KEYWORDS Software Project Portfolio Selection Modern Portfolio Theory Project Selection Project Correlation

A Technique to Calculate Correlations between SOFTWARE PROJECTS in a Portfolio Selection Setting Based on the MODERN PORTFOLIO THEORY

Hélio R. Costa
COPPE / UFRJ – System Engineering and Computer Science Department, Federal University of Rio de Janeiro, Caixa Postal: 68511 – CEP: 21945-970 – Rio de Janeiro – Brasil
heliorcosta@fgvmail.br

Marcio de O. Barros
PPGI / UNIRIO – Applied Computer Science Department. University of Rio de Janeiro, Av. Pasteur 458, Urca – CEP: 22290-240 – Rio de Janeiro – Brasil

Ana Regina C. da Rocha
COPPE / UFRJ – System Engineering and Computer Science Department, Federal University of Rio de Janeiro, Caixa Postal: 68511 – CEP: 21945-970 – Rio de Janeiro – Brasil
darocha@centroin.com.br

ABSTRACT

Selecting software projects to build a portfolio is a risky task. The Modern Portfolio Theory (MPT) has been proposed by many authors to support selecting the best group of projects that maximize the expected return x risk ratio of a portfolio. However, one major issue has to be resolved to allow the application of the theory when migrating its concepts to a project selection context, i.e, the way correlations between projects are calculated. We present a new technique to calculate them objectively, based on a risk scenario approach.

INTRODUCTION

Project Portfolio Management *(PPM)* has gained attention in recent years, as organizations have become increasingly project, program, and portfolio-oriented *(Killen, Hunt, & Kleinschmidt, 2007)*. PPM is the management of the project portfolio, aiming to maximize the contribution of projects to the overall welfare and success of the company *(Project Management Institute – PMI, 2008)*. Cooper, Edgett and Kleinschmidt *(2001)* highlight three main goals for PPM: *(i)* maximizing portfolio value; *(ii)* selecting the right projects to comprise the portfolio; and *(iii)* linking the portfolio to the organization's business strategy. In their extensive research, and according to a set of metrics created to evaluate the achievements of enterprises adopting PPM, the authors have found that selecting the best projects to form the portfolio is the least precisely-defined goal and the one with the lowest performance in industrial settings. Thus, processes, methodologies, and techniques to support project selection in software development companies are required. Besides, international standards and models, such as the Capability Maturity Model Integration *(Chrissis, Konrad, & Shrum, 2006)*, the ISO/IEC 12207:2008 Software Life Cycle Processes *(ISO/IEC12207, 2008)*, and the Standard for Project Portfolio Management *(PMI, 2008)* have introduced or updated their PPM processes, providing further evidence on the demand for portfolio-related decision-making support.

The selection of projects to build up a company's portfolio is, thus, an important activity in PPM processes. It aims to define an optimal, or close to optimal, subset of candidate projects, considering their characteristics and the relationships between them *(Ding & Cao, 2008)*. Many approaches for selecting project portfolios can be found in recent technical literature reviews *(Cooper et al., 2001; Dye & Pennypacker, 2003; ISO/IEC12207, 2008)*. These approaches can be classified into three major categories: *(i)* scoring and ranking; *(ii)* mapping models; and *(iii)* financial methods. According to Biffl, Boehm, Aurum and Grümbacher *(2006)*, financial methods are the most commonly used approaches and most suitable to deal with the project selection problem, as methods from the two former categories are usually based on subjective assessments, which may be unreliable due to lack of explicit information and the pressure and biases imposed by hidden agendas. Among the financial methods, the Discounted Cash Flow models, Return on Investment, Internal Rate of Return, and Payback Analysis are commonly used by organizations, as they are relatively easy to understand and implement. More complex approaches, such as Productivity Index, Expected Commercial Value, Expected Monetary Value, and Real Options may also be found in more sophisticated settings, but their acceptance in companies is limited.

After reviewing the technical literature for project portfolio selection, we observed the lack of techniques that take into consideration the interdependencies between candidate projects. Project interdependencies are relationships between projects *(usually a pair of projects, say A and B)* that may increase or decrease the value or risk of a given project *(A)* when it is conducted in parallel to another one *(B)*. Recognizing the importance of these relationships and taking them into account while setting a portfolio is the basis of Harry Markowitz's Modern Portfolio Theory *(MPT) (Markowitz, 1952)*. This theory was initially proposed to support the creation of investment portfolios and aims to select a group of financial assets that maximize the return x risk profile of such a portfolio. Markowitz states that portfolio selection does not only involve selecting particularly well-performing assets, but assets that best combine with each other.

Authors like Boehm and Sullivan *(2000)*, Hubbard *(2007)* and Sullivan, Chalasani, Jha, & Sazawal *(1999)* have introduced the notion that software development is a value-driven activity and, thus, the construction of software systems can be analyzed in the same way as other types of investments. The use of MPT concepts to support project portfolio selection has been formerly addressed in the technical literature *(Appari & Benaroch (2010)*; Ball & Savage, *(1999)*; Blau, Penky, Varma, & Bunch, *(2004)*; Chien *(2002)*; Graves, Ringuest, & Case *(2000)*; Levine *(2005)*; Marchewka & Keil *(2005)*). In a careful analysis of the feasibility of applying these concepts to software projects, Zimmermann, Katzmarzik and Kundisch *(2012)* discuss the limits for its application. Nonetheless, one fundamental aspect with regards to MPT application is still not properly implemented in any of the approaches that we have analyzed: estimating correlation between a pair of projects.

In this paper we depart from the foundations laid out by Zimmermann et al. *(2012)* and present a new technique to calculate correlations between projects in order to support diversification while selecting project portfolios using MPT. The technique is based on the creation and analysis of risk scenarios to which the projects are exposed, taking into account the behavior of different projects while subjected to the same risks. We use a subset of the information required by the former approach to calculate such correlations.

Besides this introduction, this paper is organized into four more sections. In the next section, we introduce the concepts underlying the Modern Portfolio Theory and the assumptions that must be asserted to apply it in a project portfolio selection context. Section 3 discusses related works, especially with regard to how they calculate correlations between projects. Section 4 presents our proposed approach, as well as an example of its application. In section 5 we present the empirical studies planned and performed to analyze the approach. Conclusions are made in the last section.

1. Modern Portfolio Theory and Software Projects

The Modern Portfolio Theory *(MPT)* is a disciplined approach to support the allocation of resources in investment portfolios comprised of financial assets. A portfolio, according to the MPT, is a weighted combination of assets, where the weight of each asset is proportional to the amount of capital invested on it. The purpose of MPT is to define which proportion of the available capital an investor should allocate to each asset to maximize the return and minimize the risk incurred by an investment portfolio. The approach calculates the expected return and risk for each possible portfolio, which can be built from the available assets. Afterwards, the potential portfolios are depicted in a chart presenting the risks incurred by the portfolios (σ_p) on its horizontal axis and their expected return (ER_p) on its vertical axis. A typical chart is presented in **Figure 1**.

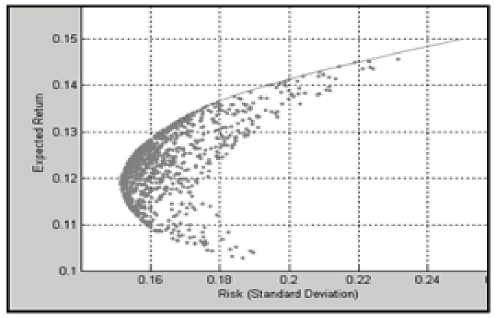


FIGURE 1. *Typical Efficient Frontier (www.mathworks.com)*

The Efficient Frontier, formed by the uppermost points set forth in the chart, presents all portfolios with maximum expected return for a given level of risk. Given how much risk the investor is willing to accept, the Efficient Frontier shows the portfolio with the greatest expected return. On the other hand, for a given expected return, the Frontier shows the portfolio with minimum risk. Thus, choosing portfolios on the Frontier is considered the rational decision for an investor. The Expected Return yielded by a portfolio (ER_p) is represented by the weighted sum of the expected returns of its assets. Considering a portfolio consisting of m assets, ER_p is represented by equation *(1)*, where w_i is the percentage of capital invested in asset i, $1 \le i \le m$ and μi is the average return provided by asset i.

$$ER_p = \sum_{i=1}^{m} w_i \mu_i$$
$$\text{subjected to } \sum_{i=1}^{m} w_i = 1 \quad (1)$$

The Risk of a portfolio (σ_p) is a function of the risks of its assets (σ_i), the proportion of capital invested in each asset (w_i) and the correlations $(\rho_{i,j})$ between them. The risk of any given asset is usually calculated as the standard deviation of its observed returns. The correlation between two assets shows how an asset affects and is affected by the presence of another asset in a given portfolio. It is a measure of the strength and direction of the relationship between two assets, represented by a number in the **[-1, +1]** interval, where -1 represents two assets moving in opposite directions with the same strength, while +1 represents two assets that tend to move in the same direction with the same strength. Correlation 0 *(zero)* means that there is no relationship between the two assets. Given the weights, the correlations, and the risks of its assets, the risk of a portfolio entailing m assets is calculated by equation *(2)*.

$$\sigma_P = \sqrt{\sum_{i=1}^{m}\sum_{j=1}^{m} w_i w_j \rho_{i,j} \sigma_i \sigma_j} \quad (2)$$

Optimum portfolios usually embed what the MPT defines as diversification, i.e., the combination of negatively-correlated assets. Such combination yields less risky portfolios, since a negative observed return on an asset is compensated by a positive observed return on another asset.

However, the MPT was originally developed to be used in a financial market context. Thus, to be applied out of this scenario, some assumptions taken by this theory must be understood and adapted for the project context. Zimmermann et al. *(2012)* provide an extensive conceptual foundation for the application of MPT within the scope of software development projects, as well as the discussion of assumptions assumed by MPT that must be considered and translated to the software development context in order to make such kind of approach feasible and reliable.

We have performed a Literature Review on how MPT is used to support project selection in portfolio build up. We focused on how MPT-based approaches for portfolio selection presented in the technical literature calculate project risk, correlations, portfolio risk, and portfolio expected return, as these are the basic concepts underlying the usage of MPT. In the next section, we demonstrate how different approaches use these concepts in the context of a project portfolio selection process.

2. Related Works

An approach found in Ball and Savage *(1999)* shows how to maximize the return x risk relation of oil and gas portfolios by means of a series of simulations. In this approach the authors calculate the risk of a project based on a series of expected values for that project instead of considering the standard deviation as preconized by the MPT. The correlations are also not considered in this approach and the risk of the whole portfolio is calculated by the sum of the risk of all projects composing a portfolio.

Based on the probabilities of success and failure of some projects and on the risks calculated by an algorithm named GINI, which calculates the mean absolute difference between two random returns from the same project, Graves et al. *(2000)* performed a series of simulations to define R&D portfolios with an optimum return x risk relation. The authors set up a simple linear program using EXCEL to minimize risk to a certain level of return. They run the linear program iteratively, increasing the required level of return after each run. The results are plotted on a chart, forming an Efficient Frontier. Finally, they screen the portfolios on this Efficient Frontier to find those preferred by risk-averse decision makers. Again, this approach neither takes the correlations between projects into account nor calculates portfolio risk.

Another demonstration of using MPT for selecting oil and gas project portfolios can be observed in Zimmermann et al. *(2012)*. In this approach the authors arbitrate values for risks, three possible returns, and correlations. All these values are estimated by the decision maker. Based on these values, the authors vary the proportion of investment on each project and try to find the best combination to maximize the return x risk relation. One major criticism to this approach lies on the difficulties of precisely estimating the risk for a project, as well as correlations between projects. The lack of support to systematically define these values undermines the practical use of the proposed approach.

An example of using MPT on pharmaceutical projects can be found in Ball and Savage *(1999)*, where the concepts are used to combine a set of candidate projects in order to help with the creation of a feasible project pipeline for the portfolio. However, the authors do not provide enough details on how to calculate the risk of a project, the correlations are not established, and the risk of the portfolio is defined as the probability of having a negative return. A fixed value for the risk of projects is also used by Ding and Cao *(2008)*, but the author does not calculate portfolio risk, because correlations are not calculated.

Another approach is a work developed by Apari and Benaroch *(2010)* in the software project context using COCOMO parameters Boehm and Sullivan *(2000)* to analyze project cost sensibility to risks. Based on cost variations, the standard deviation is calculated and project risk is defined. The approach does not calculate the correlations and as some of the research works discussed above, the authors define the risk of a portfolio as the sum of the risks of the projects belonging to the portfolio.

Marchewka and Keil *(2005)* present an MPT-based portfolio selection approach in which risk probability is a weight that expresses the comparative importance of each risk for a given project. In this approach project risks and returns are represented as percentages, which we believe are less intuitive for a project manager to interpret than nominal values. Marchewka and Keil *(2005)* calculate the correlation between a pair of projects according to a linear relation based on risk weights *(representing relative importance, but not probability of occurrence)* and percentage over/under the initial estimate for returns as a measure of risk impact *(a percentile of the original value instead of an amount of monetary value)*.

Other MPT-based approaches to project portfolio selection assume that past data exists and estimate portfolio risk and expected return from this data. Buhl and Heinrich *(2008)* present a technique to select a customer portfolio using risk and expected return. Their approach aims to calculate an optimum portfolio applying MPT, but correlations and project risks are calculated according to previous projects earnings. By means of historical data from all possible customers forming a segment *(group of similar projects)*, they form a portfolio with different correlations to minimize the risks and maximize the return yielded by clients of different segments. By means of average incomes of customer segments *(last ten years)* they calculate the standard deviation of each segment and then the correlation between two segments. The ideas underpinning the proposed approach are perfectly aligned with the concepts of MPT, but it requires data on previous returns observed for each segment. Such data is not available when selecting a project portfolio as projects are, by definition, unique enterprises, *(PMI, 2008)*.

Ozdemir and Kersten *(2004)* also use MPT to create an optimum product portfolio. By taking the monthly returns of eighteen IT products of a given seller company, they calculate the expected return of each product *(mean)*, risk *(standard deviation)*, and correlation between the products. Then, an Efficient Frontier is assembled and an optimization of the portfolio of products is made. Once again, the usage of MPT concepts in the proposed approach is valid, but it also required previous historical data about the products, data which is not available when trying to form a portfolio with new, yet to run projects.

Finally, another interesting approach to select projects based on MPT was developed by Zimmermann et al. *(2012)*. The model underlying this approach establishes and takes into account the assumptions to support a correct implementation of MPT in a project portfolio selection context. This approach forms a portfolio of projects located at different sites maintained by an IT service provider company. It considers the costs, returns, and risks of each project, along with interdependencies *(correlations)* amongst these projects and the sites where they are executed. The authors present a case comprised of three different sites and three different projects to illustrate the use of the model. They estimate the costs and define different scenarios for each project considering the risks they are prone to. For each scenario they calculate the expected changes on project costs and the *(subjective)* probabilities of occurrence for each scenario. Based on these independent scenarios, the authors build a decision tree representing all possible scenarios and their respective occurrence probabilities driving to different costs. Having such data, the risk *(standard deviation)* of each project is calculated.

As regards the estimates of interdependencies *(correlations)* between the projects, the authors affirm that there are two kinds of correlations: *(i)* projects being conducted at the same site, and *(ii)* projects being conducted at different sites. For projects being conducted at the same site the correlations are calculated as follow: based on three similar resources, the authors define the interdependencies according to the amount of resources needed by the projects. The more resources needed simultaneously by two projects, the stronger the correlation between them. Four different correlations are defined: High *(1)*, Mediocre *(0.67)*, Low *(0.33)* and 0 *(No correlation)*. For a project conducted at different sites, different factors such as political, economical, legal, and cultural differences were considered. Economical theories well described in the paper are used to estimate interdependencies between the sites.

Some comments must be made about the approach proposed by the authors: *(i)* they only handle positive correlations, thus inhibiting diversification between negatively correlated projects; *(ii)* sharing resources may affect correlation between projects, but it is probably not the single source for dependencies; *(iii)* it might be the case that resources from different sites are shared and such is not considered in the model; *(iv)* the approach does not consider positive risks.

Given the limitations on how previous research works have addressed the estimation of correlations between projects, the next section presents our proposed approach for such estimation.

3. Calculating Correlations between Projects

At financial markets, the basic inputs for MPT are the historical series of the observed asset returns, based on which risk *(standard deviation)*, expected return *(mean)* and correlations between assets can be calculated. However, due to their uniqueness, historical series for project returns does not exist and the differences must be calculated differently. In our approach, this is achieved by analyzing the estimated frequency of occurrence of risk scenarios and their impact upon the expected return of each candidate project, allowing the creation of a time series of possible expected returns for each project. The observed returns under distinct scenarios for the selected projects under risk scenarios are used instead of historical information on the original MPT application.

One of the most important variables used on MPT is the correlation calculated between each pair of assets. This variable takes up this importance by allowing the assembly of portfolios with negatively correlated assets and thus allowing diversification to minimize the risk of a portfolio. In Section 3 we have demonstrated the gap that exists in the technical literature when considering the estimation of correlation. Based on the expected return of each project and the risks that may affect them *(considering their probabilities of occurrence and impacts)*, we present a new technique to calculate the correlation between candidate projects, likely to take part in a portfolio. The idea is to fill the gaps left by other approaches and improve the technique presented by Zimmermann et al. *(2012)*, turning it to a more generic and complete technique.

The first step is to establish the list of candidate projects to form the portfolio. Assuming that an investment project has an NPV *(difference between the sum of the discounted cash flows that will be generated by the project and the discounted cash flow of money invested to fund project execution)*, we take this value as the basis of our calculation. The NPV is used in detriment of other project valuation methods, such as ROI, IRR or Payback because it considers the cash flow of a project in a time horizon and corrects the value of money over time through a given discount rate *(Ross, Jordan, & Westerfield, 2008)*. Moreover, regardless of the length of different projects, they can be compared according to their NPV, since their cash flows are brought to the present time. However, despite the fact that NPV is the most used financial method to select projects *(Chien, 2002)*, being the only source of information to support this selection may lead to incorrect decisions: used separately, NPV cannot account for uncertainties that may affect future cash flows *(Cooper et al., 2001)*. This is a typical situation faced by software development organizations, due to risks incurred while executing the projects. **Table 1** presents a hypothetical project portfolio selection situation with 5 projects along with their NPV. These projects will be used throughout this section as an example of how correlations are calculated by our approach.

Project	NPV
P1	$13,500
P2	$14,575
P3	$25,000
P4	$35,500
P5	$10,025

TABLE 1. *Project Characterization for Portfolio Selection*

The second step collects information about risks incurred by the candidate projects that may form the portfolio. Risks that affect more than one project are of special interest, because they allow observing how these projects behave if exposed to the same uncertainties, providing the basis to measure dependencies between them. In our approach, risks which may affect more than one project must be identified and information about their probability of occurrence and expected impact upon each project *(positive for opportunities or negative for threats)* must be collected. Probabilities are represented as percentiles, varying from 0% *(no risk)* to 100% *(the risk will occur throughout project execution)*. Expected impact is measured in money, to be comparable to the project's NPV. In software projects, examples of risks that may affect more than one project include creeping user requirements, implementation of new technologies, possibility of components reuse, human resources issues, support from senior management, and low productivity. **Table 2** presents a set of five risks that might affect the five projects shown in **Table 1**, along with their probability of occurrence and impact *(positive or negative)* upon each project.

In the next step, all possible risk scenarios are created by combining subsets of the formerly identified risks. These scenarios can vary from no risk to all the risks occurring simultaneously. Given *n* risks, the total number of scenarios will be 2_n. Each scenario is characterized by the probability of its occurrence and its impact upon the NPV of each candidate project. This kind of risk scenario estimation is similar to the approach proposed by Aubert, Patry and Rivard *(1998)*. Assuming that all risks are independent, the probability of occurrence for each scenario is given by multiplying the probability of occurrence of all the risks that take part in the scenario, times one minus the probability of the other risks. **Table 3** presents an example of such calculation, showing an excerpt of the 32 scenarios that can be composed from the five hypothetical risks presented in **Table 2**. It must be highlighted that the sum of all occurrence probabilities for these scenarios

Risk	Probability	P1	P2	P3	P4	P5
#1	50%	$8,500	$40,000	$10,000	$00,00	$12,000
#2	30%	$3,800	-$11,750	$6,000	-$8,500	-$2,850
#3	50%	$3,000	-$550	$4,000	-$4,200	-$250
#4	90%	$3,500	$2,250	$2,500	-$5,000	$2,250
#5	10%	$2,000	-$4,000	$1,500	-$4,800	-$100

TABLE 2. *Risk Characterization for Portfolio Selection*

Scenario	Risk 1	Risk 2	Risk 3	Risk 4	Risk 5	%
#1	no	no	no	no	no	0.01%
#2	yes	no	no	no	no	0.01%
#3	no	yes	no	no	no	0.01%
.	.	.	.	.	.	.
.	.	.	.	.	.	.
#31	yes	yes	yes	yes	no	0.06%
#32	yes	yes	yes	yes	yes	0.07%
					Total	100%

TABLE 3. *Risk Scenarios and Probabilities of Occurrence*

Scenario	P1	P2	P3	P4	P5
#1	$0	$0	$0	$0	$0
#2	$8,500	$40,000	$10,000	$00,00	$12,000
#3	$3,800	-$11,750	$6,000	-$8,500	-$2,850
.	.	.	.	.	.
.	.	.	.	.	.
#31	$18,800	$29,950	$22,500	-$17,700	$11,150
#32	$20,800	$25,950	$24,000	$22,500	$11,050

TABLE 4. *The Impact of Risk Scenarios upon Each Project*

must be equal to 100%, indicating that they can support an analysis of the candidate projects under all possible situations to which they may be subjected.

To estimate the impact of each risk scenario upon a candidate project, we sum the impact of each risk making up the scenario upon the desired project. **Table 4** presents the impacts imposed by the scenarios presented in **Table 3** upon the NPV of projects presented in **Table 1**.

Each scenario will yield a different return for each project, generating all possible Expected Returns of a project, simulating historical returns. The weight of each Expected Return is, thus, proportional to the probability of occurrence of the scenario under which it was calculated. Based on these Expected Returns, we can calculate the Average Expected Return *(ER)* of each project, the Risk *(σ)* of each project, which is represented by the standard deviation of the series, as well as the Correlations *(ρ)* between projects shown in **Table 5**. Observe that some project pairs are positively correlated, while some are negatively correlated. According to the MPT, an efficient portfolio is made up by negatively correlated assets, because the risk of an asset is partially compensated by the risk of the second one. On the other hand, portfolios with strongly, positively correlated assets are exposed to higher risk.

Thus, based on these possible returns, the Average Expected Return *(ER)* of each candidate project can be calculated by the weighted average return that each risk scenario yields for the project, where the weights are given by the scenario's probability of occurrence. Similarly, the risk of each project *(σ)* can be calculated by the weighted standard deviation of the return that each risk scenario yields for the project. Finally, the correlation *(ρ)* between two projects is calculated using Spearman Rank Order Correlation *(Spearman, 1904)*, since the conditions necessary to use the Pearson Correlation *(Pearson, 1907)*, usually employed in a financial context *(normally-distributed data)*, may not be met by project data *(Triola, 2009)*.

With the Average Expected Return *(ER)* and Risk *(σ)* of all projects, as well as the correlations *(ρ)* between them, it is possible to calculate the Risk of any given portfolio (σ_p) and its Expected Return (ER_p), to form the Efficient Frontier proposed by the MPT and choose the best return/risk ratio portfolio. The Risk of all possible portfolios (σ_p) is calculated by the equations of the MPT stated in equation *(2)* in section 2. The Expected Return of the Portfolio (ER_p) is calculated by adding the Average Expected Return *(ER)* of each project making up a portfolio. The result is an Efficient Frontier depicting the risk x return relation of all possible portfolios that can be assembled by the candidate projects, allowing the decision maker to chose which portfolio best fits his or hers interests. An example of the Efficient Frontier is presented in **Figure 2**. Others examples can be found in Zimmermann et al. *(2012)*.

This process provides significant differences between our approach and the ones previous listed in the literature: *(i)* correlations are objectively calculated, since some approaches use subjectively estimated correlations; *(ii)* positive and also negative correlation values in the continuous **[+1, -1]** interval can be determined. This is fundamental for diversification. This advantage was obtained due to the consideration of possible positive and/or negative influence of risks in the projects. Most of the approaches consider risks only as a negative event; *(iii)* it is a full scenario risk-based technique, i.e., all possible impacts of risks and expected returns are taken into account. Some of the approaches consider only some of them, which render them incomplete; *(iv)* it is independent of the type and size of the projects, because, regardless of these factors, the risk scenarios will submit candidate project to their impacts; and *(v)* all possible correlations between the interval **[-1, +1]** can be calculated. This is what we consider the most significant contribution, because none of the approaches listed before can calculate them, and thus, the correct calculation of portfolio risks are not properly defined, what may lead to incorrect decisions, due to a non efficient diversification process.

4. Empirical Studies

To evaluate the feasibility of our technique and to compare its results to those produced by other project portfolios selection techniques found in the literature or used by practitioners, four empirical studies were planned and executed. The main goal of these studies was to analyze the approach in typical situations faced by a software development company trying to maximize the overall return x risk profile of its portfolio. The studies were planned so that our proposed approach could be tested in different scenarios, so as to observe its applicability and differences from other approaches, specially the efficiency of diversification provided by the composition of portfolios combining inverted correlated projects.

Study Designs

The first study consisted in forming a portfolio with five candidate projects and a budget constraint of $100,000. In the second study, besides the previous candidate projects, other four projects were added to the list, in order to form a new

	P1	P2	P3	P4	P5
P1	1.00	0.56	0.99	-0.33	0.56
P2	0.56	1.00	0.54	0.53	0.99
P3	0.99	0.54	1.00	-0.31	0.54
P4	-0.33	0.53	-0.31	1.00	0.50
P5	0.56	0.99	0.54	0.50	1.00

TABLE 5. *Correlations between Projects*

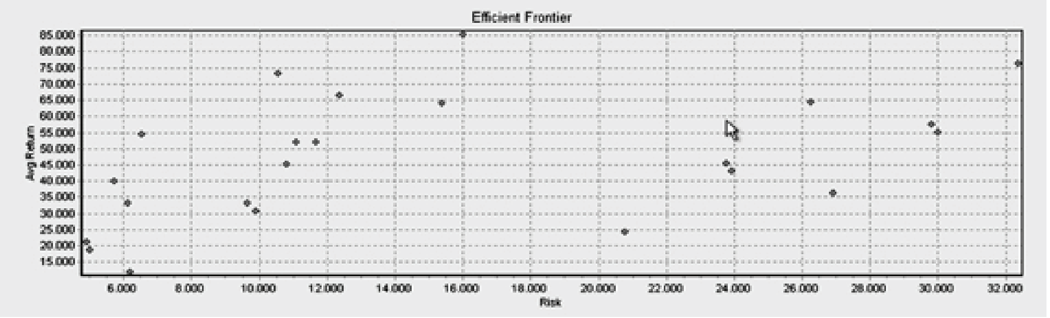


FIGURE 2. *Efficient Frontier*

portfolio with a budget constraint of $150,000. The goal of this second study was to increase the complexity of the decision, since five projects *(first study)* allow forming 32 different portfolios, while 512 different portfolios can be formed with nine projects. In the third study the same nine candidate projects were provided, but the budget constraint was raised to $200,000. The goal was to give the participants more flexibility to select different project combinations and observe which decision would be made if more resources were available. Finally, the fourth study was structured as follow: given the projects selected in the first study and the four new projects provided in the second and third ones, the participants should select a new portfolio with a budget of $200,000 *(that is, which of the four new projects should be added to a portfolio already entailing the projects selected in the first study)*. The goal here was to verify the applicability of the approach in a situation in which a new portfolio shall be formed based on an ongoing one, without cancelling the projects already under development.

Five participants were chosen to take part in the studies. The participants were chosen according to convenience, because it was understood that selecting experienced participants, who make decisions regarding the selection of software projects as part of their jobs, would be fundamental to inhibit major conclusion validity threats. Therefore, we contacted entrepreneurs *(mostly, C-level software company managers)*, and some of them volunteered to take part in the study. Each participant chose a project portfolio selection technique *(the technique that the participants customarily apply when such a decision is required in their day-to-day business)* without intervention from the researcher. To facilitate comparing the proposed approach with these techniques, the participants were instructed to use the same technique throughout the four studies.

The first participant used a Bubble Diagram in which the return x risk relation was used to prioritize and balance the portfolio. The participant used a set of criteria *(not disclosed to the researcher)* to estimate the risk of each project. The second participant used a group of risks to calibrate the most likely return and cost of the projects using Monte Carlo simulations. Next, a return x cost relation was defined and the projects were selected to maximize this relation until the budget limit was reached. The third participant used a multicriteria technique on which projects were evaluated and prioritized according to a set of criteria *(also not disclosed to the research)*. Then, projects were selected up to the available budget limit. Participant number four chose a return x cost relation to identify the most attractive projects. Then, the projects were evaluated and ranked according to a set of criteria *(such as project duration, type of technology and benefits provided by the projects)*. Finally, the fifth participant selected the candidate projects by their return x risk relation, calculating risks through the evaluation of a set of previously defined criteria. As the participants were C-level managers for software companies, most of them opted for non-disclosure regarding the complete rationale for the selection process, arguing that this rationale conveyed trade secrets.

The instruments used in the studies included information about a fictitious company *(goal, investment areas, budget availability, and internal policies)*. Candidate projects *(characterized by a brief description, their requirements, expected costs returns, and the risks they were exposed to)* were provided by the participants and were real projects developed by their companies. Risks affecting these projects *(description, probabilities and impacts)* were also provided by the participants. The projects were typical to a software development company such as web-based system for a financial enterprise, a warehouse control system developed for a city hall, and a truck fleet surveillance system. Average cost of the project was about $35,000.00. The average return was about $8,600.00. Finally, each risk was characterized by its probability of occurrence and its impact upon the projects. The risk to which the projects were submitted were *(i)* possibility to generate new business; *(ii)* degree of technology domain; *(iii)* human resource availability to develop the project; *(iv)* possibility to reuse components; and *(v)* need to hire external personnel. Each risk could, depending on the project, turn into a threat or an opportunity. A pilot study was conducted to ascertain whether the procedures and instruments had been properly prepared before submitting them to the participants of the study.

Study Results

After each study, the portfolios selected by the participants *(from now on identified as P1, P2, P3, P4, and P5)* were compared to the results of the proposed approach in relation to their position in the Efficient Frontier and their Portfolio Performance Index *(PPI)*, since the goal was to maximize the portfolios' return x risk profile. As the proposed technique is able to analyze all possible portfolios which can be formed and determine their returns, risks and performance indexes,

Participant	Portfolio	Cost ($)	Return ($)	Risk ($)	PPI
P1, P3, P4, P5	1, 3 and 4	90,850	55,530	29,935	1.86
P2	3, 4 and 5	98,125	46,851	29,036	1.61
A1	1, 4 and 5	90,525	53,132	26,247	2.02
A2	1, 3 and 5	83,875	41,025	12,565	3.26

TABLE 6. *Results of the First Study*

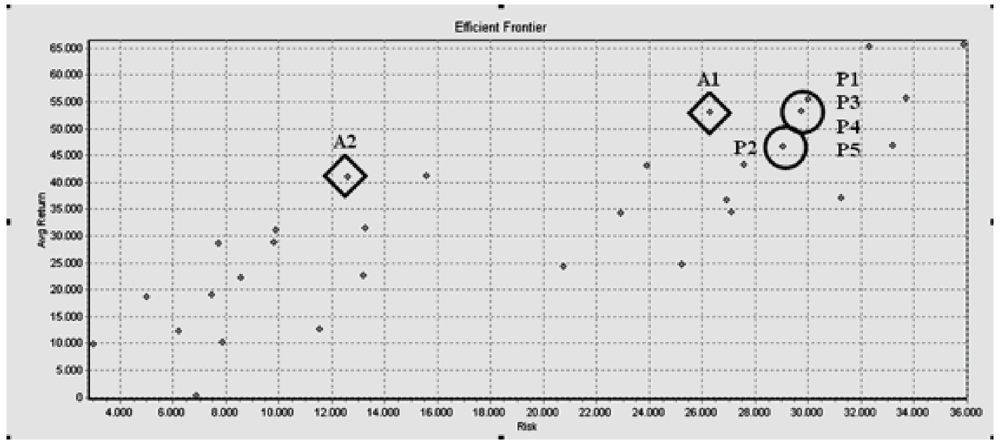


FIGURE 3. *Efficient Frontier of the First Study*

regardless of the results obtained by the participants, they could be compared to the results proposed by our approach.

Table 6 and **Figure 3** present, respectively, a summary of the results and the Efficient Frontier of the first study, where five candidate projects were provided and the budget constraint was set to $100,000.00. The Candidate projects that contain the selected portfolios are represented by numbers such as *(1, 2, 3, 4, and 5)*, and the Alternative Portfolios suggested by our proposed approach are identified by *(A1 and A2)*.

It can be observed that the portfolios suggested by our approach are better positioned in the Efficient Frontier than those selected by the participants, since they yield a similar expected return, as they are exposed to lower risk. It should be noted that despite the lower expected return of portfolio A2, its risk is about 60% smaller than the risk of the other portfolios, leading to a higher PPI and thus, to a potentially attractive option for a conservative decision maker.

In the second study, nine candidate projects *(numbered from 1 to 9)* were presented and the budget constraint was set to $150,000.00. **Table 7** and **Figure 4** present, respectively, a summary of the results and the Efficient Frontier of the second study.

Taking as example the portfolios selected by P2 and P4, it is seen that both of them have roughly the same expected return of A1, but are exposed to a considerably higher risk than the latter, making them non-rational choices in the MPT sense. The same situation is observed for P1, P3, and P5, compared to A2.

The third study was a variation of the second one, but the participants had a budget constraint of $200,000. **Table 8** and **Figure 5** present, respectively, a summary of the results and the Efficient Frontier of the third study.

As can be observed, the PPI presented by A1 is almost the double of the PPIs of portfolios suggested by the participants, and is positioned at the Efficient Frontier, which turns it into an optimal choice, despite its lower *ER*.

Finally, the fourth study was designed to ascertain how the techniques behaved facing the problem of forming a new portfolio by complementing an ongoing one. **Table 9** and **Figure 6** present, respectively, a summary of the results and the Efficient Frontier of the fourth study.

Portfolio A1 has almost the same expected return as the portfolio suggested by participants P1, P3, P4, and P5, but has a lower risk, thus representing a better choice. Also observed is that A2 is a much more attractive option than the portfolio proposed by P2, which is more conservative than the former ones, but which does not exploit the full return that can be achieved by accepting that level of risk.

Study Discussions

The studies have shown that, in all situations proposed, our approach presented more efficient portfolios than the techniques selected and used by the participants. These results provided indications that the proposed technique is feasible and might be efficient if the required information is available. Some topics can be highlighted from our studies, and justify the results obtained.

Participant	Portfolio	Cost ($)	Return ($)	Risk ($)	PPI
P1	1, 3, 4, 6 and 7	145,850	99,257	36,599	2.71
P2	3, 4, 5 and 7	135,125	79,996	35,751	2.44
P3	1, 3, 4, 6 and 7	145,850	99,257	36,599	2.71
P4	1, 3, 4, 6 and 9	137,325	78,207	38,510	2.03
P5	1, 4, 6, 7 and 9	143,725	98,925	36,585	2.70
A1	1, 6, 7 and 8	142,000	84,748	12,103	6.92
A2	1, 4, 5, 6 and 7	145,525	96,721	33,285	2.91

TABLE 7. *Results of the Second Study*

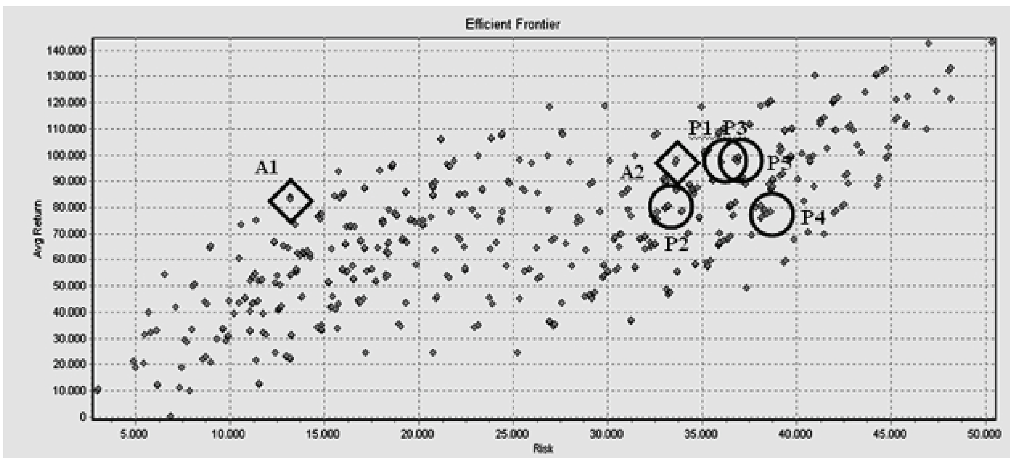


FIGURE 4. *Efficient Frontier of the Second Study*

First, correlation calculation is fundamental to an optimal choice, because it allows considering diversification while building up portfolios. It was observed that the more negatively correlated projects are used in a given portfolio, the higher its PPI, exactly as advocated by MPT. In this sense, our correlation calculation approach brings a substantial contribution to the whole process, since this variable directly influences the portfolio risk definition and consequently the risk x return relation.

Second, as the participants' techniques did not allow the calculation of overall portfolio risk, as recommended by the MPT, they can be misled to sub-optimal decisions. Similarly, the choice of a portfolio only by its expected return may lead to sub-optimal choices.

Also, the Efficient Frontier supports the analysis, since it provides a visual comparison of the return x risk profile of all portfolios. Risk scenarios allow for calculating expected returns, risks, correlations, and the creation of the Efficient Frontier, since it provides the possibility to submit the project to different situations and observe how they behave.

The more candidate projects are available, the more difficult it is to apply portfolio selection techniques, since analysis complexity is increased exponentially, and correlations cannot be studied. This fact was cited by the participants during the studies. On the other hand, the proposed approach is not influenced by these factors, because all correlations are calculated and all possible portfolios are analyzed in each scenario, and visualized in the Efficient Frontier.

We observed that even in conditions where it was not possible to choose a portfolio in the Efficient Frontier *(due to budget restrictions)*, the proposed approach has produced better results than the other techniques. We recognize that to properly apply our proposed approach a software company must have mature processes so that it can provide good cost, return and risk estimates.

Another important discussion is the fact that a project's risk is calculated by using the standard deviation of its returns. It is a consensus in the financial market that the standard deviation is not the perfect way to calculate the risk of an asset. We acknowledge that in a project selection context, risk might be influenced by intangible factors such as reputation *(gain/loss)*, user satisfaction, and schedule variation. In our approach we assume that the influence of these intangible factors is converted into monetary value and fed as part of the impact imposed by risks upon projects. We acknowledge that such conversion may not be simple, but proposed solutions for pricing these issues are out of the scope of the present paper. On the other hand, standard deviation is a globally accepted strategy to estimate risk. Thus, we decided to estimate project risk based on the observed fluctuation of its returns under different risk scenarios created to evaluate project portfolios and follow the original concept underlying MPT.

Considerations made by Wohlin et al. *(2000)* highlights that in every empirical study there are threats that may jeopardize the validity of the study. With respect to the internal validity, we can declare that all participants worked in real software development companies and performed this type of activity in their daily routine. Besides, each participant had the opportunity to apply his own technique and form his portfolios in the time and place judged appropriate. Thus, the formation of the portfolios was based exclusively on the participants' experience and their selected techniques, without direct influence from the researcher.

We agree that a higher number of participants and techniques *(actually, we involved only 5 participants)* would increase reliability of the results drawn from the studies, but time and budget restrictions prevented this possibility. Moreover, we must consider that the techniques selected by the participants belonged to different categories of portfolio selection techniques *(Cooper et al., 2001)*, which increases the coverage of the results and reduces the risk of external validity.

In order to minimize the construction validity threat, all information about the projects and the risks were provided by the participants from real projects they developed. Thus, the researcher did not embed biases, which could benefit one technique or another. Also, after each study, the participants returned their results to the researcher, who compared them to the results provided by the proposed approach. These results were not returned to the participants until the end of the fourth study, so that one previous result would not influence performing the following studies. Finally, a pilot study was conducted to improve the instruments provided to the participants, minimizing the risk that any given participant would benefit from a particularly fitted set of materials. Moreover, adjustments on the instruments due to criticism received during the pilot study helped clarify and improve the instruments.

The conclusion validity threat was reduced when four experimental studies were performed allowing us to analyze the approach at different situations compared with different categories of portfolio selection techniques. Also, the same instruments were provided to all participants. Besides, the fact that different techniques were used by different participants led to preliminary conclusions that the approach can be considered a better option in distinct scenarios when compared to other techniques.

Participant	Portfolio	Cost ($)	Return ($)	Risk ($)	PPI
P1, P3, P4, P5	1, 3, 4, 6, 7 and 9	173,925	111,531	42,845	2.60
P2	3, 4, 5, 7 and 8	198,625	101,277	35,127	2.88
A1	1, 5, 6, 7, 8 and 9	199,850	105,996	21,143	5.01

TABLE 8. *Results of the Third Study*

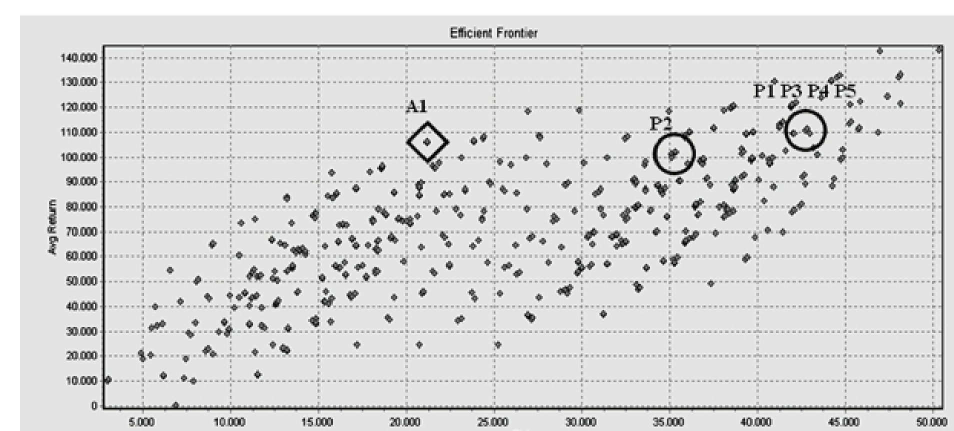


FIGURE 5. *Efficient Frontier of the Third Study*

Participant	Portfolio	Cost ($)	Return ($)	Risk ($)	PPI
P1, P3, P4, P5	1, 3, 4, 6, 7 and 9	173,925	111,531	42,845	2.60
P2	3, 4, 5, 6 and 8	179,625	78,802	33,919	2.32
A1	1, 4, 5, 6, 7 and 9	173,600	109,035	39,351	2.77
A2	1, 5, 6, 7 and 8	171,775	93,932	15,744	5.97

TABLE 9. *Results of the Fourth Study*

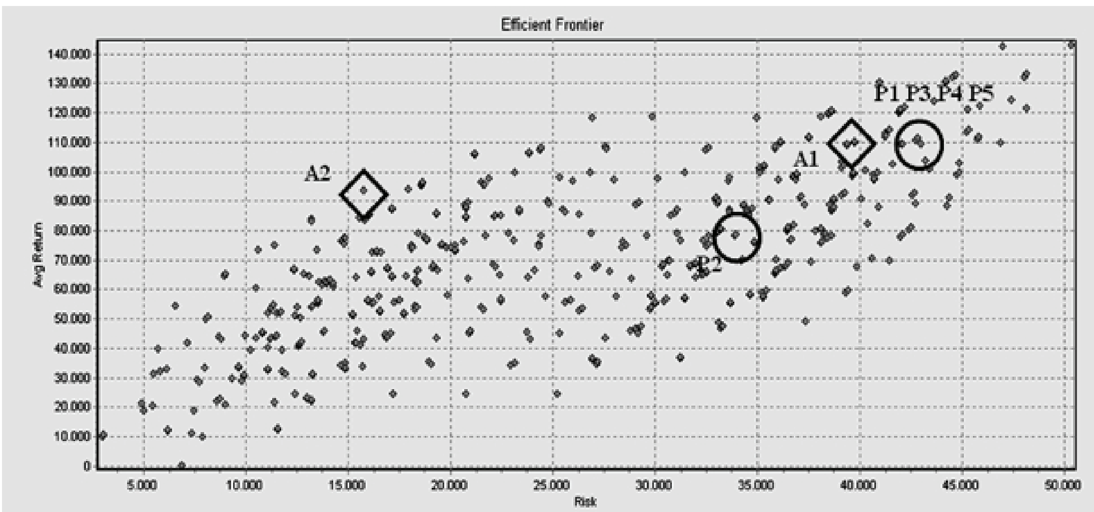


FIGURE 6. *Efficient Frontier of the Fourth Study*

5. Conclusion

In this paper, we have presented the problem of selecting projects to build up a portfolio. Next, we have introduced the concepts underlying the Modern Portfolio Theory and cited some related works that use this theory to support project portfolio selection. We have highlighted the problem of the correlation calculation in these works.

We have presented a new way to calculate the correlation and an MPT-based project portfolio selection approach, as well as an example of its application. Finally, we have demonstrated the applicability of our approach by conducting some empirical studies and the advantages of our correlation calculation approach.

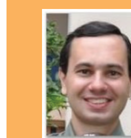
Some advantages observed in our approach in comparison to those formerly presented in the literature include: *(a)* the use of NPV as an economical parameter to compare all projects, regardless their duration; *(b)* taking into account both positive and negative risks, which enables observing all uncertainties which may affect the projects; *(c)* taking into account interdependencies between projects; *(d)* using future estimations for revenue under risk to calculate correlations between projects; and *(e)* calculating the overall risk of a portfolio. The approach can be used not only in the original formation of a portfolio, but also to analyze new portfolios based on an ongoing set of projects. The definition of the relationship between the projects *(inclusion or exclusion)* allows decision makers to impose restrictions and reduces the effort in analyzing portfolios that cannot be formed due to such restrictions.

But the most important contribution of our approach, in our point-of-view, is calculating the correlations between projects non-subjectively and covering the complete range of correlation, i.e., the **[-1; +1]** interval, which represents a solution to an issue left open in the literature, according to our review.

During our research we had planned to apply our correlation calculation technique with some of the approaches listed on section 3, but the absence of rough data used for the approaches rendered this attempt an unfeasible task.

authors

Hélio R. Costa has a Ph.D. degree in Computer Science and System Engineering. Former PMO leader of strategic softwares projects at Brazilian Air Force. Guest MBA Professor at Fundação Getúlio Vargas and Fundação Dom Cabral. Research Areas: Project and Portfolio Management, Risk Management, Decision Making and Simulation.



Marcio de O. Barros has a Ph.D. degree in Computer Science and System Engineering. Associate Professor at Universidade Federal do Estado do Rio de Janeiro. Reviewer of journals like Software Engineering and Knowledge Engineering, Empirical Software Engineering. Research Areas: Project Management, System Dynamics, Modeling and Simulation.

Ana Regina C. da Rocha has a Ph.D. degree in Computer Science and System Engineering. Associate Professor at Universidade Federal do Rio de Janeiro. She has coordinated the development of the Brazilian Software Process Model. Research Areas: Software Quality, Software Process and Knowledge Management.

We consider this a fruitful future research and put ourselves and our data available to perform empirical studies to compare the efficiency of the MPT approaches, previously listed, especially what concerns the correlations between projects.

references

Appari, A., & Benaroch, M. *(2010).* Monetary pricing of software development risks: a method and empirical illustration. The Journal of Systems and Software, 83, 2098-2107.

Aubert, B. A., Patry, M., & Rivard, S. *(1998).* Assessing the risk of IT outsourcing. Proceedings of the Annual Hawaii International Conference on System Sciences, Hawaii, USA, 31.

Ball, B. C., & Savage, S. L. *(1999).* A new era in petroleum exploration and production management. Notes on Exploration and Production Portfolio Optimization.

Blau, G. E., Penky, J. F., Varma, V. A., & Bunch, P. R. *(2004).* Managing a portfolio of independent new product candidates in the pharmaceutical industry. Journal of Product Innovation Management, 21, 227-245.

Biffl, S., Boehm, B. W., Aurum, A., & Grümbacher, P. *(2006).* Value-based software engineering. Berlin, Germany: Springer-Verlag.

Boehm, B. W., & Sullivan, K. *(2000).* Software economics: a roadmap, in the future of software engineering. International Conference on Software Engineering, 22.

Buhl, H. U., & Heinrich, B. *(2008).* Valuing customer portfolios under risk-return-aspects: a model-based approach and its application in the financial services industry. Academy of Marketing Science Review, 12*(5)*.

Chien, C. F. *(2002).* A portfolio evaluation framework for selecting R & D projects. R&D Magazine, 32*(4)*.

Chrissis, M. B., Konrad, M., & Shrum, S. *(2006).* CMMI: guidelines for process integration and product improvement *(2nd ed.)*. Addison-Wesley Professional.

Cooper, R. G., Edgett, S. J. & Kleinschmidt, E. J. *(2001).* Portfolio management for new products *(2nd ed.)*. Cambridge, MS, USA.

Ding, W., & Cao, R. *(2008).* Methods for selecting the optimal portfolio of projects. IEEE International Conference on Service Operations and Logistics, and Informatics.

Dye, L. D., & Pennypacker, J. S. *(2003).* Project portfolio management: selecting and prioritizing projects for competitive advantage. Glen Mills, PA: Center for Business Practices.

Graves, S. B., Ringuest, J. L., & Case, R. H. *(2000).* Formulating Optimal R&D Portfolios. Project Management Journal, 24*(4)*, 5-13.

Hubbard, D. *(2007).* How to measure anything. Finding the value of intangibles in business. Hoboken, NJ, USA: John Wiley & Sons.

ISO/IEC12207. *(2008).* Systems and software engineering — software life cycle processes *(2nd ed.)*.

Killen, C. P., Hunt, R. A., & Kleinschmidt, E. J. *(2007, August).* Managing the new product development project portfolio: a review of the literature and empirical evidence. PICMET Proceedings, Portland, Oregon, USA.

Levine, H. A. *(2005).* Project portfolio management: a practical guide to selecting projects, managing portfolios, and maximizing benefits. San Francisco, CA, USA: John Wiley.

Marchewka, J. T., & Keil, M. *(2005).* Portfolio theory approach for selecting and managing IT projects. Information Resources Management Journal, 8*(4)*.

Markowitz, H. M. *(1952).* Portfolio selection. Journal of Finance, 7*(1)*, 77-91.

Ozdemir, S., & Kersten, B. *(2004).* Portfolio optimization. Amsterdam: Vrije Universiteit.

Pearson, H. W. *(1907).* Deformation and variation in the sea-level. Geological Magazine *(Decade V)*, 4, 115-121.

Project Management Institute. *(2008).* The standard for portfolio management. Newtown Square: Project Management Institute.

Ross, S. A, Jordan, B. D., & Westerfield, R. W. *(2008).* Essencials of corporate finance *(6th ed.)*. McGraw-Hill.

Spearman, C. *(1904).* The proof and measurement of association between two things. Amer. J. Psychol, 72-101.

Sullivan, K., Chalasani, P., Jha, S., & Sazawal, V. *(1999).* Software design as an investment activity: a real options perspective. In L. Trigeorgis *(Ed.)*. Real options and business strategy: application to decision making. Risk Books.

Triola, M. F. *(2009).* Elementary statistics *(11th ed.)*. Addison-Wesley.

Wohlin, C., Runeson, P., Höst, M., Ohlsson, M., Regnell, B., & Wesslén, A. *(2000).* Experimentation in software engineering – an introduction. Kluwer Academic Publishers.

Zimmermann, S., Katzmarzik, A., & Kundisch, D. *(2012).* IT sourcing portfolio management for IT services providers – an approach for using modern portfolio theory to allocate software development projects to available sites. Data Base for Advances in Information Systems, 43*(1)*, 24-45.