)

Cost Estimate at completion (CEAC) also known as Estimate at completion (EAC) is the expected value of total cost of the project on its completion. It is usually based on the actual costs incurred for the completed work (AC), plus the expected cost to incur for the remaining project work, called Estimate to Complete (ETC).
There are two approaches for developing an EAC value: original estimate approach and revise estimate approach. Moreover, a pessimistic approach is also analyzed.

- Original estimate approach

The first original estimate approach is based on the assumption that future cost will be in line with the budget, regardless the performance of the project until the time of the analysis. It is calculated in this way:

$$EAC = ACWP + (BAC - BCWP)$$

Rearranging the equation, the EAC can thus be found subtracting the CV from the BAC:

$$EAC = BAC - CV$$

The criticality of this approach is the assumption that the trend applied until the end of the project is the one given by the BAC, making this approach optimistic since it assumes that cost overruns are problems of past history and will not incur in the future.

- Revise estimate approach

A better way to calculate EAC is through the revise estimate approach. It is based on the assumption that if there are no changes the future of the project will, at least, reflect the past performance. It can be calculated adding to the previous formula a correction with the CI as follows:

$$EAC = ACWP + \frac{BAC - BCWP}{CI}$$

Rearranging the equation, the EAC can thus be found dividing the BAC by the CI:

$$EAC = \frac{BAC}{CI}$$

A value of CI less than one (project is overrun) lead the EAC to be greater than the BAC; on the contrary a value of CI greater than one (project is underrun) lead the EAC to be less than the BAC.

- Pessimistic approach

This approach is important to use when, as well as the cost performance, also the schedule performance has a strong impact on the status of the project. Therefore, in this case, a combination of both Cost Performance Index and Schedule Performance Index has to be used in the calculation of the estimate at completion.
A way to combine both CI and SI is to make a variation in the revise estimate approach, including the SI_t in the ratio to calculate the EAC:

$$EAC = \frac{BAC}{CI * SI_t} = \frac{BAC}{CR}$$

In the formulae above the critical ratio (CR) represents the maximum final cost ceiling.
The grouping of the two indexes considers both the effect of a time delay or advance and a gain or loss of costs. In fact, if both CI and SI_t are less or greater than one their effect on the EAC is amplified, while if one is less and the other is greater than one their effects compensate each other.
Another way to combine CI and SI_t is through a weighted sum, giving different values to weight of the CI and SI_t according to the project manager's decision and such that the sum of the weights add to one. It is called Composite Index and the values usually given to the weighs are 0,8 for the CI and 0,2 for the SI, meaning that the CI and SI weights respectively the 20 percent and 80 percent in the overall EAC. The formula used for the calculation is the following:

$$EAC = \frac{BAC}{w_1 * CI + w_2 * SI}$$

Time Estimate at Completion (TEAC)

Time Estimate at completion (TEAC) also known as Actual Completion date (AC) is the expected total duration of completing a project. It is usually based on the actual time spent for the completed work, plus the expected time to spend for the remaining project work.
Like the calculation of the EAC, there are two approaches for developing an AC value: original estimate approach and revise estimate approach.

- Original estimate approach

The first original estimate approach is based on the assumption time overruns are problems of past history and will not incur in the future, leading this to be an optimistic approach. It is calculated in this way:

$$AC = AT + (BAC - BCWP) * \frac{(BC - AT)}{(BAC - BCWS)}$$

Where AT (time now) is the current date when the analysis is made, and BC is the schedule completion date.

- Revise estimate approach

The revise estimate approach is based on the assumption that if there are no changes the future of the project will, at least, reflect the past performance. It can be calculated adding to the previous formula a correction with the SI as follows:

$$AC = AT + (BAC - BCWP) * \frac{(BC - AT)}{(BAC - BCWS) * SI}$$

Rearranging the equation, the AC can thus be found dividing the BC by the SI:

$$AC = \frac{BC}{SI}$$

In this approach, the SI is also taken into account, and a value of SI less than one means that the project is ahead of schedule and, as a result, time for the completion of the project will be longer.

CHAPTER 2.

LITERATURE REVIEW

A literature review is *“an objective, thorough summary and critical analysis of the relevant available research and non-research literature on the topic being studied”*.[6]

It provides an overview of the state of knowledge on a precise topic, allowing to be up to date with the current themes, identify significant theories, methodologies, limitations, gaps, and challenge future research in that area.

Conducting a literature review involves 5 main steps, in order to define the principal topic, search and collect the publications relevant for it, organize the publications and finally examine and describe those studies.

In this chapter is briefly described the methodology used to conduct a literature review of project Cost Estimate at Completion forecasting methods.

2.1 STEP 1. PRELIMINARY STUDIES

The identification of the best method for the calculation of the Estimate at Completion, became one of the major challenges since the Navy’s A-12 project of the Department of Defense failed. Until now, lot of scholars attempted to improve existing models and to develop new methodologies for the EAC forecasting. The objective of this literature review is to answer the following primary research questions:

RQ1. What are the traditional methodologies used for the Estimate at Completion forecasting?
RQ2. Are there any limitations and gaps in the traditional methodologies used?
RQ3. Are there any new methodologies with the aim of overcoming these limitations?
RQ4. Are there any methodologies with the aim of improving existing ones?

To answer these questions different research has been done and the most relevant papers has been selected and analyzed in this thesis.

[6] Cronin, P., Ryan, F. & Coughlan, M. (2008). *Undertaking a literature review: a step-by-step approach.*

2.2 STEP 2. SELECTION PROCEDURE OF PUBBLICATIONS

In the field of the Estimate at Completion, lot of papers has been written by different authors in the literature. In this thesis, were taken into account books, book chapters, thesis, reviews and articles written in English.

To collect the most relevant academic articles the following research tool were used:

- Google Scholar
- Scopus
- Semantic Scholar
- ScienceDirect
- ResearchGate
- Spinger Link
- Academia.edu

The most important keywords considered in the research process are: *Earned Value Management*, *Cost Estimate at Completion*, *Forecasting Methods*, their derivative, and their abbreviations.

To better determine whether the papers are relevant to the research, the abstract and/or the introduction is read. Moreover, some Inclusion and Exclusion criteria were used to restrict the number of the papers.

Publications are included only if the Inclusion Criteria (IC) matches:

IC1. Publications that propose methodologies for the Estimate at Completion forecasting are selected.

IC2. Publications that propose comparative studies in the Estimate at Completion forecasting methods are selected.

IC3. Publications that propose new methodologies for the Estimate at Completion forecasting are selected.

IC4. Publications that propose extension of the existing Estimate at Completion forecasting methods are selected.

IC5. Publications that propose tools and techniques for the improvement in the Estimate at Completion are selected.

IC6. Publications that highlight gaps in the Estimate at Completion forecasting methods and propose methodologies to fill them are selected.

Publications are excluded if the Exclusion Criteria (EC) matches:

EC1. Publications where the most important keywords considered in the research process do not appear in the title and/or abstract and/or text of the publication are not selected.
EC2. Publications where the acronym EVM does not mean "Earned Value Management" are not selected.
EC3. Publications where the acronym EAC does not mean "Estimate at Completion" are not selected.
EC4. Publications that describe improvements in EVM not in the field of EAC forecasting are not selected.
EC5. Literature review papers about the EAC forecasting methods are not selected but are used as the basis for this study, to cross-check the results, and are cited in this thesis.
EC6. Books intended to provide descriptions and explanations of the basics about EVM, included EAC analysis, are not selected, but are used in this thesis for the basic EAC description.

The literature review considered for this study are: "*A review of estimate at completion research*" by Christensen, Antolini, and McKinney, for the traditional EAC forecasting methodology; "*Estimation at Completion in Civil Engineering Projects: Review of Regression and Soft Computing Models*" by Araba, Alhawat, Sackey, Ali, Memon and Milad, for the new EAC forecasting methodology. References of those articles were also checked and some papers from there, which meet the Inclusion/Exclusion criteria were considered.
The considered thesis and papers, written by authors of US Department of Defense (DoD), were downloaded from the Defense Technical Information Center. Some of the original documentations of these works were lost or not publicly available, thus their analysis was done by looking at their description in other authors studies.
The books intended to provide the basic EAC description are: *"A guide to the Project Management Body of Knowledge: (Pmbok Guide)"* and *"Practice standard for earned value management"* by Project Management Institute and "*Project management for Facility Constructions"* by De Marco.

2.3 STEP 3. ORGANIZATION OF INCLUDED PAPERS

To organize the papers which matched the inclusion criteria, a table containing the main paper's information was created within an Excel file. The papers were classified by title, authors, year of publication and the methodologies used for the EAC forecasting.
The summarizing table of the papers considered is represented in Table 1.

Table 1: Summarizing table of the papers considered in the existing literature

TITLE	AUTHORS	YEAR OF PUBLICATION	METHODOLOGY
A Product Improved Method for Developing A Program Management Office Estimate Cost at Completion	Holeman, J.B.	1975	Index-based
Cost Performance Analysis Program for Use on Hand-Held Programmable Calculators	Lollar, J.L.	1980	Index-based
C/SCSC and C/SSR Cost Performance Analysis Programs	Parker, C.W.	1980	Index-based
A Comparative Analysis of Two Cost Performance Forecasting Models	Land, T.J., & Preston, E.L.	1980	Index-based
Weapon System Cost Control: Forecasting Contract Completion Costs	Bright, H.R., & Howard, T.W.	1981	Index-based
A Study to Determine Indicators and Methods to Compute Estimate at Completion (EAC)	Covach, J., Haydon, J.J., Reither, R.O.	1981	Index-based
Methods of Estimating Contract Cost at Completion	Haydon, J.J., & Reither, R.O.	1982	Index-based
Validation of ASD/ACCM's Cost Performance Analysis Algorithm	Blythe, A.L.	1982	Index-based
An Evaluation of CPRA Estimate at Completion Techniques Based Upon AFWAL Cost/Schedule Control System Criteria Data	Price, J.B.	1985	Index-based
Evaluation of Weighted Indices on Algorithms Utilized for Calculating Independent Estimates at Completion	Cryer, J.M. & Balthazor L.R.	1986	Index-based
HQ Air Force Systems Command Estimate at Completion Formula Justification	Wallender, T.J.	1986	Index-based
A Logical Approach to Estimate at Completion Formulas	Totaro, J.A.	1987	Index-based
Estimates at Completion (EAC): A Guide to Their Calculation and Application for Aircraft, Avionics, and Engine Programs.	Reidel, M.A., & Chance J.L.	1989	Index-based
A Cost Performance Forecasting Concept and Model.	Karsch, O.A.	1974	Regression-based
A Production Study Sequel to the Cost Performance Forecasting Concept and Model.	Karsch, O.A.	1976	Regression-based
Forecasting Techniques Employed in a Line Organization.	Olsen, D., & Ellsworth, R.W.	1976	Regression-based
Space and Missile Systems Organization Cost Performance Forecasting Study.	Heydinger, G.N.	1977	Regression-based
A Cost Performance Forecasting Model.	Busse, D.E.	1977	Regression-based
A General Technique for R&D Cost Forecasting.	Weida, W.J.	1977	Regression-based
An Application of Rayleigh Curve Theory To Contract Cost Estimation and Control.	Watkins, H.	1982	Regression-based
Forecast of Schedule/Cost Status utilizing Cost Performance Reports of the Cost/Schedule Control Systems Criteria: A Bayesian Approach.	El-Sabban, M.Z.	1973	Probability-based
An evaluation of a Bayesian approach to compute Estimate at Completion for Weapon System Programs.	Hayes, R.A.	1977	Probability-based
Probabilistic Model Development for Project Performance Forecasting: Semantic scholar.	Naeini, M. E., & Heravi, G.	2011	Probability-based
Combination of growth model and earned schedule to forecast project cost at completion.	De Marco, A., & Narbaev, T.	2013	Index-Regression based
An earned schedule-based regression model to improve cost estimate at completion.	De Marco, A., & Narbaev, T.	2014	Index-Regression based
Nonlinear cost estimates at completion adjusted with risk contingency.	De Marco, A., Narabaev, T. & Rosso, M.	2016	Contingency-Adjusted
A New Estimation at Completion of Project's Time and Cost Approach Based on Particle Filter.	Hajialinajar, M. T., Mosavi, M. R. & Shahanaghi, K.	2016	Particle Filter
Application of Artificial Neural Network to Forecast Actual Cost of a Project to Improve Earned Value Management System.	Iranmanesh, S.H., & Zarezadeh, M.	2008	Artificial Neural Network
Web-based conceptual cost estimates for construction projects using evolutionary fuzzy neural inference model.	Cheng, M. Y., Tsai, H. C., & Hsieh, W. S.	2009	Evolutionary Fuzzy Neural Inference Model (EFNIM)
Conceptual cost estimates using evolutionary fuzzy hybrid neural network for projects in construction industry.	Cheng, M. Y., Tsai, H. C., & Sudjono, E.	2010	Evolutionary Fuzzy Hybrid Neural Network (EFHNN)
Evolutionary fuzzy decision model for construction management using support vector machine.	Cheng, M. Y. & Roy, A. F. V.	2010	Evolutionary Fuzzy Support Vector Machine Inference Model (EFSIM),
A novel time-depended evolutionary fuzzy SVM inference model for estimating construction project at completion.	Cheng, M. Y., Hoang, N. D., Roy, A. F. V. & Wu, Y.W.	2012	Time-depended evolutionary fuzzy support vector machine inference model (EAC-EFSIMT)
A fuzzy neural network to estimate at completion costs of construction projects.	Feylizadeh, M. R., Hendalianpour, A. & Bagherpour, M.	2012	Fuzzy Neural Network (FNN)
Implementation of Genetic Algorithm Integrated with the Deep Neural Network for Estimating at Completion Simulation.	Kamoona, K. R. K. & Budayan, C.	2019	Deep Neural Network (DNN)
Estimation at Completion Simulation Using the Potential of Soft Computing Models: Case Study of Construction Engineering Projects.	AlHares, E. F. T. & Budayan, C.	2019	Extreme Learning Machine (ELM)
Multiple Linear Regression Model for Improved Project Cost Forecasting.	Ottaviani, F. M. & De Marco, A.	2021	Multiple Linear Regression Model

2.4 STEP 4. OUTLINE OF THE STRUCTURE

There are several approaches in which the findings from the literature can be reported:

- Thematic
- Chronological
- Methodological
- Theoretical

For this research, the methodological approach has been chosen, the authors were listed based on their proposed methodology or comparative study for the EAC forecasting. The identified methodologies are divides as follows: index-based methods, regression-based methods, probability-based methods, and new methods.

2.5 STEP 5. ANALYSIS AND DISCUSSION TO EXAMINE THE STUDIES

The final step consists of writing the literature review. In the following chapter will be described and analyzed the methodologies used for the EAC forecasting, during the years. The review is divided into four sections. First, the index-based methods and authors proposing related studies are listed. Second, the regression-based methods and authors proposing related studies are listed. Third, the probability-based methods and authors proposing related studies are listed. Fourth, new methods to fill the previous methodologies' gaps are described and authors proposing related studies are listed. Final considerations about the above-mentioned methodologies are finally written.

CHAPTER 3.

ESTIMATE AT COMPLETION

In this chapter, it is depicted a list of the most important studies for the Estimate at Completion forecasting, conducted by authors in the literature, among years, until today. The methodologies for the EAC forecasting are identified, described and analyzed.

The first one is the is the index-based methods, which include formulas based on the Past Performance Factors. Authors proposed different methods and conducted comparative studies among existing models, starting from the late seventies in the field of military projects. Then regression-based methods are described, in which the EAC is estimated with the use of linear and/or nonlinear regression equations. The index-based method and the regression-based method are the two major categories in the EAC forecasting and the list of most of the authors considered come from the literature review made by Christensen, Antolini, and McKinney in their paper "*A review of estimate at completion research*". However, during the years, other studies has been conducted with the aim of overcoming the limitations found in the traditional models and thus of finding a better forecasting model. Some of these other methodologies are described in this section. Few probability-based methods are described. A combined index-regression methodology based on the Gompertz growth model is considered with the aim of overcoming the absence of the time factor in the regression-based models and of taking into account the risk factor. A method based on the use of the Particle Filter, as a solution for the Bayesian filter execution, is described. In the last years, Soft Computing and Artificial Intelligence approaches, such as Neural Networks (NN), Fuzzy Logic (FL), Genetic Algorithms (GA) and other derivative ones have been applied in the EAC forecasting methods with the aim of improving the accuracy of results. The list of most of the authors, considered come from the literature review made by Araba, Alhawat, Sackey, Ali, Memon and Milad in their paper "*Estimation at Completion in Civil Engineering Projects: Review of Regression and Soft Computing Models*", are described. Finally, the last and the newest method considered, developed in 2022, is a multiple linear regression model with the aim of improving the project cost estimate approaches.

3.1 INDEX-BASED METHODS

The first method to calculate the Estimate at Completion is the index-based one. It considers the so-called Past Performance Factors (PPF) which are: Budget Cost of Work Scheduled (BCWS), Actual Cost of Work Performed (ACWP) and Budget Cost of Work Performed (BCWP). As showed in

Chapter 1, starting from these factors other performance indicators can be derived: Cost Performance Index (CI), Schedule performance index (SI), Cost-Schedule Index (CSI) also known as Critical Ratio (CR) and Composite Index. These represents the four major indices used in this method, which formulas are reassumed as follow.

$Cost\ Performance\ Index\ (CI) = BCWP/ACWP$
$Schedule\ performance\ index\ (SI) = BCWP/BCWS$
$Cost - Schedule\ Index\ (CSI) = SI * CI$
$Composite\ index = w_1 * CI + w_2 * SI$

To make these values as accurate as possible it is necessary to carefully monitor the progress of the project during its course, accurately collecting data on a monthly basis, usually three, six and 12 months are used as time periods for the calculation of the EAC. As a result, indices can be based on monthly data, or cumulative and average indices can be derived. The labeling conventions adopted are the subscript "m" for the representation based on the most recent month, "c" for a cumulative representation and "x" for a representation averaged over x number of months, beginning with the most recent month, and going backwards.

The generic index-based formula for the calculation of the EAC is:

$$EAC = ACWP + \frac{BAC - BCWP}{Index} = \frac{BAC}{Index}$$

It is similar to the Equation seen in Chapter 1, with the only difference in the denominator, the $Index$, which is one of the several performance indices listed above in order to determine the EAC.

The following lists the main authors in the literature who described a method or conducted comparative studies on the use of the index-based formulas for the calculation of the EAC.

3.1.1 Holeman J.B., 1975

One of the earliest studies regarding the calculation of EAC is given by J.B. Holeman, in 1975, proposing three different methods.

The first method is the one considering the Cost performance Index as the index, thus calculating the EAC as the ratio between the BAC and CI.

The second method uses a subjective past performance factor as index, which include a linear combination of variables subjectively determined by judgments of project managers. These variables are contract changes, inflation, schedule variations, overhead changes and unexpected technical problems.
The formula for the calculation of the EAC takes in consideration these factors and assumes that technical problems are reflected in the CI.

$$EAC = ACWP + \left[\frac{(BAC - BCWP)}{PPF}\right] + Contract\ Changes + Schedule\ Variation$$

Where:
$PPF = CPI + Inflation + Overhead$

The third method involves a series of steps to determine a possible range for the Performance Factor. To do so a three-point method can be used, having a high, most likely and worst value of CI for each task. A large number of estimations are produced by performing a Monte Carlo simulation, these are grouped into ranges with relative and cumulative frequencies. Following this, the data found are plotted on a graph and are analyzed to obtain a range of EACs or an average EAC.

It is possible to say that the use of subjective variables makes Holeman's model mostly flexible, and as a yardstick against contractors' EAC since the estimates usually turned out to be optimistic. Thus, basing the calculation of the EAC, not only on historical data but also on the experience of experts, represents on one hand a strength of this model, but on the other hand a limitation. In fact, this model has unfortunately not been validated with data from real projects in any study, but it is nevertheless particularly interesting because it is one of the first to introduce variables different from the past performance factors into the calculation of the Estimate at Completion.

3.1.2 Jakowski, 1977

Jakowski proposed a method which alternate the CI and the SCI as index depending on the progress of the project. In the first stage of the project the CI is used, until it decreases significantly in the most recent months. At this point, it is used an “optimally weighted” composite index, where the weights

are calculated based on the standard deviation of the two indicators and the optimal one is which results in the lowest historical standard deviation in the composite index. In the last phases of the project, specifically after the 60 percent completion point, it is used again the CI as index.
The original documentation of Jakowski's model has been lost and traces of it are present only in other works, such as the one of Covach, which will be described afterwards.

3.1.3 Lollar J.L., 1980

James Lollar, in his method, uses the Composite Index as index, thus the formula for the EAC is the following one.

$$EAC = \frac{BAC}{w_1 * CI + w_2 * SI}$$

However, differently from Holeman, proposed a less flexible model for the calculation of the Estimation at Completion. In fact, the assignment of values to the CI and SI weights is not left to the analyst's judgment, but it is derived from the contribution that the Cost Variance and Schedule Variance values make to their total. Thus, in Lollar's model the coefficient of the Cost Performance Index is calculated as follows:

$$w_1 = \frac{CV}{CV + SV}$$

While the coefficient of the Schedule Performance Index is calculated as follows:

$$w_2 = \frac{SV}{CV + SV}$$

From a mathematical point of view, the use of the variances is done in absolute values, excluding the possibility for the variances to have negative values, which can happen in the realty, and so it can not appear to be a viable way to determine the two weights. The logic behind this procedure is not described in Lollar's studies, who do not give any justifications or explanation of how he developed this procedure. However, this method was then validated by Blythe in his comparative studies, even if this formula did not well against the others in terms of accuracy.

3.1.4 Parker C.W., 1980

Like the Lollar's method, also the one proposed by Parker uses the Composite Index as index, thus the formula for the EAC is the following one.

$$EAC = \frac{BAC}{w_1 * CI + w_2 * SI}$$

To determine the appropriate composite index, the weights are varied from 0 to 1, increasing their values from time to time by 0.1. The various combinations of w_1 and w_2 give rise to a range of composite indexes from which the project manager must choose the best combination among those examined, according to the characteristics of the project. This is the basic difference from Lollar's method since the choice of the weights is subjective to the analyst judgment.

3.1.5 Land T.J. and Preston E.L., 1980

Land and Preston conducted a comparative study evaluating index-based linear models and the regression model of Karsch, a nonlinear model reviewed in the regression section in the next paragraph. The linear formulas come from the Automated Financial Analysis Program at ESD, which was a computer program used to conduct the CPR and C/SSR analysis and it allowed the analysts to choose three different proposed EAC formulas leading to a range of values for an EAC.

The general equation present in the Automated Financial Analysis Program is the following:

$$EAC = ACWP + ETC$$

In the formulae above, the ETC is a performance factor that can be calculated as follows:

1) $ETC = (1 - CVI) * BCWR$
2) $ETC = (1 - CVI_c) * BCWR$
3) $ETC = (1 - CVI_3) * BCWR$
4) $ETC = \left[1 - (BCWP_3 - \frac{ACWP_3}{BCWP_3}\right] * BCWR$
5) $ETC = (1 - CVW * CVI + SVW * SVI) * BCWR$

Where CVI and SVI are two new parameters calculated as the ratio between the related variances and the Budget Cost of Work Performed:

$$CVI = CV/BCWP;\ SVI = SV/BCWP$$

and CVW and SVW are respectively the Cost Variance Weight and the Schedule Variance Weight.

The documentation about the data origins was very poor and the database was limited to Air Force programs which represents a weakness of this study. The research used a sample of 30 contracts for the evaluation of the linear formulas, of which 20 are aircraft programs (a mix of R&D and Production contracts), and five more contracts were needed for the calculation of a parameter required for the Karsch model.

Overall, the results showed that the linear formulas were not less accurate the nonlinear one, at different percent completion point, with CI_c the most accurate of the index-based formulas.

3.1.6 Bright H.R. and Howard T.W., 1981

Bright and Howard conducted a comparative study evaluating at various contract stages some index-based linear models and two regression-based models (one linear, one nonlinear), with ACWP regressed against CI, allowing the monthly calculation of EAC to be compared. The linear formulas come from the Automated Contractor Performance Measurement System (ACPMS) which is another computer program used by the Army Missile Command (MICOM).

An "improved method" method was developed, which has the composite index as the index in the general formula:

$$EAC = ACWP + \frac{BAC - BCWP}{SI * CI}$$

This equation was then added to the other ACPMS formulas for comparison.

The general equation present in the Automated Contractor Performance Measurement System is the following:

$$EAC = ACWP + (BAC - BCWP) * P$$

In the formulae above, factor P is determined by the analyst and can be calculated as follows:

1) $P = \frac{1}{CI_3}$
2) $P = \frac{1}{CI_6}$

3) $P = \frac{1}{CI_{12}}$

4) $P = \frac{1}{CI}$

In case of taking in consideration the weights the P factor can be calculated as follows:

$$P = \frac{1}{\frac{w_1 * CI + w_2 * SI}{w_1 + w_2}}$$

Where: $w_1 + w_2 = 1$

and different weight combinations where considered:

1) $w_1 = 0{,}5;\ w_2 = 0{,}5$
2) $w_1 = 0{,}25;\ w_2 = 0{,}75$
3) $w_1 = 0;\ w_2 = 1$

   In this third case, having weighted as 0 the CI is the same as calculating the P considering only the Schedule Index: $P = \frac{1}{SI}$

The documentation about the data was very poor, not including the calculations to support the combination of the CI and SI. The research used a sample of 11 Army R&D contracts and the estimates were compared at different percent completion points, normalizing for other factors such as rebaselining, further scope, and not completed contracts.

As results of the study, Bright and Howard demonstrated that the success of a specific formula for the calculation of the EAC respect to the others depends on the percent completion point of the project. Overall, in the early stages of a project, namely in the first 30%, the regression model did better against the linear one. Instead, if comparing only the linear formulas, the ones considering the composite index were more accurate. Getting into the specifics, the formulas with the SI provided better results in the early stages of a project since the information content, thus the importance, of the Schedule Performance Index decreases as the project progresses. In the middle stages, after the 30% completion, the average CI performed better and if the Cost Variance has significant growth, it is recommended to use indices averaged over short periods rather than over long ones. Finally, the composite index performed in an "excellent" way in the time fraction ranging from 31% to 80% of the status of the project and in a "good" way after the 80% completion of the project.

From these results, it is highlighted that the choice between the different formulas for the calculation of the EAC depends not only on the type of the project but also on its progress status when they are applied.

3.1.7 Covach, 1981

Covach conducted a comparative study evaluating 24 formulas and models of which 12 were index-based linear formulas and 12 regression-based formulas, using data from 17 contracts (14 development, 3 production) managed by the Navy. The two models were not compared with each other but were analyzed individually, comparing the formulas in different time frame.

The 12 index-based formulas consist of linear equations having as index CI_m, CI_3, CI_6, CI_{12}, CI_c and SI_c, and some of them had divided an averaged index into BAC. Using a database of six Navy programs, these formulas were evaluated on monthly basis with three methods: 10% method, BAC method and LRE method that are briefly described below:

- The 10% method consists of comparing the EAC (calculated using the formulas) with the final AWCP. If the calculated EAC is within the range $\pm$ 10% of the ACWP more than 75% of the time, the formula used is successful. Otherwise, if it is within the range between the 75% and 50% the formula used is classified as indifferent and if it is less than 50% the formula used is unsuccessful.
- The BAC method consists of comparing the EAC (calculated using the formulas) with the final AWCP and BAC looking if the EAC is between the BAC and the ACWP for determining if the formula used is successful or not.
- The LRE method consists of comparing both the EAC (calculated using the formulas) and the contractor's EAC with the ACWP. If the EAC calculated was better than the contractor's EAC the formula used is considered a success.

Moreover, an evaluation of the formulas was done by phases, dividing the life of the project into quintals (20% increment) and three phases were defined: the early phase (two quintals), the middle phase (three quintals), and the late phase (two quintals). In doing so, it was established a numerical scoring system for the attribution of a score of 1 for successes, 0 for indifference, and -1 for failures.

So finally, the formulas were evaluated by scoring them by both the three methods and the three phases mentioned above.

Also in this study, the result was that there is not a general formula for the calculation of the Estimate at Completion but, depending on the time frame in which they are used, some of them give most accurate results than the others.

The 12 regression-based models used $ACWP_c$, $BCWP_c$, or CI_c as the dependent variable, and $BCWP_c$ or Time (months) as the independent variable, resulting in three regression approaches against four different curves (linear, power, exponential, log) that are briefly described below.

- The first method consists of regressing the ACWP against BCWP which is equal to the BAC with a linear curve.
- The second method consists of regressing the CI against Time, matching the final CI to the BAC with a power function equation.
- The second method consists of regressing the ACWP and BCWP against Time, with an exponential equation and then using this found final Time in the ACWP regression equation to calculate the EAC.

The SAMSO nonlinear model (which will be discussed later in the regression section) was also considered for evaluation, but it was rejected because too unstable.

As results, the comparison of the regression models was less clear since one model not always performed better than others. Moreover, the three approaches and curves were best relative to the regression equations only.

A particular aspect from the Covach studies is the existence of some interviews conducted to the different leaders of the projects, which revealed the following points:

- the trends, positive or negative, of the project performances rarely reverse themselves over time;
- managers prefer to use simple and quick forecast methods for the estimation of time and cost;
- manpower loading data were under-utilized by analysts;
- problems can start to appear in the scheduling phase.

Finally for the evaluation of the formulas and models were considered also the different levels of the Work Breakdown Structure and it was found out that EACs computed with data from the level one in the WBS were as accurate as EACs computed at lower levels and summed to level one.

3.1.8 Haydon J.J. and Reither R.O., 1982

An extension study of the previous comparison analysis conducted by Covach, was done by Haydon and Riether, extending the database to 21 programs, 15 of which were complete and 6 were at the final stages. It was evaluated a different method for the calculation of the EAC called the "Range Method", which consists of developing a point estimate from a range of EACs. So, the first step entails finding this range of EACs, using index-based formulas computed previously by Covach. Second, the range of results obtained has been extended by 2.5 percent and the median of this expanded range is taken for the estimation of the EAC.
From this application, as a result, has emerged that if the contractor's EAC was less than the one found with the method of Haydon and Riether, this one was more accurate in 79 percent of cases.

3.1.9 Blythe A.L., 1982-1984

In 1982 Blythe conducted a comparative study evaluating the Lollar and Parker model, as well as the contractor's EAC at six percent completion points, using data from 26 contracts (7 Research and Development, 19 production) managed by the Air Force, without distinctions between development and production ones. The selection process for determine the most accurate model was not clear but it seems to consider the smallest standard deviation (nearest to BAC) and the estimates range. The results of the study were basically two. First, Lollar method was compared to the contractor's EAC, and it appeared to be the less accurate against other methods, as previously said in his study. Second, Parker model was then compared using different combinations of weights in the Composite Index and the most correct one was a value of 0.2 for the Schedule Performance Index and a value of 0.8 for the Cost Performance Index at any stage of completion.

In view of these studies, Blythe exposed that the contractor EAC had a small standard deviation and was underestimating. Starting from these considerations, in 1984, he extended his research, and he developed an adjustment factor based on the contractor's EAC and obtained through regression analysis. The formula proposed is called Adjusted Contractor EAC (AEAC) and it is shown below:

$$AEAC = \frac{EAC}{0.9108 + 0.0892 * \frac{BCWP}{BAC}}$$

The results obtained showed that with this new approach, the Adjusted Contractor EAC was more accurate than any other index-based EAC.

Blythe then proposed a formula for the calculation of the standard error (SE) of the contractor's EAC shown below:

$$SE = EAC * (0.1289 - 0{,}0925 * \frac{BCWP}{EAC})$$

Then a confidence interval for the Adjusted Contractor EAC was obtained:

$$AEAC \pm z(SE)$$

Where z is the value from a normal distribution table.
A limitation of this model is the fact that the adjusted factor was determined based only on the 26 programs mentioned before and the model was not validated among other projects. This problem was not considered by Blythe in his studies who affirmed that this factor could be used universally, but this is risky since the sample had contracts of few different types.

3.1.10 Price J.B., 1985

Price conducted a comparative study evaluating five index-based formulas and one linear regression model using data from 57 on-going Research and Development contracts managed by the Air Force. The linear formulas come from the Cost Performance Report Analysis (CPRA), another computer program used financial analysts in the Air Force Wright Aeronautical Laboratories (AFWAL) at Wright-Patterson AFB and only five out of six formulas were considered.
The first three formulas consist of general index-based formula seen for the calculation of the Estimate at Completion, where, as index, are used the CI, CI_c and CI_3:

1) $EAC_1 = ACWP + \frac{BAC-BCWP}{CI}$
2) $EAC_2 = ACWP + \frac{BAC-BCWP}{CI_c}$
3) $EAC_3 = ACWP + \frac{BAC-BCWP}{CI_3}$

The other two has unusual composite indices and are the following:

4) $EAC = ACWP + ETC$

In the formulae above ETC is calculated by the summation of the CV to the 75% of the SV as follows:

$ETC = [100 - (CV\%) + 0.75 * (SV\%)] * BCWR/100$

5) $EAC = EACC + ETCS$

In the formulae above the EACC is calculated with a weighted sum of combination of three EAC (before found), as follows:

$EACC = (0.12 * EAC_1) + (0.64 * EAC_2) + (0.24 * EAC_3)$

And in the calculation of the ETCS a new performance indicator is used, namely the ACWP Rate, and it is calculated as follows:

$ETCS = (months\ behind\ schedule) * ACWP\ Rate * 0.75$

Where $ACWP\ Rate = ACWP/Total\ contract\ completed\ months$

To find the most accurate predicting formula it was made a regression analysis, where each formula was evaluated by regressing the ACWP against the calculated EAC, the one with the highest coefficient of determination R^2 was considered the most accurate formula.
Results of this study pointed out that the best formulas were the second one with CI_c and the fourth one with the composite index followed by the third one with CI_3 and the regression-based model.
Limitation of this study are in the data coming from a database of uncompleted projects, in fact even if they had different percent completion, they were treated as if they were finished at that percent.

3.1.11 Cryer J.M. and Balthazor L.R., 1986

Cyrer e Balthazor conducted a comparative study replicating the Blythe's one. They used the same models and methods but stratified the database that is dividing the projects among the two categories Research and Development and Production, instead of analyzing them all together since they hypothesized that for different type of contracts different models were more precise than others.
As Blythe, Cyrer e Balthazor came to the conclusion that, overall, the contractor's EAC performed better than the index-based formulas, although when calculating the Adjusted Contractor EAC different outcome were found out for different type of projects, thus demonstrating that the formula cannot be generalized.
Regarding the best combination of weights for the Composite Index, differently from Blythe who said that the most accurate one was a value of 0.2 for the Schedule Performance Index and a value of

0.8 for the Cost Performance Index at any stage of completion, Cyrer e Balthazor findings were different depending on the type of the project.
The evaluation criteria for the determination of the best weights were not stated, thus represents a non-clear understanding of the model since depending on it different results ate obtained. Looking at the lowest coefficient of variation and the best weights given to SI and CI were respectively 0.1 and 0.9 for the combined and production contracts but not for research and development ones. Second were respectively 0.2 and 0.8, with the best coefficient of variation in research and development projects, followed then by respectively 0.3 and 0.7 values for the weights.
While, looking at lowest standard deviation the best combination is 0.4 and 0.6.
Moreover, another problem of this study was that the Lollar model was incorrectly stated, with errors in the formula, which could lead to improper comparison results.

3.1.12 Wallender T.J., 1986

Wallender wrote a paper on the evaluation of index-based formulas using data from 44 contract managed by the Ballistic Missile Office (BMO) which was considered by Hq AFSC as having the best EAC analysis techniques. Moreover, the Wallender paper was the only documentation that could be found for the AFSC studies. The objective of Wallender was to justify the so called "Hq AFSC EAC Formula" showed below:

$$EAC = \frac{BAC}{0.2 * SI + 0.8 * CI}$$

Wallender confirmed what was already stated by Blythe in 1982, that the Composite Index with a value of 0.2 for the Schedule Performance Index and a value of 0.8 for the Cost Performance Index give the most accurate results.
So, the Blythes' study, the AFSC study led by Wallender, and an Armament Division (AD) study done by Rutledge and Dinato give the evidence that the combination 0.2 – 0.8 for the composite index can be considered valid and accurate for many projects, even if not universally the best for all types of projects.

3.1.13 Totaro J.A., 1987

Like the Lollar's and Parker's method, also the one proposed by Totato uses the Composite Index as index. The difference from the previous models consists of considering the weights not in a definitive way and established since the beginning of a project, but as a function of percent complete, so letting them to change as the project progresses. The initial choice of the weights for the SI and CI is subjective to the analyst judgment depending upon the project characteristics. The formula used by Totaro is the following:

$$EAC = ACWP + \frac{BAC - BCWP}{[0.25 - 0.25(\%complete)] * SI + [0.75 + 0.25(\%complete)] * CI}$$

Where
$\%complete = BCWP/BAC$

The Totaro approach had a clearly stated purpose, which was to reduce the importance of the Schedule Performance Index and to increase that of the Cost Performance Index as the project progresses. This is done in the formula reducing the weight given by the analyst by a factor based on percent complete for the Schedule performance Index and on the contrary increasing the weight given by the analyst by a factor based on percent complete for the Cost Performance Index.
The description of this model was clear and well stated; however, a limitation of this model is that the sample size of data considered was not sufficient for its validation and no other studies have supported this model.

3.1.14 Reidel M.A. and Chance J.L., 1989

Reidel e Chance proceeded Prince's research and conducted a comparative study evaluating six index-based formulas, using data from 56 contracts (16 Research and Development, 40 production) managed by the Air Force. The formulas are the general ones of index-based method already seen before, and they use as indexes: CI, CI_c, CI_3, SCI, Composite index with a value of 0.2 for the Schedule Performance Index and a value of 0.8 for the Cost Performance Index similarly to other studies, while the last formula uses a Composite index which depends on the percent complete and represented as follows:

$$EAC_1 = ACWP + \frac{BAC - BCWP}{x * CI + (1 - x) * SI}$$

Where x is the percent complete.

These six formulas were evaluated at four different percent complete points (25%, 50%, 75%, 100%) and the final ACWP was compared against the found EAC. Moreover, it was also considered the typology of the project and the typology of weapon system. The evaluation criteria for the determination of the most accurate formula was the lowest Average Absolute Percent Deviation and Average Rank Order as a check.

As results, the EACs found for production projects were more accurate than the EACs found for development ones, more specifically, formulas with the SCI or CI were more appropriate for Research and Development projects, while formulas with the Composite Index having weights as 0.2 and 0.8 were more appropriate for production projects.

So, depending on the type of project, the sector and other variables, the most reliable methods for calculating the Estimate at Completion could be different.

3.2 REGRESSION-BASED METHODS

The index-based methods for the calculation of the Estimate at Completion, seen so far, rely on some assumptions that lead to some restriction in the accuracy of the estimates. In fact, these formulas assumes that the cumulative project work is not represented with a S-Curve line, even if it represents better how the project progresses in time, as seen in Chapter 1, but it is approximated to a straight line. The linear model relies on past data and makes the assumption that the performance measured last would remain constant until the end of the project, ignoring any other possible changes.
To overcome these limitations in the linear model, growth models which reflect the shape of S-Curve are used for the fitting process, through the regression analysis.
The regression-based models, that will be described below, are complicate and are based on a deep regression knowledge and skills, thus it is necessary to introduce the basic elements of a regression analysis to better understand the reasoning behind these methods.

Regression analysis attempt to model the statistical relationship between two variables, an independent variable denoted with Y and a dependent variable denoted with X. The final aim is to estimate the casual effect on Y of a unit change in X, namely the slope parameter of the population regression line. A regression equation can have one or more independent variables. It is also present the so-called regression error, denoted with Ɛ, which includes unobserved variables (factors different from X which can affect Y), and the errors made while measuring Y.
The relationship between the dependent and independent variable can be linear or nonlinear. Below, the different forms of the regression equation are defined.

- Linear regression function with one regressor

  The regression equation presents a linear relationship between the dependent variable Y and only one independent variable X. The formula is the following:

  $$Y = \beta_0 + \beta_1 X + \mathcal{E}$$

  Where:

  β_0 is the intercept of the regression line, that is the expected dependent variable Y when the regressor X is equal to zero.

  β_1 is the slope of the regression line, that is the expected change in the dependent variable Y for a unit change in the independent variable X.

- Linear regression function with multiple regressor

  The regression equation presents a linear relationship between the dependent variable Y and multiple independent variables X_i. An example formula, with only two dependent variables for simplicity reason, is defined below:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \mathcal{E}$$

  Where:

  β_0 is the intercept of the regression line, that is the expected dependent variable Y when the regressors X_1 and X_2 are equal to zero.

  β_1 is the expected change in the dependent variable Y for a unit change in the independent variable X_1, holding constant the independent variable X_2.

  β_2 is the expected change in the dependent variable Y for a unit change in the independent variable X_2, holding constant the independent variable X_1.

- Nonlinear regression function of a single independent variable with polynomials in X

  The population regression function is approximated by a quadratic, cubic or higher-degree polynomial. The form is similar to the linear multiple regression model, except that the regressors are powers of the same independent variable. An example formula of a quadratic regression function is defined below:

$$Y = \beta_0 + \beta_1 X + \beta_2 X^2 + \mathcal{E}$$

  Differently from the linear model, the coefficients have difficult interpretations.

- Nonlinear regression function of a single independent variable with Logarithmic transformations

  In the population regression function the dependent variable Y and/or the independent variable X is transformed by taking its logarithm. This transformation permits modeling relations in percentage terms, which can give a better interpretation in many applications rather than linearity. Three basic model exists, where the interpretation of β_1 is different case to case:

1) Linear-log regression function: the independent variable X is transformed into its logarithm. An example formula is defined below:

$$Y = \beta_0 + \beta_1 \ln(X) + \mathcal{E}$$

A 1% increase in X (0,1 increase in $\ln(X)$) lead to a $0,01\beta_1$ change in Y.

2) Log-linear regression function: the dependent variable Y is transformed into its logarithm. An example formula is defined below:

$$\ln(Y) = \beta_0 + \beta_1 X + \mathcal{E}$$

A unit increase in X lead to a $100\beta_1\%$ change in Y.

3) Log-log regression function: both the dependent variable Y and the independent variable X are transformed into their logarithm. An example formula is defined below:

$$\ln(Y) = \beta_0 + \beta_1 \ln(X) + \mathcal{E}$$

A 1% change in X lead to a $\beta_1\%$ change in Y. Thus, in this particular case, β_1 has the interpretation of an elasticity.

When evaluating a regression-based model, an important factor is the measurement of the fit, meaning verify if the estimated regression equation represents the data in a good way. For this purpose, a statistical tool called Coefficient of Determination, indicated with R^2, can be used. It is unitless and it can have values between 0, which means that there is no fit at all, and 1, which means a perfect fit. Thus, higher is the R^2 better is the fit. The Coefficient of Determination will be used in most of the studies to determine the most accurate formula for the calculation of the EAC.

Differently from the index-based models which require at most a simple calculator, the regression-based models require the use of computer programs to find important data for the analysis of the formulas, such as the R^2 seen before. However, the greater complexity of regression analysis compared with the use of the indexes for the EAC calculation does not imply a greater effectiveness. In fact, sometimes regression-based models are not completely validated because of, for example, the limitation of the sample size of data considered.

Below are described the main regression-based models for the EAC calculation proposed in the literature over the years.

3.2.1 Karsch O.A., 1974 – 1976

In 1974, Karsch conducted a comparative study evaluating one index-based formulas and two nonlinear regression model using data from a development contract managed by the Air Force. The index-based formula is called "Cumulative Current Percent Cost Variance" and it takes into consideration the EAC formula with the CI_c as index. The nonlinear models developed by Karsch has been included in the computer analysis programs for Cost/Schedule reports and it was based on the nonlinear long-term growth of the ACWP and the BCWP. Karsch stated that if the samples of completed programs are similar in their characteristics, thus, the following formula represents a *"reasonable behavioral relationship"*[7]:

$$Y = b_1 * X^{b_2}$$

Where b_1 and b_2 are non-dimensional parameters which represent the growth characteristics of a sample, reflecting the similarities stated above.
Karsch identified two different variants of this formula:

1) Unconstraint exponential: the parameter b_2 vary and it is calculated each month.
   This model solves for the two parameters "*without the benefit or intelligence of other information dealing with growth characteristics"*[8]. In fact, they are derived from very few samples and not from the entire population, resulting as inadequate estimates.
2) Constraint exponential: the parameter b_2 is held constant for the entire project life.
   This model overcome the limitations of few samples taken in the unconstraint one "*by superimposing an intelligence source that is available from completed programs"*.[9]

As results, looking at three time periods of the project life, it was demonstrated that in the first 93% of the project completion the constrained exponential model gave the most accurate EAC against the unconstrained exponential model and the Cumulative Current Percent Cost Variance model. From 93% to 98% completion none of the methods did better against the other, giving almost equal results. Beyond 98% completion the Cumulative Current Percent Cost Variance model did a little better than the other methods.
One limitation in this study is that the data are based only on one project, thus it cannot be generalized to different type of projects and there is no validation of this model.

[7] Karsch, O.A. (1974). *A Cost Performance Forecasting Concept and Model*.
[8] Karsch, O.A. (1974). *A Cost Performance Forecasting Concept and Model*.
[9] Karsch, O.A. (1974). *A Cost Performance Forecasting Concept and Model*.

In 1976, Karsch conducted a subsequential study in which he evaluated the methods considered in the studies of the 1974 using not just one contract as before but 13 production aircraft and missile contracts managed by the Air Force. The aim of this study was to determine a range for the value of the parameter b_2 in the Constrained Exponential model.

Again, results were distinguished in three time periods of the project life and emerged differently for the aircraft and missile contracts. Moreover, the range of the values for b_2 was evaluated to be between 0.94 and 1.13, with the most common values between 1.0 and 1.1 without differing meaningfully between aircraft and missile contracts.

For the aircraft contract a value of $b_2 = 1.053$ was considered and it was demonstrated that in the first 88% of the project completion the Constrained Exponential model gave the most accurate EAC against the unconstrained exponential model and the Cumulative Current Percent Cost Variance model. From 88% to 98% completion both Constrained and Unconstraint Exponential model performed better. Beyond 98% completion the Cumulative Current Percent Cost Variance model did better than the other methods.

For the missile contract a value of $b_2 = 1.041$ was considered and it was demonstrated that in the first 75% of the project completion the Constrained Exponential model was the most accurate against the other models. Beyond 75% completion the Cumulative Current Percent Cost Variance model did better than the other methods.

Some limitations in this study are that the evaluation criteria were not explained, and results were not supported with justifications or literature review.

Finally, in both studies of 1974 and 1976, Karsch recommended to continue the studies for the range estimation of the parameter b_2, in order to establish a generalization of this model for different projects.

3.2.2 Olsen D. and Ellsworth R.W., 1976

Olsen and Ellsworth described the time series forecasting technique, a method used by the B-1 SPO (System Program Office) for the calculation of the Estimate at Completion. The time series analysis is a collection of data defined sequentially over time intervals and it uses the regression analysis as a forecasting method. The forecasts of the EAC were made by "GETSA", a computer program developed by General Electric and leased by the B-1 SPO. The formula for this method is the following:

$$F_t = \sigma S_{t-1} + (1 - \sigma) F_{t-1}$$

Where σ is a smoothing constant, F_t is the forecast for time t, F_{t-1} is the forecast for time t-1 and S_{t-1} is the cost for period t-1.

In addition to this formula, Olsen and Ellsworth briefly described other models used by the B-1 SPO, such as regression analysis, trend extension analysis and modified exponential smoothing.

Finally, these studies proved that there is not just a specific formula for the calculation of the EAC but acknowledged how important it is the use of multiple methods with the related documentation and judgment of experts.

3.2.3 Heydinger G.N., 1977

Heydinger conducted a comparative study evaluating 7 formulas and models of which 4 were index-based linear formulas and 3 nonlinear regression-based formulas using data from a development contract managed by the Air Force. The 4 index-based formulas are the one seen previously which use as index the CI, CI_c and CI_3. The 3 regression-based models include the Constraint exponential model of Karsch and two models proposed by The Space and Missile Systems Organization (SAMSO), the first one assumed linearity while the second one assumed a modified Erlang equation.

The first model proposed by the SAMSO consists in regressing the BCWP and ACWP of the current month and the previous 5 months against time by Least Squares regression. The value of the Estimate at Completion is the value of the ACWP read in the point in which BCWP equals the BAC. This method was compared with the Karsch Constraint exponential model and the formulas coming from the Automated Financial Anlysis Program at ESD, also analyzed later by Land and Preston in 1980. As results, the Karsch model did the better in the early (3^{rd} to 7^{th}) and late (27^{th} to 42^{nd}) months, while the ESD model did better in the middle (8^{th} to 26^{th}) stage of the project.

The second model of SAMSO is for the first time proposed by Heydinger, defined as the "New SAMSO" model. The formula is based on the Erlang equation and is the following:

$$Y = a * X^b * e^{cX}$$

Where:

$Y = ACWP \ or \ BCWP \ or \ BCWS$

$X = number \ of \ cumulative \ months$

The regression procedure is the same as for the first model, with the only difference that ACWP, BCWP and BCWS are regressed with this formula and that also BCWS can be regressed, thus also schedule can be estimated at completion.
Overall, this "New SAMSO" model results to be the most accurate for the EAC calculation throughout the project's life.
A limitation of this model is that the sample size of data considered was limited. For this reason, the model cannot be generalized for other type of contracts and so Heydinger recommended to conduct further studies to demonstrate the real effectiveness of this model.

3.2.4 Busse D.E., 1977

During the same year of Heydinger studies, also Busse analyzed the Karsch nonlinear regression-based model and proposed a different way to develop its coefficients. Busse based its model on the theory for which there exist a relationship between ACWP and BCWP which is indicated by the so-called sensitivity factor (e). The formula for this method is the following:

$$EAC = Z * BAC^e$$

Where e is the sensitivity factor as explained before and it can be calculated as follows:

$$e = \frac{\frac{ACWP_c}{ACWP}}{\frac{BCWP_c}{BCWP}}$$

and Z can be calculated as follows:

$$Z = \frac{ACWP}{BCWP^e}$$

As results, after the comparison between Busse and Karsch models at several project completion stages, the calculated EACs with the Karsch model were the most accurate.
A limitation of this study is that there was no comparison with other formulas and the model was not validated. In fact, Busse advised not to use this model alone.

3.2.5 Weida W.J., 1977

Weida found out that the Research and Development projects followed the form of S-shaped curve both in terms of cumulative expenditures and milestone completed in time. This because the R&D

process reflects the three phases of a growth process expressed with the S-curve: 1) a slow initial development, 2) a period of expansion and 3) maturity. Weida started his studies finding a way to fit the shape of a S-curve in such a way that could be used for an accurate Cost/Schedule forecasting. To obtain this result, Weida divided the S-curve into two parts at the inflection point, being the highest point for the lower curve and the lowest point for the upper curve. These two curves can be expressed with two possible regression curves described below.

a) quadratic equation: $Y = \beta_0 + \beta_1 X + \beta_2 X^2$

b) logarithmic equation: $Y = aX^b$

One limitation of the use of these formulas is the statistical problems of heteroscedasticity and autocorrelation of the R&D data which must be eliminated. After adjusting these data problems, Weida verified that these equations fit the curves by taking the actual costs incurred throughout the life of each R&D project for 22 development programs.

Then, for the determination of the curve which has the best fit, the econometric fitting techniques are utilized and the quadratic form for both upper and lower curve resulted to be the best equation.

The EAC can be calculated as follows:

$$EAC = BAC * Y$$

Where Y is calculated through a mathematic procedure which consist in the following passages. First, the lower curve equation is taken and the Y is determined, then it is substituted in the second equation to calculate the y-intercept. This y-intercept is then used in the upper curve equation and schedule variance is determined. Finally, for the calculation of the value of the Y multiplied by the BAC, the schedule variance replaces X, and the y-intercept replaces β_0 in the upper curve equation.
The most important aspect of this method is that the forecast obtained is considered as a range of cost values, thus leading to three different estimates for the EAC, namely the best, worst and most likely value, with the relative confidence interval.
The results were finally checked with the data coming from other 15 weapon systems.

3.2.6 Chacko G. K., 1981

Chacko proposed a time series forecasting technique called "adaptive forecasting" in which the model adapts to the data and not vice versa. This model needs at least five months of data to produce accurate estimates and it varies whenever new data become available, therefore, it predicts better in the short run. Calculations are done through MESGRO, a computer program and it is not available any formula. Original documentation of this method is lost, and this study is only quoted and briefly described in other authors studies.

3.2.7 Watkins H., 1982

Watkins proposed a model based on the application of the Rayleigh-Norden curve to contract cost data, using the MINITAB computing system. The choice of adapt the Rayleigh-Norden curve come from the Watkins knowledge about studies that demonstrated *"regular patterns of manpower buildup and phase-out in complex projects"*[10]. Other techniques used to support the Rayleigh-Norden curve are the scatter plots and linear regressions. The scatter plot was used for representing the variables in time on a graph, while the regression analysis was needed to evaluate the empirical relationship between the historical data about the costs of the project and the Rayleigh-Norden model.

Through mathematic procedure, the Rayleigh-Norden equation was modified into a linear form and a natural logarithmic transformation was applied, leading to the following equation:

$$\ln\left(\frac{\dot{Y}}{t}\right) = \ln\left(\frac{K}{t_d^2}\right) + \left(-\frac{1}{2t_d^2}\right) * t^2$$

Where $\dot{Y}$ is the ACWP during each report period, K is the total cumulative cost by the end of the project, t is the time passed from start of contract and t_d is time of peak ACWP.

This model was applied to a helicopter engine contract and to a cruise missile contract and the fit of the data to the model was evaluated through the R^2 and the following results were obtained. In the helicopter engine contract, the values of R^2 are slightly high, with a maximum of 86.8%, in the 10th period, which indicates a relatively good fit. In the cruise missile contract instead the lowest value of

[10] Watkins, H. (1982). *An Application of Rayleigh Curve Theory To Contract Cost Estimation and Control.*

R^2 found is 3.3% in the 6th period, indicating a very poor fit and very few are the periods in which R^2 has a higher value with a maximum of 87.2% in the 13th period.
One limitation of this study is that the data are not adjusted for the autocorrelation and the model is not completely validated.

Differently from other regression-based models, Watkins considered the importance of the inflation which is one of the causes of the increasing costs according to project duration. The uncertain value of the inflation is a variable that can distort the cost and time estimation at completion and this model resulted in being potentially able to measure the impact of inflationary uncertainty.

3.3 PROBABILITY-BASED METHODS

In this section are contained some forecasting models found in the literature which are based on probability approaches. The first two come from the categorization made by Christensen, Antolini, and McKinney in their paper "*A review of estimate at completion research*", where also the other methods where listed, as previously discussed. The last model, in addition to being a probability-based method, is interesting because it uses the Monte-Carlo simulation in the process for the estimation of the cost at completion, and for this reason considered in this elaborate. As for the index-based and the regression-based methods, these are just few methods described for explicative reasons.

3.3.1 El-Sabban M.Z., 1973

El-Sabban proposed a forecasting method for the calculation of the EAC based on the probabilities and the Bayes Theorem, which calculates a posterior probability from an assumed prior probability.[11] In fact, this method determines first initial probabilities, namely prior probability, which are then updated with the data reported by the Cost Performance Report (CPR) as the project progresses, namely the posterior probability.

For the development of this method, El-Sabban made two main assumptions:

1) the cost and schedule follow a normal distribution
2) the prior distribution at point 0 has a mean $c\mu_a$ and a standard deviation $c\sigma_a$

The proposed model includes the calculation of the EAC with its mean value and its variance as follows:

$$E(\mu) = \frac{c\mu_a\sigma_0^2 + \mu_0 c^2\sigma_a^2}{\sigma_0^2 + c^2\sigma_a^2}$$

$$V(\mu) = \frac{c^2\sigma_0^2\sigma_a^2}{\sigma_0^2 + c^2\sigma_a^2}$$

Where:

[11] El-Sabban, M.Z. (1973). *Forecast of Schedule/Cost Status utilizing Cost Performance Reports of the Cost/Schedule Control Systems Criteria: A Bayesian Approach.*

$\mu_a = ACWP$, reported by the CPR at point "a" (prior to point 0) of the project

$\mu_0 = BAC$, at point "0" of the project

$\sigma_0 = 0.05\mu_0$ and $\sigma_a = 0.1\mu_a$

$$c = \frac{\mu_0}{BCWP}$$

According to the author, the use of the Bayes Theorem has the following advantages:

1) easiness of the formulas;
2) formulas are equally valid at each stage of the project;
3) updating the forecast is not a difficult process, it just involves a reapplication of the formulas with the updated data;
4) it is useful in early stages of the project, when the information available is scarce, since it does not depend upon a long history of data.

However, there are some limitations in this model, such as there was no explanation about the association of the performance data to specific formula variables and some terms and were not sufficiently described. The model was not validated.

3.3.2 Hayes R.A., 1977

Hayes continued the study about the probabilities and the Bayes Theorem started by El-Sabban in 1973, using data from five contracts (3 research and development, 2 production) managed by the Air Force. He proposed to use the BCWS for the prior probability instead of the ACWP, using it instead for the posterior probability, as a correction to El-Sabban methodology. Then he conducted a comparative study evaluating his method, the Karsch regression-based model and the Cumulative Current Percent Cost Variance, the index-based formula with the CIc as index previously considered by Karsch.
The evaluation criteria for the determination of the most accurate formula were based on the speed with which the formula can estimate the cost at completion.
As result, the Karsch model performed better three out of five times, followed by the modified El-Sabban model proposed by Hayes.

3.3.3 Naeini M. E. and Heravi G., 2011

Naeini and Heravi proposed a forecasting method, called "Beta S-curve method", for the calculation of the EAC based on the probabilities and the Monte-Carlo simulation. In this method, the activities of the project are grouped in "sub-groups" and for each one is obtained S-curve. The summation of these sub-group S-curve gives the SS-curve of the entire project. As probability function in this model is considered the Beta distribution, which is a continuous distribution defined by two parameters α and β on a finite interval [0,1]. As in the previous method, also this one determines original estimation of project cost, which are then updated with new data as the project progresses. Then a Monte-Carlo simulation is used for the forecast of the EAC. It is based on an algorithm that generates a series of uncorrelated numbers that follow the probability distribution that the phenomenon to be investigated is supposed to have. Typically, 1000 iterations are needed to produce a range of equally likely potential outcomes. In the model, for each activity cost, a distribution function is assigned and at time intervals iterations generate a one value for cost. Therefore, it is simulated a cumulative project cost, which is normally distributed according to the Central Limit Theorem.

3.4 NEW METHODS

In this section, further gaps and new methodologies, proposed in the literature, with the aim of filling them are described.

3.3.1 Index-Regression based and Contingency-Adjusted Method

Within the EVM framework, the computation of the Estimate at Completion with the index-based approach are known for their limitations given by the reliance on past data and the unreliability of forecasts in the early stages of the project due to limited information, as seen in the previous chapters. Therefore, the regression-based methods are used to overcome these limitations. However, the available regression-based methods do not consider schedule progress in the estimation of the cost at completion. To fill these gaps and give importance to the time progress which might affect the estimate of the project cost, Narbaev and De Marco, in their papers "*Combination of Growth Model and Earned Schedule to Forecast Project Cost at Completion*" and *"An Earned Schedule-based regression model to improve cost estimate at completion"*, proposed a combined index–regression methodology based on the Gompertz growth model, whose parameters are found through regression analysis, and which integrates in the formula the contribution of the time through the ES method.

To have better understandings, before describing this model, it is necessary to introduce the basic elements of the Gompertz growth model (GGM).
The general equation of the Gompertz growth model is represented as follows:

$$GGM(X) = \alpha e^{[-e^{(\beta-\gamma x)}]}$$

Where:
x is the time
α is the asymptote for the GGM model, which represents the project final cost as time tends to infinity
β is the y-intercept, which represents the initial budget size
γ is a scale parameter, which represents the cost growth rate (GR).

The typical characteristics of the Gompertz growth model are described below and graphically shown in Figure 7.
In the early stages the work progress starts gradually, and the project's growth is slow. As the project moves forward, close to the middle stage, the progress begins to become more rapid with a position

of the inflection point at approximately 1/3 of the final cost with cumulated growth $GGM(x) = \alpha/e$, at time $x = \beta/\gamma$, when the growth rate reaches its maximum level $GR_{max} = \alpha\gamma/e$. After the point of inflection, as the project reaches its completion, the growth starts to decline.

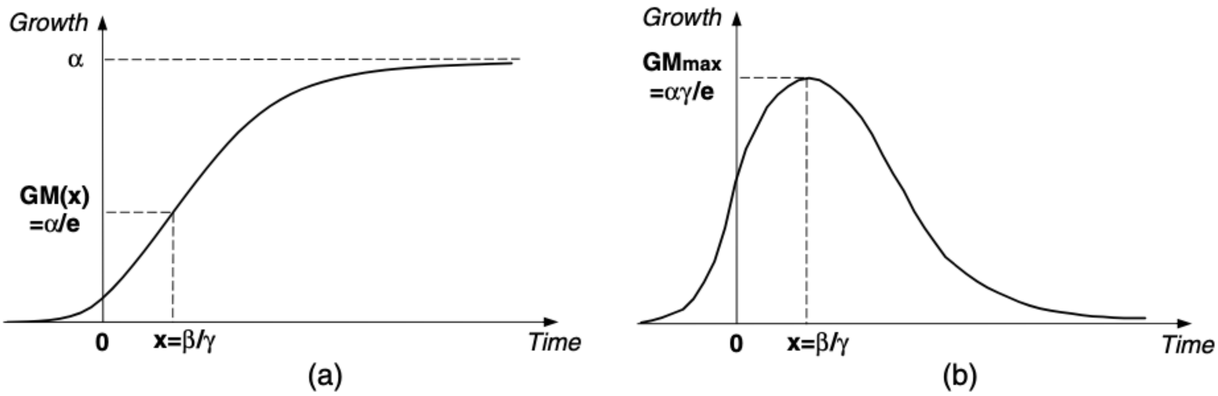


Figure 7: Gompertz growth model: (a) cumulative-growth curve; (b) growth rate curve

This model was developed using data from 9 construction projects of small to medium scale, of which five faced cost overruns and six faced schedule delays.

Three steps are necessary for its implementation: first, the parameters of the growth model are found through regression analysis; second, the EAC formula is developed; third, the ES is integrated into the EAC formula. These three steps will be described below.

I. Development of the Growth Model

The first step is to find the parameters of the Gompertz growth model through regression curve fitting. The time points are normalized to unity (1.00) assuming the maximum value, namely the Planned Duration (PD), equal to 1.00, which correspond to the 100% project time complete. The next time points are a cumulated potion of this unit. These values denote the predictor variable (x) of the GGM. The Actual Cost of Work Performed (AC) values are normalized to unity from time 0 to the Actual Time (AT) and the Budget Cost of Work Scheduled (PV) values are normalized to unity from AT to PD, assuming the maximum value, namely the BAC, equal to 1.00. The combination of the normalized AC and PV denote the response variable (y) of the GGM. Therefore, to each time point (x) of the GGM equation correspond a cost point (y).

Finally, having determined the x and y values, the best fit values of the three parameter α, β, γ of the model can be found through nonlinear regression analysis, using the Minitab® software tool.

II. Estimate at Completion formula

The proposed equation for the calculation of the EAC is a modified version of the general index-based formula and is defined as follows:

$$EAC = AC(x) + [GGM(1.00) - GGM(x)] * BAC$$

GGM(1.00) is the value of GGM when $x = 1.00$, meaning that the project is 100% time complete on time, tending to is α-asymptote.

GGM(x) is the value of GGM when the project is at the Actual time.

The difference between GGM(1.00) and GGM(x) is the estimated part of BAC necessary to complete the project.

III. Integration of ES concept

The last step for the implementation of this model consists of integrating the influence of the schedule progress over the cost estimation. If the schedule progress is favorable and efficient then the final cost will improve, while if the schedule progress is poor then the final cost will increase. To take this assumption into account, the modified equation is the following:

$$EAC = AC(x) + [GGM(CF(x)) - GGM(x)] * BAC$$

In the formula above, the GGM(1.00) is replaced with GGM(CF(x)) where CF(x) is the value of Completion Factor. The Completion Factor indicates forecasted time completion yielded to unity and is the reciprocal of the time Schedule Performance Index ($\mathrm{SI_t}$):

$$CF = \frac{1}{\mathrm{SI_t}} = \frac{1}{ES/AT}$$

If the CF is equal to 1.00 then the project duration is the planned one, thus the project is on schedule, if the CF is less than 1.00 the project is ahead of schedule and if the CF is greater than 1.00 the project is behind of schedule.

Therefore, integrating the CF into the formula entails that the schedule is a cost performance factor and its influence, such as the project delay and early finish, is included in the estimation of the cost at completion.

The evaluation criteria of the model are accuracy and precision. Accuracy is measured through the Percentage Error (PE) and the Mean Absolute Percentage Error (MAPE) for different stages of the project (early, middle, and late).
The percentage error (PE) is the difference between the final estimated (EAC) and actual (CAC) cost expressed as a percentage:

$$PE\% = \frac{EAC - CAC}{CAC}100\%$$

The Mean Absolute Percentage Error (MAPE) is the average of the absolute values of the PE over the number of valid projects tested n:

$$MAPE\% = \frac{1}{n}\sum_{i=1}^{n}|PE_i|$$

Precision is defined as the narrowness of a forecast error and it is measured by the Standard Deviation (SD) as follows:

$$SD\% = \sqrt{\frac{\sum_{i=1}^{n}(PE_i - MPE)^2}{n}}$$

Where MPE is the Mean Percentage Error.
The model that produces the more accurate EACs is the one which has a small value of SD, which means that the cost estimates calculated are closer to its Mean Percentage Error.

A comparative study between the EAC formula with and without the integration of CF was conducted and it resulted that the formula which takes into account the schedule factor through the CF produced both a more accurate and a more precise estimation in all the three stages (early, middle, and late) of the project.

Moreover, Narbaev and De Marco, conducted a comparative study evaluating their model and four index-based methods. These four methods use as index the Cost Performance Index, the Critical Ratio, the Composite Index, and the Moving Average, which is an average CI considering the three periods starting with the most recent one and can be calculated as follows:

$$MA = \sum_{i=1}^{n=3} \text{BCWP} \Big/ \sum_{i=1}^{n=3} \text{ACWP}$$

As results, the proposed method is statistically valid and resulted to be the one which produced both a more accurate and precise EAC respect to the four index-based methods and that the one integrated with the CF improved both accuracy and precision of the model.
In addition, this combined approach is able to obtain more accurate and precise results also thanks to the property of timeliness, i.e., the accuracy of the method is reliable over a certain period of time or over the entire life of the project Finally, the proposed method can be applied to a huge variety of projects in different industries and at different progress time stages.

For a dynamic project environment, it's important also to perform uncertainty and risk analysis in order to detect major risks and undertake corrective actions which might affect the forecast of the cost at completion. In fact, the estimation of cost at completion and the management of risk contingencies are often two separated processes and this represent another gap in the traditional EAC forecasting methodologies.

Narbaev and De Marco, in 2016, proposed a Contingency-Adjusted EAC formula with the aim of integrating the cost contingency management in the cost estimates at completion. The basis for this integration is the GGM model previously described due to its good level of accuracy and computational simplicity.

A contingency budget, include all management contingency reserves and it is considered as the part of the budget within the cost baseline that is allocated for identified risks, thus intended to address the known-unknowns that can affect a project.[12] At the beginning of the project, it is added to the cost baseline, then as the project moves forward, it is progressively used to undertake risk corrective actions, until the majority of the contingency cost account is spent. This behavior leads to a reversed S-curve line, graphically represented in Figure 8, which can be modelled through GGM.

[12] Project Management Institute. (2017). *A guide to the Project Management Body of Knowledge: (Pmbok Guide)* (6th edition).

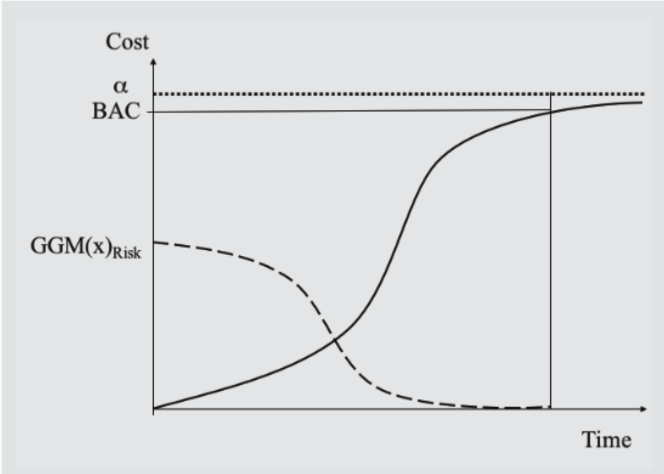


Figure 8: Cumulative BAC and Contingency Budget behaviors

So, under the simplified assumption that the contingency budget is a preset k percentage of the BAC, it can be calculated as follows:

$$GGM(X)_{Risk} = \alpha - GGM(X) * K$$

In this way, GGM(x) describe both the incurred cumulated actual cost and the contingency budget spent, so that the actual cost and residual contingency are summed, leading to a risk-adjusted budget at completion, calculated as follows:

$$BAC_{adj} = BAC * \{1 + k[\alpha - GGM[CF(x)]]\}$$

By replacing the BAC with BAC_{adj} in the EAC formula based on GGM, the new model is the following:

$$EAC = AC(x) + \{GGM[CF(x)] - GGM(x)\} * BAC_{adj}$$

$$EAC = AC(x) + \{GGM[CF(x)] - GGM(x)\} * BAC * \{1 + k[\alpha - GGM[CF(x)]]\}$$

The proposed model is tested on eight projects of different nature to better investigate its applicability and cost estimates are calculated at early (10-25%), middle (45-65%) and late stage (70-95%) of the considered projects.

Then a complete application of the proposed model was performed to one of the 8 projects considered. Accuracy and stability of the estimates at the various stages of the project were considered as evaluation criteria of the model and, for the comparison with the base model, the accuracy is measured through the percentage error (PE%).

A sensitivity analysis was conducted to understand if the k parameter of the contingency could affect the resulting EAC. This model resulted to be stable to the k value, thus it can be applied in any project regardless the contingency value.

Results show that as the k increases, the EAC becomes more accurate which proves that the larger is the risk contingency budget estimated at the beginning of the cost baseline, and so when risk impact the future performances behavior, the more accurate is this model. Moreover, the proposed risk-adjusted model performed better than the basic one especially at the early stages of the project: this happened for six out of eight projects and 75% of the estimations get closer to the actual cost at completion. While, as the project progresses, the accuracy decreases gradually: for mid stage estimates in only five out of eight projects the risk-adjusted model performed better. Thus, at the early stages of a project, when few data are available, this model has the ability to give very accurate estimation, which is an important advantage, in addition of considering the contingency costs, that it has against other models.

3.3.2 Particle Filter Methodology

The estimation of nonlinear systems is one of the most important issues in developing a project. Among the probability-based methos the Bayesian filter is the most prevalent one and it is based on the estimation of the so-called Posterior Density Function (PDF). One practical solution for the execution of the Bayesian filter are the nonparametric methods, among which the Particle filter is the most important and most used one. Hajialinajar, Mosavi and Shahanaghi proposed a method for the estimation of the cost at completion based on the Particle filter. The Particle filter smooth the learning data and for the estimation of the next states of the system it uses the smoothed data by itself, based on system fitted autoregressive (AR) model. In this is present the innovation of the method proposed. The procedure starts with the system, which is initially designed using an AR model and the system's description is then applied to the particle filter for smoothing and learning. The following step consists of estimating the next state of the system based on the estimation, through the smoothed data, of the PDF of next probability of last state. Thus, the next states are estimated in series form, resulting in the prediction of how the system will behave in later seconds.
In the proposed model, the calculation of the EAC is based on the Actual Cost estimated through the Particle filter. The CV_t is used as input of the particle filter, in each second, to derive the estimation of it in the next second, denoted with $CV_{(t+1)}$. Thus, the AC in the next time is estimated as follows:

$$AC_{(t+1)} = EV_{(t+1)} + CV_{(t+1)}$$

Where the prediction of EV is similarly done, using SV_t as input of Particle filter for the estimation of the Actual Time AT.

Another important concept to expose is the state space of a system since the CV is considered as a system state variable.

The considered state space of a system is the following:

$$\begin{cases} x_t = A.x_{t-1} + w_t \\ z_t = H.x_t + v_t \end{cases}$$

Where in the model: x_t and z_t are respectively the state and measuring variables;

H is a vector, assumed equal to unit;

v_t is the observations error caused by lack of information or delay factors;

w_t is the progress noise, which is uncertainty index and, as v_t, is caused by uncomplete knowledge about progress.

The importance of explaining these factors lies in the fact that, using the particle filter any kind of noise and uncertainty with non-Gaussian distribution and any amount of complexity can be considered and this filter has the powerful ability in denoising and smoothing the measuring data and decreasing the uncertainty effect. For these reasons this model has the priority respect to other models based on other type of filtering.

3.3.3 Soft Computing Methodology

For the estimation of the cost at completion, in addition to the use of mathematical and statistical methods, the judgement of expert represents an important factor on which rely on. Within the models seen so far, the consideration of the human perception and intuition on the progress of a project, and its related factors, represents a limit for the management and control of a project. Another fundamental element in the determination of the estimates at completion, as widely discussed in the previous paragraphs, is the reliance on past information and due to the restricted availability of data on the duration and costs during the early stages of a project, criticalities can arise during the project progress. Throughout the years, scholars attempted to fill these gaps developing models able to improve the accuracy of the cost estimation up to the use of Soft Computing and Artificial Intelligence (AI). In fact, the ability of the algorithm to quickly learn from historical information is a great advantage in the estimation of the cost at completion, respect to the other models.

Soft computing is a branch of computing systems which deals with the development of artificial intelligent systems able to mimic human brain and able to adapt to the possible changes in the project evolution. Lot of improvements regarding AI approaches, such as Neural Networks (NN), Fuzzy Logic (FL), Genetic Algorithms (GA) and other derivative approaches, have been done in the research domain of project management, including the application into cost estimation techniques.
The most important methods used will be discussed below.

3.3.3.1 Artificial Neutral Networks

The Neural Network, also known as Artificial Neural Network (ANN), are mathematical models inspired by the functioning of the human brain, imitating the way signals are sent by biological neurons.
The Artificial Neuron Network is composed of layers of nodes, defined as Artificial Neurons, located in the same structure. The typical Artificial Neuron Network is the Multiple Layer Perceptron (MLP) and it is composed by the Input Layer which consists of a number of neurons equal to the number of the inputs that enter in the network, the Hidden Layer which can contain any number of neurons and sends the inputs from the Input Layer, to which it is fully or partially connected, to the Output Layer, with which it is totally connected, and the Output Layer which must contain as many neurons as the outputs of the system.

The Artificial Neuron, also called Perceptron, is a MISO (Multiple-Input-Single-Output) system, where the input of a neuron $\{x_1, x_2, \dots, x_n\}$ are connected to the it through synaptic strength, called weights $\{w_1, w_2, \dots, w_n\}$, based on their importance and the level of contribution to the output. The single neuron is a function with the aim of combining the input and the output, and it is called activation function or transfer function and it is chosen according to the application for which the neural network will be used. The input value is transformed by the transfer function, and it's repeated between layers until a final output is determined. The output of a neuron is the weighted sum of the inputs modified by the activation function: $y = \sigma(\sum w_i x_i)$. Moreover, the output is also characterized by a constant weight w_0 called threshold or bias (b).

$$y = \begin{cases} 1, & if\ the\ output\ is\ above\ the\ bias\ value \\ 0, & otherwise \end{cases}$$

If the output of any individual node is above the specified bias value, the node is activated and it sends data to the next layer in the network, becoming the input of the next node. This process of passing data from one layer to the next, in which the outputs of one layer are inputs of the next layer only, defines this model a Feed-forward Neural Network. An exemplificative of a MLP with a feed-forward architecture is graphically represented in the Figure 9 below.

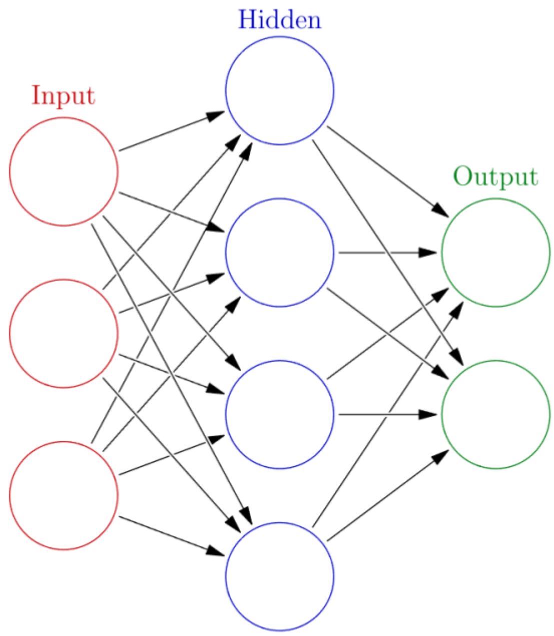


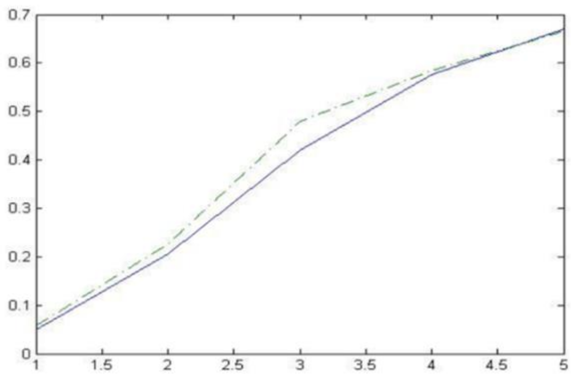
Figure 9: Artificial Neural Network

Two steps are involved in the neural network forecasting, which are training and learning. The procedure is normally performed with the supervised learning, a learning paradigm of the Machine Learning in which the learning task is to produce the desired output for each input. For the training procedure of feedforward networks, a training set needs to be available, which is a dataset of historical data containing the inputs and the desired output. In the learning procedure an input–output mapping is formed, providing repetitively an output and a variance (error measure) between it and the desired one. The objective is to minimize this error, adjusting the weights and biases, and this minimization process is repeated until the desired output is obtained.

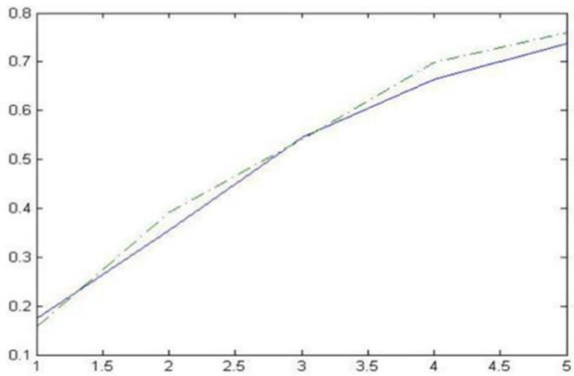
Iranmanesh and Zarezadeh, in 2008, used the ANN model to forecast the Actual Cost of Work Performed (ACWP) and EAC with the aim of improving the Earned Value Management System. They applied the Multiple Layer Perceptron MLP but with a backpropagation technique, which consist of adjusting the weights by feeding the error of a forward propagation backward through the layers of the neural network.

For their studies, they utilized 100 sample projects, simulated randomly. The project data were generated at 5 time periods: 20%, 40%, 60%, 80% and 100% of project duration.

The Feed-forward neural network used is composed by three layers with: one hidden layer, five input layers, which inputs are the BCWS in the time periods stated above and one layer output which is ACWP in the similar five input periods. A training process is performed in about 3 months, similar for all of the projects. Five types of neural networks were evaluated respectively with 2, 3, 5, 7 and 10 neurons in hidden layer and the best architecture among them had to be selected. For this purpose, the mean absolute percentage error (MAPE) is the criterion chosen for evaluating the accuracy of the neural network. In view of the computed errors for each neural networks, the selected ANN is the one with 5 neurons in the hidden layer since is the one with the minimum MAPE. After this training section, the neural network was tested among two randomly selected projects and then the ACWP and EAC were forecasted. For each project, the network was run, and the results can be seen in the following Figure 10.



(a)



(b)

Figure 10: Forecasts and actual results: (a) project 1; (b) project 2

In the figures, the continuous line is real value of ACWP, and the dash line is the forecasted one.

The results show an acceptable level of forecasting error, in fact it can be seen a strong relation between the forecasted and actual costs, demonstrating the success of the Artificial Neural Network in forecasting the ACWP and the EAC.

According to the authors, the reason why the Artificial Neural Network model is recommended for the ACWP and EAC estimations is need for project manager to estimate these values at short time and predict the EAC value at the beginning of the projects, since is a faster and more accurate model than traditional ones.

3.3.3.1 Fuzzy Neural Network

A Fuzzy Neural Network, also known as a neuro-fuzzy system, is a learning machine algorithm which uses neural network approximation techniques to determine the parameters of a fuzzy system (fuzzy set and rules). Thus, it can be seen as combination of fuzzy logic and artificial neural networks. The fuzzy logic is based on the belief that people's decisions are based on non-numerical and imprecise information. Fuzzy set is mathematic model of set, composed of vague and imprecise qualitative or quantitative data, usually generated by means of the natural language. A fuzzy subset A is defined by a membership function μ_A, mapping the 'universum' set U into a unit interval [0,1]. Given $x \in U$:

$$\mu_A(x) = 0 \; if \; x \notin A$$

$$\mu_A(x) = 1 \; if \; x \in A$$

$$\mu_A(x) \in (0,1) \; if \; x \; possibly \; belongs \; to \; A, but \; is \; not \; sure, depending \; on \; a \; vague \; opinion$$

This type of model, which incorporate the fuzziness analysis, lead to estimations based on the uncertainty of already existing factors, differently from the traditional probability-based models which forecasts the development of defined factor.

A fuzzy neural network consists of a feed-forward neural network with three layers: the first layer represents the input variables, the second one the fuzzy rules and the third one the output variables, and it is graphically represented in the Figure 11 below.

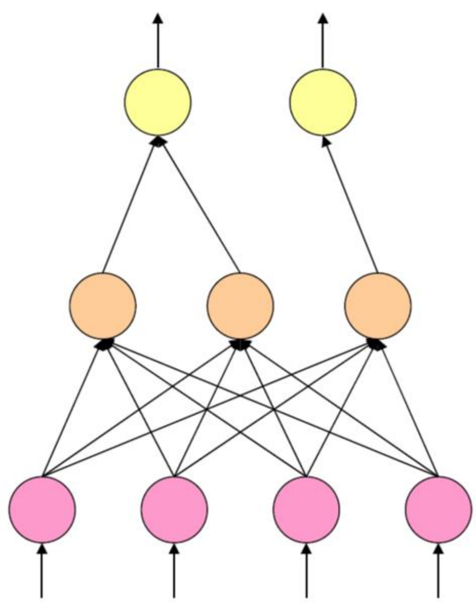


Figure 11: Fuzzy Neural Network

The fuzzy neural network training is based on the neural network learning methods and fuzzy rules can be considered as vague prototypes of the training data. In the learning process it can be represented as a set of fuzzy rules at any time, and for this reason the initialization of the system can be done with or without a prior knowledge of fuzzy rules.
There are two types of fuzzy neural network: Cooperative Fuzzy Neural Network and Hybrid Fuzzy Neural Network.
In the Cooperative Fuzzy Neural Network, the neural network and fuzzy system work autonomously from each other. While the fuzzy system is applied, the ANN learns the data from the fuzzy system.
In the Hybrid Fuzzy Neural Network, since the fuzzy system is considered as a special kind of neural network, the neural network and fuzzy system form together a fully merged entity, and therefore they do not have to communicate with each other. The fuzzy sets are considered as weights, while the rules, the input variable and the output variable are modeled as neurons.

Some authors who used the Fuzzy Neural Network and derivatives of such, for the estimation of the project cost at completion in the literature, will be briefly described below.

In 2009, Cheng, Tsai and Hsieh proposed an AI model called Evolutionary Fuzzy Neural Inference Model (EFNIM) to improve the accuracy of cost estimation, and a process of developing construction cost estimators called Evolutionary Web-based Conceptual Cost Estimators (EWCCE), obtained by integrating EFNIM, WWW, and historical construction data.
The EFNIM combines the Neural Network (NN), used in fuzzy input–output mapping, the Fuzzy Logic (FL) to deal with vagueness and the Genetic Algorithm (GA) for the optimization, in such a way that the model maximizes the strength of each and try to compensate the weaknesses of the three used alone.
In this study the NN model used is the MLP with error back-propagation. Hidden layers number ranged from 1 to 4 and neurons in each hidden layer varied from 1 to 5.
The Genetic algorithm (GA) is a method for solving constrained and unconstrained optimization problems, where a population of candidate solutions, characterized by a set of properties (chromosomes), is mutated toward better solutions. Typically, in the simplest algorithm, a chromosome is represented as a bit string with a fixed length.
The proposed EWCCE procedure for the determination of the cost estimates in construction projects is graphically represented in Figure 12.

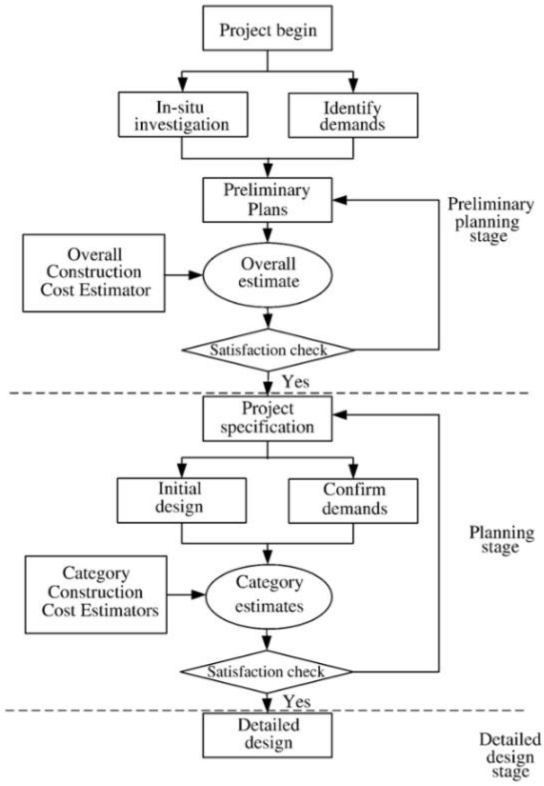


Figure 12: EWCCE procedure in construction projects

For this study were considered 28 construction projects data from 1997 to 2001, among which 23 were used for training purposes, and 5 for testing the approach.

As the basis for EWCCE, two different cost estimators were developed: overall and category estimator.

The overall estimator was designed for the total cost estimation in the absence of categorized engineering plans. In this study, six quantitative (e.g., total floor area, number of households) and four qualitative factors (e.g., interior decoration) has been identified in the planning stage of project. These factors represent the EFNIM inputs. The advantage of using the overall estimator is the ability to provide cost estimates in the early stages of a project even with a small number of data available.

Once a project design has been drafted, the category estimator is designed to evaluate each engineering work generalized within construction categories according to particular factors (in this study 7 categories were considered e.g., geotechnical construction, structural construction, decorative construction). The overall cost estimation can then be found by the summation of all category cost estimates. Respect to the overall estimator, the category estimator is recommended for a more detailed construction cost management, and it provides more reasonable and practicable results.

EWCCE result to be a web accessible platform that, with specific regional/national factors modifications, could be applicable to various construction management worldwide.

In 2010, Cheng, Tsai and Eudjono proposed an AI model called Evolutionary Fuzzy Hybrid Neural Network (EFHNN), with the aim of improving cost estimate precision. This model combines the Neural Network (NN), the high order neural network (HONN), the Fuzzy Logic (FL) and the Genetic Algorithm (GA).
The combination of the neural network and the high order neural network forms the so-called Hybrid Neural Network (HNN) which comprise the inference engine. This proposed method uses a high order neural network constructed of three layers with a high order connection and a linear connection between the 1st and 2nd layers and 2nd and 3rd layers, respectively.[13]

For this model, the same 28 projects of the study conducted by Cheng, Tsai and Hsieh in 2009, were considered and overall and category estimator are defined using in this case the EFHNN and not the previous developed EFNIM.
As results, the two type of cost estimators performed with high levels of precision. Moreover, as results of a comparative study, the proposed EFHNN, which uses both linear and non-linear connection layers, performed better against the EFNIM, which only uses traditional NN connections. Finally, also the EFHNN could be applicable to construction management worldwide with specific regional/national factors modifications.

In 2010, Cheng and Roy proposed an AI model to overcome the problems found in the uncertain and complex field of construction project. One of these is the accuracy of the EAC estimation, since the factors involved vary due to the differences present in the construction project stages. The proposed model is the Evolutionary Fuzzy Support Vector Machine Inference Model (EFSIM), an artificial intelligence hybrid system that fuses together Fuzzy Logic (FL), a Support Vector Machine (SVM) and fast messy Genetic Algorithm (fmGA).[14]

Support vector machine (SVM) is a supervised learning tool for solving classification and regression problems, which performs better against other traditional learning tool.
Fast messy genetic algorithm (fmGA) is a machine learning and optimization tool based on the genetic algorithm approach (GA). If the genetic algorithm approach uses fixed length strings, the fmGA uses strings of various lengths formed by messy chromosomes.

---

[13] Cheng, M. Y., Tsai, H. C., & Hsieh, W. S. (2009). *Web-based conceptual cost estimates for construction projects using evolutionary fuzzy neural inference model.*

[14] Cheng, M. Y. & Roy, A. F. V. (2010). *Evolutionary fuzzy decision model for construction management using support vector machine.*

The major steps involved in the EFSIM are 8 and are briefly described below and its structure is represented graphically in the Figure 13.

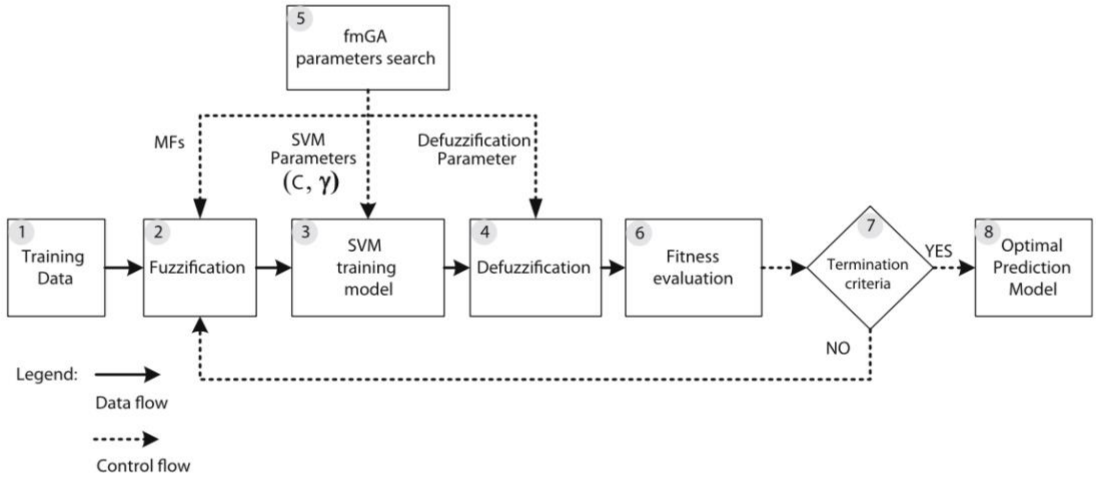


Figure 13: EFSIM structure

1. Training data: data from past cases are used by EFSIM as training data.
2. Fuzzification: this process consists of converting the values of each normalized input variable, from the preceding step, into the corresponding membership grade related to the specific membership function set generated and encoded by fmGA.
3. SVM training model: the SVM train the dataset to obtain the prediction model, using as input the output of the fuzzification process.
4. Defuzzification: this process consists of converting the fuzzy set output numbers into a single real number, after the SVM has finished the training process.
5. fmGA parameter search: this process consists of using the fmGA to search the fittest shapes for membership functions, and other parameters needed for the method. The chromosome as the model variable in EFSIM is encoded into a binary string.[15]
6. Fitness evaluation: Every chromosome is used to train the dataset, when a prediction model of the train dataset is gained the model accuracy is obtained.
7. Termination criteria: if the termination criteria are satisfied the process is terminated, if not a loop start and the process of fmGA parameter search occurs until termination criteria are satisfied.

[15] Cheng, M. Y. & Roy, A. F. V. (2010). Evolutionary fuzzy decision model for construction management using support vector machine.

8. Optimal prediction model: when the termination criteria are satisfied the optimal input/output mapping relationship with optimized parameters are identified by the prediction model. The training process is finished at this point and afterwards the model is ready for predicting new facts.

For the validation, the calculation of the estimate at completion for construction projects was performed using the EFSIM model. Data are taken from 15 different projects and consists of 10 inputs with 289 training data and 17 testing data.

The evaluation criteria used for the determination of the validity of the model are RMSE and Average Deviation. The desired and the estimated EAC with the proposed model were compared among the projects and EFSIM accuracy resulted to be acceptable and moreover, after a comparison study between this model and SVM and EFNIM results, EFSIM performed better.

The proposed model, since it combines the ability of the Fuzzy Logic to deal with vagueness, the ability of the supervised learning SVM handle fuzzy input–output mapping and the ability of the fmGA to optimize FL and SVM parameters, the level of human intervention is significantly limited respect to conventional models where these parameters rely most on expert knowledge. Therefore, this model can be used without AI knowledge.

In 2012, Cheng, Hoang, Roy and Wu proposed an AI model called EAC-EFSIMT which is a time-depended evolutionary fuzzy support vector machine inference model specifically developed for the EAC forecasting. It combines the Fuzzy Logic (FL), the weighted Support Vector Machine (wSVM), and the fast messy Genetic Algorithm (fmGA).

In Cheng previous model, the EFSIM, the SVM had still a limitation, which is that, for the determination of the regression function, it considers equally all training data points of a given class. On the contrary, in real-world, the datasets are not always equally relevant, resulting in unbalanced time series. For this reason, in these cases, the SVM is not able to generate the desirable results. Therefore, for this model, it is considered a modification of the SVM, called fuzzy SVM (fSVM) or weighted SVM (wSVM), because different types of time functions are used to weight the training data points.

In wSVM, to each input data point in the SVM is addressed a weighting value, which enables the input point to contribute differently to the learning process of the regression function. With regard to time series prediction problems, the influence of older training data point is reduced by assigning a lower weight to them.

The major steps involved in the EAC-EFSIMT are 10 and are similar to the EFSIM steps. They are briefly described below, and its structure is depicted in the Figure 14.

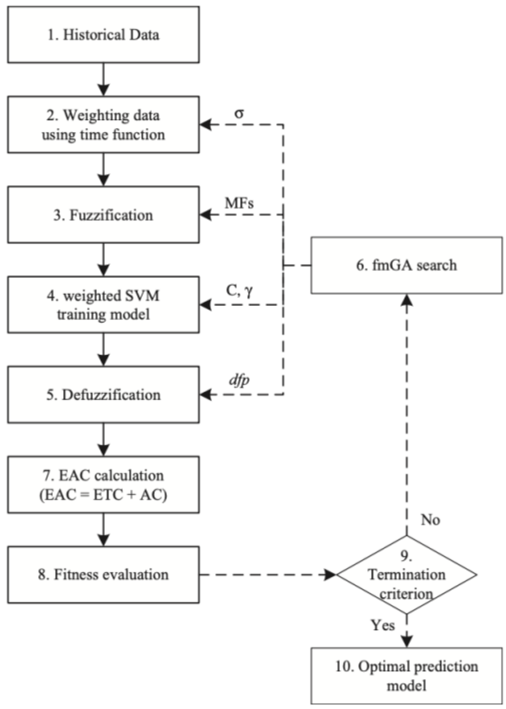


Figure 14: EAC-EFSIM steps

1. Historical data: for this study were considered 13 building projects data from 2000 to 2007, divided into 269 training data set and 37 testing data set.
2. Data weighting: as stated above, due to time importance and degree of noise corruption, a weighting value to each input point is addressed to prioritize the data point, considering the last data point the most important and the first data point the less important.
3. Fuzzification: this process consists of converting the values of each normalized input variable, from the preceding step, into the corresponding membership grade related to the specific membership function set generated and encoded by fmGA. In this study, trapezoidal triangular MF shapes are used.
4. Weighted Support Vector Machine (wSVM): the wSVM train the dataset to obtain the prediction model, using as input the output of the fuzzification process.
5. Defuzzification: this process consists of converting the fuzzy set output numbers into a single real number, after the wSVM has finished the training process.

6. Fast messy GA searching: this process consists of using the fmGA to search the fittest shapes for membership functions, and other parameters needed for the method.
7. EAC calculation: the output of defuzzification, represent the Estimate to Completion percentage (ETC) and by adding the Actual Cost of Work Performed, the EAC value is obtained, which is the final output of the model.
8. Fitness evaluation: the fitness function integrates the model accuracy and complexity, describing the fittest shape of MF, the optimized defuzzification and wSVM parameters. The fitness function represents a trade-off between accuracy and generalization of the model.
9. Termination criterion: if the termination criteria are satisfied the process is terminated, if not a loop start and the process of fmGA parameter search occurs until termination criteria are satisfied.
10. Optimal prediction model: when the termination criteria are satisfied the optimal input/output mapping relationship with optimized parameters are identified by the prediction model. The training process is finished at this point and afterwards the model is ready for predicting new facts.

For the validation of the proposed model, a cost database of projects was founded with planned values, actual values and the difference between them. Ten influencing factors were identified. After the training process, the calculation of the estimate at completion was performed by putting the monthly cost information into the optimal prediction model.

The evaluation criteria used for the determination of the validity of the model are RMSE and average deviation. The desired and the estimated EAC with the proposed model were compared among the projects and it resulted to be more accurate and reliable than traditional methods, reducing the maximum error margins in each forecasting period. Even after a comparison study with other AI models, including the EFNIM, the proposed model performed better. These results demonstrate how this model has contributed a significant improvement in EAC forecasting.

In 2012, Feylizadeh, Hendalianpour and Bagherpour, proposed a methodology which considers not only quantitative factors that might affect the forecasting of the EAC but also qualitative factors such as the weather condition, the employer cash status, and the degree of experience of the project team, thus also negative impact are taken into account. Moreover, in addition to the EAC calculation, also the forecast of any future period can be generated.

A comparative study was conducted between the use of the proposed methodology with the Fuzzy Neural Network and traditional EAC forecasting methods, in a real case study, and its results are shown in Figure 15. The proposed approach resulted to perform better against the traditional models.

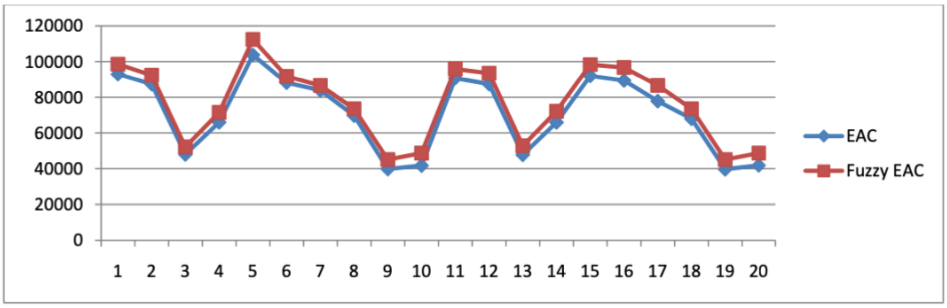


Figure 15: EAC comparative results of traditional based method and Fuzzy Neural Network

3.3.3.2 Deep Machine Learning

The traditional Machine Learning techniques, as ANN and SVM, have some limitations such as significant amount of requirement data, overfitting, and time consumption, compared to the Deep Machine Learning techniques. The introduction of deep learning algorithm dates to the late twentieth century and rapid development in these techniques occurred over the past decade.
In 2019 Kamoona and Budayan, proposed an AI model based on Deep Machine Learning with the aim of overcoming the before mentioned limitations of the traditional ones for the prediction of the EAC and its control during the project progress. It is called Deep Neural Network (DNN) which is an ANN with three or more hidden layers, and these levels of nonlinearities lead the DNN to be used in the representation of highly nonlinear and/or varying functions. A graphical representation of a DNN with three layers is in the Figure 16.

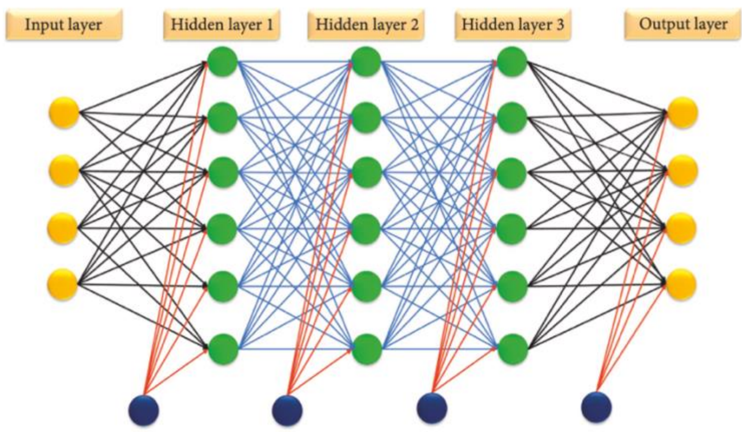


Figure 16: Deep Machine Learning

The Deep Neural Network is integrated first with the Genetic Algorithm, forming the GA-DNN and then with the Brute Force, forming the BF-DNN.

In the GA-DNN, the Genetic Algorithm is used as an optimization approach, and it involves two phases: the Genetic algorithm phase and the DNN Prediction phase which are represented in the Figure 17 below:

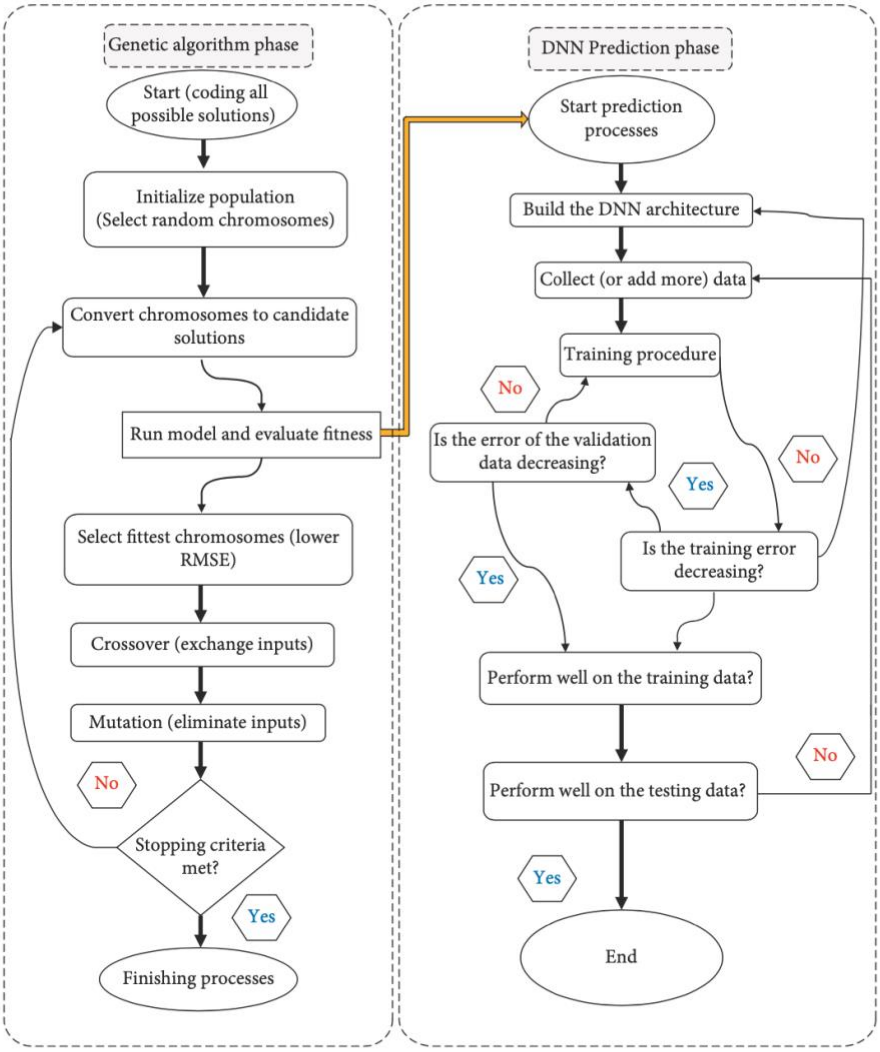


Figure 17: GA-DNN

In the BF-DNN, the DNN is combined with the Brute Force (BF) which is a selecting approach that finds all the possible solutions to select the most adequate solution to a given problem. In this study it is used as a benchmark input variable abstraction approach for the genetic algorithm.[16] Since the BF search cost is directly related to the number of choices considered, BF is suitable when there is a

[16] Kamoona, K. R. K. & Budayan, C. (2019). *Implementation of Genetic Algorithm Integrated with the Deep Neural Network for Estimating at Completion Simulation.*

limited problem's size or when there isn't a particular tool that can be employed to significantly minimize the number of answers, even if it is considered simple to use.

For the validation of the proposed model, fifteen construction projects with a duration between nine to fourteen months were considered. They covered 174 periods, 75% of which were used for the training phase and 25% for the testing phase. A cost database of projects was founded with planned values, actual values and the difference between them. Nine influencing factors were identified: cost variance (CV), schedule variance (SV), cost performance index (CI), schedule performance index (SI), subcontractor billed index, owner billed index, climate effect index, change order index, and construction price fluctuation (CCI). The estimation at completion is predicted in the learning process. The proposed model was used to compute the mathematical relationship between the nine features identified and the targeted EAC. After the training process, the calculation of the estimate at completion was performed by putting the cost information into the optimal prediction model.

A comparison study was made between the DNN, GA-DNN and BF-DNN and RMSE and R^2 were two of the used evaluation criteria. First, positive results were gained by the stand-alone proposed DNN. With the hybridization with GA and BF the outcomes were improved. For the GA-DNN, the optimal input combination for the prediction matrix was composed by CV, SV, and CI variables. This highlights the robustness of the model and its ability in achieving reliable and accurate prediction with few input variables. For the BF-DNN, the optimal input combination for the prediction matrix was composed by CV, SV, CI, SI, subcontractor billed index, change order index, and CCI. Its accuracy was higher than the one performed by GA-DNN, with a minor value of RMSE, resulting to be the most capable model among the others. However, to attain this level of accuracy the BF-DNN needed more input variables than the GA-DNN. Therefore, the GA-DNN is recommended for project cost forecasting with low availability of data.

3.3.3.3 Extreme Learning Machine

In 2019, AlHares and Budayan, proposed new version of Artificial Neural Networks called Extreme Learning Machine (ELM), for the forecast of the project construction estimation at completion, with the aim of overcoming the before mentioned limitations of the ANN model. Over the past years, the ELM model has been improved and it has been applied for solving non-linear and stochastic complex problems among multiple engineering fields.

In this study, the ELM was proposed for its ability of mimic in short time the human brain, for this reason the term "extreme" is used. Its advantage, respect to the traditional data intelligent models, is the single layer feedforward network (SLFN) which lead to a rapid rate of learning. In fact, the ELM has a simple and unique learning process because the hidden neurons do not require any tuning process during the learning phase.[17] The ELM learning process has the aim of obtaining the least error. The input and output variables are connected by the hidden neurons in such a way that the features are determined by input and output weights randomly generated. The internal weights are generated through an M-dimensional mapping feature space, which is used by the ELM to process the input data. The output is generated with this mathematical procedure:

$$F_{(x)} = \sum_{i=1}^{M} \beta_i h_i(x)$$

Where: x are the input variables, M is the dimension of the mapping feature space, β_i is the weight and h_i is the output of the hidden nodes.
The structure of the ELM is denoted in the Figure 18.

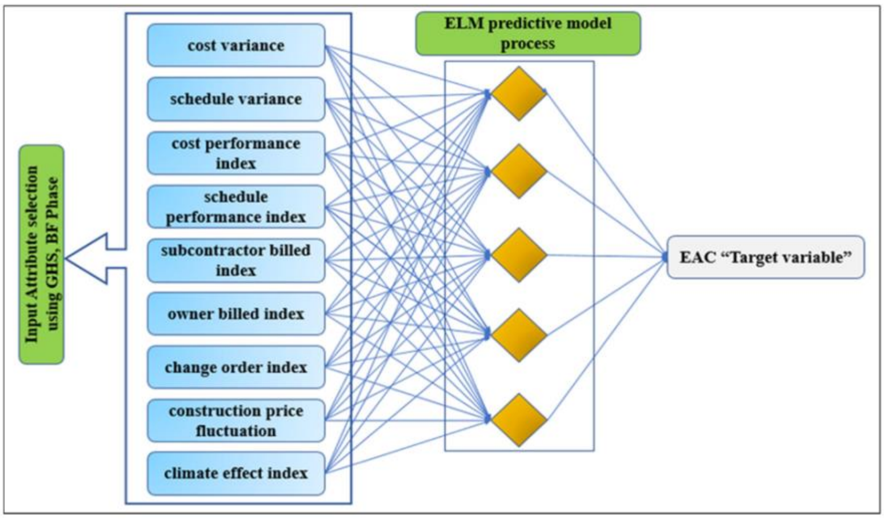


Figure 18: ELM structure

The ELM is integrated first with the Global Harmony Search, forming the GHS-ELM and then with the Brute Force, forming the BF-ELM.

[17] AlHares, E. F. T. & Budayan, C. (2019). *Estimation at Completion Simulation Using the Potential of Soft Computing Models: Case Study of Construction Engineering Projects*

In GHS-ELM, the ELM is combined with the Global Harmony Search (GHS), which is a search algorithm that was developed based on the improvisation process of musicians in finding a pleasant harmony. It is easy to implement, and it is used for the optimization process in finding the optimal solution for problems with continuous and discrete variables. For this study purpose, the GHR algorithm is applied as selection approach of the input variables for the ELM model. This model involves two phases: first the GHS algorithm is used to select the attribute-based variable and the second phase in which the ELM model is implemented for the prediction of the EAC.
In the BF-ELM, the Brute force is used for feature selection process and as a benchmark to the GHS algorithm performance.

For the validation of the proposed model, 11 construction projects with a duration between 11 to 12 months were considered. They covered 132 periods, 75% of which were used for the learning process and 25% to initiate the testing phase. In this study, 9 input variables has been identified: cost variance (CV), schedule variance (SV), cost performance index (CI), schedule performance index (SI), subcontractor billed index, owner billed index, change order index, construction price fluctuation (CCI) and climate effect index. The proposed model was used to compute the mathematical relationship between the nine features identified and the targeted EAC. The estimation at completion is predicted in the learning process.

A comparison study was done between the hybridization with the GHS and BF, the ELM model and the simple ANN model. RMSE and R^2 were two of the used evaluation criteria. First, positive results were gained by the stand-alone proposed ELM, and it resulted to reach a better prediction performance compared to the ANN model. Even with the hybridization of the ELM and ANN models with GHS and BF, the ELM performed better than the ANN. Then it was performed a comparison between the GHS-ELM and BF-ELM. For the GHS-ELM the optimal input combination was composed by CV, SV, and SI variables. For the BF-ELM, the optimal input combination was composed only by CV and SV parameters. This analysis shows that BF-ELM did better against the GHS-ELM both in terms of performance accuracy and in its ability of achieving reliable and accurate prediction with less input variables. However, to abstract the internal relationship, the BF-ELM model resulted to require more execution time against the GHS-ELM model.

3.3.4 Multiple Linear Regression Model for Improved Project Cost Forecasting

Several studies have been done during the last years to improve the methods used for the forecasting of the project Cost Estimate at Completion. Some methods, resulted to be difficult to applicate to various types of projects and sometime, like in some soft computing models, the focus is in finding the exact accuracy of the forecasts and not the lead to understand the processes and the relationship between the variables used.

In 2022, Ottaviani and De Marco, proposed a multiple linear regression model to improve the project cost estimate. The objective of this study was to develop a model by looking at the trade-off between the variance, accuracy, and application difficulty. Moreover, also the relationship between the variables used are considered to find the best forecasting model. For these reasons, to improve the traditional EAC formula and evaluate its performance, the multiple linear regression analysis has been used.
The proposed model was developed using EVM data from 29 different real-life projects with a total of 805 observations and involved three steps. First, three regressors, $fEAC$, CI and WP, were chosen in the generalized linear model selection procedure. Second, a correlation analysis was performed, where it was verified the absence of the correlation between the variables. Third, multiple linear regression analysis was implemented, and the definitive model was developed:

$$r\widehat{EAC} = \beta_0 fEAC + \beta_1 CI - \beta_2 WP + \varepsilon$$

The first regressor, $fEAC$ is the forecast that need to be adjusted and is computed as follows:

$$fEAC = \frac{[AC + (1 - WP)/CI]}{BAC}$$

The second regressor, CI, is the Cost Performance Index and shows how it is not only considered in the forecasting of the remaining costs as the $fEAC$ formula indicates, but in the forecasting of the whole project at completion costs.
The third regressor, WP, is the percentage of work performed; the negative sign indicates that the $fEAC$ and CI contributions to the output decrease as the project tends to its completion.
Finally, the coefficients are defined, and the fitted model is the following:

$$r\widehat{EAC} = 0.70612fEAC + 0{,}45998CI - 0{,}07931WP$$

This new EAC forecasting model is benchmarked against the traditional index-based formulas and the evaluation criteria of the model are accuracy and variance.

As results, the proposed model performed better than the traditional ones, providing a lower variance and higher accuracy. This confirms the need to adjust the simple EAC computation as it alone would fail to capture the correlation between the EVM variables' evolution over time. It is also confirmed the need for additional EVM variables to account for the trends in the cost and time performance indexes.

3.5 FINAL CONSIDERATIONS

From the study of the literature regarding the most used methodology for the Estimate at Completion forecasting, some important considerations have been arisen.

First, after examining the computation of the Estimate at Completion with the Index-based methods, it can be stated that the formulas implemented are relatively easy to use. This can represent an advantage of this methodology over more difficult ones since, most of the time, managers prefer to use simple and quick forecast methods for the estimation of time and cost. In fact, an accurate and precise time and cost forecasting need to be balanced also with the easiness to use and practical applicability of the adopted method. The index-base models are also known for their three primary limitations given by the unreliability of forecasts in the early stages of the project due to limited information and the reliance on past data, assuming that the remaining budget is adjusted by a performance index. Moreover, the assumption of linearity of the project work, approximating it to a straight line, leads to the belief that no further performance changes or risk could occur until the project ends.

To overcome the limitations in the index-based methods, growth models which reflect the shape of S-Curve are used for the fitting process, through the regression analysis. Regression-based models are thus proposed by authors to better describe the nonlinear relationships between the variables. They resulted to improve the accuracy of the EAC through their curve fitting process and perform better in the early stages of a project, even with the limited availability of information. However, this superiority is not demonstrated in each study conducted in the literature and sometimes the index-base formulas were not less accurate the regression-based one. Moreover, differently from the index-based models which require at most a simple calculator, the regression-based models require the use of computer programs to find important data for the analysis of the formulas.

To overcome the limitation of the dependency on past data and availability of few data at the beginning of a project, some probability-based models were proposed by authors in the literature, according to which, these models are useful in early stages of the project since they do not depend upon a long history of data.

Another gap found towards the literature is that most EMV methods do not consider schedule progress in the estimation of the cost at completion. To overcome this limitation the index-regression based model was proposed, integrating the ES concept in the EAC formula to give importance to the time progress.

For a dynamic project environment, it is important also to perform uncertainty and risk analysis to detect major risks and undertake corrective actions which might affect the forecast of the cost at completion. In fact, the estimation of cost at completion and the management of risk contingencies are often two separated processes and this represent another gap in the traditional EAC forecasting methodologies. To overcome this gap, a Contingency-Adjusted EAC was proposed and at the early stages of a project, when few data are available, this model can give very accurate estimation, which is an important advantage, in addition of considering the contingency costs, that it has against other models.

The consideration of the human perception and the judgement of expert represents an important factor on which rely on but, at the same time, a limitation within the EVM models seen in the literature. At this purpose, different models based on Soft Computing, Machine Learning and Artificial Intelligent approaches were proposed to limit the level of human intervention, respect to conventional models where these parameters rely most on expert knowledge. Moreover, also the ability of the algorithm to quickly learn from historical information fill the gap of the limited availability of data at the early stages of a project. However, the main criticality related to the use of Machine Learning for the calculation of the EAC is that most of them are "black-box" models. They are able to provide very accurate and precise results in short time, but the logic by which they are obtained is unknown, because the process is based on the aggregation of too many mathematical functions. This can be a problem since during the project progress it's necessary to inform the stakeholders about the status of the project and sometimes they also require information about how the forecasted EAC has been reached. In this case are preferred "white-box" models, which process is instead easy to understand, but they provide less accurate predictions. For this reason, it's important to understand which are the priorities in a project, thinking about the trade-off between precision and interpretability of results.

Finally, after having studied all these models for the Estimate at Completion forecasting and its related limitations, the most important thing that can be stated is that it does not exist a general method for the calculation of the Estimate at Completion that can be universally used. On the contrary the choice of the most accurate method will depend on a multiplicity of factors of a project such as the type, characteristics, industry, sector, and the percent at completion when they are applied. The best choice would be to have a tailored method to the specific context of the project. However, implementing a model for each specific project would be complex and time consuming.

CONCLUSION

The estimation at completion of the time and cost of a project is a critical and essential element for a project to be successful. It represents an important tool for a project manager in the phase of monitoring and control of the project progress. In fact, the project manager needs an accurate estimation of cost and time of work completion to be able to undertake corrective actions in time, if some problems arise, to bring the project back to the optimal conditions.

In this thesis, the focus is on the importance of the cost at completion and the way at which it can be estimated. In the light of what emerged from the conducted literature review, numerous forecasting methodologies for the Estimate at Completion where identified. The need to find a way to obtain very accurate estimations arises in the late seventies in the field of military projects among the US Department of Defense. Since then, scholars tried to find the most appropriate and precise methodology for the EAC forecasting. The Earned Value Management was the most important management tool for the EAC calculation. However, the field in which the project management was being applied evolved in time, passing from the military to many other areas of applications. Thus, since a project can present peculiar characteristics, that may not be present in other projects, also the factors that can affect them are different. For this reason, the need to use different EAC forecasting methodologies appropriate for each project occurs and limitations of the EVM methodologies were identified. Until today, a lot of studies has been conducted to find accurate and precise estimations and thanks to the progress of the technology also Soft Computing methodologies are used. Their ability to mimic human brain and to adapt to the possible changes in the project evolution represents an important step foreword in the search of very accurate estimations. However, the complexity of these models and their unknown logic behind the results obtained are critical elements of these models.

Finally, a standard EAC forecasting methodology, applicable in every project typology, does not exist and a good project manager should be able to choose the most accurate method to use according to the multiplicity of factors that characterize a specific project. The fact that a generalized method is almost impossible to achieve does not mean that research with the aim of improving existing models or ideating new ones should be limited. On the contrary, further studies regarding this topic should continue to make further contribution to the existing literature.

References

Abba, W. F. (1997). *Earned value management - reconciling government and commercial practices.*

AlHares, E. F. T. & Budayan, C. (2019). *Estimation at Completion Simulation Using the Potential of Soft Computing Models: Case Study of Construction Engineering Projects*.

Araba, A. M., Alhawat, M., Sackey, S., Ali, A., Memon Z. A. & Milad, A. (2021). *Estimation at Completion in Civil Engineering Projects: Review of Regression and Soft Computing Models.*

Blythe, A.L. (1982). *Validation of ASD/ACCM's Cost Performance Analysis Algorithm*.

Bright, H.R., & Howard, T.W. (1981). *Weapon System Cost Control: Forecasting Contract Completion Costs.*

Busse, D.E. (1977). *A Cost Performance Forecasting Model.*

Cheng, M. Y., Hoang, N. D., Roy, A. F. V. & Wu, Y.W. (2012). *A novel time-depended evolutionary fuzzy SVM inference model for estimating construction project at completion.*

Cheng, M. Y. & Roy, A. F. V. (2010). *Evolutionary fuzzy decision model for construction management using support vector machine.*

Cheng, M. Y., Tsai, H. C., & Hsieh, W. S. (2009). *Web-based conceptual cost estimates for construction projects using evolutionary fuzzy neural inference model.*

Cheng, M. Y., Tsai, H. C., & Sudjono, E. (2010). *Conceptual cost estimates using evolutionary fuzzy hybrid neural network for projects in construction industry.*

Christensen, D. S., Antolini, R. C., & McKinney, J. W. (1992). *A review of estimate at completion research*.

Covach, J., Haydon, J.J., Reither, R.O. (1981). *A Study to Determine Indicators and Methods to Compute Estimate at Completion (EAC).*

Cronin, P., Ryan, F. & Coughlan, M. (2008). *Undertaking a literature review: a step-by-step approach.*

Cryer, J.M. & Balthazor L.R. (1986). *Evaluation of Weighted Indices on Algorithms Utilized for Calculating Independent Estimates at Completion.*

De Marco, A. (2018). *Project management for Facility Constructions* (2nd edition). SPRINGER.

De Marco, A., & Narbaev, T. (2013). *Combination of growth model and earned schedule to forecast project cost at completion.*

De Marco, A., & Narbaev, T. (2014). *An earned schedule-based regression model to improve cost estimate at completion*.

De Marco, A., & Narbaev, T. (2014). *Earned value-based performance monitoring of Facility Construction Projects*.

De Marco, A., Narabaev, T. & Rosso, M. (2016). *Nonlinear cost estimates at completion adjusted with risk contingency.*

El-Sabban, M.Z. (1973). *Forecast of Schedule/Cost Status utilizing Cost Performance Reports of the Cost/Schedule Control Systems Criteria: A Bayesian Approach*.

Feylizadeh, M. R., Hendalianpour, A. & Bagherpour, M. (2012). *A fuzzy neural network to estimate at completion costs of construction projects*.

Hajialinajar, M. T., Mosavi, M. R. & Shahanaghi, K. (2016). *A New Estimation at Completion of Project's Time and Cost Approach Based on Particle Filter*.

Haydon, J.J., & Reither, R.O. (1982). *Methods of Estimating Contract Cost at Completion.*

Hayes, R.A. 1977. *An evaluation of a Bayesian approach to compute Estimate at Completion for Weapon System Programs*.

Heydinger, G.N. (1977). *Space and Missile Systems Organization Cost Performance Forecasting Study.*

Holeman, J.B. (1975). *A Product Improved Method for Developing A Program Management Office Estimate Cost at Completion.*

Iranmanesh, S.H., & Zarezadeh, M. (2008). *Application of Artificial Neural Network to Forecast Actual Cost of a Project to Improve Earned Value Management System*.

Kamoona, K. R. K. & Budayan, C. (2019). *Implementation of Genetic Algorithm Integrated with the Deep Neural Network for Estimating at Completion Simulation.*

Karsch, O.A. (1974). *A Cost Performance Forecasting Concept and Model*.

Karsch, O.A. (1976). *A Production Study Sequel to the Cost Performance Forecasting Concept and Model.*

Land, T.J., & Preston, E.L. (1980). *A Comparative Analysis of Two Cost Performance Forecasting Models.*

Lipke, W. (2003, January). *Schedule is different*.

Lollar, J.L. (1980). *Cost Performance Analysis Program for Use on Hand-Held Programmable Calculators.*

McKinney, J.W. (1991). *Estimate at Completion Research – A Review and Evaluation.*

Naeini, M. E., & Heravi, G. (2011). *Probabilistic Model Development for Project Performance Forecasting: Semantic scholar.*

Nystrom. T.D. (1995). *A Comparative Study of Linear and Non-Linear Estimate at Completion Methods.*

Olsen, D., & Ellsworth, R.W. (1976). *Forecasting Techniques Employed in a Line Organization.*

Ottaviani, F. M. & De Marco, A. (2021). *Multiple Linear Regression Model for Improved Project Cost Forecasting.*

Parker, C.W. (1980). *C/SCSC and C/SSR Cost Performance Analysis Programs*.

Price, J.B. (1985). *An Evaluation of CPRA Estimate at Completion Techniques Based Upon AFWAL Cost/Schedule Control System Criteria Data*.

Project Management Institute. (2011). *Practice standard for earned value management* (2nd edition).

Project Management Institute. (2017). *A guide to the Project Management Body of Knowledge: (Pmbok Guide)* (6th edition).

Reidel, M.A., & Chance J.L. (1989). *Estimates at Completion (EAC): A Guide to Their Calculation and Application for Aircraft, Avionics, and Engine Programs*.

Totaro, J.A. (1987). *A Logical Approach to Estimate at Completion Formulas*.

Wallender, T.J. (1986). *HQ Air Force Systems Command Estimate at Completion Formula Justification.*

Watkins, H. (1982). *An Application of Rayleigh Curve Theory To Contract Cost Estimation and Control.*

Weida, W.J. (1977). *A General Technique for R&D Cost Forecasting.*

Sitography

https://www.pmi.org/

https://www.ibm.com/it-it/cloud/learn/neural-networks#toc-cosa-sono--nIYJFM88

http://www.scholarpedia.org/article/Fuzzy_neural_network

https://en.wikipedia.org/wiki/Genetic_algorithm