TELKOMNIKA Indonesian Journal of Electrical Engineering
Vol.12, No.6, June 2014, pp. 4617 ~ 4623
DOI: 10.11591/telkomnika.v12i6.5438

A New Statistical Model to Estimate Information System Contingency Budget

Junwei Zeng*[1], Fajie Wei[1], Haitao Xiong[2], Anying Liu[1]
[1]School of Economics and Management, Beihang University,
Beijing 100191, China
[2]School of Computer and Information Engineering, Beijing Technology and Business University,
Beijing 100048, China
*Corresponding author, e-mail: zengjunwei@yeah.net

Abstract

Development of an information system is a complex process, which expose to a great number of risks.Hence, the high failure rates long associated with information system projects, despite advances in techniques for information technology development, suggest that organizations need to improve their ability to identify and to manage associated risks. To improve the risk management in information system development projects, a pragmatic procedure is suggested to determine the size of a project's contingency plan budget at any specified level of certainty. Considering the interaction among risk factors, a method based on common risk factors and copula functions is used to model and quantify positive dependence between risks.

Keywords: *information extraction, copula, information system development project, contingency plan budget, risk dependency*

1. Introduction

Even before the widespread use of “the mythical man-month” [1], information system (IS) project failure was described as a common phenomenon and many development projects did not achieve previously cost, schedule, and performance goals. It has been continuing to the present. Over last three decades, there has been considerable interest in exploring and explaining the reasons for the abnormal high failure rates. MacFarlan points out that failure to assess individual project risk is a major source of the IS development problem [2]. Then, many researchers, therefore, try to identify the various risks associated with the IS development [3]. Jiang et al. argue that the high failure rates associated with IS projects suggest that organizations need to improve not only their ability to identify, but also to manage the associated risks [4]. Also, Project management researchers state that risks in information system development projects are key factors affecting project success [5]. For example, Dillon et al. show the importance of risk management and contingency reserves for successful development of projects in complex and uncertain environment [6]. That research also recognizes the challenges of managing these reserves. More and more researches reveal that, to manage the inherent risks associated with dynamic environments and uncertainties (e.g. requirement changes in IS projects), enough resources should be assigned to projects [7-9].

A contingency plan is designed for the identified uncertainty, which means that people can have a relatively controllable environmental setting. This feature is very meaningful for the IS development projects because most of the changes can be predictable and project managers have a lots of alternative technological tools to reduce or constraint the impact of those predictable changes. Some project managers try to use a fixed deterministic percentage of project budget to capture and quantify the degree of confidence that the contingency plan should cover. Touran criticizes the use of this approach [10].

Budgeting for project contingency plan, an efficient way to reserve resources, has been studied by scholars [10-12]. Recently, Khamooshi and Cioffi propose a pragmatic procedure to determine the size of contingency plan budget for a project, a programs, or a system at any specified level of certainty confidence [8]. That method, like many other risk management approaches, considers risks as independent events.

Received December 29, 2013; Revised February 28, 2014; Accepted March 15, 2014

However, Kwan and Leung argue that hypothesis is counterintuitive as it is more likely that one risk would impact anther risks [13]. This is reasonable, especially in risk management of IS development projects. So, the independence assumption between these risks can be specious [14]. The long-standing issue of dependence between risks has recently been discussed in project risk analysis [15-17]. These researches unanimously model dependence using the copula approach [18, 19].

In this research, the combination of risk tolerance and statistical dependence will be analyzed and then be used to allocate the contingency plan budget in IS development projects. A applied procedure is proposed to manage the contingency plan budget of IS development projects, considering dependencies among risk events in IS projects. The procedure demonstrates how to formulate a project contingency plan and to allocate contingency plan budget for risks defined over the duration of an IS development project. Taking the coupling of blocks in a information system into account, the copula is introduced to model relationship between dependent risks to make the estimation more practical. The Monte Carlo methods are also used in the calculation of the joint distribution of risk events.

In the rest of this paper, a statistical dependence model for IS projects is firstly described. And then, a framework that enable project managers to work better with inevitable risks in projects is presented. The last section is the conclusion.

2. A Statistical Dependence Model for Information System

2.1. Methods Currently Used

Cioffi et al. [20] present a tool that helps managers to deal with “tactical risks” which are defined in [21]. Given a specified level of certainty, e.g. 99\%, their pragmatic procedure can be used to determine the size of a project's risk contingency budget. According to their description, the key ideal of the procedure can be summarized as: use binomial distribution to estimate the probability of occurrence of any specific number of risk events happening and then calculate the size of the potential damage corresponding with a given confidence level. More specifically, considering each individual risk event may or may not occur, people can represent all possible scenarios by combinations of these individual risk events. Finally, they suggest two different ranking schemes, sorting risk events by their impacts or their expected values, to set the contingency plan budget.

They assume that each risk event is totally independent of the other risk events and the increase in impact because of possible compounding effects could be included by aggregating connected, subsequent impacts into fewer, large impacts. So, the variables that should be considered are the total number of possible risk events, the average risk probability, the number of risk events that will occur, the given confidence level of any specific risk outcome, and the impact of the risk events.

Although extremely novel projects face risk events far out in the tails of a probability distribution, they point out that most organizations do not face these risk events on most projects and work on projects only slightly different from the ones they worked on before. This is especially true for IS development projects. For most of the IT companies, code reuse is very popular. In some large companies, almost all new information systems are developed by using the existing software architects and/or software development platforms.

The probability of occurrence of each risk event as well as its impact can be estimated by many risk management procedures. These estimation procedures are common practice and the results are often given as point estimates. People also can calculate the expected value of the loss by multiply the probability and the loss (usually written as, [22]). Both “P” and “I” are the basic inputs of the procedure and the accuracy is low because good risk models and hard data are rarely available.

Then, [20] try to use statistic methods to make approximation of an average occurrence probability of risks after they showing the probability of occurrence of many risk events at the same time is low. Only after getting that, can the binomial distribution be used to estimate the number of risks that can occur at any given confidence level.

There are two main drawbacks to get the accuracy and reliable results by using the above procedure in the IS development projects:

(1) A critical assumption in [20], the independence between risk events which is a common practice of current risk management approaches, is discussed by more and more

researches. Among these researchers, [17] states that it is intuitively obvious that the assumption is highly suspect for many large projects which may contain multiple similar activities or several different kinds of activities that can be influenced by common risk factors. [13] argues that the assumption is counterintuitive because it is more likely that risks in one area would impact risks in another area.

(2) Although it is much easier to estimate the probability of occurrence and damage separately, the limitation of the static estimation (most of the, point estimates) of damage/loss can not be ignored because the calculation of the potential damage is based on the value from risk identification and quantification. In practice, the damage caused by risk events can be several different levels of severity with different occurrence probability.

In the below subsection, a statistical dependence model is proposed to deal with the first drawback and a new procedure will be described in next section to overcome the second drawback.

2.2. A Copula-based Statistical Dependence Model

The important of relaxing the independent assumption has been clearly recognized by scholars.The copula approach is used to solve this issue by many researchers [15-17], [23]. A copula is a function that links the marginal distributions to the joint distribution, which is a statistical concept that relates random variables. Reference [17] discusses a simulation-based model to quantify positive dependence between uncertainty distributions of activities in a project network.Their model can provide a less cumbersome method to elicit dependency information from experts.Their research is useful in uncertainty analysis where dependence between random variables with a bounded support is present due to common factors. Later, [16] puts forward a model for building statistical dependence between marginal distribution.Wu et al. [23] employ multivariate copula to model the dependence among risk factors.

The core idea of using copula method to cope with dependency is the procedure to calculate the multivariate distributions between subsets of uncertainty distributions.After analyzing the methods to deal with two extremes caused by assuming the marginal distributions to be specified separately, the multivariate can be constructed by an intermediate method. The method assumes independence between marginal distributions and allows to use joint distributions for subsets of uncertainty distributions which share common risk factors. Actually, the ideal of latent variable models has been used when specifying the independence ([24, 25]).People can use many skills, e.g. brainstorming sessions, to identify the common risk factors in a specific project. The dependence diagrams can also be introduced into project risk analysis.Considering the construction of a rank correlation by project experts is impractical.

The Diagonal Band distribution introduced by [26] is suggested to be used under this environment [17]. A bivariate Diagonal Band distribution $D(U,V)$ of two uniform on [0,1] distributed random variables U and V is shown in Figure 1. To model the statistical dependence between risk events in a specific project, a multivariate distribution of risk events need to be modeled. A copula-based statistical dependence model is proposed in this paper.

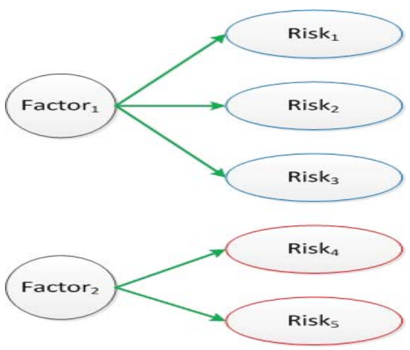

Figure 1. A Model for Statistical Dependence of Risks due to Common Risk Factors

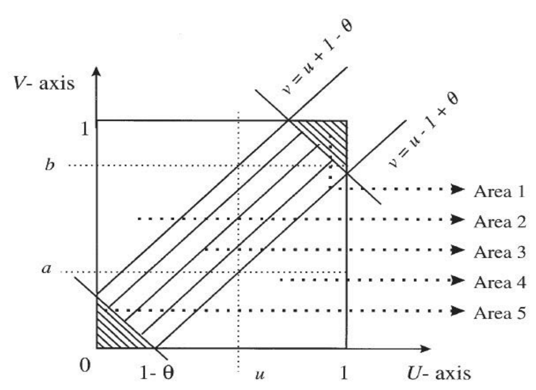

Figure 2. A Diagonal Band Distribution Seen from Above [17]

In Figure1, Risk_i is used to denote a specific risk event whose potential loss follows some kinds of distribution and $Factor_j$ represents a common factor that can impact on several risk events in a IS development project.

To estimate the potential loss of a IS project, a multivariate joint distribution, $F_0(\cdot)$, can be specified between $Risk_1, Risk_2, Risk_3, Risk_4$ and $Risk_5$.In most case, the marginal distributions, denoted by $F_{r_i}(Risk_i)$,of $Risk_1, Risk_2, Risk_3, Risk_4$ and $Risk_5$ are available (from estimation of project engineers and experts). By introducing new common influential factors (common risk factors or so-called latent variables), the independent common risk factor $Factor_1$ and $Factor_2$ (whose marginal distribution can be denoted by $F_{f_j}(Factor_j)$), joint distribution of five risk events, $F_0(\cdot)$, can be simplified.

It can be represented by the product of two independent distribution $F_{123}(Risk_1, Risk_2, Risk_3)$ which means the joint distribution for $Risk_1, Risk_2, Risk_3$ and $F_{45}(Risk_4, Risk_5)$ which means the joint distribution for $Risk_4$ and $Risk_5$.So, to get the two joint distributions, $F_{123}(Risk_1, Risk_2, Risk_3)$ and $F_{45}(Risk_4, Risk_5)$, the joint distribution between $Factor_j$ and $Risk_i$ (j=1, i=1, 2, 3 or j=2, i=4, 5) should be specified. The conditional independence can be used to simplify the joint distribution to a combination of several bivariate distributions [16]. According to the definition of copula, the joint distribution between $Factor_j$ and $Risk_i$ such as $Factor_1$ and $Risk_1$ can be uniquely determined by its associated copula.

In Figure 2, the DB copula, $D(U,V)$, is a bivariate Diagonal Band distribution of two uniform on [0, 1] distributed random variables (U and V) with one parameter (θ). The DB copula can be used to describe the relationships of risk events in Figure 2. Uncertainty information can be elicited and the marginal cumulative distribution function of $Factor_j$, $F_{f_j}(Factor_j)$ and the marginal distribution function of $Risk_i$, $F_{r_i}(Risk_i)$ also can be figured out. $F_{f_j}(Factor_j)$ and $F_{r_i}(Risk_i)$ contain all information on the marginal distribution. From standard distribution theory, all marginal distribution which is absolute continuous may be derived from the uniform marginal by an appropriate transformation.So, both $F_{f_j}(Factor_j)$ and $F_{r_i}(Risk_i)$ are related to uniform random variables on [0, 1].Hence, in Fig.2, U can be associated with $F_{f_j}(Factor_j)$ and V can represent $F_{r_i}(Risk_i)$.

Generally speaking, if considering the dependence of risks, the joint distribution of risk events, $Risk_i$, should be obtained.The problem can be simplified by introducing some independent common risk factors, $Factor_j$, and using a theory called conditional independence [27]. So, only the bivariate distributions of $Factor_j$ and $Risk_i$ need to be considered.

The Monte Carlo method for assessing variability and uncertainty in project risk analysis has become more common [28]. The reverse of procedure deriving copulas can be used to generate pseudo-random samples from general classes of multivariate probability distributions.That is, given a procedure to generate a sample from the copula distribution, the required sample can be constructed.More details can be found in [29].

3. New Framework for Budgeting

The procedure proposed in this paper inherits the theoretically sound foundations of the rank correlation method allowing the marginal distributions to be specified separately and is practical enough to be used by IS project analysts.It contains five steps, as shown in Figure 3.

Risk Identification: It has become the consensus that the requirement of risk identification is to determine which risk events are likely to affect the project and to document the characteristics of each potential risk event. Beside the normal requirements, the common risk factors, e.g., common code blocks, development framework et al., are also needed to be identified by experts and then risk events are grouped according to the common risk factors.The final result of this step can be illustrated by figures similar to Figure 1.

Risk Quantification: In this stage, risks and interactions among them need to be evaluated to assess the range of possible project outcomes. The potential impact and the probability of occurrence of each risk should be evaluated. Usually, a risk events can cause damage at several different levels with corresponding probabilities. The more scenarios with probabilities can be listed in this step, the more accurate the final result can be.

Risk Fitting: It is noticeable that, in this new procedure, people can take all uncertainty factors into consideration instead of dealing with impact and occurrence separately and deterministically. In other words, this procedure can eliminate the second drawback of the currently used method mentioned in section 2.1. Based on the scenarios from previous step, the distributions of the risk events can be simulated. There are a lots of tools can help people to translate the scenarios of a risk event into a suitable distribution.The uniform, triangular, binomial, Poisson, exponential, Student's t and normal are very popular distributions in project management and most of project managers are familiar with these distributions. The quality assurance (QA) teams in most of IT companies are using some efficient bug management software which can help to generate the reliable assessment of the distributions.

Risk Updating: This step is used to update the distribution of each risk in a IS project. The information extraction and the fitting of risk events do not contain the quantifiable dependency information. This design is reasonable because the impact of common risk factors are not easy to be accurately valued and the best practice of this work is to build a special team with experts in many fields.With the model in section 2.2, the description of distributions of the common risk factors can be avoided. So, the joint distribution of each risk events group can be calculated by considering the dependency structure.

Risk Cumulation: The amount of damage that the IS project will experience depends on risk events that will occur, which people cannot know specifically in advance. But, the expectation or a specific value given any level of certainty, e.g., 99\% confidence, can be calculated after analyzing the joint distribution of all risk events.

This procedure is very pragmatic. It simplifies the extraction process of expert opinion without loss of accuracy.

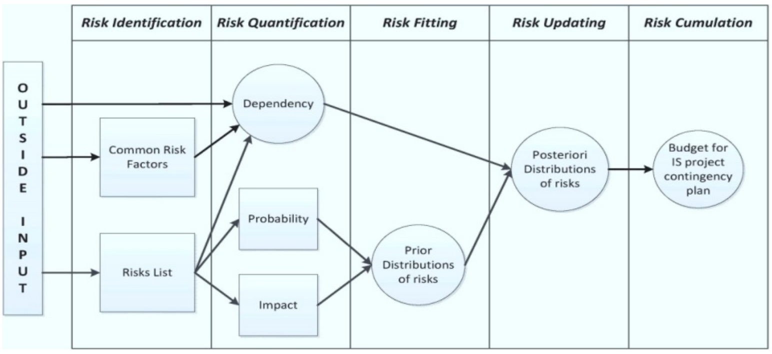

Figure 3. Procedure for Getting Contingency Plan Budget

4. Conclusion

Failures of IS development projects (cost overruns, schedule delays, poor quality) can cause enormous losses. Unfortunately, the rate of occurrence of failures is still high. Many scholars try to solve this problem by software engineering and project risk management. At the same time, contingency plan budget for IS development project, which is a very import tool for project managers to reduce the risk exposure, attracts more and more researchers. However, there are two weak points in existing methods: (1) The assumption about independence of risk events, which is obviously counterintuitive; (2) The deterministic description of risk events, which may cause data loss. The statistical model and procedure proposed in this paper can improve the accuracy of the estimation of contingency plan budget for IS development projects by overcoming the above two disadvantages. The five steps can help stakeholder to grab and extract information about risk events in the IS projects and translate them into a specific number when given a level of certainty. Diagonal Band distribution and Monte Carlo methods are used to quantify the dependency among risks.

Acknowledgements

This research was partially supported by the National Natural Science Foundation of China (No. 71201004, 71101153) and the Scientific Research Common Program of Beijing Municipal Commission of Education (No. KM201310011009).

References

[1] FP Brooks. The Mythical Man-Month. Boston: Addison-Wesley. 1975: 2-15.
[2] FW McFarlan. Portfolio approach to information systems. *Harvard Business Rev*. 1981; 59(5): 142–150.
[3] J Jiang, G Klein. Software development risks to project effectiveness. *Journal of Systems and Software*. 2000; 52(1): 3–10.
[4] JJ Jiang, G Klein, R Discenza. Information system success as impacted by risks and development strategies. *IEEE Transactions on Engineering Management*. 2001; 48(1): 46–55.
[5] PR Garvey, DJ Phair, JA Wilson. An Information Architecture for Risk Assessment and Management. *IEEE Software*. 1997; 14(3): 25–34.
[6] RL Dillon, ME Pate-Cornell, SD Guikema. Optimal Use of Budget Reserves to Minimize Technical and Management Failure Risks During Complex Project Development. *IEEE Transactions on Engineering Management*. 2005; 52(3): 382–395.
[7] JE Ramirez-Marquez, BJ Sauser. System Development Planning via System Maturity Optimization. *IEEE Transactions on Engineering Management*. 2009; 56(3): 533–548.
[8] H Khamooshi, DF Cioffi. Program Risk Contingency Budget Planning. *IEEE Transactions on Engineering Management*. 2009; 56(1): 171–179.
[9] U Ojiako, M Ashleigh, M Chipulu, S Maguire. Learning and teaching challenges in project management. *International Journal of Project Management*. 2011; 29(3): 268–278.
[10] A Touran. Calculation of contingency in construction projects. *IEEE Transactions on Engineering Management*. 2003; 50(2): 135–140.
[11] M Ranasinghe. Contingency allocation and management for building projects. *Construction Management and Economics*. 1994; 12(3): 233–243.
[12] KT Yeo. Risks, Classification of Estimates, and Contingency Management. *Journal of Management in Engineering*. 1990; 6(4): 458–470.
[13] TW Kwan, HKN Leung. A Risk Management Methodology for Project Risk Dependencies. *IEEE Transactions on Software Engineering*. 2011; 37(5): 635–648.
[14] J René van Dorp. *A Distribution for Modeling Dependence Caused by Common Risk Factors*. ESREL Conference Proceedings, Netherlands. 2003: 551–558.
[15] MR Duffey, JR Van Dorp. Risk analysis for large engineering projects: Modeling cost uncertainty for ship production activities. *Journal of Engineering Valuation and Cost Analysis*. 1998; 2: 285–301
[16] JR.van Dorp. Statistical dependence through common risk factors: With applications in uncertainty analysis. *European Journal of Operational Research.* 2005; 161(1): 240–255.
[17] JR van Dorp, MR Duffey. Statistical dependence in risk analysis for project networks using Monte Carlo methods. *International Journal of Production Economics*. 1999; 58(1): 17–29.
[18] C Genest, J Mackay. The Joy of Copulas: Bivariate Distributions with Uniform Marginals. The American Statistician. 1986; 40(4): 280-283.
[19] RB Nelsen. An Introduction to Copulas. Springer. New York. 2006: 226-232.
[20] DF Cioffi, H Khamooshi. A practical method of determining project risk contingency budgets. *Journal of the Operational Research Society*. 2009; 60(4): 565–571.
[21] S Chatterjee, RM Wiseman, A Fiegenbaum, CE Devers. Integrating Behavioural and Economic Concepts of Risk into Strategic Management: the Twain Shall Meet. *Long Range Planning*. 2003; 36(1): 61–79.
[22] T Williams. The two-dimensionality of project risk. *International Journal of Project Management*. 1996; 14(3): 185–186.
[23] D Wu, H Song, M Li, C Cai, J Li. *Modeling risk factors dependence using Copula method for assessing software schedule risk*. 2nd International Conference on Software Engineering and Data Mining SEDM 2010. Chengdu. 2010: 571–574.
[24] PW Holland, PR Rosenbaum. Conditional Association and Unidimensionality in Monotone Latent Variable Models. *The Annals of Statistics*. 1986; 14(4): 1523–1543.
[25] DJ Bartholomew, M Knott, I Moustaki. Latent Variable Models and Factor Analysis: A Unified Approach. West Sussex: John Wiley and Sons. 2011: 58-59.

[26] RM Cooke, R.\ Waij, Monte Carlo Sampling for Generalized Knowledge Dependence with Application to Human Reliability. *Risk Analysis*. 1986; 6(3): 335–343.
[27] V Musolino, A Pievatolo, E Tironi, A statistical approach to electrical storage sizing with application to the recovery of braking energy. *Energy*. 2011; 36(11): 6697–6704.
[28] CN Haas. On modeling correlated random variables in risk assessment. *Risk analysis*. 1999; 19(6): 1205–14.
[29] F Durante, C Sempi. Copula Theory: an Introduction. *Copula Theory and Its Applications*. 2010; 198(1): 3–31.