A PROBABILISTIC PROJECT CONTROL TOOL FOR PROJECTS WITH HIGH RISKS AND UNCERTAINTY

YÜKSEK RİSK VE BELİRSİZLİĞE SAHİP PROJELER İÇİN OLASILIKSAL PROJE KONTROL ARACI

YASEMİN SÜ

ASSIST. PROF. BARBAROS YET
Supervisor

Submitted to
Graduate School of Science and Engineering of Hacettepe University
as a Partial Fulfillment to the Requirements
for the Award of the Degree of Master of Science
in Industrial Engineering.

2020

İthaf

Canım Anneme…

ÖZET

YÜKSEK RİSK VE BELİRSİZLİĞE SAHİP PROJELER İÇİN OLASILIKSAL PROJE KONTROL ARACI

Yasemin SÜ

Yüksek Lisans, Endüstri Mühendisliği Bölümü

Tez Danışmanı: Dr. Öğr. Üyesi Barbaros YET

Ağustos 2020, 68 sayfa

Proje izleme ve kontrol, projelerin başarılı olması için çok önemlidir. En yaygın olarak kullanılan proje kontrol yöntemlerinden biri Kazanılan Değer Yönetimi'dir (KDY). KDY, projelerin maliyet, zaman ve iş kapsamı açısından kontrol edilmesini sağlar ve projenin ilerlemesine göre tamamlanma süresi ve maliyeti hakkında tahminlerde bulunabilir. Ancak projelerin ortak özelliği risk ve belirsizlik içermeleridir ve KDY belirsizlik ve risk faktörlerini dikkate almadığı için, yüksek risk ve belirsizliğe sahip projelerde etkili değildir.

Bu çalışma, belirsizlik ve risk altında etkin proje kontrolü yapabilen bir proje kontrol aracı geliştirmeyi amaçlamaktadır. Bu araç, projeyi maliyet, zaman ve kapsam açısından çok boyutlu olarak kontrol edebilir ve bu parametrelerle ilgili olan belirsizlikleri ve nedensel risk faktörlerini hesaplayabilir. Araç, proje parametrelerindeki belirsizliği ve risk faktörlerini modellemek ve bunlarla ilgili istatistiksel hesaplamalar yapmak için Bayes Ağları'nı kullanır. Bayes Ağları, olasılıksal ilişkilerin modellenmesi ve hesaplanması için uzman bilgi ve verilerin birleştirilmesine izin veren güçlü bir modelleme tekniği sunar.

Geliştirilen aracın farklı proje alanlarına uygulanabilirliğini incelemek için, üç farklı alanda vaka çalışmaları incelenecektir. Bu vaka çalışmalarından ikisi farklı sektörlerden gerçek proje verilerine dayanmaktadır. Geliştirilen aracın olumlu ve olumsuz yanları değerlendirilecektir.

Anahtar Kelimeler: Proje Yönetimi, Bayes Ağları, Kazanılan Değer Yönetimi, Risk Yönetimi

ABSTRACT

A PROBABILISTIC PROJECT CONTROL TOOL FOR PROJECTS WITH HIGH RISKS AND UNCERTAINTY

Yasemin SÜ

Master of Science, Department of Industrial Engineering
Supervisor: Assist. Prof. Barbaros YET
August 2020, 68 pages

Project monitoring and control are essential for project success. One of the most commonly used project control methods is Earned Value Management (EVM). EVM ensures that the projects are controlled in terms of cost, time and scope of the work and can make estimates about the completion time and cost according to the progress of the project. However, the common feature of the projects is that they contain risk and uncertainty and since EVM does not take into account uncertainty and risk factors, it is not effective in projects with high risk and uncertainty.

This study aims to develop a project control tool that is capable of effective project control under uncertainty and risk. The tool can control the project in multiple dimensions in terms of cost, time and scope, and it can calculate the uncertainty and causal risk factors related to these parameters. The tool uses Bayesian Networks (BNs) to model uncertainty and risk factors in the project parameters and to make statistical calculations related to them. BNs offer a powerful modeling technique for modeling and calculating probabilistic relationships, allowing expert knowledge and data to be combined.

In order to examine the applicability of the developed tool to different project areas, case studies will be examined in three different areas. Two of these case studies were based on real project data from different sectors. The positive and negative sides of the developed tool will be evaluated.

Keywords: Project Management, Bayesian Networks, Earned Value Management, Risk Management

ACKNOWLEDGMENT

I would like to express my special thanks to my Academic Supervisor “Dr. Barbaros YET” for his guidance and support in completing my thesis.

I would also like express my gratitude to my family and my friends who supported me always with love and understanding.

I am also grateful because this work was supported by the Scientific and Technological Research Council of Turkey (TUBITAK – Grant no: 216M483).

CONTENTS

FIGURES

TABLES

SYMBOLS AND ABBREVIATIONS

Symbols

pa	Parents
E_i	Approximate relative entropy error
f_{max}	Maximum density
f_{min}	Minimum density
$\bar{f}$	Mean density
w_j	Discretization interval

Abbreviations

AC	Actual Cost
BN	Bayesian Network
CPT	Conditional Probability Tables
CPI	Cost Performance Index
CV	Cost Variance
DAG	Directed Acyclic Graphs
DD	Dynamic Discretization
EAC	Estimate at Completion
ESM	Earned Schedule Method
ETC	Estimate to Complete
EV	Earned Value
EVM	Earned Value Management
HBN	Hybrid Bayesian Network
IEDAC	Independent Estimate of Duration at Completion
JT	Junction Tree
PDWR	Planned Duration for Work Remaining
PV	Planned Value
SPI	Schedule Performance Index
SV	Schedule Variance

1. INTRODUCTION

A *project* can be defined as activity series that aims delivering a targeted system, product or service. Although each project has different features, some elements are common to all projects. Projects consist of the sum of activities whose beginning, and end are clearly defined to achieve a goal. An output is obtained at the end of the project. Therefore, it is a result-oriented approach. In project management, monitoring is one of the major requirements and control of the projects as measuring is essential for success.

Projects generally contain high uncertainty and risk. Unexpected events are inevitable in projects generally. These events can cause negative effect on the project performance and they rarely resolve on their own. The project manager should make correctional interventions when negative risk events occur. Monitoring and controlling the projects by project managers are really important for success of projects. Therefore, the project manager needs project control tools that monitors the project progress. They need measure the performance of projects by these tools. Also these tools should support the decision making on how to intervene in the project. *Earned Value Management (EVM)* is the tool that preferred for using mostly in project measurements. *EVM* was developed to be used in US Department of Defense projects in the 1960s. Today, it is widely used by public and private organizations in many countries working in various fields.

EVM evaluates the projects in point of time, performance and cost and makes estimates about completion cost and time of the projects. It aids the project manager about whether the project is being carried out within budget limits, whether it is progressing in accordance with the planned schedule and when and at what cost the project can be completed. *EVM* makes it possible to detect deviations in performance expectations, allowing the project manager to compensate for the remaining time.

EVM methodology compares the performed tasks and planned tasks and also evaluate the actual expenditures. It uses three parameters *i.e*. planned value that is budgeted for the evaluated time in the project baseline plan, earned value for the finished tasks and actual cost incurred for calculating performance indices. However, planned value parameters are often subject to uncertainty and risk because of the projects' nature. Moreover, the completion rates of tasks may not be precisely defined therefore they may involve uncertainty as well. In *EVM*, there is not a structure for evaluating the parameters' uncertainty and the risk factors affecting the parameters, and this is one of its main limitations. This study focuses on integrating project uncertainty and risk with the *EVM* approach.

1.1. Aim of Study and Method Used

In this study, developing an *EVM* based project control tool that enables effective control of projects with high uncertainty and risk is aimed. The proposed tool uses *Bayesian Network (BN)* modeling technique to model parameters and risk factors related to project performance and make statistical calculations related to them. *BN*s offer a powerful modeling technique for modeling and calculating probabilistic relationships, and it enables the integration of expert knowledge and data in these models. The proposed tool aims to control the project dimensions of cost, schedule and scope. In addition calculating the uncertainty and causal risk factors related to these parameters are also aimed.

The underlying *BN* model of the tool contains discrete and continuous values. It was built and computed by using the AgenaRisk software. Case studies are conducted in three different project examples and the results were evaluated to examine the applicability of the developed tool to different project areas, One of the case studies is real world project which belongs to a company in Turkey.

1.2. Outline of Thesis

The rest of this study is organized in the following order. The second chapter gives a definition of Bayes Theorem and describes the properties of *BNs*. *Hybrid BNs* are described by comparing the continuous and discrete *BNs*. Finally, previous applications of *BN*s in project management are reviewed. The third chapter examines *EVM* method in detail including its definition, parameters and properties. It demonstrates how the *EVM* method can be used for measuring project performance and shows the application process of *EVM* step by step. The *Earned Schedule Method (ESM)*, that is a derivative form of *EVM,* is also described in this section, and the previous *EVM* and *ESM* studies from the project management literature are examined. The fourth chapter describes the *BN* model developed within the scope of this study. The structure and properties of the proposed model, and the steps of building the model are shown. The fifth chapter applies the developed model to three different project case studies, than analyzes the results. The sixth chapter presents the analysis and conclusions about study and discusses its results, advantages, disadvantages and relevant future studies.

2. BAYESIAN NETWORKS

Bayes Theorem is one of the fundamental building blocks of probability and was introduced by Thomas Bayes and Pierre-Simon Laplace in the 18th and 19th century. This theorem used for determining the inverse conditional probability of two events and formulated as follows:

$$P(B|A) = \frac{P(A|B) * P(B)}{P(A)} \tag{2.1}$$

Bayes theorem enables us to revise our prior beliefs about a hypothesis when we observe data. In this equation, we can interpret *A* as data or observations, and *B* as our prior belief about the hypothesis. The theorem enables us for calculating the revised probability of *B* given *A*, i.e. *P(B|A)*, based on *P(A)*, *P(B)* and *P(A|B)*. This way of revising beliefs based on data is called Bayesian inference.

BNs are also called belief networks. *BNs* are member of the probabilistic graphical models. BNs are models that is used for representing the joint probability distribution and conditional independence assumptions of a set of variables. *BNs* enable us to represent and calculate more complicated Bayesian inference problems with multiple and interrelated variables. A *BN* contains two main parts: a graphical structure that called *Directed Acyclic Graphs (DAG)* and conditional probability tables. *DAGs* contains nodes that shows the variables and directed edges, which shows the probabilistic dependency relations among the variables.

In the *BNs*, there is parent node that is at the beginning of the edge and child node that is at the end of the edge when two nodes in the network are connected to each other by a directed edge. In addition, nodes on the path from a node are called descendant, nodes on the path coming to a node are called ancestors.

Causal relations for any subject can be shown like a graphical model by using the graphical structure of the *BN*. We can see an example of *BN* with six variables in *Figure 2.1*. In this model, *A* and *B* are C's parent while *C* is a child node. *F* and *C* are D's parent while *D* is a child node. *E* is also child node. *D* is a descendant of *A* and *B* while

A and *B* are ancestor of *D*. The parameters of this *BN* are conditional probability tables that show the strength of the relationships modeled in graphical structure. In this model, the conditional probability distributions of variables, *P(A), P(B), P(F), P(C|A,B), P(E|B) and P(D|C,F)* must be specified. Probabilistic calculations can be made about the variables using BN inference algorithms (Section 2.2) after the graphical structure and parameters are determined.

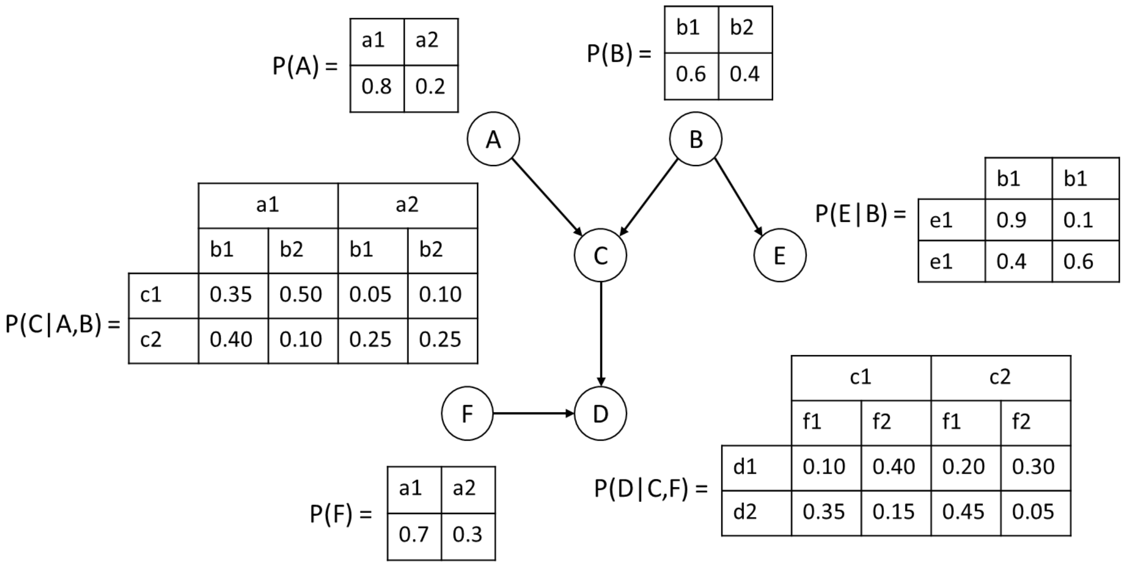


Figure 2.1. BN example with six variables

2.1. Relations in BNs and D-Separations

A set of three nodes in a *BN* can be connected by serial, diverging and converging connections. There is a demonstration of this in *Figure 2.2.*

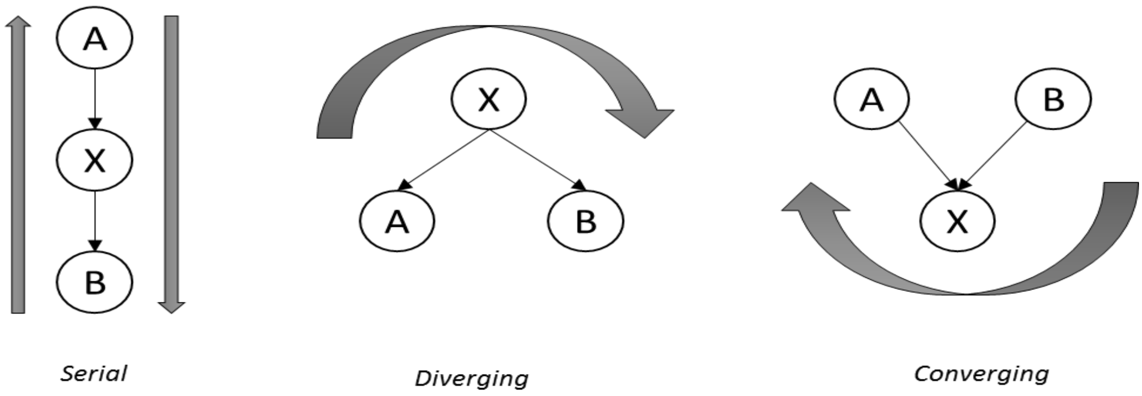


Figure 2.2. Types of Bayesian Connections

The independence status of the variables in the *BN* can be determined by using the d-separation principle. "Two distinct variables *A* and *B* in a causal network are d-separated if for all paths between *A* and *B*, there is an intermediate variable *X* (distinct from *A* and *B*) such that either

- The connection is serial or diverging and *X* is instantiated

  or

- The connection is converging, and neither *X* nor any of *X* descendants have received evidence." (Jensen, F. V. 1996)

In a *BN* structure for the variables of *A* and *B*, if we assume that they are d-separated given *X*, they can be called conditionally independent given *X* in every joint probability distribution can be modelled in this *BN*. If they are not d-separated, they can be called d-connected.

In Figure 2.1, nodes *A* and *B* are marginally independent, but given variable *C*, these variables are conditionally dependent. When variable *B* is given, variables *C* and *E* become conditional independent. When *C* is given, *D* variable becomes conditional independent from *A* and *B*.

For a variable in *BN*, if we have a set that consist the parents and children of that variable and the other parents of children of that variable, it is called Markov Blanket of that variable. We can give an example to this from *Figure 2.1*; the Markov Blanket of node *C* is {*A*, *B, D, F*}. In a *BN*, when the Markov Blanket of a variable is observed, it makes the variable d-separated from all other variables in the *BN*.

2.2. Inference in BNs

In a *BN*, conditional probability distributions defined for every node. The set of these distributions are the parameters of *BN*. They show the probability distribution for each variable conditioned on its parents. Conditional probability distributions are shown in tabular form when variables are discrete, and they are called "*conditional probability tables (CPT)*". In *Figure 2.1,* we can see *CPT* for all nodes is provided next to the relevant node.

In a *BN*, the joint probability distribution computed by multiplying these conditional probability tables by the Chain Rule. Let A_1, $A_2, \ldots. A_n$ be variables in the *BN*, the joint probability distribution can be found as in the formula 2.2:

$$P(A_1, A_2, \ldots. A_n) = \prod_{i=1}^{n} P(A_i \mid pa(A_i) \qquad \textbf{(2.2)}$$

where $pa(A_i)$ is the parents of A_i in the *BN*.

BNs can be used to answer any probability-based queries related to its variables, as it defines the joint probability distribution of its variables in *BNs*, it is used to find updated information about the status of other variables.

A commonly used method when inferencing in *BNs* is the *Junction Tree (JT*) Algorithm. According to this algorithm, *BN* is first converted into a tree structure. In the tree, each node defines a cluster of certain variables. *JT Algorithm* can calculate discrete BNs and the reader are referred to Lauritzen and Spiegelhalter (S. L. & D. J., 1988) for more detailed information.

2.3. Hybrid Bayesian Networks

Variables in *BNs* can be discrete or continuous, and until recently, the main limitation of BNs has been the continuous variables. Popular *BN* algorithms like variable elimination (Zhang and Poole, 1994) and JT (Lauritzen and Spiegelhalter, 1988) are designed for discrete variables. Therefore, most of the *BN* models developed in the project management literature contains only discrete variables. However, the use of continuous variables in BN models is important especially to model project time and cost in the project management domain.

BNs that consist of discrete and continuous variables together, they are named *the Hybrid BN (HBN).* In recent years, there have been improvements in algorithms to solve HBN including the *Dynamic Discretization (DD)* algorithm (Neil et al., 2007). In our study, *DD* algorithm will be used to solve *HBNs*. In the remainder of this chapter, a brief information about *DD* algorithm will be given.

DD algorithm works by minimizing the relative entropy among true and discretized marginal probability densities and thanks to this, iteratively discretizes the continuous variables in a *HBN* model and approximate relative entropy error for a discretized interval is calculated as follows:

$$E_j = \left[\frac{f_{max} - \bar{f}}{f_{max} - f_{min}} f_{min} \log\frac{f_{min}}{\bar{f}} + \frac{\bar{f} - f_{min}}{f_{max} - f_{min}} f_{max} \log\frac{f_{max}}{\bar{f}}\right] |w_j| \qquad \textbf{(2.3)}$$

where *Ej* is the approximate relative entropy error, f_{max}, f_{min}, $\bar{f}$ are respectively the maximum, minimum and mean density values for a given discretization interval w_j.

The *DD* algorithm increases the states by adding in high-density areas. In zero-density areas, it decreases the states by merging. *DD* algorithm discretizes every continuous variable in highest density area at every iteration. After that, it use a propagation algorithm such as *Junction Tree (JT)* algorithm for calculating posterior marginal. The *JT* calculates the posteriors in a discrete *BN* and turns *BN* structure to a tree structure using clusters. If a new evidence entered to the *BN*, the *DD* algorithm updates the all discretized continuous variables in *JT*.

In *DD* algorithm, the convergence threshold adjusts an upper bound relative entropy for stopping the algorithm. If the sum of total entropy errors for all intervals in a node is smaller than the convergence threshold, discretization stops. When the discretization has more states, the entropy error decreases. If the number of states approaches to infinity, the relative entropy approaches zero. The *DD* algorithm aims to calculate the optimal discretization for probabilistic distributions and their functions. *DD* has been implemented in the AgenaRisk software.

2.4. BN Models in Project and Risk Management

BNs are efficient tools to make probabilistic calculations. They are also give decision support for complex areas (Pearl, 1988, Fenton and Neil, 2012). Their graphical structure is applicable to model causal relations determined by experts or learned from data. Inference and learning algorithms discussed above enables them to compute Bayesian calculations and learn from large datasets efficiently.

BNs also provide a suitable infrastructure to help modeling data and expert knowledge in combined form. (Yet et al.,2014a ; 2014b). The expert knowledge is used for determining the graphical structure for constructed *BN* which contains the variables and causal relations between variables. Then, the data is used to learn the parameters of graphical structure. Dynamic *BNs* models are extensions of *BNs.* In the dynamic systems of time-varying and control systems, dynamic *BNs* models provide a suitable structure.(Murphy, 2002).

BNs offers a suitable modelling environment for analyzing risks and uncertainties in project management as they can represent complex probabilistic relations and incorporate the parameter uncertainty of their variables. Moreover, their ability to combine expert knowledge and data is an advantage for project evaluation. There is a wealth of knowledge in the field of project management, but because the projects are not like each other, the relevant data are often not available in large quantities. This section is revised of the previous *BN* studies in this area and risk management and examine their contributions and limitations.

Yet et al., (2016) designed a *BN* model which predict the efficiency of the project and performs risk analysis on it. The model is developed for selecting project and this model can calculate both uncertainty of parameter and causal risk factors for this task. However, this model assumes fixed project time and does not include project time estimates.

Khodakarami et al., (2007) used *BNs* for modeling and calculating the uncertainty of projects timelines. Khodakarami et al. built a *BN* of critical path method that also modeled the causes of delays. This model is designed to plan durations of project activities at micro level; it is not intended to be used for evaluating projects at macro level. Moreover, it does not analyze or control the costs of project activities.

Luu et al., (2009) has developed a *BN* structure that is discrete. This simple model estimates project delays for the construction sector. However, this model does not consist parameter uncertainty and does not consider the changes that are dynamic throughout the project.

Fineman et al., (2009) developed a version of *BN* for modeling "the trade-off" among cost, schedule and quality in the projects. Lee et al., (2009) designed a structure of *BN* model for ship building area that estimates the budget, time and insufficient requirements risks in the project. Khodakarami & Abdi (2014) developed models for cost analysis of projects by using BNs.

BNs have been more commonly applied for software engineering projects. Fan and Yu (2004) used a *BN* model which estimates and manages the risks at different stages of the software development process. Fenton et al., (2004) has developed a type of *BN* model to decide which software should be invested by considering the among quality, schedule and cost.

De Melo & Sanches (2008) has designed a *BN* model which estimates the delays in maintenance of software by using discrete *BNs*. Perkusich et al., (2015) identified the problematic processes in software development using *BNs*. Hu et al., (2013) used a model for analyzing risks in projects of software by using *BNs* with causality constraints.

3. PROJECT CONTROL METHODS

Risk and uncertainty are common features of all projects, and they may cause a project to be worse than planned in terms of cost, time or performance. For example, the "US Government Accountability Office Report (2010)" stated that 72% of the government's technology projects are under high budget overruns and delays (Mishra et al., 2016). When risk events related to budget or schedule can be detected, project manager can take actions to mitigate them. Otherwise, they may cause the project to fail completely. A major duty of the project management is to control a project for minimizing the risk of failure, and to make corrective decisions to mitigate unexpected events (Khamooshi and Golafshani, 2014). This requires effective measuring of project progress.

One of the widely used project control tools for measuring progress and predicting outcome is *Earned Value Management (EVM)* (Anbari, 2003; Fleming and Koppelman, 2002; Project Management Institute Inc, 2017). *EVM* evaluates the projects in dimensions of cost, scope and schedule *(Figure 3.1.)*, computes variances of performance and indexes for helping projects managers to notice over budget and lags. Additionally, *EVM* estimates the completion cost for the project and completion date (Pajares and López-Paredes, 2011).

This chapter provides an overview of *EVM* (Section 3.1) and its derivation *ESM* (Section 3.2) and discusses their limitations. The previous studies that aimed to incorporate uncertainty to *EVM* and *ESM* is reviewed (Section 3.3).

3.1. Earned Value Management

EVM controls the project progress in three main ways. These are *Planned Value (PV), Earned Value (EV)* and *Actual Cost (AC). PV* is the approved budget allocated to the planned work for the evaluated time period and it is also named the *Budgeted Cost of Work Scheduled (BCWS). EV* is the value of the confirmed budget that planned for the work done in the evaluated time period and it is also named the *Budgeted Cost of Work Performed (BCWP).* Finally, *A*C is the virtually cost for done work in the evaluated time period and it is also named *Actual Cost of Work Performed (ACWP).*

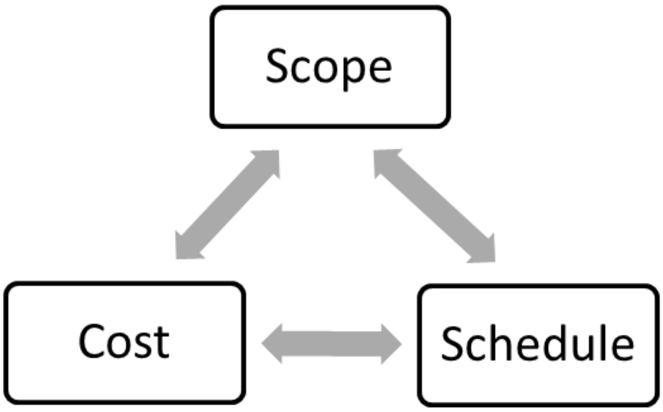


Figure 3.1. Project Triangle Model at EVM

EVM provides some variances and indices using the values described above to measure project performance. When we extract *AC* from *EV*, we find the *Cost Variance (CV)* and it means how much was actually spent through the work performed. In this case, *CV* is defined as:

$$CV = EV - AC \quad \textbf{(3.1)}$$

From here, a negative value means that, in the project money has spent more than planned for the work done in the evaluated time. Just the opposite, a positive value shows that money has spent less than planned for the work done. So, positive value of *CV* is preferred in *EVM*.

When we extract *PV* from *EV*, we find the *Schedule Variance (SV).* It shows whether the project schedule is as planned. In this case, *SV* is defined as:

$$SV = EV - PV \quad \textbf{(3.2)}$$

Similar to *CV*, a negative value of *SV* means that the project is behind the planned calendar and positive value shows that the project is ahead of planned calendar. However these two performance measures may not be reliable in some cases. When the critical activities are delayed, because their cost is part of other activities this delay does not change *CV* much. Similarly, when the project is over and the all activities are completed, *EV* and *PV* are equal and *SV* becomes zero value. This result does not indicate whether the activities exceed the scheduled time during the project.

There are other performance indices that measure the project efficiency. These include, *Cost Performance Index (CPI)* and it means cost efficiency of the budgetary resources. *CPI* is defined as:

$$CPI = EV/AC \qquad \textbf{(3.3)}$$

Other index is *Schedule Performance Index (SPI)* and it is measure of how efficiently the project team's time is used. *SPI* is defined as:

$$SPI = EV/PV \qquad \textbf{(3.4)}$$

EVM measures and indexes are shown graphically in *Figure 3.2.*

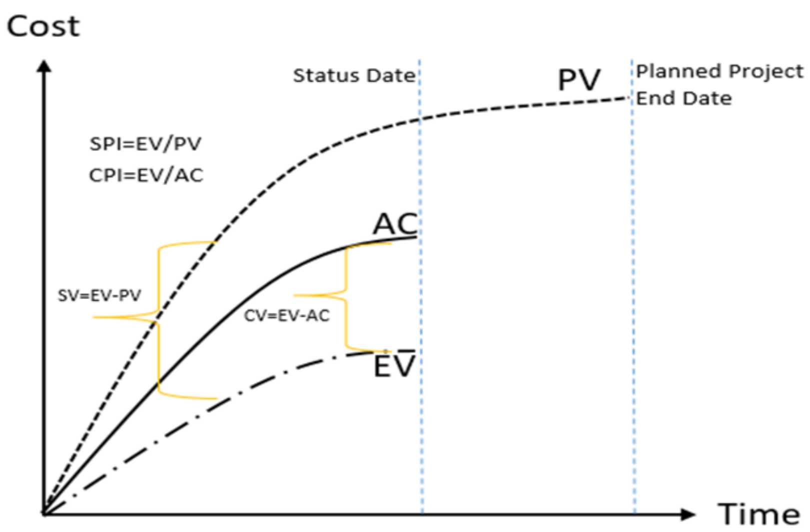


Figure 3.2. EVM measures and indices

3.1.1. Limitations of EVM

EVM estimates the completion time and cost for the project according to the *SPI* and *CPI* values. This method models the performance of the project for the future as a deterministic mathematical function based on the cost and time spent so far. Caron et al., (2016) likened this to driving a car only by looking to rearview mirror. This is not a realistic assumption to consider that events of future will look alike to the past events, especially in a risky and uncertain area such as project management. The events which cause troubles in the past, and the risk factors which may cause problems in the future should be considered to make more accurate estimates.

Therefore, in the projects that are risky and uncertain, the advantage of *EVM* is limited (Caron et al., 2016). The first reason for this is that the *EVM* largely ignores the uncertainty in the projects. When calculating the performance indices and variances in the *EVM*, point values of *EV*, *PV* and *AC* are used. Also the uncertainty of parameters is not considered. Naeni et al., (2011) said that although the progress of activities in a project is uncertain, these progresses are considered deterministic in all existing *EVM* techniques. Every project has some uncertainty because human decisions are also effective in the data of project. Knowing or estimating the point values of variables properly is really hard because of complexion and uncertainty of the projects. In the completed projects, determining the point value of *EV* is difficult for the managers of project. This limits the use of *EVM*. For example, in the projects of software, it is hard to specify percentage of completed work exactly. Pinto (2016) pointed out that specifying the exact *E*V value is not very possible in most of projects, and demonstrated that setting different values for *E*V by different assumptions come up with different results.

The second reason restricting the use of *EVM* is that it does not allow to model and analyze individual risk events. *EVM* shows the problems about cost or time to the project managers, but it does not provide any solution to make numerical analysis at problem reasons. In order to make the correct decision for the problems, project manager should examine whether the problem persists or not, whether it is caused by a structural problem or not. For doing this, the method of project control should also allow analysis at problems reasons.

The third reason is that *EVM* does not provide guidance for specifying the *EVM* parameters. When the projects are not similar and have limited data, we need a project control model which use both expert knowledge and numerical data.

3.2. Earned Schedule Method with EVM Relations

There have been some studies to overcome the limitations of the *EVM*. Lipke (2003) observed that the time indices and deviations in the *EVM* did not make good measurements in the final stages of the project and suggested a "*Earned Schedule Method (ESM)"* to better measure the time performance. Schedule analysis in *EVM* is less comprehensive and consistent than cost analysis. Even if some activities are finished after the planned time, the *SPI* of all finished activities is equal to 1, and in this case *SPI* gives incorrect results when evaluating the schedule performance of these activities. The *ESM* is an extension to *EVM* aiming to overcome this issue. The *ESM* uses time units instead of cost units for measuring schedule performance. The time point of the *PV* value which is equal to *EV* value at the given time is used for *ESM*. Shortly, *ESM* finds the time the *EV* should actually be, the time actually scheduled for that *EV*. *Figure 3.3.* shows this equality state on *ESM* with *EVM*.

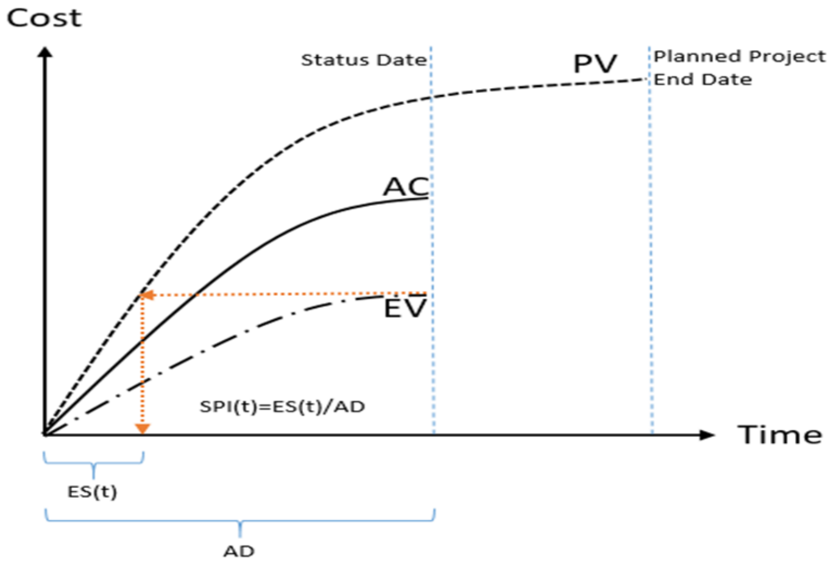


Figure 3.3. ESM graph with EVM

ES of a project is calculated as (Khamooshi & Golafshani, 2014, Henderson & Lipke, 2006) :

$$ES(t) = t' + \frac{EV - PV_{t'}}{PV_{t'+1} - PV_{t'}} \qquad \textbf{(3.5)}$$

where t' is such that $EV \geq PV_{t'}$ and $EV < PV_{t'+1}$. Then, *ESM* defines *SV* and *SPI* like *SV(t)* and *SPI(t)* as:

$$SV(t) = ES(t) - AD \qquad \textbf{(3.6)}$$

$$SPI(t) = ES(t)/AD \qquad \textbf{(3.7)}$$

where *AD* is the *Actual Duration* that we evaluated.

EVM and *ESM* which derived from *EVM* are widely used and evaluates projects based on schedule and cost performance. These methods can make schedule and cost estimates about future of the project. After *PV, EV* and *AC* determined, the performance indexes can be calculated. *Budget at Completion* (*BAC)* is total planned project budget before starting the project. *Estimate to Complete* (*ETC)* represents the cost needed of completion for the remaining work if the project performance continues as it is. It is calculated rate of difference *BAC-EV* and *CPI. Estimate at Completion (EAC)* represents total needed cost for the completion of project if the project performance continues as it is. It is calculated as sum of *ETC* and *AC. To Complete Performance Index (TCPI)* expresses the performance rate required for completing the remaining work within the planned budget. It is computed as rate of difference *BAC-EV* to difference *EAC-AC*.

ES(t) represents the time the *EV* should actually be, the time actually scheduled for that *EV. AD* means the time spent on project. *AD* in *ESM* is similar *AC* in *EVM*. Planned Duration (*PD)* means planned time for the project. *PD* in *ESM* is similar *PV* in *ESM*.

SV(t) is variance of schedule like *SV* in *EVM. SV(t)* is calculated difference between *ES(t)* and *AD. SPI(t)* is performance index of schedule like *SPI* in *EVM. SPI(t)* is calculated rate of *ES(t)* to *AD*. In *ESM* there is no cost variance and performance index because *ESM* is based on time, not cost. This is explained in detail Section 3.2. Planned Duration for Work Remaining) *PDWR* means the planned time for the completion of remaining work of project. It is similar to *ETC* in *EVM* and calculated difference between *PD* and *ES*. Independent Estimate of Duration at Completion (*IEDAC)* means prediction of total project duration from present schedule status.

3.3. Previous EVM and ESM Studies

Several previous studies that aimed to incorporate uncertainty into *EVM* and *ESM* by using statistical approaches. Earlier studies aimed to integrate *EVM* and *ESM* with statistical process control techniques. Anbari (2003) proposed "*Critical Ratio (CR)"* which is product of the *SPI* and the *CPI* as a project control index. It has been suggested that the *SPI, CPI* and the newly proposed index *(CR)* should be applied by applying statistical process control principles. Colin and Vanhoucke (2014) suggested a new statistical project control approach based on *EVM/ES* by setting tolerance limits for *EVM/ES* metrics in order to examine the possibility of the ongoing status of the project is less than or equal to the planned schedule.

Forecasting techniques and various modelling approaches have been combined with *EVM*. Batselier and Vanhoucke (2017) defined an approach named "*XSM"* which is an acronym for "eXponential Smoothing-based Method". In the *XSM, EVM* metrics are incorporated into the exponential smoothing formulas.

Narbaev and De Marco (2014) suggested a different "*CEAC(Cost Estimate at Completion)"* methodology based on a modified index-based *EVM*. The new *CEAC* formula is integrated into the Gompertz growth model *(GGM)* for doing better prediction for future of project. *GGM* (Gompertz, 1825) describes the growth, generally animals and plants.

Naeni et al. (2011), proposed a fuzzy-based *EV* approach which aims the improving and evaluating the *EV* indices and doing estimations of cost and time under uncertainty. In fuzzy-based *EVM*, planned values of works are defined as linguistic variables and transformed into fuzzy numbers when total work required to finish activities are uncertain and out of control. For this, one membership function has to be defined for expressing the relations among the "linguistic term" and "fuzzy numbers". It can be like "very low", "low", "medium", "high" and "very high".

"Monte Carlo Simulation" has also been widely used to incorporate risk and uncertainty in *EVM*. Pajares and López-Paredes (2011), suggested integrating the uncertainty to *EVM* with using Monte Carlo simulation for computing statistical distribution of project cost and duration for finishing project. The authors defined two buffers which are

"*CPBf (the Cost Project Buffer)"* and "*SPBf (the Schedule Project Buffer)"*. The *CPBf* is difference between maximum project cost at the confidence level which satisfies that the probability of the project cost to be lower than the maximum cost and the mean project cost. In the same way the *SPBf* is difference between maximum project duration at the confidence level which satisfies that the probability of the project duration to be lower than the maximum duration and the mean project time. Project manager decides the confidence levels. By using Monte Carlo simulation, maximum cost and duration at confidence level are found. They also defined the project risk baseline for schedule *(SRB)* and for cost *(CRB).* And computed the risk reduction at time t for finding weights "($wc_t = CRB_t\text{-}1 - CRB_t$ and $ws_t = SRB_t\text{-}1 - SRB_t$)". For determining the size of buffer for every period, total buffers are split proportionally to the risk reduction during that time interval such as:

$$CBf_t = wc_t * CPB_f/\sigma^2{}_{pc} \quad \textbf{(3.8)}$$

$$SBf_t = ws_t * SPB_f/\sigma^2{}_{ps} \quad \textbf{(3.9)}$$

where σ^2_{pc} and σ^2_{ps} are total project cost and schedule variances computed by Monte Carlo. After that *Accumulative Cost Buffer* ($ACBf_t = CBf_t + ACBf_t\text{-}1$) and *Accumulative Schedule Buffer* ($ASBf_t = SBf_t + ASBf_t\text{-}1$) are computed at time t. The authors suggested two new metrics for *EVM*: *The Cost Control Index* ($CCoI_t = ACBf_t{=}ES + CV$) and the *Schedule Control Index* ($SCoI_t = ASBf_t + SV$) . They combined *EVM* and Risk Management to analyse the project over-runs under risk condition and performed Monte Carlo analysis with 100,000 simulations.

Acebes et al. (2014), used Monte Carlo simulation to obtain the distribution of possible project realizations. They aimed to know cost and time deviations are planned or not while project is running. The authors defined triads "x, T_{xj}, C_{xj} where x is the percentage of completion, $C_{xj} = x*C_j$ is the money spent at *comletion x%*, T_{xj} is the time when cost C_{xj} is spent for every possible project j (every simulation) and C_j is the total cost of project for simulation j". These triads are splitted into two graphs which are "*time vs. x*" and "*x vs. cost*". For every realization of a Monte Carlo simulation, the final cost and duration are provided. If thousands of project simulations are considered, an area of possible cost and time values is computed in this graph. After that, distributions of cost and time are determined confidence interval for time and cost can be calculated.

Acebes et al.,(2015) developed a slightly different method based on triad method described above. They made a change in the triad variables and used the terms "(*EV, t, c*)" instead of "*(%, t, c*)". This was considered more useful as calculating the accurate percentage of completion is more difficult than the *EV*. Similar to their previous study, Acebes et al. used Monte Carlo simulation for generating realizations of the project and making expectation about future of the project. Simulations are used to detect if project deviations are like planned or not using anomaly detection algorithms. Anomaly detection is a technique used for identifying data points, items, observations or events that do not conform to the expected pattern of a given group and also known as novelty detection or outlier detection (Ding et al.,2014). They used a probabilistic approach- multivariate density estimation- to deal novelty techniques whose aim is identifying the data that are not consistent in normal expectations. Miljković, (2016) explained novelty techniques in detail.

Finally, a few studies aimed to estimate the confidence intervals of *EVM* parameters for specific statistical distributions. Lipke et al. (2009) used the lognormal distribution to determine the confidence interval of performance indices in *EVM*. Similarly, Caron et al.(2013), modeled these indices with a lognormal distribution and used Bayesian model and Gibbs sampling method to update these distributions with the observed new data. Both Lipke et al. and Caron et al. used only lognormal distribution and have not suggested solutions for situations where indices are distributed with other distributions. Moreover, they set the confidence interval only for indices. However, *SPI* and *CPI* are deterministic functions of *EV, PV* and *AC*. Therefore, modeling the *EVM* variables with statistical distributions and calculating the confidence intervals and distributions of the indices according to these variables will facilitate the determination of parameter distributions and provide a more understandable modeling approach. Lipke et al. and Caron et al. did not use the individual risk factors that might affect *EVM* parameters and indices. In the following chapter, we present a novel approach that uses *BNs* to overcome these issues. Our model offers a flexible framework to incorporate risk factors and uncertainty with the *EVM* approach.

4. METHOD

4.1. Model Framework

In this study, a hybrid *BN* model that includes both continuous and discrete variables is used for developing a project control tool. We used *EVM* structure and modeled the *EVM* metrics which are *PV*, *EV*, *AC*, *SPI*, *CPI* and *ES* in this model. The relations between these metrics at the same and different time steps and their parameter uncertainties are modeled as reusable model fragments, also called as idiom.

Idioms are basic and fragmented structures of object-oriented *BNs*. Neil et al.,(2000) proposed idioms for modeling large and complex problems by *BNs*. When there are repeated patterns of probabilistic reasoning in a problem, the related idiom can be used repeatedly.

Our idiom can be used in the project of all areas. *Figure 4.1.* shows the structure of our idiom. In our idiom, the relations between the variables are mathematically as follows:

$$EV = PV * Completion\,\% \tag{4.1}$$

$$SPI = EV / PV \tag{4.2}$$

$$CPI = EV / AC \tag{4.3}$$

PV is determined before starting project, *Completion %* and *AC* are determined at the time of evaluation of project progress by project managers. After these, the posterior distributions of other metrics are calculated by using the *BN*.

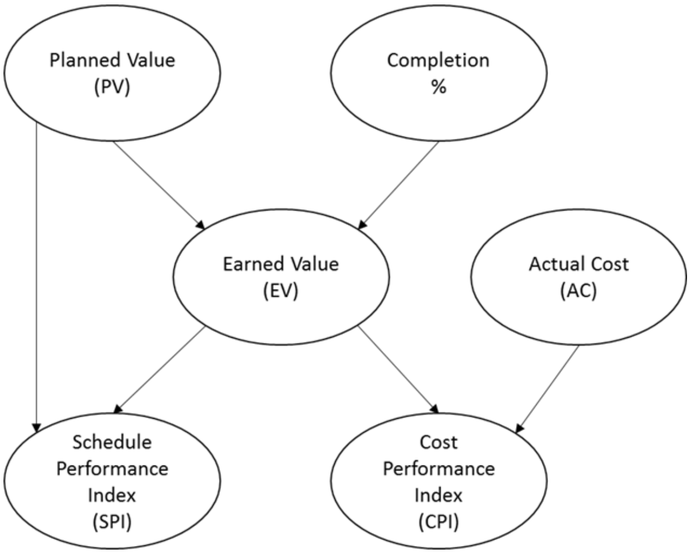


Figure 4.1. Structure of idiom

We modeled and run our framework in AgenaRisk program. AgenaRisk program calculates with Dynamic Discretization and Junction Tree algorithms. We can evaluate different risk factors and use different distributions in the model. *Figures 4.2.* and *4.3*. show brief information about AgenaRisk program. *Figure 4.2.* shows the modelling interface of AgenaRisk. Each *BN* fragment can be modelled as an risk object shown at the left side of the interface. The content of a *BN* fragment is shown on the main part of the interface. Continuous nodes are defined as ‘simulation nodes’ and nodes with ordinal states can be defined as ‘ranked nodes’. *Figure 4.3.* shows how probability distributions can be defined for continuous ‘simulation’ nodes. Note that, a continuous node can be conditioned on a discrete node, and a mixture of continuous distribution can be defined. For example, in *Figure 4.3.* ‘Actual Cost’ is conditioned on ‘Risk 3’, and the user has defined a mixture of Normal distributions.

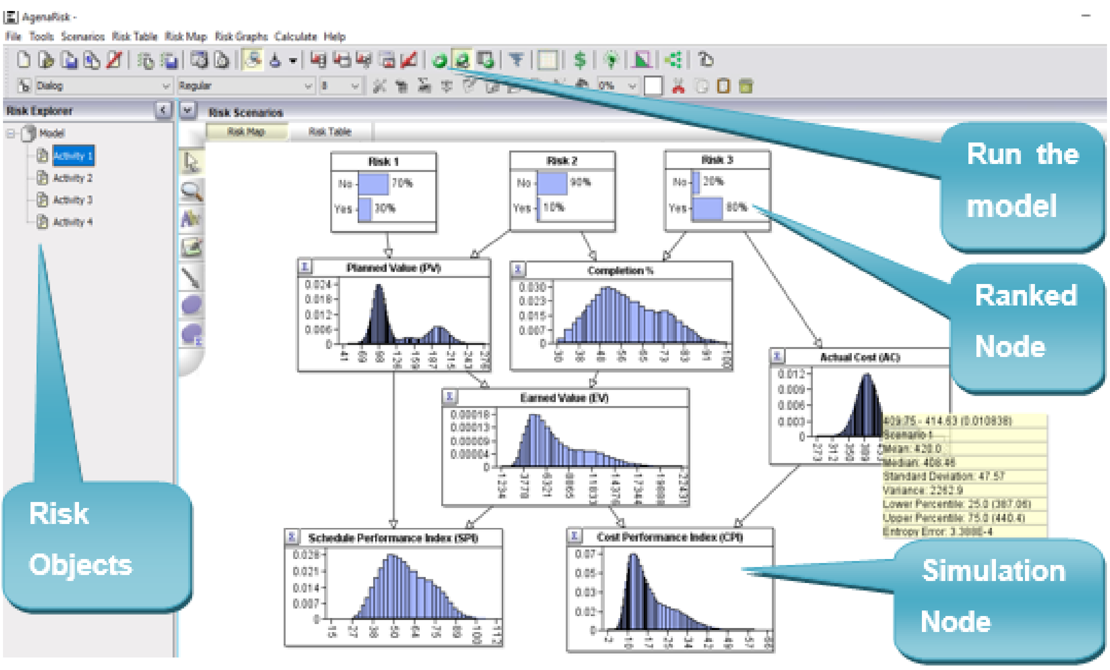


Figure 4.2. Risk Objects of AgenaRisk

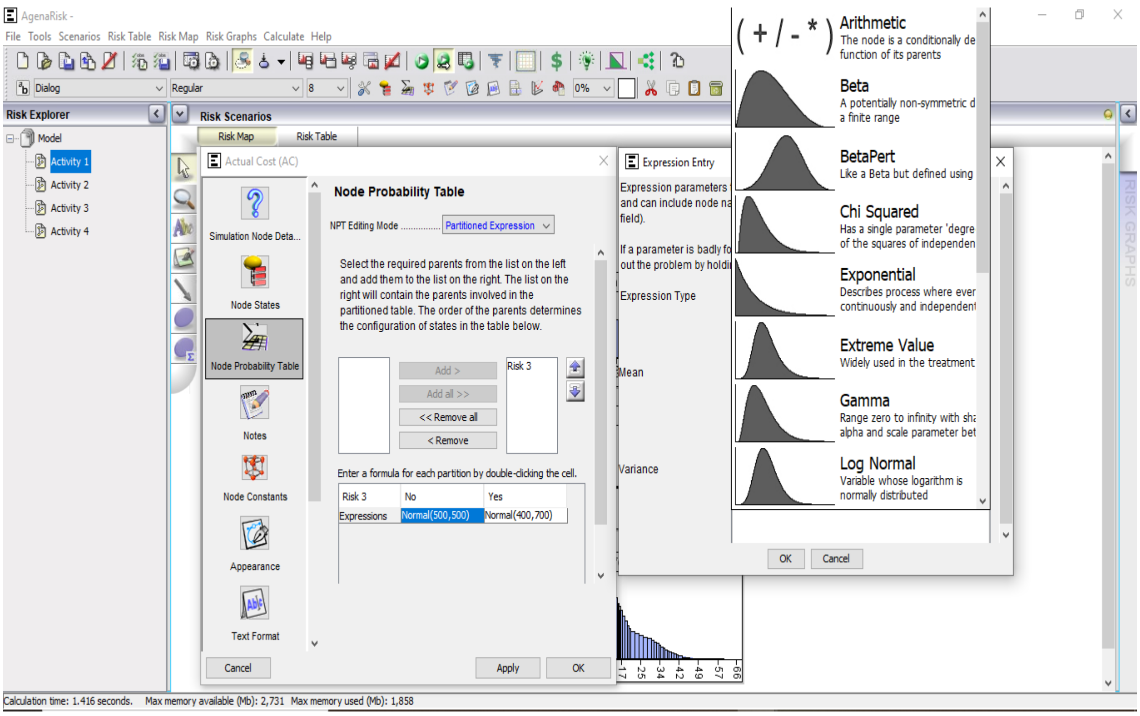


Figure 4.3. Defining Distributions for variables

4.2. Determining Distributions of Metrics

Statistical distributions of parameter uncertainties for the *PV* and *Completion %* parameters are need to be determined in the idiom. Project manager decide which distribution is suitable for the *PV* before start of the project. Distributions of *Completion %* and *AC* are determined at the time of evaluation of project performance again by project manager. While determining distributions of Completion %, project manager should choose a distribution between 0 (zero) and 1 (one). Beta distribution can be given as an example for *Completion %*. *PV*, *EV* and *AC* must be bigger than 0 (zero), so attention should be paid to this while determining distribution. *Figure 4.4.* shows the structure of model with distributions.

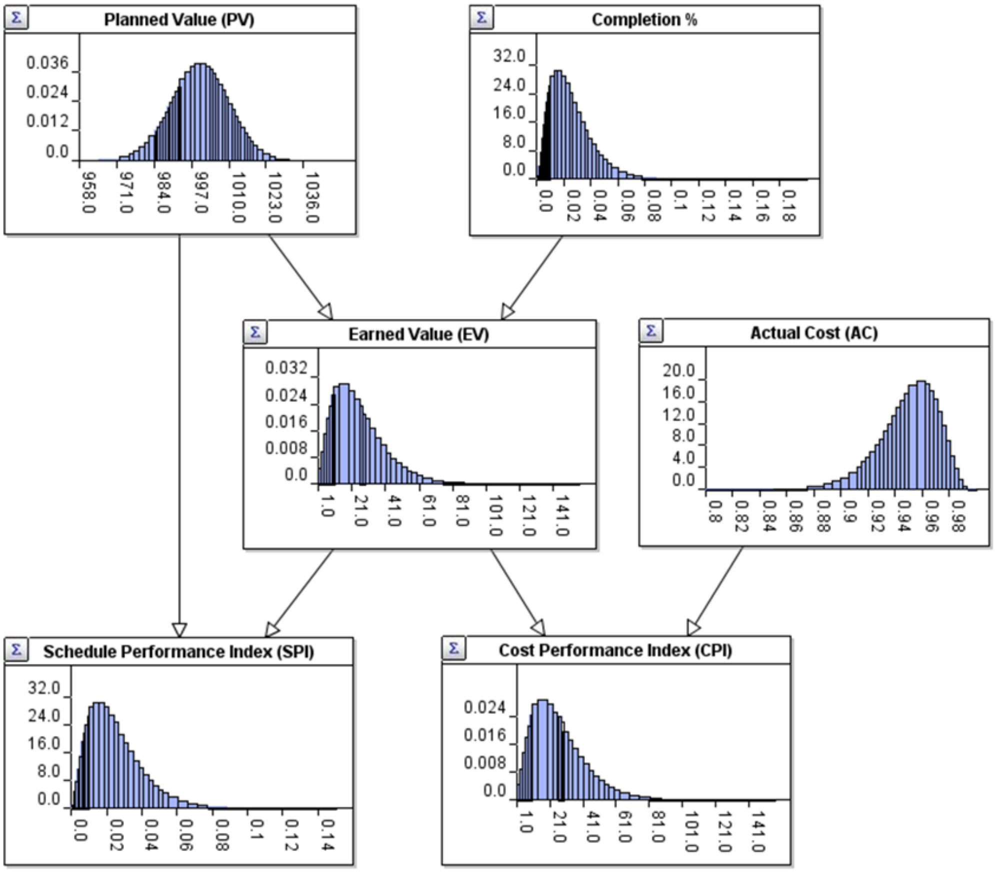


Figure 4.4. Structure of model with distributions

4.3. Determining Risk Factors

After determining of suitable distributions, the risk factors that can affect the project performance and their relationship with the variables of the *EVM* are determined. In our model, risk factors are connected to the desired *EVM* variables and effect of their distributions. *Figure 4.5.* shows an example that how risk factors are connected and affect the variables of *EVM*. We define risk factors and their parameters. Then we determine possibility of parameters and how effect the distributions of variables of *EVM*. Parameters of risk factors are discrete variables in general, but depending on the type of projects, these parameters can be continuous. In this study, we ignore the situation of continuous risk parameters and built our model with discrete risk parameters. We can connect risk factors to multiple parameters. Risk factors are determined mostly by the expert or project manager.

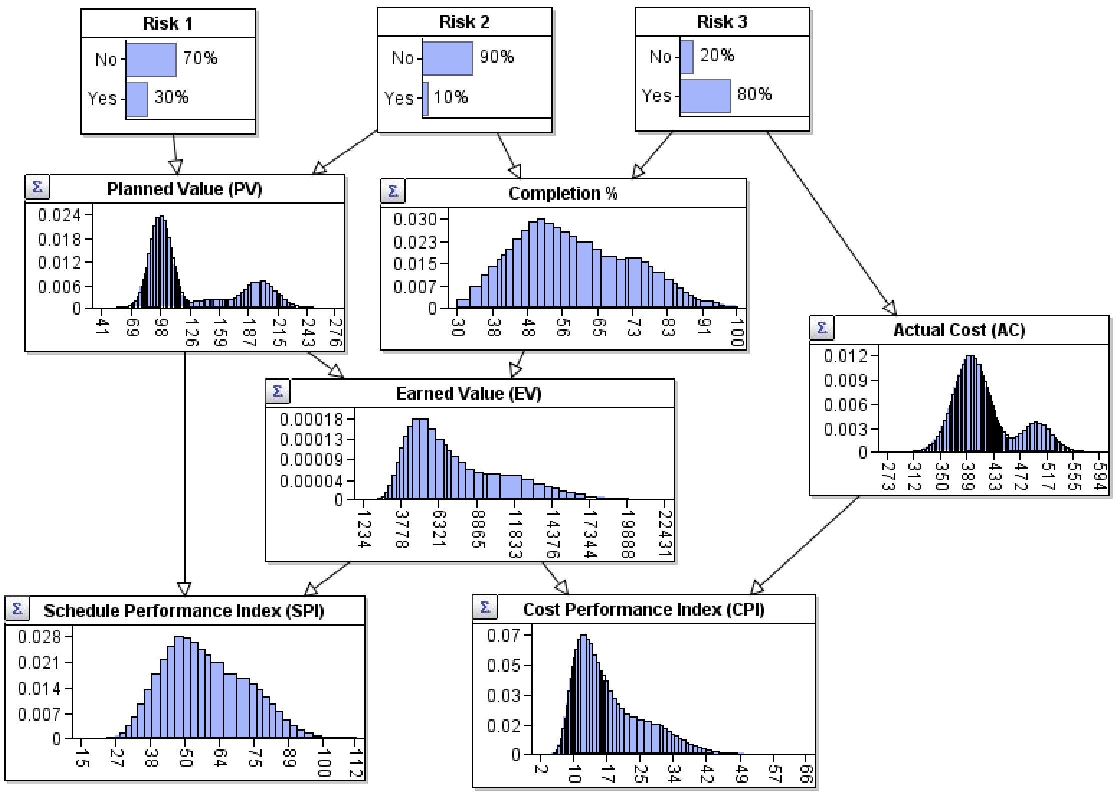


Figure 4.5. Model with Risk Factors

4.4. Combined BN Structure of Activities

The same idiom model with the risk factors shown in Section 4.3. is used for every activity of project. While building the whole structure of project, we built Total *PV* metric for each activity for total project time with planned time intervals like months, days etc. After that, at the time of evaluation of project performance, we add another PV metric for the evaluation time. *PV, % Complete* and *AC* metrics determined by the project team, other metrics (*EV, SPI, CPI*) automatically calculated. After building structure for each activity, we create an object which contains all values. We can explain this subject with a small example for understanding clearly.

We have 3 activities which are Activity 1, Activity 2 and Activity 3. The project total time is 3 months, every activity is also for 3 months and we planned evaluate the project for months. For all of them, we built the idiom models like *Figure 4.6.,4.7.* and *4.8.*

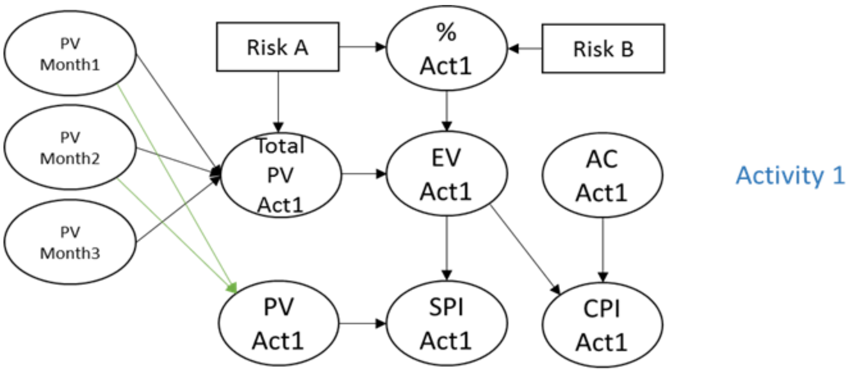


Figure 4.6. The idiom structure of Activity 1

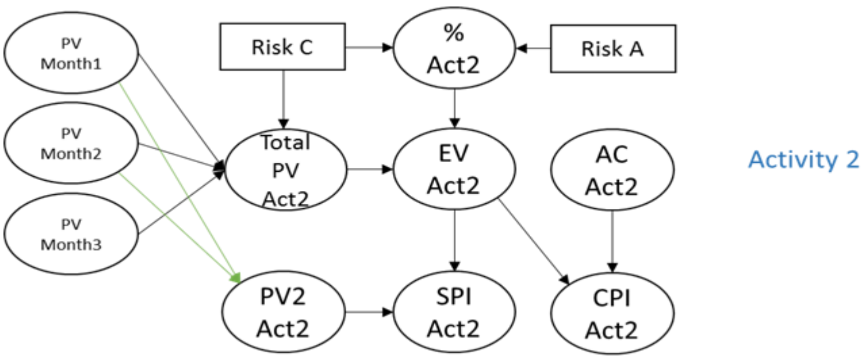


Figure 4.7. The idiom structure of Activity 2

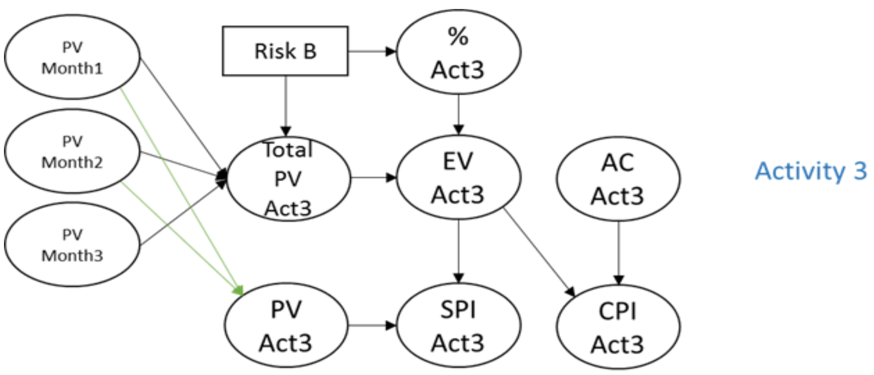


Figure 4.8. The idiom structure of Activity 3

Suppose we evaluate the project progress after 2 months from the project start. Therefore, *PV_Act1*, *PV_Act2* and *PV_Act3* values are the sum of *PV_Month1* and *PV_Month2* for each activity. For the idioms shown figures above, the risk factor can be different and can affect different metrics. After determining this structure, we create an object for all activities as shown in *Figure 4.9.*

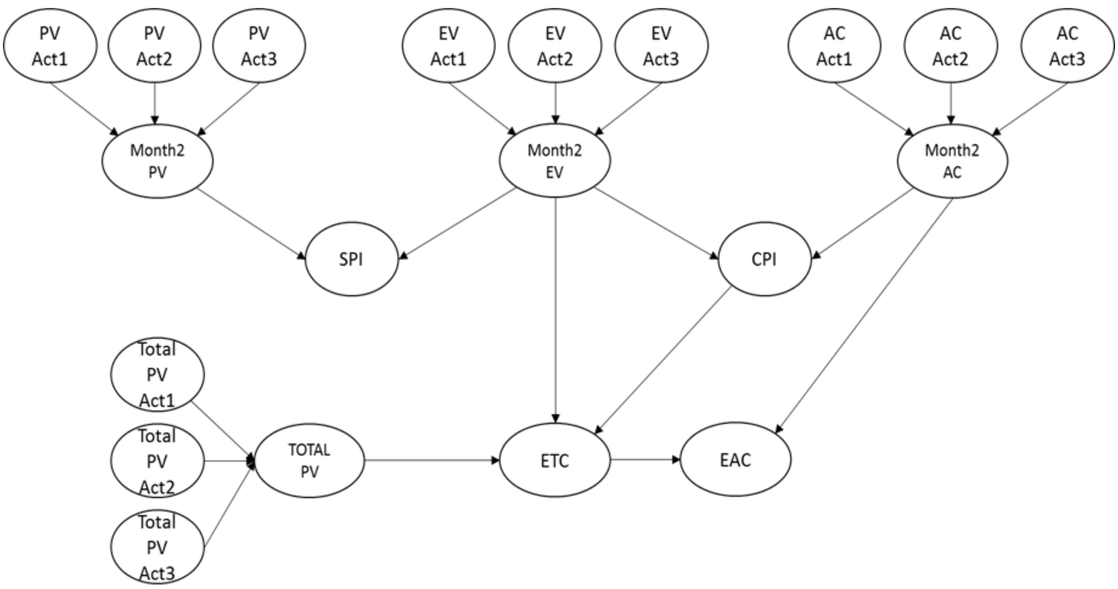


Figure 4.9. EVM object structure

AgenaRisk allows linking between different metrics. In the *Figure 4.9.; PV_Act1, PV_Act2, PV_Act3, EV_Act1, EV_Act2, EV_Act3, AC_Act1, AC_Act2, AC_Act3, Total PV_Act1, Total PV_Act2* and *Total PV* are export from the idioms. *Month2_PV, Month2_EV and Month2_AC* are the sum of values which connected to them. In the same way, *TOTAL PV* are the sum of values which connected to it. *SPI* and *CPI* are calculated by the formula defined to AgenaRisk which was shown in Section 4.1. *ETC* and *EAC* are calculated according to the formula mentioned in Section 3.2. These mathematical equations are defined as the parameters of the relevant variables in AgenaRisk. Once the AgenaRisk model is calculated based on the Dynamic Discretization algorithm, the posterior distributions of the performance indexes and project budget and duration predictions can be obtained. In the following section, we will apply our methodology to three different case and evaluate the results.

5. CASE STUDIES

5.1. Case 1

The first case study is a simple ‘toy-example’ about a research study. This study is planned for 6 months.

5.1.1. Data of Case 1

The project is being evaluated at the end of April. There are four activities in the plan and planned values of each activity are given in hours. *Table 5.1.* shows total and monthly planned values for each activity.

Table 5.1. Planned case study (in hours)

ACTIVITIES	JAN	FEB	MAR	APR	MAY	JUN	TOTAL
Literature Research	80	60	60	50			250
Modelling of Structure		50	60	30	10		150
Case Study				20	80		100
Analysis and Evaluation of results				30	30	20	80
TOTAL	80	110	120	130	120	20	**580**

The planned values in *Table 5.1.* are certain values but in real life it is almost impossible to make a point estimate of planned value of any work as discussed in Chapters 3 and 4. Therefore, we have applied our model to this example.

We have four different risk factors that may affect these activities, which are “*Exams”, “Technical Errors”, “Insufficiency of data”* and *“Unexpected results*”. “*Exams”* effects all activities. “*Technical errors*” effects only activity of “*Modelling of Structure*”. “*Insufficiency of data*” effects only activity of “*Case Study*”. “*Unexpected results*” effects only activity of “*Analysis and Evaluation of results*”. The probabilities of these activities are shown in *Table 5.2*. We used Beta distribution for activity of “*Literature Research”,* Normal distribution for activities of “*Modelling of Structure”* and “*Case*

Study", TNormal distribution for activity of "*Analysis and Evaluation of results".* The probability distributions of PVs of project activities conditioned on these risk events are also shown in *Table 5.2*.

The completion percentages of project activities are also modelled by using Beta distributions as shown in *Table 5.3*. Finally ACs regarding this project by the end of April is shown in *Table 5.3.*

Table 5.2. Distributions and risks of activities

ACTIVITIES	**RISK FACTORS AND PROBABILITIES**	**STATUS OF RISKS**	**PV EXPRESSIONS OF ACTIVITIES**
Literature Research	Exams (Yes:0.8, No:0.2)	Exams: Yes	Beta (280,310,0,500)
		Exams: No	Beta (250,300,0,550)
Modelling of Structure	Exams (Yes:0.8, No:0.2) Technical Errors (Yes:0.3, No:0.7)	Exams: Yes Technical Errors: Yes	Normal (200,100)
		Exams: Yes Technical Errors: No	Normal (160,100)
		Exams: No Technical Errors: Yes	Normal (170,100)
		Exams: No Technical Errors: No	Normal (150,100)
Case Study	Exams (Yes:0.8, No:0.2) Insufficiency of data (Yes:0.4, No:0.6)	Exams: Yes Insuffiency of data: Yes	Normal (220,100)
		Exams: Yes Insuffiency of data: No	Normal (190,100)
		Exams: No Insuffiency of data: Yes	Normal (200,100)
		Exams: No Insuffiency of data: No	Normal (180,100)
Analysis and Evaluation of results	Exams (Yes:0.8, No:0.2) Unexpected results (Yes:0.3, No:0.7)	Exams: Yes Unexpected Results: Yes	TNormal (120,100,1,150)
		Exams: Yes Unexpected Results: No	TNormal (90,100,1,120)
		Exams: No Unexpected Results: Yes	TNormal (100,100,1,120)
		Exams: No Unexpected Results: No	TNormal (80,100,1,120)

Table 5.3. Completion data

ACTIVITIES	Actual Cost by the end of April	Completion Percentage by the end of April
Literature Research	230 hours	Beta (90,10)
Modelling of Structure	100 hours	Beta (70,30)
Case Study	35 hours	Beta (20,80)
Analysis and Evaluation of results	30 hours	Beta (40,60)

5.1.2. Model of Case 1

The *Figures 5.1, 5.2, 5.3* and *5.4* show the distributions of variables of *EVM* for each activity. In this case study, we simulated the scenario where there was not "Insuffiency of data" for the activity "Case Study" and the presence of other risk factors were unknown. We defined the distributions of *Planned Value (PV)* and *Completion %* based on *Table 5.2* and *5.3* in this model. The posterior distributions of other variables were calculated by our model. There are total *PV* and monthly *PV* for each activity. Because we evaluate the performance at the end of April, there are *PV* for every activity for end of April. We defined the *AC* for every activity for end of April. In the case, we connect the risk factors that we determined and identified the expressions according to risk options which are Yes and No. When we run the model, we found *EV, SPI* and *CPI* for every activity for end of April. We will discussed the results in section 5.1.3.

In Figure 5.1., the model for activity of "*Literature Research*" is shown. This activity takes five months so, there are five month and total planned value distributions. The risk of *Exams* effects total *PV* of activity. *PV1* shows that *PV* for end of April because the evaluation is end of April. As a result, we have *SPI* and *CPI* for activity of *Literature Research*.

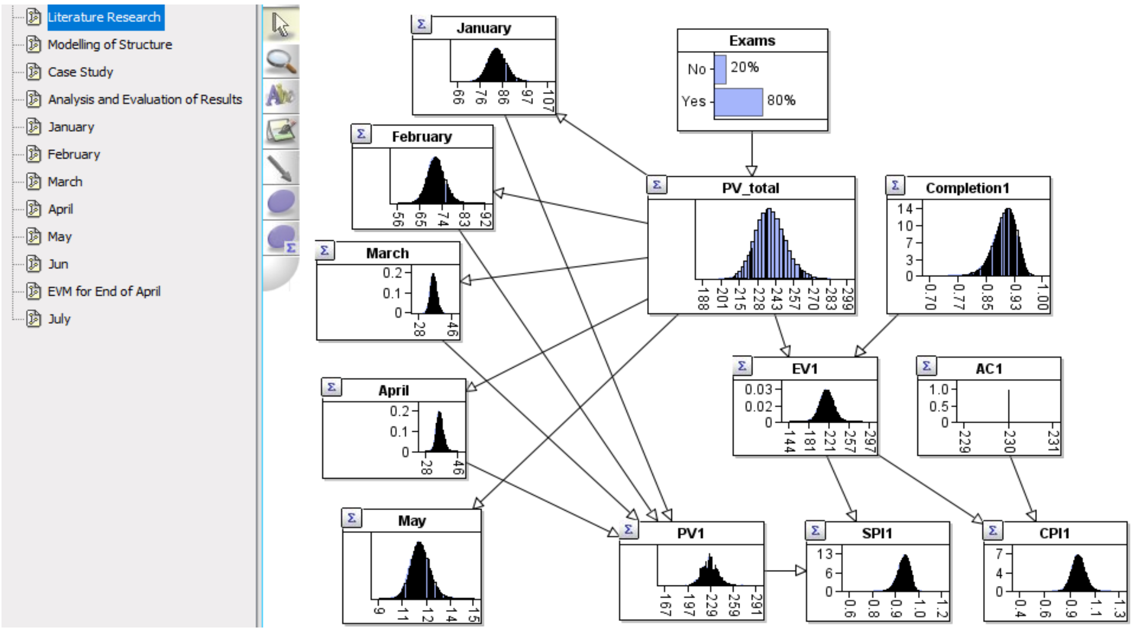


Figure 5.1. AgenaRisk Model of Literature Research activity

Figure 5.2. shows the model for activity of "*Modelling of Structure*". This activity takes four months so, there are four month and total planned value distributions. The risk of *Exams* and *Technical Errors* effects total *PV* of activity. *PV2* shows that PV for end of April. As a result, we have *SPI* and *CPI* for activity of *Modelling of Structure*.

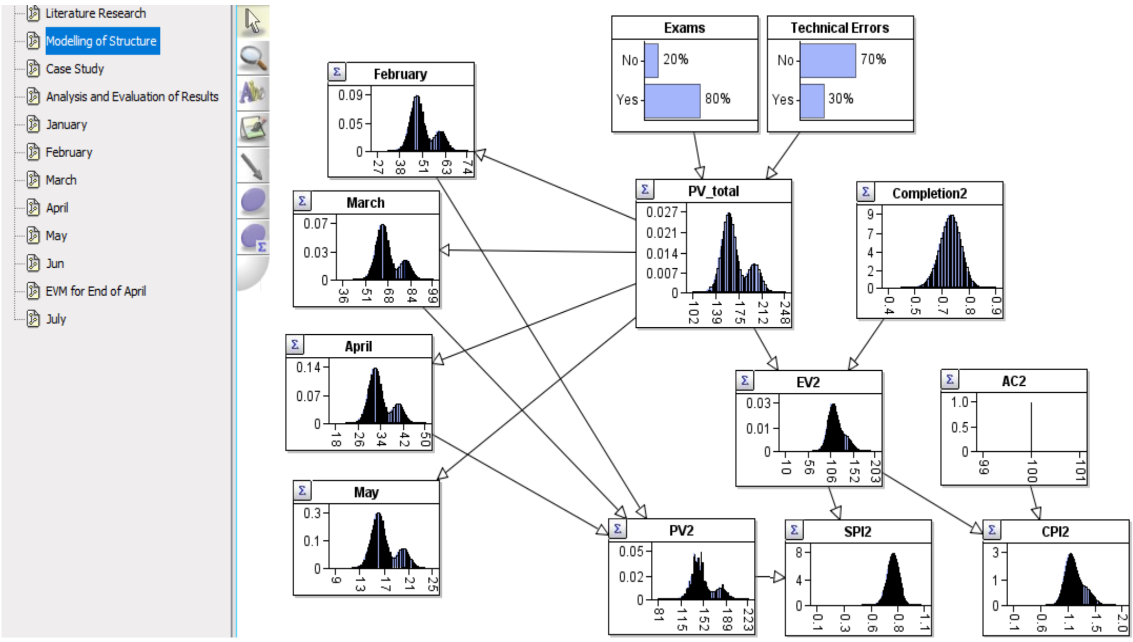


Figure 5.2. AgenaRisk Model of Modelling of Structure activity

Figure 5.3. shows the model for activity of "*Case Study*". This activity takes three months so, there are three months and total planned value distributions. The risk of *Insufficiency of data* and *Exams* effects total *PV* of activity. Here we simulated that the risk of *Insufficiency of data* did not happened. So we observed it 100% for option of *No*. *PV3* shows that *PV* for end of April. As a result, we have *SPI* and *CPI* for activity of *Case Study*.

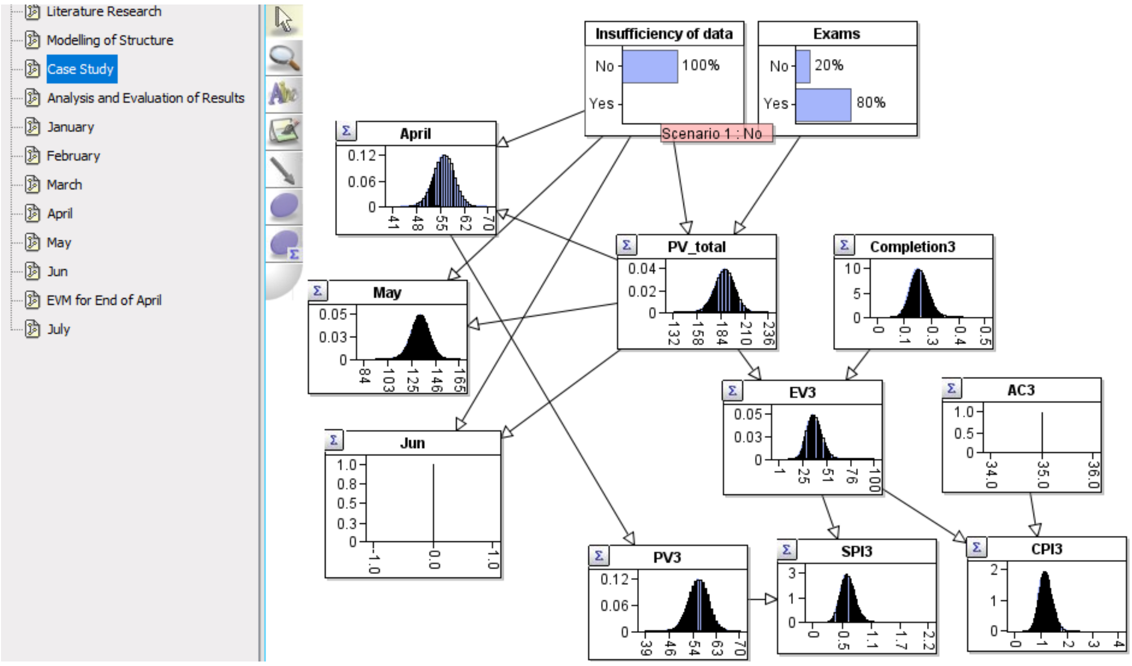


Figure 5.3. AgenaRisk Model of Case Study activity

Figure 5.4. shows the model for activity of "*Analysis and Evaluation of Results*". This activity takes three months so, there are three month and total planned value distributions. The risk of *Exams* and *Unexpected results* effects total *PV* of activity. *PV2* shows that PV for end of April. As a result, we have *SPI* and *CPI* for activity of *Analysis and Evaluation of Results*. Here we add one more month after June. This was made against the possibility of continuous the activity to July due to mishaps that were not taken into account.

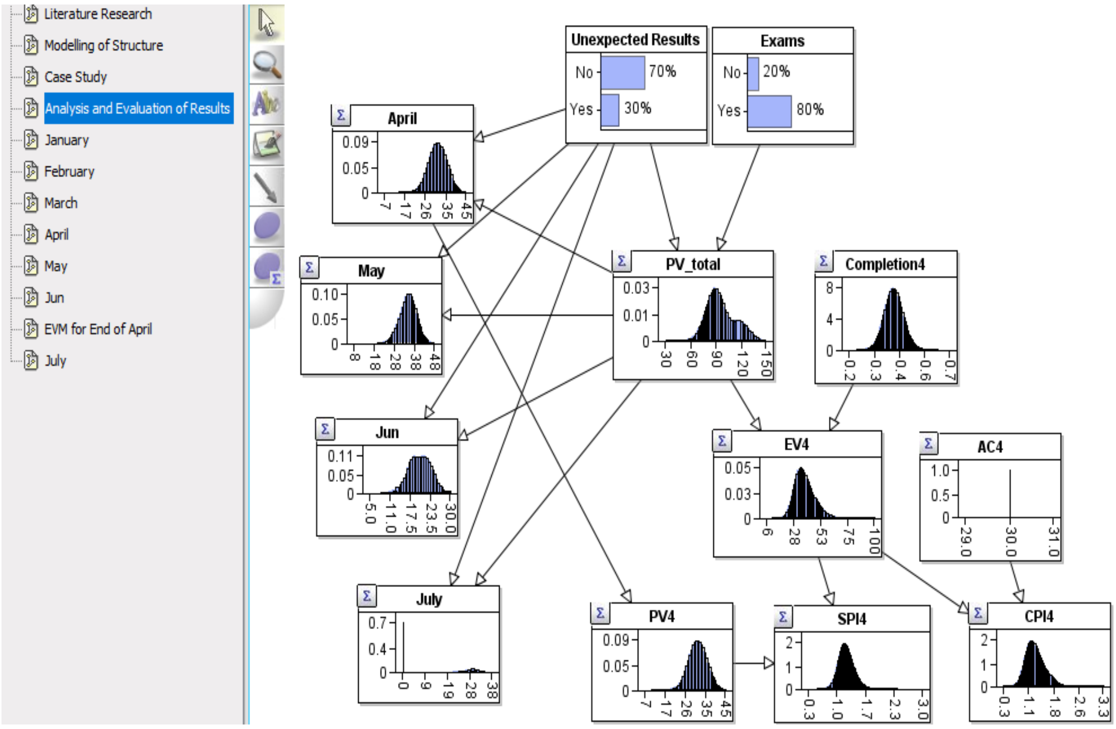


Figure 5.4. AgenaRisk Model of Analysis and Evaluation of Results activity

Figure 5.5. shows the *EVM* object for case 1. Here all planned values, earned values and actual costs of each activities end of April are transferred and summed up. Also total *PV* of activities are summed up. As a result we have performance indexes which are *SPI* and *CPI*, estimations which are *ETC* and *EAC* which are calculated by the model.

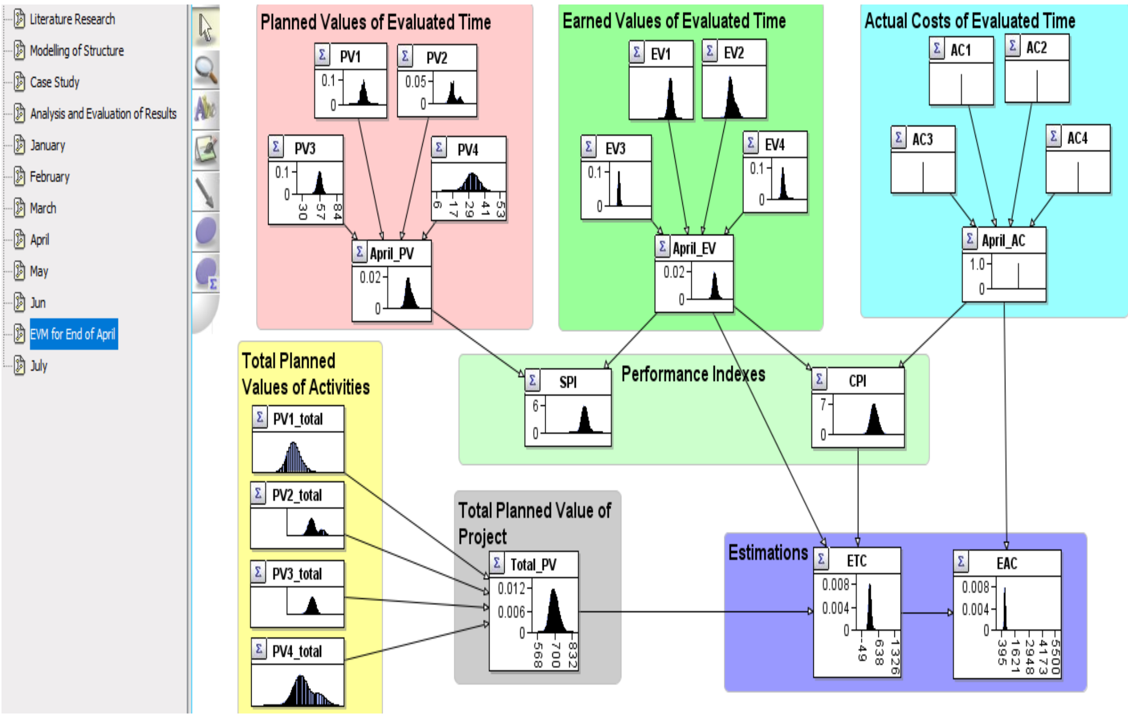


Figure 5.5. EVM object of Case 1

Figure 5.6. is the map of project model. It shows which value is transferred from where to where. We linked the interested values for transforming and we can see that how links are built.

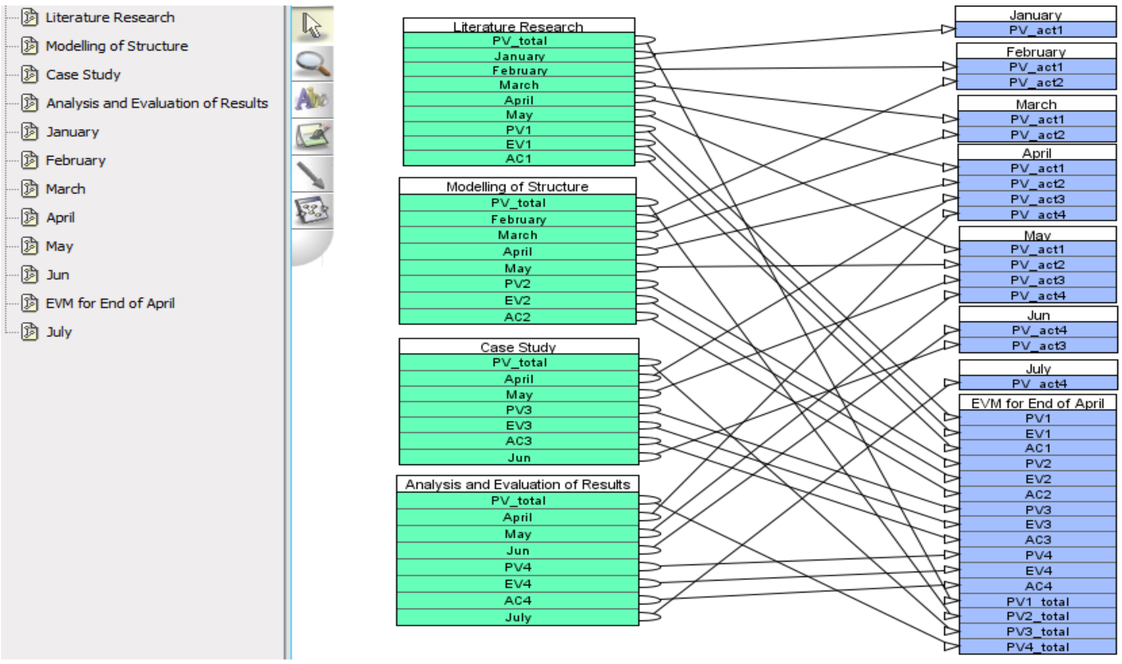


Figure 5.6. Map of all model for Case 1

5.1.3. Result of Case 1

After run the model, the results for end of April are shown below. Here we can see the performance of every activity in *Table 5.4.* and performance of all project in *Table 5.5.*

In the *Table 5.4*, we can say that we are behind the schedule for activities Literature Research, Modelling of Structure and Case Study at the end of April because *SPI* values of them are smaller than 1. On the other hand, we are ahead of schedule for Analysis and Evaluation of results because *SPI* value is bigger than 1. When we look at *CPI* values, except Literature Research, all *CPI* values are bigger than 1, so we can say that we spent less hours than planned for all activities except Literature Research.

At *Table 5.5.*, we can see that all performance values for all project end of April. Here *SPI* is smaller than 1 and this shows that we are behind of schedule planned but we spent less hours than planned for the project end of April because *CPI* is bigger than 1. *ETC* shows that the time for complete the project and here it is about 274.71 hours. *EAC* shows that total hours project will complete if we progress with this performance and here it is about 669.78 hours. When we look at the results, we can say that our *PV* for the all project is about 468.21 hours but at the end of April, the total hours for the project completion increased to 669.78 hours. So we progressed slower than we planned.

Note that our method calculated probability distributions for all uncertain values. This provides a richer decision support for decreasing the uncertainty of the evaluation.

Table 5.4. EVM for project activities

ACTIVITIES		PV	EV	AC	SPI	CPI
Literature Research	Mean	227.87	215.86	230.0	0.94729	0.9385
	Median	227.18	215.6	230.0	0.95003	0.93739
	SD	11.19	12.835	0.0	0.032213	0.055848
	Variance	125.23	164.73	0.0	0.00104	0.00312
Modelling of Structure	Mean	152.93	118.17	100.0	0.7779	1.1817
	Median	147.33	115.38	100.0	0.77921	1.1538
	SD	18.628	16.464	0.0	0.05189	0.16473
	Variance	346.98	271.07	0.0	0.00269	0.0271
Case Study	Mean	56.401	37.609	35.0	0.66689	1.0746
	Median	56.446	37.143	35.0	0.65995	1.0614
	SD	3.2497	7.8448	0.0	0.13395	0.22445
	Variance	10.56	61.541	0.0	0.01794	0.0504
Analysis and Evaluation of results	Mean	31.979	38.539	30.0	1.2007	1.2848
	Median	31.976	37.518	30.0	1.1896	1.2507
	SD	4.2124	8.3766	0.0	0.17356	0.27953
	Variance	17.744	70.167	0.0	0.0301	0.0781

Table 5.5. Project EVM for end of April

	PV	EV	AC	SPI	CPI	ETC	EAC
Mean	468.21	410.19	395.0	0.8781	1.0384	274.71	669.78
Median	465.4	408.97	395.0	0.8761	1.0354	272.91	667.91
SD	22.555	24.02	0.0	0.066426	0.060827	49.894	50.348
Variance	508.74	576.98	0.0	0.004	0.004	2489	2534.9

5.2. Case 2

Our second example is created by working on the Biofuel Refinery Project data which is from the project data repository of the “Operations Research & Scheduling Research Group” at University of Gent. (*http://www.projectmanagement.ugent.be*)

5.2.1. Data of Case 2

Table 5.6. shows the data we from the Biofuel Refinery Project. Here we have 23 activities in total but there 6 activities which have 0 (zero) value of variable cost. While evaluating project performance, we didn’t include the activities that in have only fixed cost and no variable cost. Because variable costs reflects activity progress in the project. As a result, we have 17 activities, their planned variable costs and their start-finish dates as shown in *Table 5.6.*

Table 5.6. Data of Biofuel Refinery Project

General		Baseline Durations			Baseline Costs			Costs Distribution Profiles		
ID	Name	Baseline Start	Baseline End	Duration	Fixed Cost	Variable Cost	Total Cost	Optimistic	Most Probable	Pessimistic
0	Biofuel Refinery	02.03.2015 7:00	15.07.2016 17:00	360d	13.205.000,00€		14.362.625,00€			
1	Tanks - Preparation	02.03.2015 7:00	11.02.2016 17:00	249d	5.800.000,00€	56.025,00€	5.856.025,00€	55.464,75€	56.025,00€	56.585,25€
2	Tanks - On Site	12.02.2016 7:00	12.02.2016 17:00	1d	0,00€	1.350,00€	1.350,00€	1.050,00€	1.350,00€	6.750,00€
3	Skids - Preparation	02.03.2015 7:00	01.05.2015 17:00	45d	0,00€	10.125,00€	10.125,00€	9.100,00€	10.125,00€	11.150,00€
4	Skids - Lead Time 1	18.05.2015 7:00	12.02.2016 17:00	195d	500.000,00€	0,00€	500.000,00€	0,00€	0,00€	0,00€
5	Skids - Lead Time 2	15.02.2016 7:00	11.03.2016 17:00	20d	50.000,00€	0,00€	50.000,00€	0,00€	0,00€	0,00€
6	Skids - On Site 1	15.02.2016 7:00	13.05.2016 17:00	65d	120.000,00€	117.000,00€	237.000,00€	111.800,00€	118.000,00€	165.000,00€
7	Skids - On Site 2	17.05.2016 7:00	13.06.2016 17:00	20d	40.000,00€	36.000,00€	76.000,00€	34.600,00€	36.000,00€	45.600,00€
8	Skids - Commissioning	16.05.2016 7:00	15.07.2016 17:00	45d	0,00€	121.500,00€	121.500,00€	104.400,00€	121.500,00€	252.300,00€
9	Utilities - Preparation	29.06.2015 7:00	28.08.2015 17:00	45d	600.000,00€	0,00€	600.000,00€	0,00€	0,00€	0,00€
10	Utilities - Lead Time	31.08.2015 7:00	11.03.2016 17:00	140d	0,00€	94.500,00€	94.500,00€	91.350,00€	94.500,00€	97.575,00€
11	Tie-inns	30.11.2015 7:00	04.12.2015 17:00	5d	85.000,00€	2.250,00€	87.250,00€	1.950,00€	2.250,00€	2.650,00€
12	Civil - Preparation	04.05.2015 7:00	28.08.2015 17:00	85d	1.300.000,00€	38.250,00€	1.338.250,00€	36.950,00€	38.250,00€	39.600,00€
13	Civil - On Site	12.10.2015 7:00	29.01.2016 17:00	80d	0,00€	360.000,00€	360.000,00€	338.000,00€	360.000,00€	477.000,00€
14	Labo Container	12.10.2015 7:00	26.02.2016 17:00	100d	150.000,00€	90.000,00€	240.000,00€	86.400,00€	90.000,00€	93.600,00€
15	Piping - Preparation	04.05.2015 7:00	01.01.2016 17:00	175d	800.000,00€	0,00€	800.000,00€	0,00€	0,00€	0,00€
16	Piping - On Site	29.02.2016 7:00	13.05.2016 17:00	55d	0,00€	99.000,00€	99.000,00€	95.000,00€	99.000,00€	119.600,00€
17	Electrical - Preparation	29.06.2015 7:00	01.01.2016 17:00	135d	400.000,00€	0,00€	400.000,00€	0,00€	0,00€	0,00€
18	Electrical - On Site	04.04.2016 7:00	13.05.2016 17:00	30d	0,00€	13.500,00€	13.500,00€	12.400,00€	13.500,00€	18.150,00€
19	Automation - Preparation	29.06.2015 7:00	29.04.2016 17:00	220d	80.000,00€	49.500,00€	129.500,00€	47.550,00€	49.500,00€	51.500,00€
20	Automation - On Site	02.05.2016 7:00	27.05.2016 17:00	20d	0,00€	4.500,00€	4.500,00€	4.300,00€	4.500,00€	5.925,00€
21	Analytics - Preparation	01.06.2015 7:00	27.05.2016 17:00	260d	80.000,00€	58.500,00€	138.500,00€	56.525,00€	58.500,00€	60.475,00€
22	Analytics - On Site	30.05.2016 7:00	01.07.2016 17:00	25d	0,00€	5.625,00€	5.625,00€	5.400,00€	5.625,00€	7.500,00€
23	Purchases	28.09.2015 7:00	26.02.2016 17:00	110d	3.200.000,00€	0,00€	3.200.000,00€	0,00€	0,00€	0,00€

Table 5.6. also shows the cost distributions of activities, which we have used to define the planned value distributions of each activity. There is also project progress status for each month at the site where the project data was taken. We applied the *EVM* for the project progress until 29.07.2015. The data of this date are shown at *Table 5.7.*

Table 5.7. The progress data of Biofuel Refinery Project on 29.07.2015

ID	Name	Duration	Variable Cost (€)	Actual Cost (€)	Percentage Completed
0	Biofuel Refinery	360d	1.157.625,00	9.637.625,00	67%
1	Tanks - Preparation	249d	56.025,00	5.824.300,00	43%
2	Tanks - On Site	1d	1.350,00	0,00	0%
3	Skids - Preparation	45d	10.125,00	10.125,00	100%
6	Skids - On Site 1	65d	117.000,00	0,00	0%
7	Skids - On Site 2	20d	36.000,00	0,00	0%
8	Skids - Commissioning	45d	121.500,00	0,00	0%
10	Utilities - Lead Time	140d	94.500,00	0,00	0%
11	Tie-inns	5d	2.250,00	0,00	0%
12	Civil - Preparation	85d	38.250,00	1.328.350,00	74%
13	Civil - On Site	80d	360.000,00	0,00	0%
14	Labo Container	100d	90.000,00	0,00	0%
16	Piping - On Site	55d	99.000,00	0,00	0%
18	Electrical - On Site	30d	13.500,00	0,00	0%
19	Automation - Preparation	220d	49.500,00	85.175,00	10%
20	Automation - On Site	20d	4.500,00	0,00	0%
21	Analytics - Preparation	260d	58.500,00	89.675,00	17%
22	Analytics - On Site	25d	5.625,00	0,00	0%

5.2.2. Model of Case 2

We identified 7 risk factors for this problem as follows:

- Managerial Problems
- Money Problems
- Supplier Problems
- Employee Loss
- Vehicle-Tool Accidents
- Civil Damage
- Financial Loss

Risk factors are determined and linked to the interested variables. If we want, we connect the risk factors each other and determined their relationships. In *Figure 5.7.,* we can see how risk factor are modeled in the BN model.

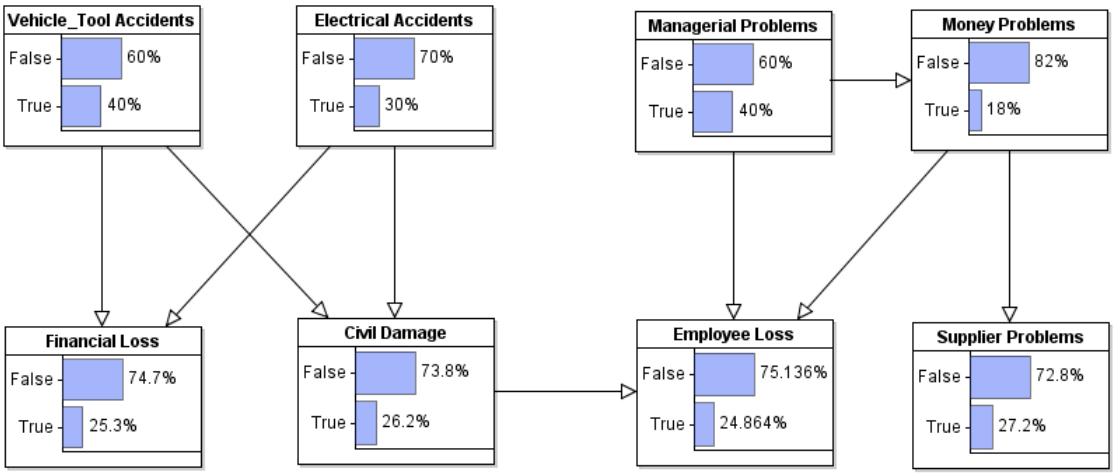


Figure 5.7. Risk Factors and Relations in AgenaRisk

The *PV* distributions from Table *5.6*. we adjusted for each risk factor as shown in *Table 5.8*. Here we used Beta distribution for activity of Tanks-Preparation and Triangle distribution for all other activities. We have three risk factors for the activities which are progressed until evaluated time. The risk of *"Supplier Problems"* effects activity of *"Tanks-Preparation"*, the risks of *"Financial Loss"* and *"Employee Loss"* are effects the activity of *"Skids-Preparation"*. The probability of Supplier Problems, Financial Loss and Employee Loss is 0.272, 0.253 and 0.249 respectively.

Table 5.8. Expressions of activities

ID	ACTIVITIES	RISK FACTORS AND POSSIBILITIES	STATUS OF RISKS	PV EXPRESSIONS OF ACTIVITIES
1	Tanks – Preparation	Supplier Problems (True:0.272, False:0.728)	Supplier Problems: T	Beta (4,4,0,8)
			Supplier Problems: F	Beta (4,4,0,8)
2	Tanks - On Site	No Risk	-	Triangle (1050, 6750, 1350)
3	Skids – Preparation	Financial Loss (True:0.253, False:0.747) Employee Loss (True:0.249, False:0.751)	Financial Loss: T Employee Loss: T	Triangle (9300, 11300, 10300)
			Financial Loss: T Employee Loss: F	Triangle (9250, 11250, 10250)
			Financial Loss: F Employee Loss: T	Triangle (9200, 11200, 10200)
			Financial Loss: F Employee Loss: F	Triangle (9100, 11150, 10125)
6	Skids - On Site 1	No Risk	-	Triangle (111800,165000,118000)
7	Skids - On Site 2	No Risk	-	Triangle (34600,45600,36000)
8	Skids – Commissioning	No Risk	-	Triangle (104400,252300,121500)
10	Utilities - Lead Time	No Risk	-	Triangle (91350,97575,94500)
11	Tie-inns	No Risk	-	Triangle (1950,2650,2250)
12	Civil - Preparation	No Risk	-	Triangle (36950,39600,38250)
13	Civil - On Site	No Risk	-	Triangle (338000,477000,360000)
14	Labo Container	No Risk	-	Triangle (86400,93600,90000)
16	Piping - On Site	No Risk	-	Triangle (95000,119600,99000)
18	Electrical - On Site	No Risk	-	Triangle (12400,18150,13500)
19	Automation – Preparation	No Risk	-	Triangle (47550,51500,49500)
20	Automation - On Site	No Risk	-	Triangle (4300,5925,4500)
21	Analytics - Preparation	No Risk	-	Triangle (56525,60475,58500)
22	Analytics - On Site	No Risk	-	Triangle (5400,7500,5625)

Since we evaluated the performance for 29.07.2015, we modeled activities that happen until this date. There are five activities happen until this date and these are "*Tanks-Preparation*", "*Skids-Preparation*", "*Civil-Preparation*", "*Automation-Preparation*". and "*Analytics-Preparation*".

Figure 5.8. shows the model for activity of "*Tanks-Preparation*". There are *PV* for until month 8, *PV* for until month 7 and total planned value distributions. The reason of exist both of *PV* for until 8 and until 7 is calculating the *ESM* beside *EVM* which is shown in Section 3.2. The risk of *Supplier Problems* affects total *PV* of activity. As a result, we have *SPI* and *CPI* for activity of *Tanks-Preparation*.

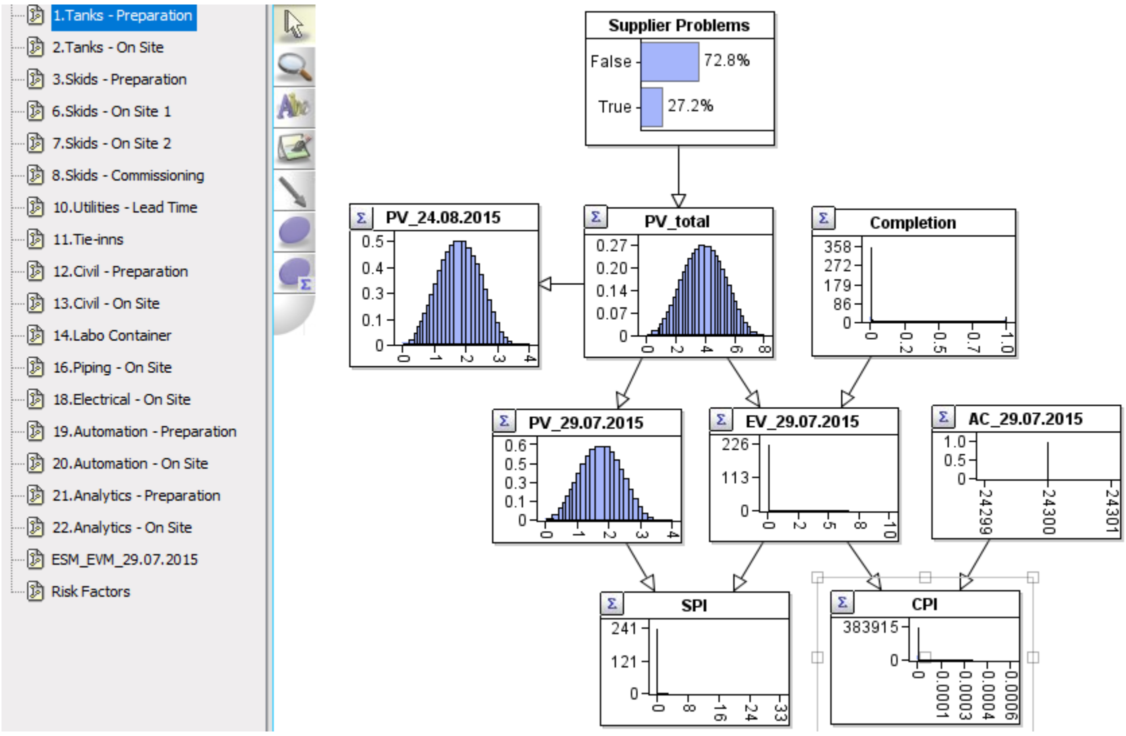


Figure 5.8. AgenaRisk Model of Tanks-Preparation

The models of other four activities are at the appendix part. They all have *PV* for until month 8, *PV* for until month 7 and total planned value distributions. The activity of "*Skids-Preparation*" has the risks of *Financial Loss* and *Employee Loss* effects total *PV* of activity. The activities of "*Civil-Preparation*", "*Automation-Preparation*" and "*Analytics-Preparation*" has no risk factor that directly affects them.

Figure 5.9. shows that all *EVM* values for Case 2. Here we can see that also *ES* value which told in section 3.2. We applied *ESM* and *EVM* both for our method. Here all planned values, earned values and actual costs of each activities for the evaluation time are transferred and summed up. Also total *PV* of activities are summed up. As a result we have performance indexes which are *ES*, *SPI* and *CPI*, estimations which are *PDWR*, *IEDAC*, *ETC* and *EAC* which are calculated by the model.

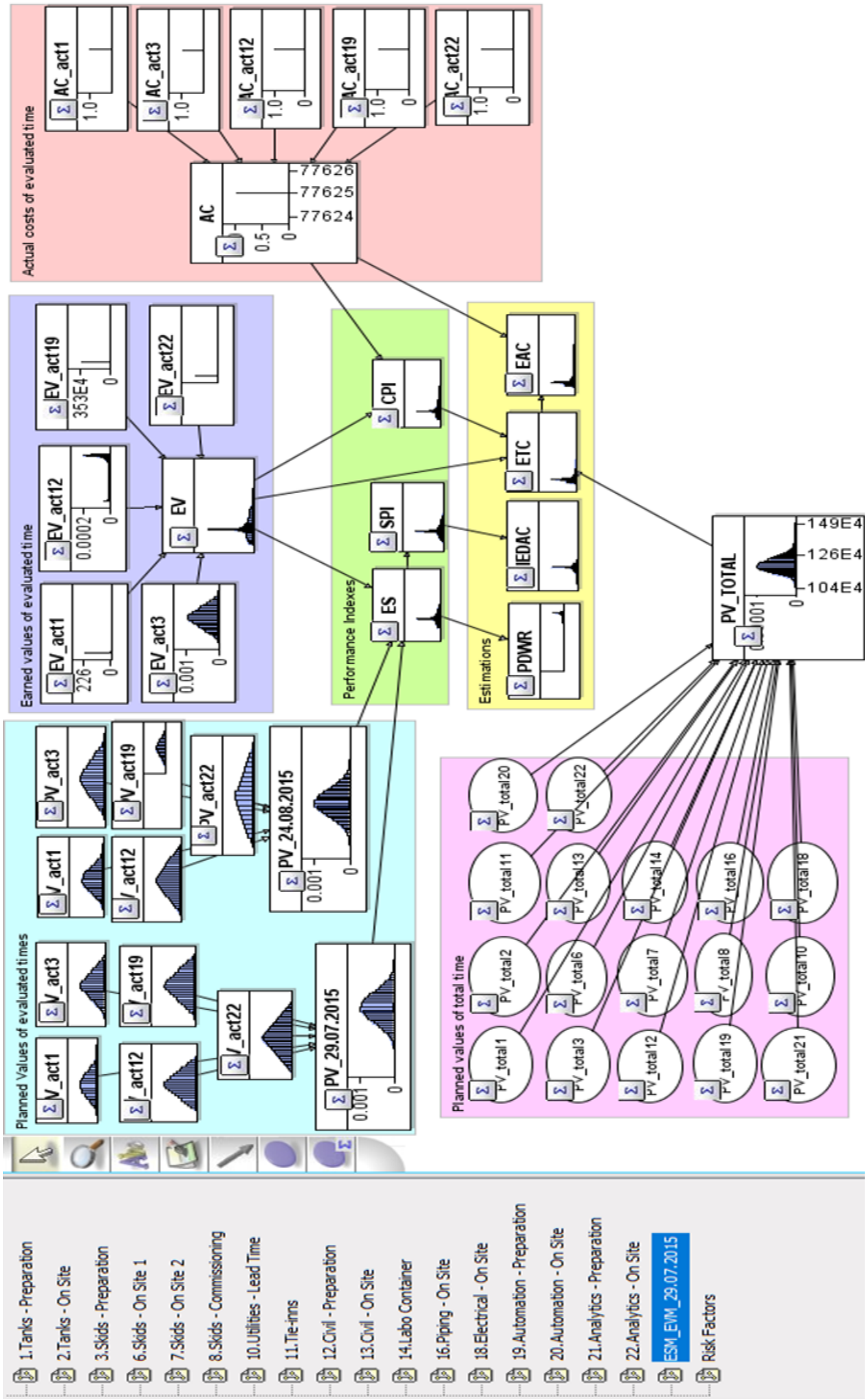


Figure 5.9. EVM Object of Case 2

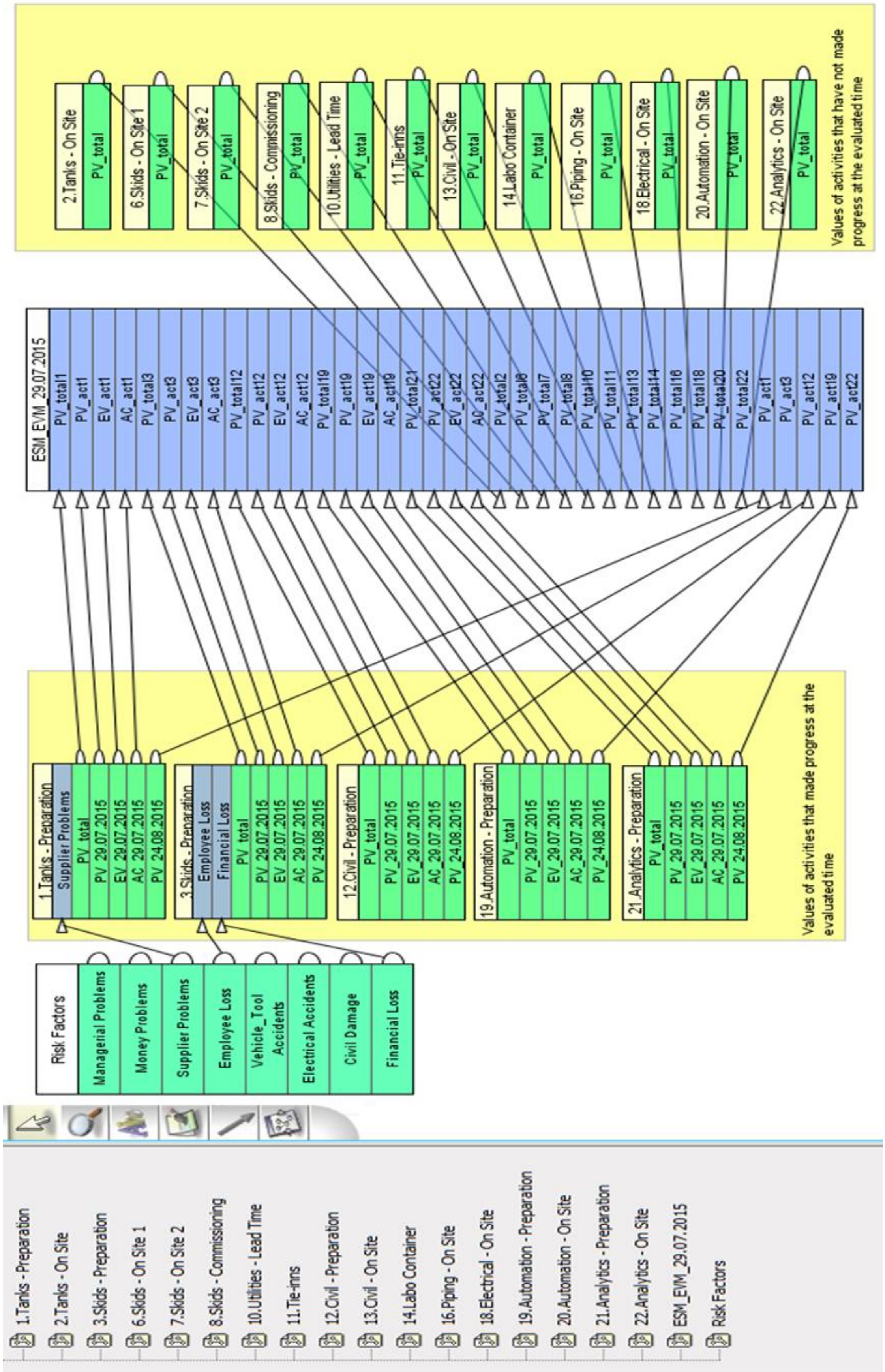


Figure 5.10. Map of All Model for Case 2

Figure 5.10. shows the links between all values. It shows which value is transferred from where to where. We linked the interested values for transforming and we can see that how links are built.

5.2.3. Result of Case 2

After run the model, the results for evaluation time are shown below. In *Table 5.9.* and *5.10.*, we can see that all performance values for all project. Here the *SPI* value is 0.99982 (SD:0.013124) which is smaller than 1 and shows that we are behind of schedule planned and we spent more hours than planned for the project. The *CPI* value is 0.68338 (SD:29239) which is also smaller than 1 and this shows that, we spent more money than we planned.

ETC shows that the cost for completing the remain of project. *EAC* shows the total cost that project will complete if we progress with this performance. Here the *ETC* value is 2.129.300 € (SD:1356800) so we need this amount of money for the finishing project. *IEDAC* shows estimated duration for completion total project with this performance and *PDWR* shows the remaining duration for complete project. Here *IEDAC* value is 360.12 days and that's mean we need 360.12 days for finishing project. These values are all probabilistic and estimation for the future of project. We can take precautions in uncertainty with our method.

Table 5.9. Results of Case 2

	PV	**EV**	**AC**	**SPI**	**CPI**
Mean	53390.0	53020.0	77625.0	0.99982	0.68338
Median	53389.0	49147.0	77625.0	0.99765	0.63328
SD	610.61	22579.0	0.0	0.013124	0.29239
Variance	372840.0	5.098	0.0	1.722	0.085494

Table 5.10. Estimations of Case 2

	ETC	**EAC**	**IEDAC**	**PDWR**	**ES**
Mean	2129300.0	2209000.0	360.12	252.02	107.98
Median	1797700.0	1875000	360.84	252.25	107.75
SD	1356800.0	1363600.0	4.7239	1.4215	1.4158
Variance	1.8408	1.8593	22.315	2.0208	2.0045

5.3. Case 3

We also applied our method to a case study from Turkish Aerospace Industries, Inc. (TAI). Due to confidentiality issues, the name of the project and its activities are masked.

5.3.1. Data of Case 3

The project has begun in 01.08.2017 and planned to be finished in 31.12.2020. The project has over 480 activities and 13 milestones. In this application, we will analyze the progress in terms of milestones of rather than specific activities. The project owner department gave each activity a weight score depending on their importance between 1-5 to measure project performance. Then they reached a value by multiplying these weight scores and durations of activities. After that normalizing the values, they used these normalized weight values in their measurements. We also used these weights for our method. In this example, there is no uncertainty expressions for *PVs*. In TAI, project managers planned all project with certain values because the size of the projects and they do Risk Management for all projects separately. The table below shows the weight values of 13 milestone for total project time. Total weight of the projects is 50320 score.

Table 5.11. PV for activities of Case 3 (Weight Score × Duration)

Milestones	**Sum of Weight Normalized**
Milestone-1	3696
Milestone-2	1069
Milestone-3	1424
Milestone-4	2964
Milestone-5	362
Milestone-6	1143
Milestone-7	1106
Milestone-8	228
Milestone-9	2102
Milestone-10	21054
Milestone-11	6260
Milestone-12	4257
Milestone-13	4653
Grand Total	**50320**

We have optimistic and pessimistic weight values for all activities and risk factors for some of them that are determined by the project managers. The *Table 5.12.* shows the weight values and related risks. According to these weight values and risks, triangle distributions are defined for the planned values in AgenaRisk.

Table 5.12. Possible optimistic and pessimistic PV of activities

Milestones	Optimistic	Pessimistic	Related Risks
Milestone-1	3562	3817	Delay in materials imported from abroad
Milestone-2	964	1112	Delay in domestic auxiliary industry companies
Milestone-3	1384	1496	-
Milestone-4	2802	3114	-
Milestone-5	320	398	Producing faulty material
Milestone-6	981	1403	Producing faulty material
Milestone-7	1092	1156	Delay in materials imported from abroad
Milestone-8	200	250	Change of political strategies
Milestone-9	2052	2200	-
Milestone-10	20987	21126	-
Milestone-11	6174	6302	-
Milestone-12	4132	4326	-
Milestone-13	4591	4944	Delay in domestic auxiliary industry companies
Grand Total	**49241**	**51644**	

5.3.2. Model of Case 3

Due to the complexity of this project, under the company systems, the cost spent on projects is recorded on a project basis rather than activity basis. So here, we do not have *AC* values for each activity. We have *AC* for total project. The *Table 5.13.* shows that determined *PV*, *% Completion* values of each activity and *AC* for total project in terms of weight by the date 21.07.2019.

Table 5.13. PV, Completion and AC values for Case 3

Milestones	PV distribution	Completion%	Total AC
Milestone-1	Normal (2253, 100)	58,4%	
Milestone-2	Normal (811, 100)	72,3%	
Milestone-3	Normal (762, 120)	51,1%	
Milestone-4	Normal (2284, 120)	74,3%	
Milestone-5	Normal (315, 100)	87,2%	
Milestone-6	Normal (627, 100)	53,2%	
Milestone-7	Normal (1106, 120)	100,0%	
Milestone-8	Normal (228, 150)	100,0%	
Milestone-9	Normal (2102, 100)	100,0%	
Milestone-10	Normal (18189, 120)	45,6%	
Milestone-11	Normal (395, 120)	1,3%	
Milestone-12	Normal (188, 150)	0,0%	
Milestone-13	Normal (3454, 100)	41,1%	
Grand Total	**32715**	**43,3%**	**28784**

All milestones are modeled in AgenaRisk with the determined distributions and risks. The models of every milestone are at the appendix part. The following figure shows that *EVM* object of case 3.

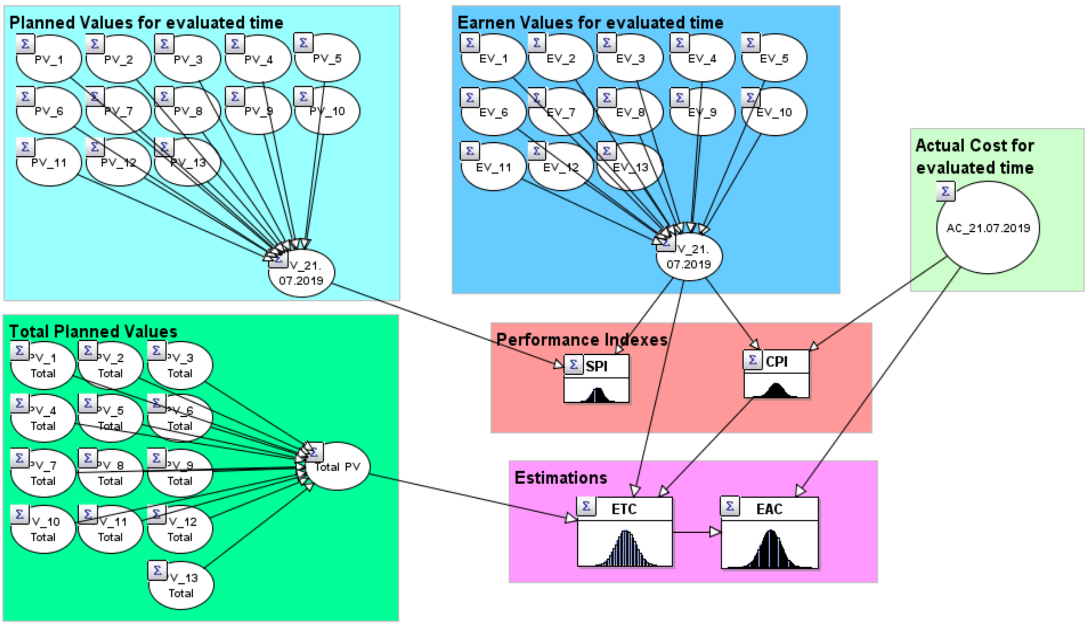


Figure 5.11. EVM object of Case 3

5.3.3. Result of Case 3

We run the model in AgenaRisk and reached the results. *Table 5.14. and 5.15.* shows the results and estimations. Here, *EV* is 21812 (SD: 105.23) and with 95% confidence *EV* is between the values of 21606 and 22017. *PV* is 32714 (SD: 39.6) and with 95% confidence *PV* is between the values of 32637 and 32791. The *SPI* value is 0.67 (SD: 0.003) and with 95% confidence *SPI* is between the values of 0.66022 and 0.67322. The *CPI* value is 0.76 (SD: 0.004) and with 95% confidence *CPI* is between the values of 0.75058 and 0.76491. Since *CPI* and *SPI* values are smaller than one, the project is expected to be delayed and over-budget.

We did not apply *ESM* here. Because, for *ESM*, we have to have data of month before the evaluated month. The company did not share these data because of confidentiality issues.

ETC shows that the cost for completing the remain of the project. *EAC* shows that total cost which project will complete if we progress with this performance. When we look at *EAC* value, we can cay that the project will cost more than planned. Because *EAC* is 66320 (SD: 405.17) and with 95% confidence it falls between the values of 65532 and 67117. We know that total *PV* of project is 52320 score. The *EAC* value is higher than *PV*. That's mean we will complete the project at a much higher cost than we planned. Also the *ETC* value says the same things. The *ETC* value is 37536 (SD: 404.5) and with 95% confidence it falls between the values of 36748 and 38333. We spent the AC value which is 28784 score until so far and the *ETC* says we need 37536 score for finishing the project. But in our *PV*, we have 23536 *(PV-AC)* score for finishing project. Because *ETC* is higher than *PV-AC* value, more money is needed for remaining jobs. So, project team have to plan the cost requirement of the project and share it with Management.

Table 5.14. Results of Case 3

	PV	EV	AC	SPI	CPI
Mean	32714.0	21812.0	28784.0	0.66674	0.75777
Median	32714.0	21812.0	28784.0	0.66675	0.75778
SD	39.596	105.23	0.0	0.00333	0.00366
Variance	1567.8	11073.0	0.0	1.1093	1.3424

Table 5.15. Estimations of Case 3

	ETC	EAC
Mean	37536.0	66320.0
Median	37535.0	66319.0
SD	404.5	405.17
Variance	163620.0	164160.0

6. CONCLUSION

6.1. General Evaluation

This thesis proposes a new approach for measuring performance of risky and uncertainty projects by integrating the *EVM* method with the proposed *BN* model. In this context, this study gave an brief information about *BNs* and *EVM*, and reviewed previous studies that used BNs for project management and that aimed to incorporate uncertainty to *EVM*. The proposed approach was applied to three different case studies. The results of the approach in these case studies were examined and compared with the traditional *EVM* approach.

The *EVM* approach disregards the uncertainty and risk associated with its parameters, and it uses exact planned value and completion percentage parameters when evaluating the project performance. However, in planned values and completion percentages often involve uncertainty, and there may be risk factors associated with them. In the proposed approach, *EVM* is modeled based on *BNs*, the risks and uncertainty associated with its parameters were included in the model. This enabled calculation of the uncertainty regarding performance indices, time and budget predictions. What-if analyses regarding risk scenarios could also be performed. In summary, the proposed approach provides a more comprehensive decision support to the project manager by providing a combined summary of the progress, risk and uncertainty associated with the project.

The proposed method was applied to 3 different case studies. The first case is a simple 'toy-example' about a research study. There are four activities which are *"Literature Research", "Modelling of Structure, "Case Study"* and *"Analysis and Evaluation of results"* in the plan. We have four different risk factors in total that affect the some of the activities. The "*Exam"* risk affects all four activities, the *"Technical Error"* risk effects *"Modelling of Structure"*, the *"Insufficiency of data"* risk effects *"Case Study",* the *"Unexpected results"* risk effects *"Analysis and Evaluation"*. The whole project was planned for 6 months and the evaluation was at the end of fourth month. The total *PV* of case study is determined for the end of fourth month 468.21 (SD: 22.555). At the end of fourth month, we saw that our *EV* is 410.19 (SD: 24.02) and our *AC* is 395.0 (SD: 0). As a result, the *SPI* value is 0.8781 (SD: 0.066) and the *CPI* value is 1.0384

(SD: 0.0608). We can say that in this case, we are behind the planned schedule because *SPI* is smaller than 1. Also we can say this by looking to *PV* and *EV* difference. We earned the less value than we planned. On the other hand, we can say that we are under the planned hours in terms of spent time because *AC* is bigger than 1. If we continue with this performance, we will need 274.71 (SD: 49.894) hours which is *ETC* value for the complete all project. Therefore, the project will be finished in total 669.78 (SD:5 0.348) hours which is *EAC* value while our plan was 468.21 (SD: 22.55) hours. In this case, the project manager can try to speed up the project at the end of fourth month to complete the project in six month as it was initially planned.

Our second case is a Biofuel Refinery Project, which have 23 activities in total, but we did not include six activities that only have fixed costs and therefore do not reflect the progress of the project. We evaluated the project based on 17 activities. The project started on March 2015 and we evaluated the results for July 2015. They are seven different risk factors in total but only three of those risk factors were associated with the activities that had started at the time of evaluation. The total *PV* of the case for July 2015 is 53390 (SD: 610.61). The *EV* for July 2015 is 53020 (SD: 22579) and as a result, *SPI* value is 0.99. The values of *PV* and *EV* are close to each other and the *SPI* value is close to the one. Therefore, we can say that we are behind the schedule slightly and this difference is acceptable for the project managers. The value of *AC* is 77625 and the *CPI* value is 0.68338 (SD: 0.29239). As a result, expenditures are over the planned budget. Planned budget was 1.157.625,00 € but the *EAC* value shows that project will be completed with 2.209.000,00 €. There is a big difference planned budget and estimated budget, so project manager can take precautions for the money issue. Project also is planned to finish in 360 days. The *IEDAC* value that shows the estimated duration for completion the project is 360.12 days as there seems to be no problem with the schedule performance at the time of evaluation.

The third case is a real-world project from a defense company in Turkey. In this case, there is no detailed data because of the confidentiality issue. This project has 13 milestones, and *PV* for the evaluated time is 32714 (SD: 39.596) and *EV* is 21812 (SD: 105.23). As a result, *SPI* value is 0.66 and we can say that, we are behind the schedule highly because SPI is smaller than one. The *AC* value is 28784 and the *CPI* value is 0.757. Because *CPI* is smaller than one, planned budget was exceeded. The expenditures of the project were over the planned budget, and the current performance indicates possible delays. In other words, there are problems regarding the progress of this project, and further precautions should be considered.

6.2. Discussion

Considering all three case studies, it was seen that the proposed approach provides a comprehensive assessment of the project performance and the associated uncertainty and risk. This method allows the project management team to manage the uncertainties in the projects, to take early precautions about emerging risks and to comment about the project results. On the other hand, there may be some difficulties for the project management teams in terms of applicability of this method. In order to apply the method, it is necessary to determine the distribution of the planned values, the risk factors and to how these risks affect the planned values. Likewise, the necessary distributions for completion rates may need to be determined. However, since it is not easy to determine these informations, the project management team must have technical knowledge in the Statistics and work carefully. Statistical knowledge is also required for the interpretation of the results. Especially for the large and complex projects, it can be very difficult to determine the distribution and risk.

6.3. Suggestions

As future research, the proposed approach can be expanded to give automatic warning about risk factors. It can give information about risk when there is data entry in activities involving risk factors. In addition, this method can be applied to existing project management tools and software by making the necessary changes and adaptations. Many companies manage a multitude of projects that share resources and risk. In those cases, the proposed approach can be expanded to manage and review multiple projects at the same time. In other words, it can be improved to control the company project portfolio in a unified way.

REFERENCES

Acebes, F.a, Pereda, M. ., Poza, D. ., Pajares, J. ., & Galán, J. M. . (2015). Stochastic earned value analysis using Monte Carlo simulation and statistical learning techniques. International Journal of Project Management, 33(7), 1597–1609. https://doi.org/10.1016/j.ijproman.2015.06.012

Acebes, Fernando, Pajares, J., Galán, J. M., & López-Paredes, A. (2014). A new approach for project control under uncertainty. Going back to the basics. International Journal of Project Management, 32(3), 423–434. https://doi.org/10.1016/j.ijproman.2013.08.003

Anbari, F. T. (2003). Earned Value Project Management Method and Extensions. Project Management Journal. https://doi.org/10.1109/EMR.2004.25113

Batselier, J., & Vanhoucke, M. (2017). Improving project forecast accuracy by integrating earned value management with exponential smoothing and reference class forecasting. *International Journal of Project Management*, *35*(1), 28–43. https://doi.org/10.1016/j.ijproman.2016.10.003

Caron, F., Ruggeri, F., & Merli, A. (2013). A Bayesian Approach to Improve Estimate at Completion in Earned Value Management. Journal of Internet Banking and Commerce, 17(1), 3–16. https://doi.org/10.1002/pmj

Caron, F., Ruggeri, F., & Pierini, B. (2016). A Bayesian approach to improving estimate to complete. International Journal of Project Management, 34(8), 1687–1702. https://doi.org/10.1016/j.ijproman.2016.09.007

Colin, J., & Vanhoucke, M. (2014). Setting tolerance limits for statistical project control using earned value management. Omega, 49, 107–122. https://doi.org/10.1016/j.omega.2014.06.001

De Melo, A. C. , & Sanchez, A. J. (2008). Software maintenance project delays prediction using Bayesian networks. Expert Systems with Applications, 34 , 908–919 .

Ding, X., Li, Y., Belatreche, A., & Maguire, L. P. (2014). An experimental evaluation of novelty detection methods. Neurocomputing, 135, 313–327. https://doi.org/10.1016/j.neucom.2013.12.002

Fan, C.-F. , & Yu, Y.-C. (2004). BBN-based software project risk management. Journal of Systems and Software, 73 , 193–203 .

Fenton, N., Khodakarami, V., & Neil, M. (2007). Title : “ Project Scheduling : Improved approach to incorporate uncertainty using Bayesian Networks .” Project Management Journal, (June), 39–49.

Fenton, N. , Marsh, W. , Neil, M. , Cates, P. , Forey, S. , & Tailor, M. (2004). Making re- source decisions for software projects. In Software engineering, 2004. ICSE

2004. Proceedings. 26th international conference on (pp. 397–406). IEEE .

Fenton, N. , & Neil, M. (2012). Risk assessment and decision analysis with Bayesian networks . Boca Raton, FL: CRC Press .

Fineman, M. , Fenton, N. , & Radlinski, L. (2009). Modelling project trade-off using Bayesian networks. In Computational intelligence and software engineering, 2009. CiSE 2009. International conference on (pp. 1–4). IEEE .

Fleming, Q. W., & Koppelman, J. M. (2002). Using Earned Value Management, 44(9).

Gompertz, B. (1825). On the Nature of the Function Expressive of the Law of Human Mortality, and on a New Mode of Determining the Value of Life Contingencies. Philosophical Transactions of the Royal Society of London, 115(0), 513–583. https://doi.org/10.1098/rstl.1825.0026

Henderson, K., Henderson, K., & Lipke, W. (2006). Earned Schedule in Action London , United Kingdom, (June).

Hu, Y., Zhang, X., Ngai, E. W. T., Cai, R., & Liu, M. (2013). Software project risk analysis using Bayesian networks with causality constraints. Decision Support Systems, 56(1), 439–449. https://doi.org/10.1016/j.dss.2012.11.001

Jensen, F. V. (1996). Bayesian networks basics. AISB quarterly, 9-22.

Khamooshi, H., & Golafshani, H. (2014). EDM: Earned Duration Management, a new approach to schedule performance management and measurement. International Journal of Project Management, 32(6), 1019–1041. https://doi.org/10.1016/j.ijproman.2013.11.002

Khodakarami, V., & Abdi, A. (2014). Project cost risk analysis: A Bayesian networks approach for modeling dependencies between cost items. International Journal of Project Management, 32(7), 1233–1245. https://doi.org/10.1016/j.ijproman.2014.01.001

Lee, E., Park, Y., & Shin, J. G. (2009). Large engineering project risk management using a Bayesian belief network. Expert Systems with Applications, 36(3 PART 2), 5880–5887. https://doi.org/10.1016/j.eswa.2008.07.057

Lipke, W. H. (2003). Schedule is Different. The Measurable News, 2, 31–34. Retrieved from http://www.pmi-cpm.org/members/library/Schedule Is Different.lipke.pdf

Lipke, W., Zwikael, O., Henderson, K., & Anbari, F. (2009). Prediction of project outcome. The application of statistical methods to earned value management and earned schedule performance indexes. International Journal of Project Management, 27(4), 400–407. https://doi.org/10.1016/j.ijproman.2008.02.009

Luu, V. T., Kim, S. Y., Tuan, N. Van, & Ogunlana, S. O. (2009). Quantifying schedule risk in construction projects using Bayesian belief networks. International Journal of Project Management, 27(1), 39–50. https://doi.org/10.1016/j.ijproman.2008.03.003

Miljković, D. (2016). Review of Novelty Detection Methods, (June).

Mishra, A., Das, S. R., & Murray, J. J. (2016). Risk, process maturity, and project performance: An empirical analysis of US federal government technology projects. Production and Operations Management, 25(2), 210–232. https://doi.org/10.1111/poms.12513

Murphy, K. (2002). Dynamic Bayesian Networks: Representation, Inference and Learning. University of California, Berkeley, 223. https://doi.org/10.1.1.129.7714

Naeni, L. M., Shadrokh, S., & Salehipour, A. (2011). A fuzzy approach for the earned value management. International Journal of Project Management, 29(6), 764–772. https://doi.org/10.1016/j.ijproman.2010.07.012

Narbaev, T., & De Marco, A. (2014). An Earned Schedule-based regression model to improve cost estimate at completion. International Journal of Project Management, 32(6), 1007–1018. https://doi.org/10.1016/j.ijproman.2013.12.005

Neil, M., Fenton, N., & Nielson, L. (2000). Building large-scale Bayesian networks, 15(May), 257–284.

Neil, M., Tailor, M., & Marquez, D. (2007). Inference in hybrid Bayesian networks using dynamic discretization. Statistics and Computing, 17(3), 219–233. https://doi.org/10.1007/s11222-007-9018-y

Nevin Lianwen, Z., & Poole, D. (1994). A simple approach to Bayesian network computations, 0–7. Retrieved from https://bitbucket.org/bmmalone/library/src/b5f06a50b629/Murphy2002.pdf%5Cnhttp://onlinelibrary.wiley.com/doi/10.1002/cbdv.200490137/abstract%5Cnhttp://higherintellect.info/texts/science_and_technology/artificial_intelligence/Dynamic Bayesian Networks Repr

Operations Research & Scheduling Research Group at University of Gent (http://www.projectmanagement.ugent.be.)

Pajares, J., & López-Paredes, A. (2011). An extension of the EVM analysis for project monitoring: The Cost Control Index and the Schedule Control Index. International Journal of Project Management, 29(5), 615–621. https://doi.org/10.1016/j.ijproman.2010.04.005

Pearl, J. (1988). Probabilistic reasoning in intelligent systems: Networks of plausible inference. San Mateo: Morgan Kaufmann.

Perkusich, M. , Soares, G. , Almeida, H. , & Perkusich, A. (2015). A procedure to detect problems of processes in software development projects using Bayesian networks. Expert Systems with Applications, 42 , 437–450.

Pinto, J. K. (2016). List of Cases by Chapter.

Project Management Institute Inc. (2017). A guide to the project management body of knowledge (PMBOK® guide).

Steffen L. Lauritzen, & David J. Spiegelhalter (1988). Local computations with probabilities on graphical structures and their application to expert systems, 43(3), 265–292.

US Government Accountability Office Report. 2010. OMB Dash- board has increased Transparency and Oversight, but Improvements Needed. Publication No. GAO-10-701. Avail- able at http://www.gao.gov/new.items/d10701.pdf

Yet, B., Constantinou, A., Fenton, N., Neil, M., Luedeling, E., & Shepherd, K. (2016). A Bayesian network framework for project cost, benefit and risk analysis with an agricultural development case study. Expert Systems with Applications, 60, 141–155. https://doi.org/10.1016/j.eswa.2016.05.005

Yet, B., Perkins, Z. B., Rasmussen, T. E., Tai, N. R. M., & Marsh, D. W. R. (2014). Combining data and meta-analysis to build Bayesian networks for clinical decision support. Journal of Biomedical Informatics, 52, 373–385. https://doi.org/10.1016/j.jbi.2014.07.018

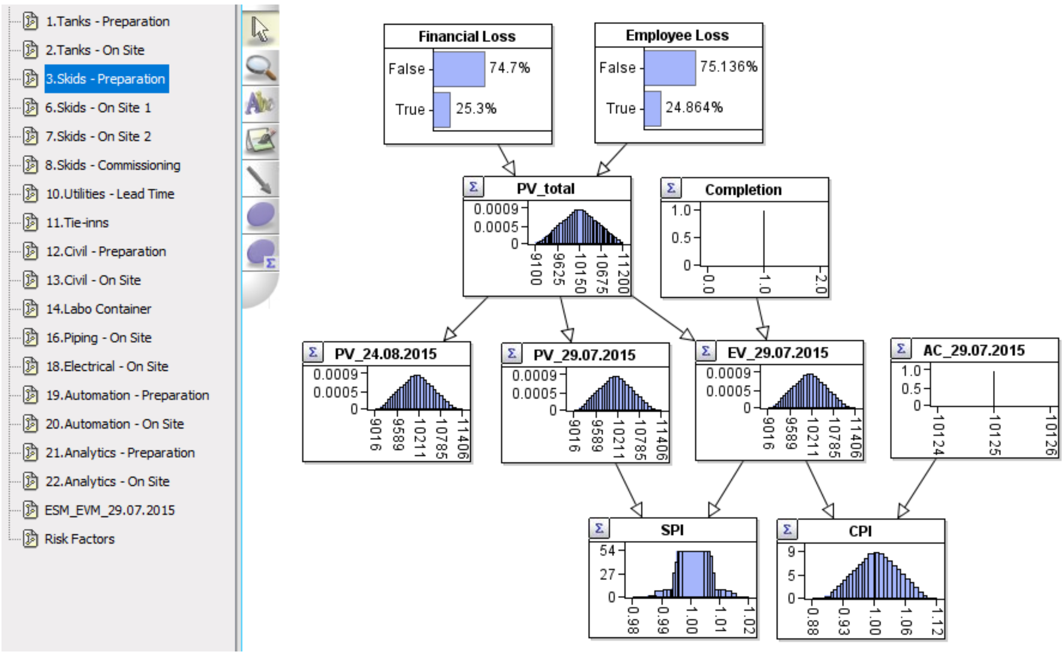
Yet, B., Perkins, Z., Fenton, N., Tai, N., & Marsh, W. (2014). Not just data: A method for improving prediction with knowledge. Journal of Biomedical Informatics, 48, 28–37. https://doi.org/10.1016/j.jbi.2013.10.012

APPENDICES

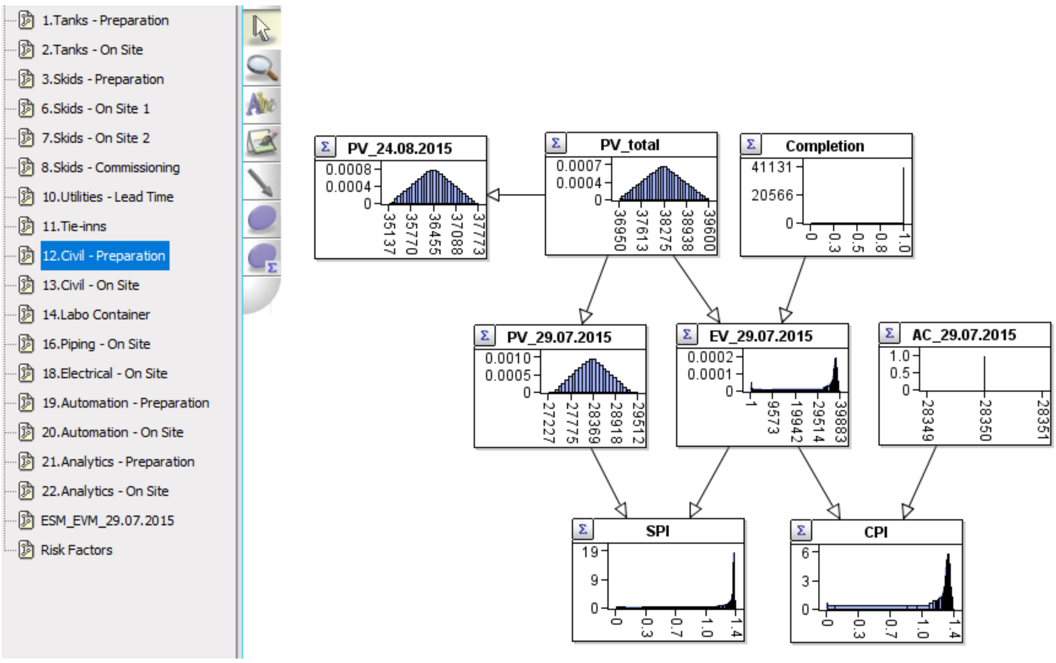
APPENDIX 1 – AgenaRisk Models

Models of Case 2 in AgenaRisk

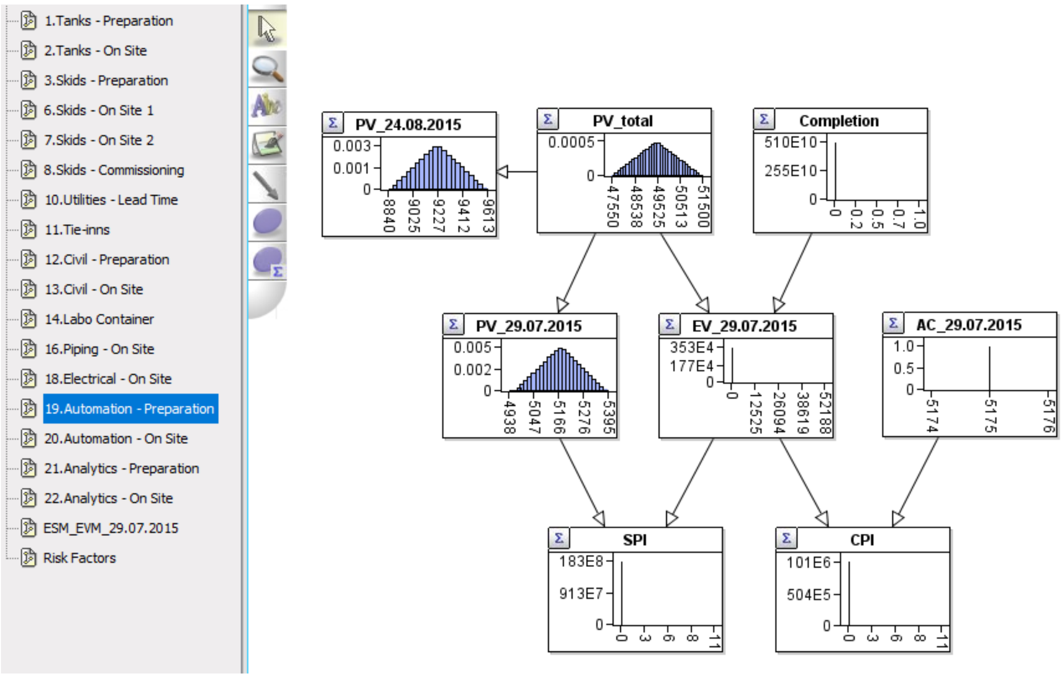
Activity of "*Skids-Preparation*"



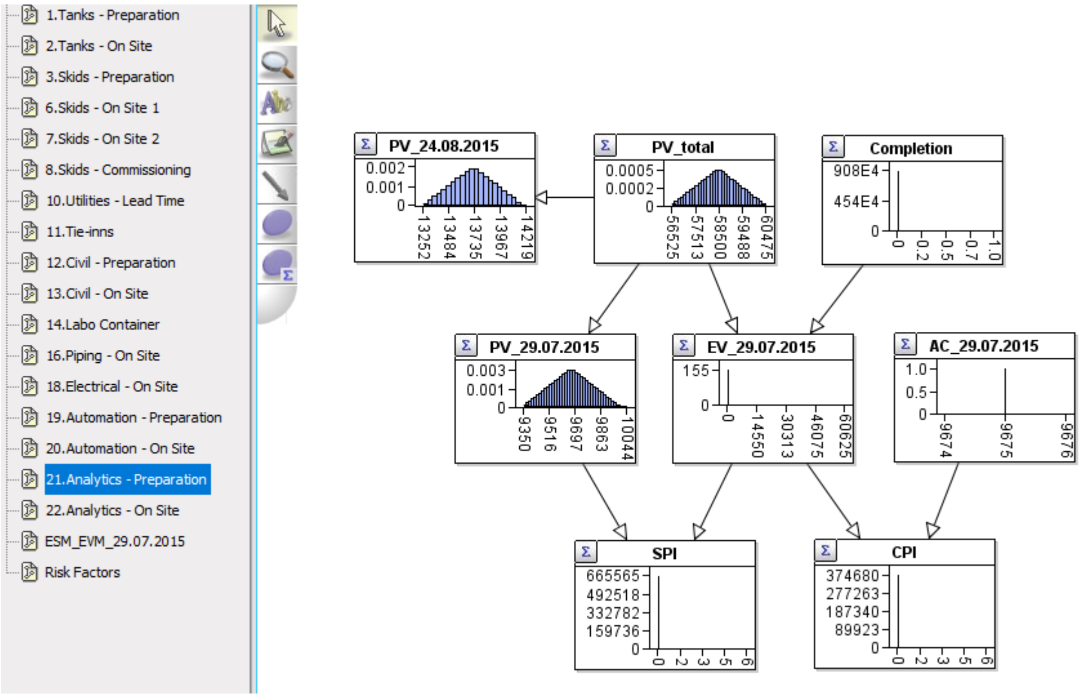
Activity of "*Civil-Preparation*"



Activity of “*Automation-Preparation*”

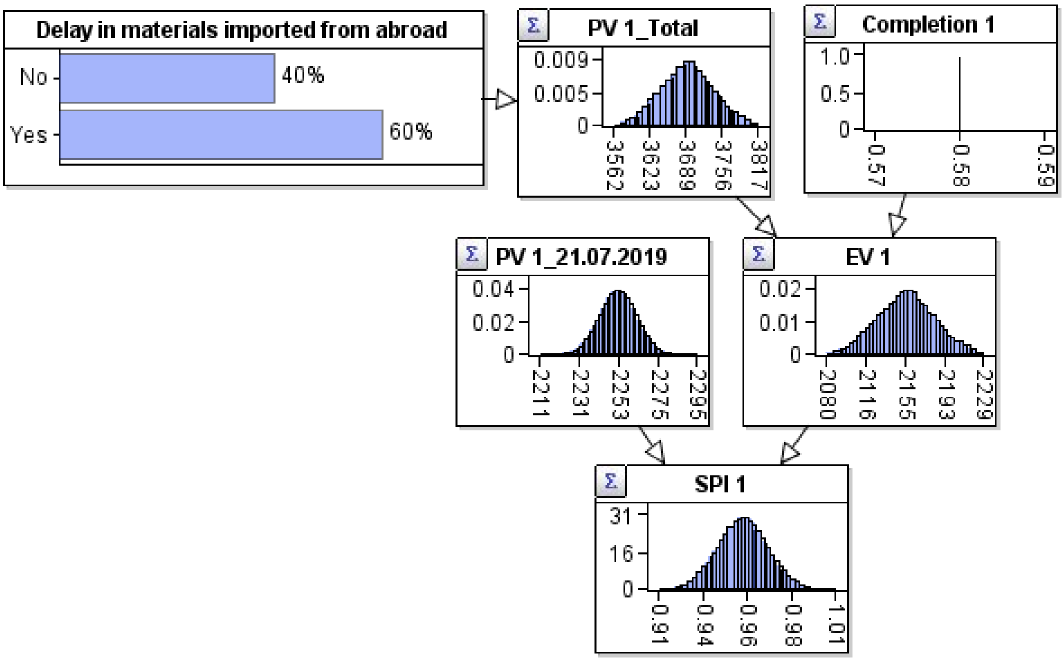


Activity of “*Analytics-Preparation*”

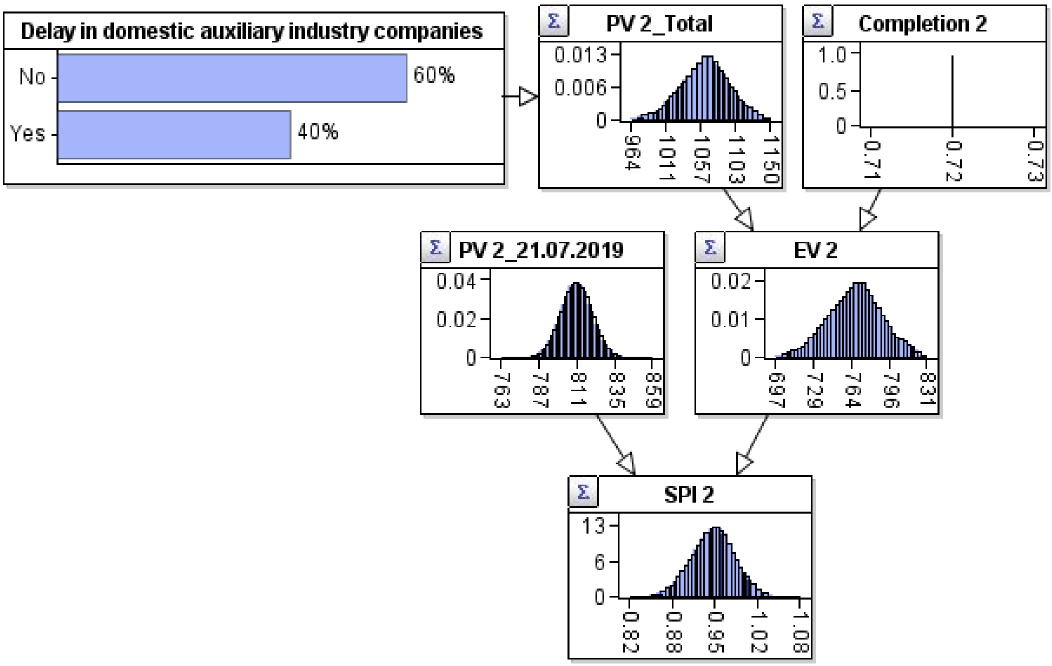


Models of Case 3 in AgenaRisk

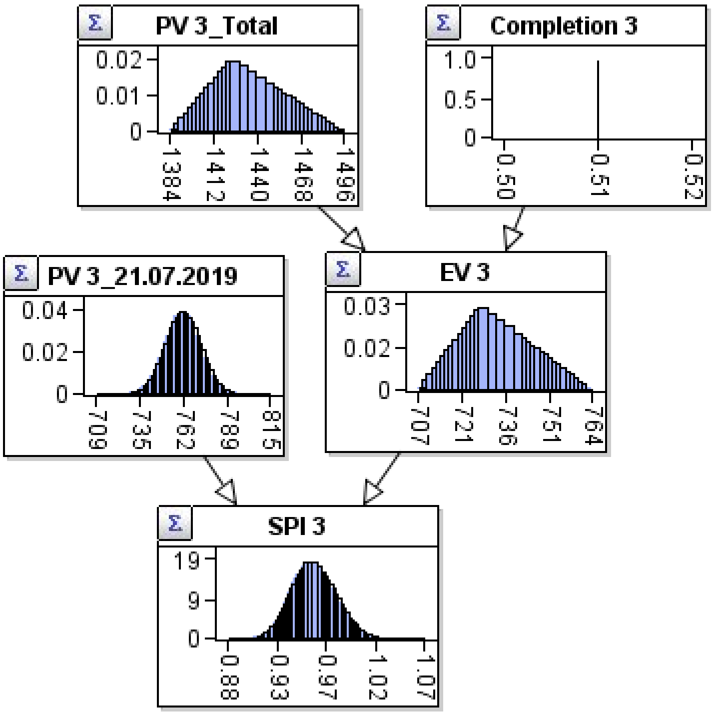
Milestone-1



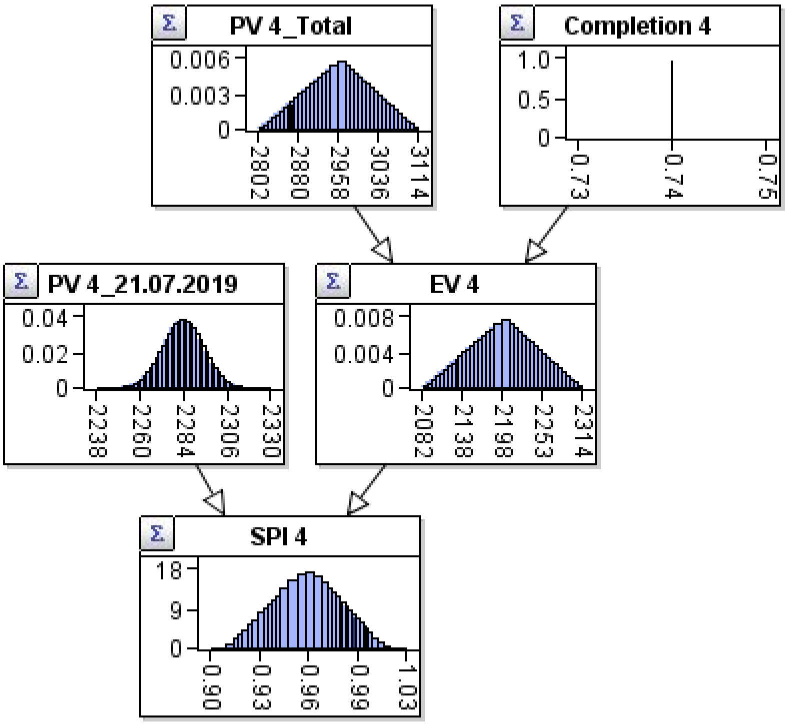
Milestone-2



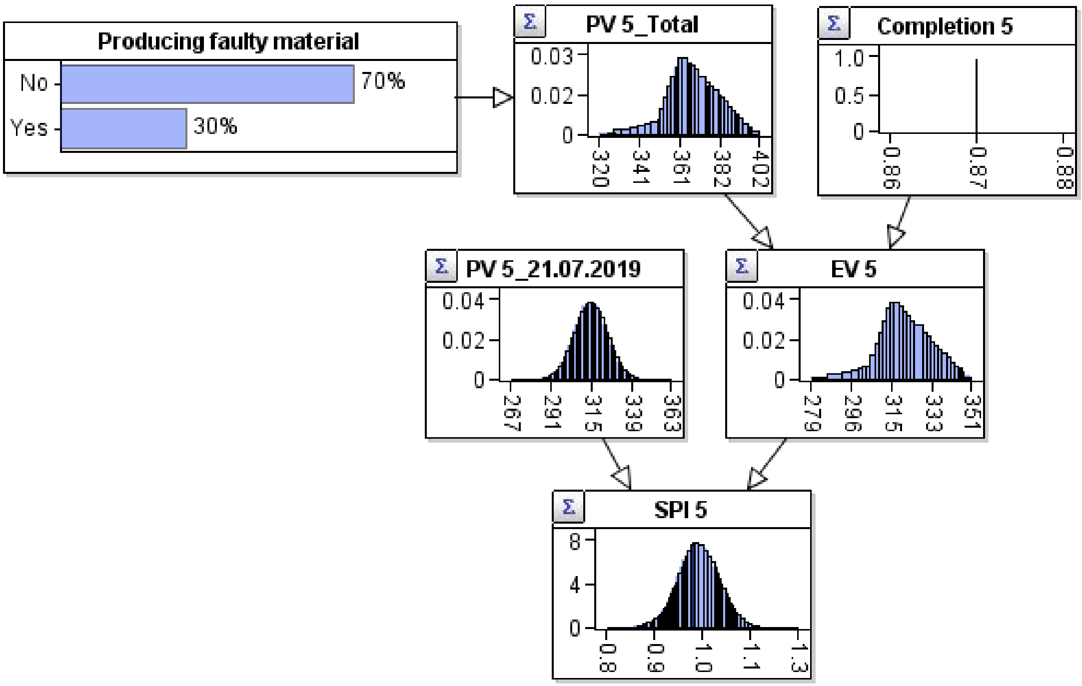
Milestone-3



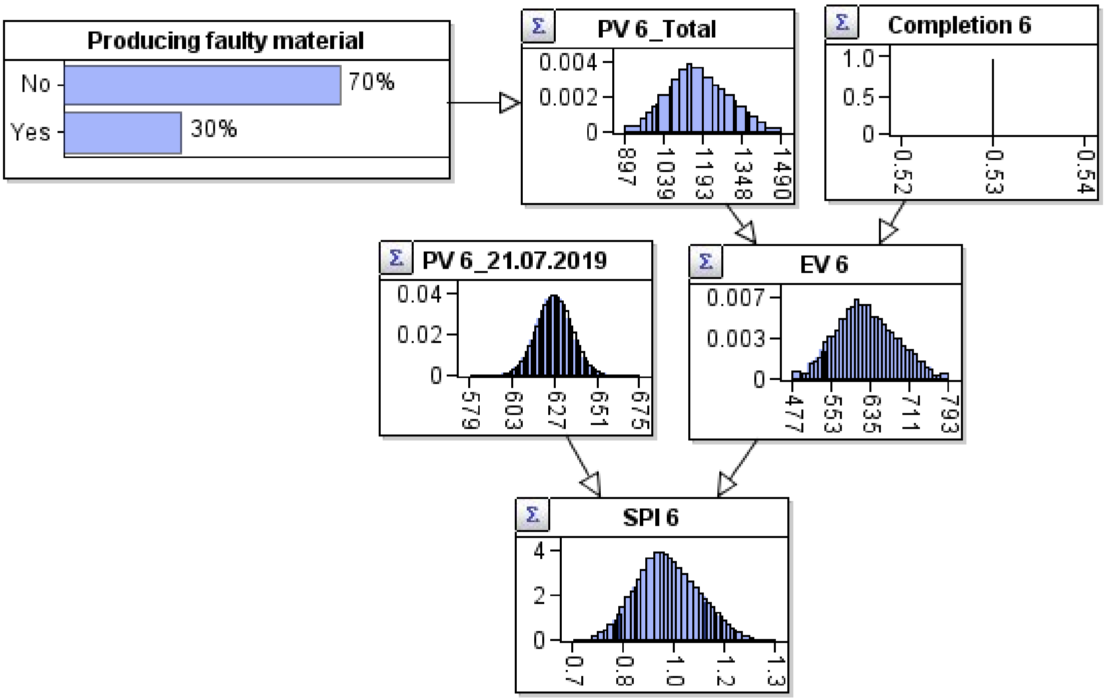
Milestone-4



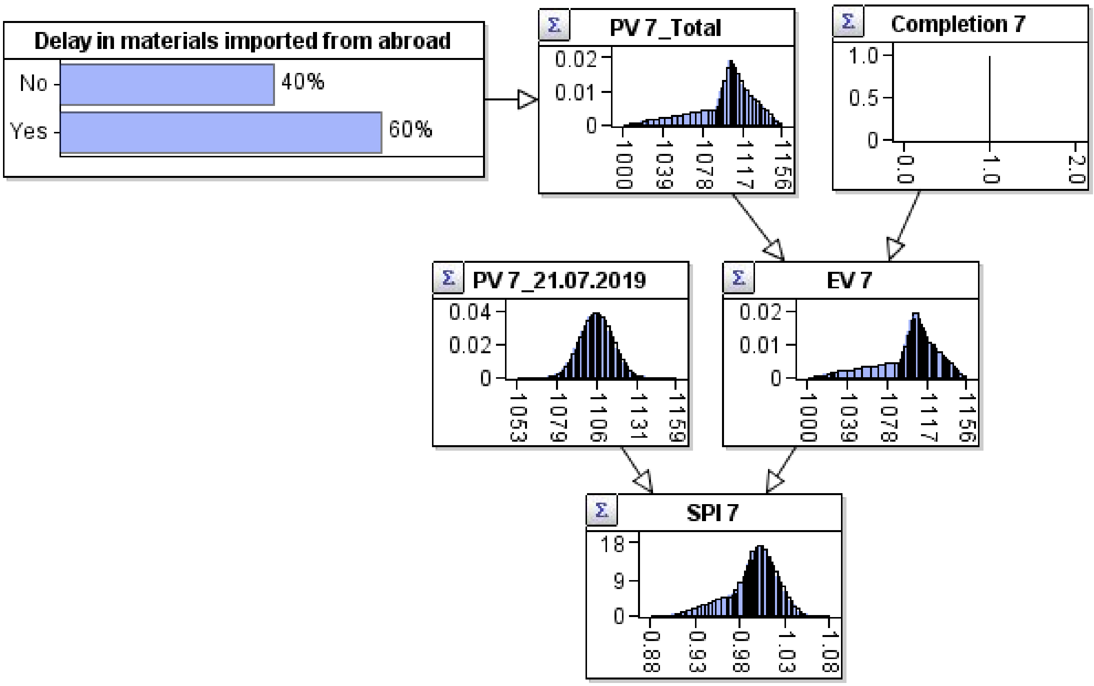
Milestone-5



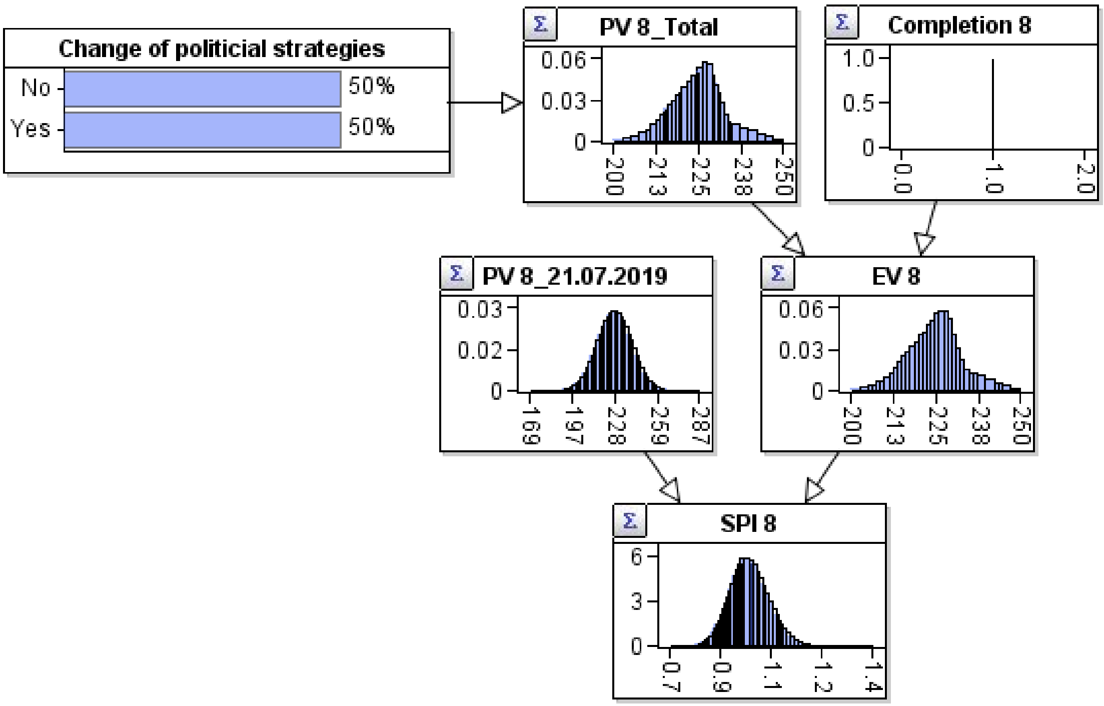
Milestone-6



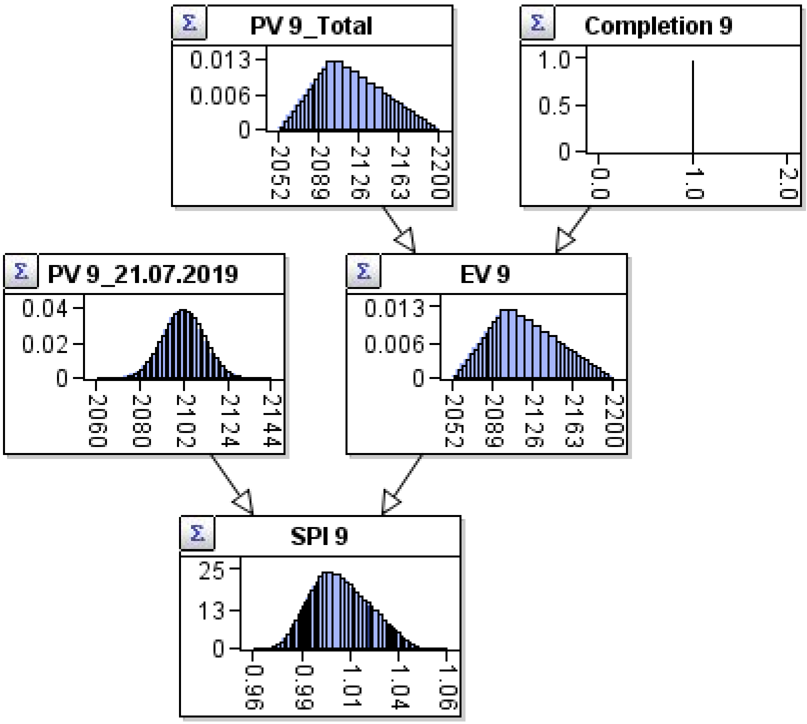
Milestone-7



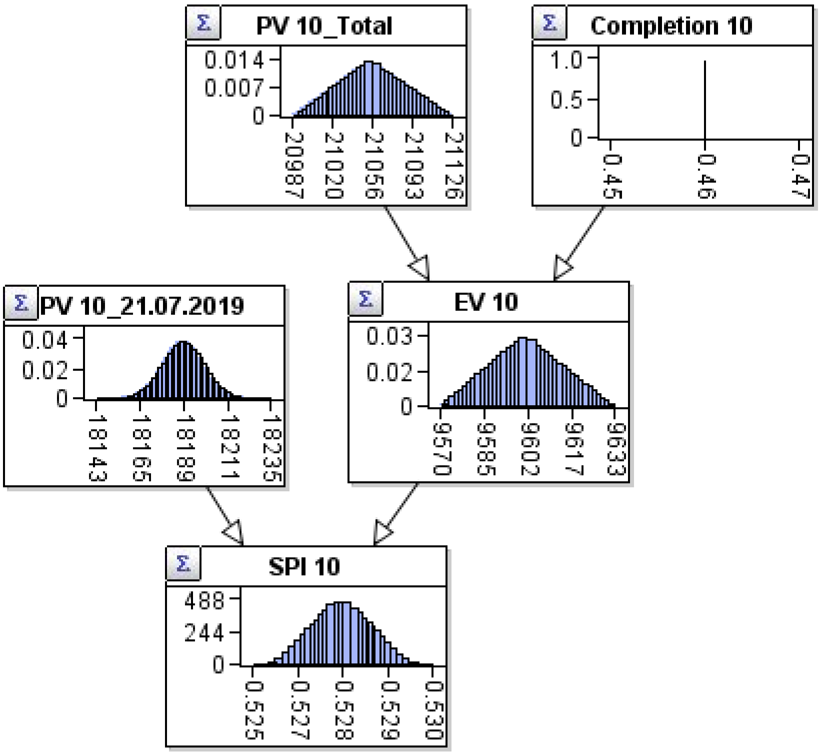
Milestone-8



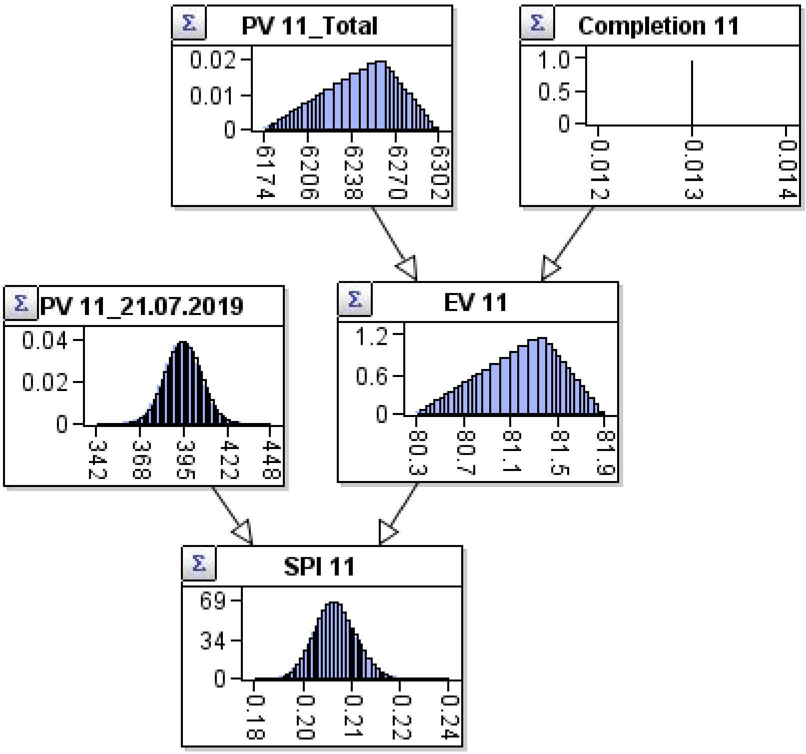
Milestone-9



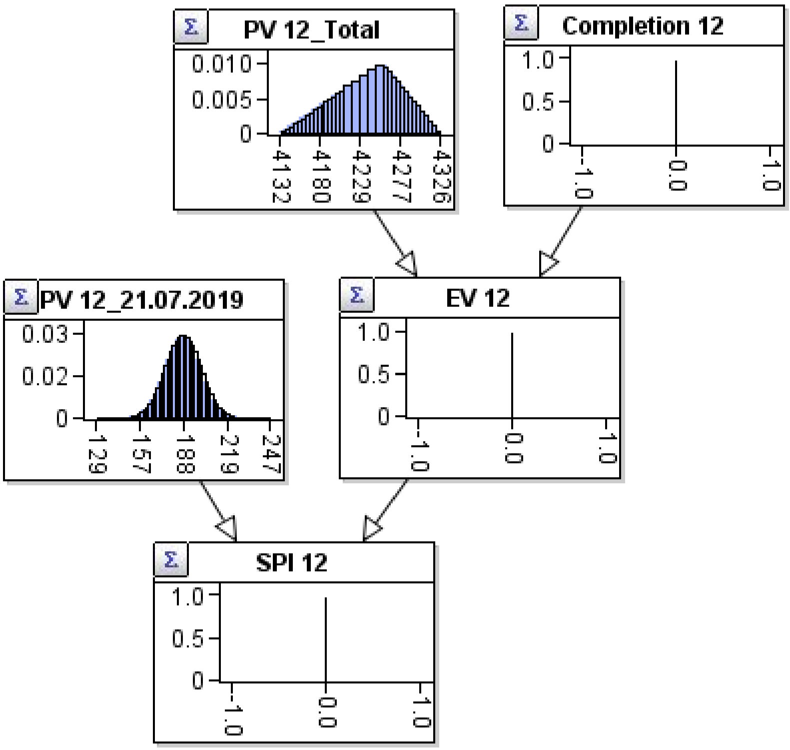
Milestone-10



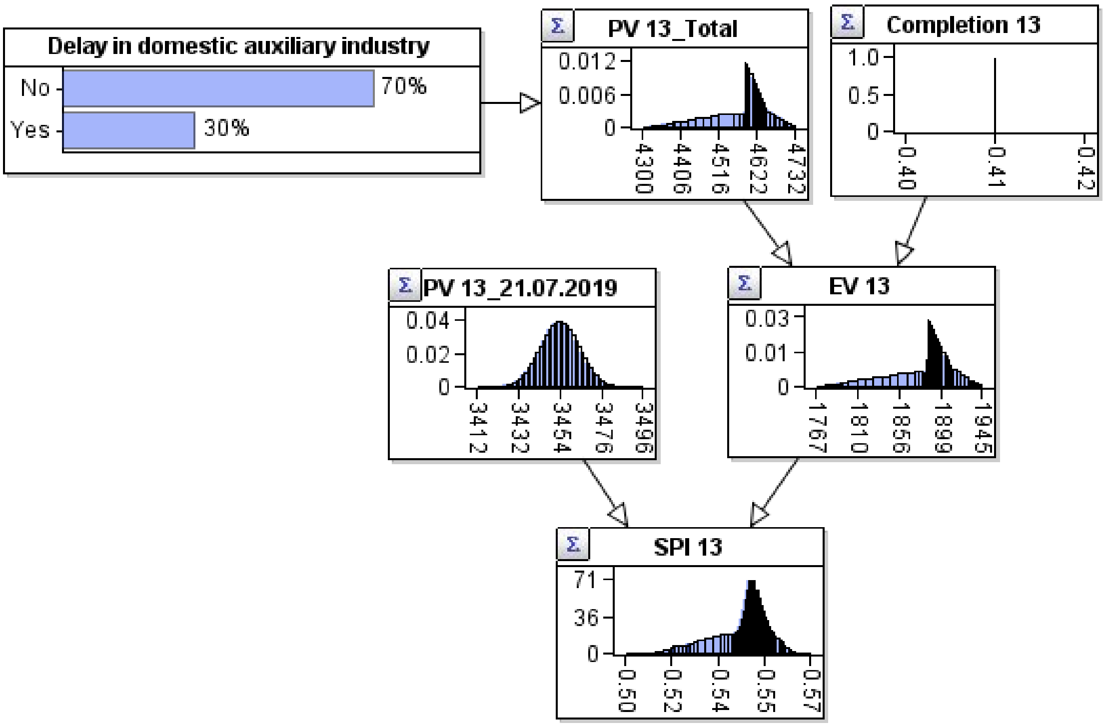
Milestone-11



Milestone-12



Milestone-13



APPENDIX 2 – Publications Based on the Thesis

Erhan Pişirir, Yasemin Sü, Barbaros Yet (2020), Integrating Risk into Project Control using Bayesian Networks, *International Journal of Information Technology & Decision Making*, DOI: 10.1142/s0219622020500315