# Estimating the Accuracy of the Return on Investment (ROI) Performance Evaluations


## Abstract

**Background:** Return on Investment (ROI) is one of the most popular performance measurement and evaluation metrics. ROI analysis (when applied correctly) is a powerful tool in comparing solutions and making informed decisions on the acquisitions of information systems.

**Purpose:** The purpose of this study is to provide a systematic research of the accuracy of the ROI evaluations in the context of the information systems implementations.

**Setting:** NA

**Intervention:** NA

**Research Design:** Literature review method was used to gather and analyze information related to the accuracy of estimating project costs and returns, distribution functions of errors. Measurements theory and error analysis, specifically, propagation of uncertainties methods were used to derive analytical expressions for ROI errors. Monte Carlo simulation methodology was used to design and deliver a quantitative experiment to model costs and returns estimating errors and calculate ROI accuracies. Spreadsheet simulation (Microsoft Excel spreadsheets enhanced with Visual Basic for Applications) was used to implement Monte Carlo simulations.

**Data Collection and Analysis:** This study reviews multiple publications to collect data on the accuracy of costs and benefits.

**Findings:** The main contribution of the study is that this is the first systematic effort to evaluate ROI accuracy. Analytical expressions have been derived for estimating errors of the ROI evaluations. Results of the Monte Carlo simulation will help practitioners in making informed decisions based on explicitly stated factors influencing the ROI uncertainties.



**Keywords:** Return on Investment, ROI, evaluation, costs, benefits, accuracy, estimation error, error propagation, uncertainty, information system, performance measure, business value, effort estimation, cost estimation, software engineering.

## 1.0 Introduction

Return on Investment (ROI) is one of the most popular performance measurement and evaluation metrics. ROI analysis (when applied correctly) is a powerful tool in making informed decisions on the acquisitions of information systems.

ROI is a performance measure used to evaluate the efficiency of investment or to compare the efficiency of a number of different investments. To calculate ROI, the net benefit (return) of an investment is divided by the cost of the investment; the result is expressed as a percentage or ratio (Erdogmus, Favaro and Strigel 2004)

There are many other ROI definitions in the literature (e.g. (Return on Investment (ROI), Glossary n.d.; Mogollon & Raisinghani 2003)). Each definition focuses on certain ROI aspects. With all the diversity of the definitions, the primary notion is the same: ROI is a fraction, the numerator of which is “net gain” (return, profit, benefit) earned as a result of the project (activity, system operations), while the denominator is the “cost” (investment) spent to achieve the result.

In general, predicting future is notoriously prone to uncertainties and errors. Estimating future project costs and returns also is a challenging endeavor (Stamelos & Angelis 2001; Daneva & Wieringa 2008; Eckartz 2009; Jorgensen & Shepperd 2007). Due to a variety of reasons actual numbers usually differ from the ones estimated in advance. The errors in estimating costs and returns will propagate through the ROI formula and result in inaccuracies of the ROI evaluations.

Estimating the accuracy of the ROI evaluations should be considered an essential part of the ROI calculations because ROI is used to make critical business decisions. Neglecting to estimate ROI accuracy may lead to wrong decisions on acquisition of information systems.

The purpose of this study is to estimate the accuracy of the ROI evaluations. The study provides estimates of the ROI accuracy in the context of the information systems implementations. Although the focus of the research is on the information systems, significant part of it can be applied to other types of systems and other fields of ROI evaluations.

The research is intended to answer the following questions:

- What factors influence the accuracy of the ROI calculations/ evaluations?
- What are the accuracies of estimating project costs of the information systems implementations?
- What are the accuracies of estimating project benefits of the information systems implementations?
- How inaccuracies of determining project costs and benefits propagate through the ROI calculations and affect ROI accuracy?
- What levels of the quantitative error estimates of the ROI evaluations can be expected for typical scenarios of the information system implementations?

Several methodologies have been used to achieve the research objectives. Literature review method was used to gather and analyze information related to the accuracy of estimating project costs and returns, distribution functions of errors. Measurements theory and error analysis, specifically, propagation of uncertainties methods were used to derive analytical expressions for ROI errors. Monte Carlo simulation methodology was used to design and deliver a quantitative experiment to model costs and returns estimating errors and calculate ROI accuracies. Spreadsheet simulation (Microsoft Excel spreadsheets enhanced with Visual Basic for Applications) was used to implement Monte Carlo simulations.

This research has the following scope and assumptions.

1. Most common definition treats ROI as a measure / metric / ratio / number (Erdogmus, Favaro and Strigel 2004). In some cases, return on investment is understood as a "method" or "approach" – "ROI analysis" (Mogollon & Raisinghani 2003; Andolsen 2004). This research is focused on the ROI as an individual measure.

2. ROI analysis can be performed with different purposes. As it was mentioned, ROI can provide rational for the future investments and acquisition decisions (e.g. project prioritization/ justification and facilitating informed choices about which projects to pursue). Evaluating future investments and making decisions on the information systems acquisitions are the processes based on the predicted data. By definition predicted data is likely to have certain level of variance from the amounts that will be really experienced later.

To avoid unnecessary complications and focus on the ROI accuracy, it has been assumed that projects are relatively short-time efforts and value of money is not explicitly considered. Also, such effects as “negative benefits” (Lim et al 2011) or decrease of productivity immediately after implementation of a new information system are not considered.

3. Software effort/costs and benefits estimation methods are out of the research scope. It is assumed that appropriate methods were used to estimate costs and benefits, and the results are available to the ROI estimators.
4. The focus of the study is on the ROI accuracy. Higher –level aspects of ROI research, e.g. its positioning in the business value of information technology (IT) and information systems (IS), IS/IT valuation or benefit valuation/management – are out of the scope.
5. Other typical performance measures such as the net present value of IS/IT projects are out of the scope.

The results of this research are intended for researchers in information systems, technology solutions and business management, and also for information specialists, project managers, program managers, technology directors, and information systems evaluators. Most results are applicable to ROI evaluations in a wider subject area.

The importance of the problem is due to a wide use of the ROI evaluations in making investment decisions. The main contribution of the study is that this is the first systematic effort to evaluate ROI accuracy. Analytical expressions have been derived for estimating errors of the ROI evaluations. Results of the Monte Carlo simulation will help practitioners in making informed decisions based on explicitly stated factors influencing the ROI uncertainties. Also, the paper contributes to more accurate ROI evaluations by drawing evaluators’ attention to the ways of minimizing evaluation errors.

The paper is structured as follows. Section 1 provides a brief introduction, outlines research objectives, defines methodology, and identifies limitations and assumptions of the study. Section 2 reviews previous work on ROI, IT cost and benefit estimations. Section 3 analyzes how uncertainties propagate through the ROI formula. The author derives mathematical approximations for the ROI accuracy by applying accepted approaches from measurements theory. In Section 4, the author applies a Monte Carlo simulation to illustrate the main implications of the study. The evidence is presented that the errors for ROI estimates are considerably high and that they should be taken account when making IT decisions. The paper concludes with a brief discussion in the sections 5 and 6.

# 2.0 Literature Review

## *2.1 ROI Accuracy Estimation*

A literature review has been conducted in support of this research. The review didn't reveal any papers specifically investigating methods of estimating ROI accuracy or case studies on this topic.

Two articles deal with the ROI accuracy (Botchkarev & Andru 2011; Andru & Botchkarev 2011). The value of these articles is in demonstrating the approach, and illustrating the level of the ROI accuracy for a typical CRM project. Accuracy assessment of the ROI calculations was performed on a specific example. Though not claiming any generic value, it was shown that even relatively low-level errors of estimating costs and returns (+/- 10%) may lead to significant ROI inaccuracies. That led to a conclusion that to make ROI number meaningful, it should be provided with an assessment of its accuracy.

Further literature review was focused on the accuracy of the components used to calculate ROI: costs and financial returns/benefits.

## *2.2 Cost Accuracy Estimation*

A cluster of publications was retrieved that deal with the accuracy of forecasting costs in various industries and project settings, e.g. (Stamelos & Angelis 2001; Daneva & Wieringa 2008; Eckartz 2009; Jorgensen & Shepperd 2007).

A subsection of this cluster deals with the software development effort estimation and its accuracy. A variety of estimation techniques are being used, which could be divided into several categories: estimation by analogy, parametric models, expert estimation, artificial intelligence methods (Morgenshtern, Raz, & Dvir 2007; Basha & Ponnurangam 2010; Nassif, Ho & Capretz 2013; Khatibi & Jawawi 2011; Buglione & Ebert 2011). Mostly often used techniques, to name a few, are: COCOMO II (Constructive Cost Model II) (COCOMO n.d.; Boehm et al 2000), Function Point Analysis (Albrecht 1979;  Lavazza & Morasca 2013), Use Case Points Method (Alves, Valente & Nunes 2013; Chemuturi 2009), a variety of artificial intelligence (machine learning) methods that are based on neural networks, fuzzy logic, regression trees, rule induction, Bayesian belief networks, evolutionary computation, grey relational models, etc. (Reddy et al 2010; Kaur & Salaria 2013; Du et al 2013; Muzaffar & Ahmed 2010; Song & Shepperd 2011).

Several authors compared the cost estimate at different stages of a product lifecycle (especially, at early stages) and the actual costs when the project was completed. The deviation/error of the estimates was documented.

A variety of estimation accuracy measures are being used (Reddy et al 2010; Basha & Ponnurangam 2010; Nassif, Ho & Capretz 2013; Keung, Kocaguneli & Menzies 2011; Zapata & Chaudron 2012; Zapata & Chaudron 2012; Zapata & Chaudron 2013; Jørgensen 2007): e.g. Balanced Relative Error (BRE), Balanced Relative Error Bias (BREbias). Although criticized (Zapata & Chaudron 2013; Jørgensen 2007), the Mean Magnitude of Relative Error (MMRE) remains the most commonly used measure. In order to present results of different papers in amore comparable form, this measure is used in the literature review (where possible).

The literature review revealed several important notions shared by many researchers:

- Cost prediction for software development projects is prone to errors.
- Estimates are mostly overoptimistic and underestimating is a problem for the software industry (McConnell 2009; Attarzadeh & Ow 2010). 60-80% of the projects experience effort or schedule average overruns of 30-40% (Molokken & Jorgensen 2003).
- A known cone of uncertainty illustrates that the variation of costs for the initial project phase could have as much as a +/-400% error (Attarzadeh & Ow 2010). The authors of the (Goh et al 2010) referring to an earlier publication indicate that cost estimates at the conceptual stage are in the range of -30% to +50%, which reduces to between -5% and +15% when the detailed design phase is entered.
- Factors, contributing to the estimation errors, include: estimation process complexity, volatile and unclear requirements, redefinition of requirements under pressure from senior management and clients, lack of experienced resources for estimation, misuse of estimates, technical complexity, requirements redefinition, business domain instability,

selection of a proper estimation technique, managerial issues (Morgenshtern, Raz, & Dvir 2007; Zapata & Chaudron 2013; Yang et al 2008).

Most authors admit limitations of the accuracy estimating studies (Du et al 2013; Toka & Turetken 2013): the first is incomplete project data affecting the accuracy of estimations and the second is limited number of projects with data on actual costs making results less reliable. These limitations pose risks on the validity of the estimation results.

Table I in Appendix 1 illustrates estimating errors collected from 15 studies. For better visualization, Fig. 1 shows cost/effort error estimates. Two outliers: 9% and 1,218% were not included. The graph demonstrates that 75% of the sample error estimates fall within the error range of 20% to 60%. This range will be used in the simulation modeling.

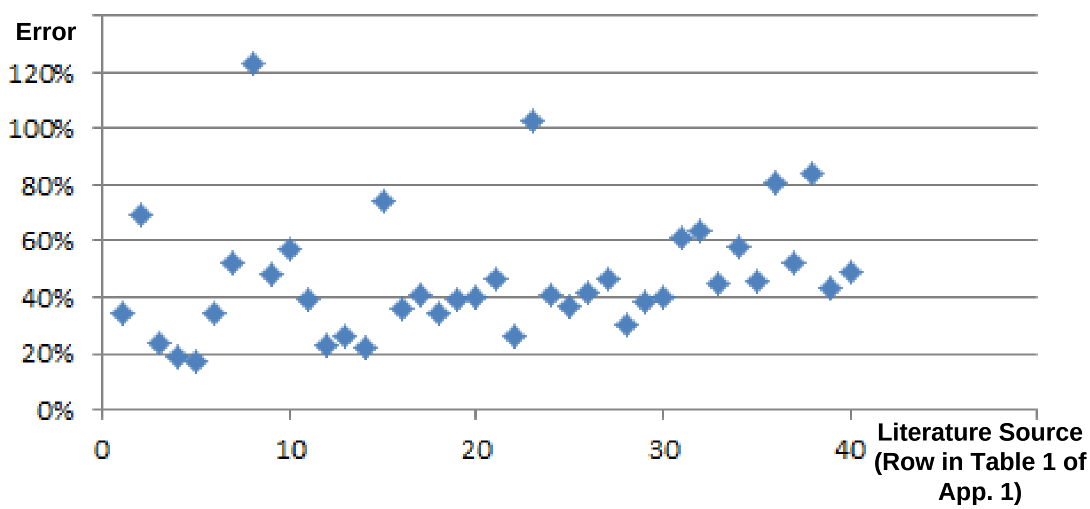


*Figure 1.* Sample graph of cost/effort error estimates

## *2.3 Financial Returns/Benefits Accuracy Estimating*

Estimation of the financial returns received much less attention in the academic literature than estimation of the costs. The main reasons for that are the difficulties indentifying, quantifying and monetizing benefits (e.g. (Menachemi et al 2006; Bojanc & Jerman-Blaži 2012; Irani & Love 2013; Uzoka 2009; Laudon & Laudon 2005; Vogel 2002; Wagner et al 2007)).

There are certain explanations for that:

- Actual costs are recorded through the project life and finalized at the end of the project. Benefits are only starting to emerge and accrue when the implementation is completed (Irani & Love 2013). Usually, there are no processes and information systems to record value of benefits. After the project has been closed, there is just nobody to collect and explore the data.
- A commonly documented type of benefit is worker productivity gain and related time and, consequently, financial savings. Obviously, these savings can be realized only if certain percent of the workforce is terminated after the system implementation (Sidorov 2006; Irani & Love 2013). However, there is no body of evidence to substantiate this being a regular practice. Hence, there is lack of data to support initial project benefit estimates or to measure variances.
- In subsection 2.2, we stated that there is lack of costs historical data. Regarding benefits we should admit that there is almost no benefits data. Companies consider benefits data even more confidential than cost information.
- The direct impact of the information system implementation project is difficult to establish (Vogel 2002).
- Measuring benefits, which may be tangible, quasi-tangible and/or intangible, is another challenge (Irani & Love 2013).
- Lack of research and commonly accepted benefits estimating methods (Wagner et al 2007). “Effective methods for modelling software benefits tend to be highly domain-specific” (Boehm & Sullivan 2000).

Another challenge is the evolution of the information systems and their respective benefits over time. This process is illustrated on Fig. 2 (based on Smith & Burnett 2003; Matthews 2013; Vogel 2002). The chart demonstrates that modern information systems tend to deliver benefits (in full accordance with the purposes they were created for) that are largely intangible and hardly can be estimated in financial terms, e.g. enhanced collaboration, more pertinent search results, etc. (Irani & Love 2013; Uzoka 2009; Laudon & Laudon 2005; Wagner et al 2007). It should be noted that the horizontal axis on Fig. 2 is not a timeline and the figure should not be undestood in the way that modern IS do not are based totally on knowledge and not on data. Th figure illustrates the innovation trend.

Identification of benefits should be closely aligned with the systems' goals/objectives. The desire to find hard-dollar benefits (inherent to older generations of the information systems) may divert researchers' attention from assessing the actual benefits of the systems.

For example, measuring benefits of the enterprise content management (ECM) system only by the volume of computer memory (and hence actual dollars saved as a result of reduced document duplication) may seem to be simple and attractive, but questionable, because it doesn't reflect the benefits the system was designed for.

The literature review didn't reveal any studies neither on the methodology of estimating accuracy of predicted benefits nor on actual numbers based on the case studies.

As the literature review reveals, methods used to estimate benefits are similar to those used to estimate costs: analogy (Driessen et al 2013), expert judgement (Vogel 2002; Wagner et al 2007), expert judgement enhanced with fuzzy models (Uzoka 2009), etc. That led us to the assumption that we can expect the same quantitative levels of benefits estimation accuracy as we experience for cost estimation accuracy. This assumption will be used in the sections below devoted to the quantitative estimation of the ROI accuracy.

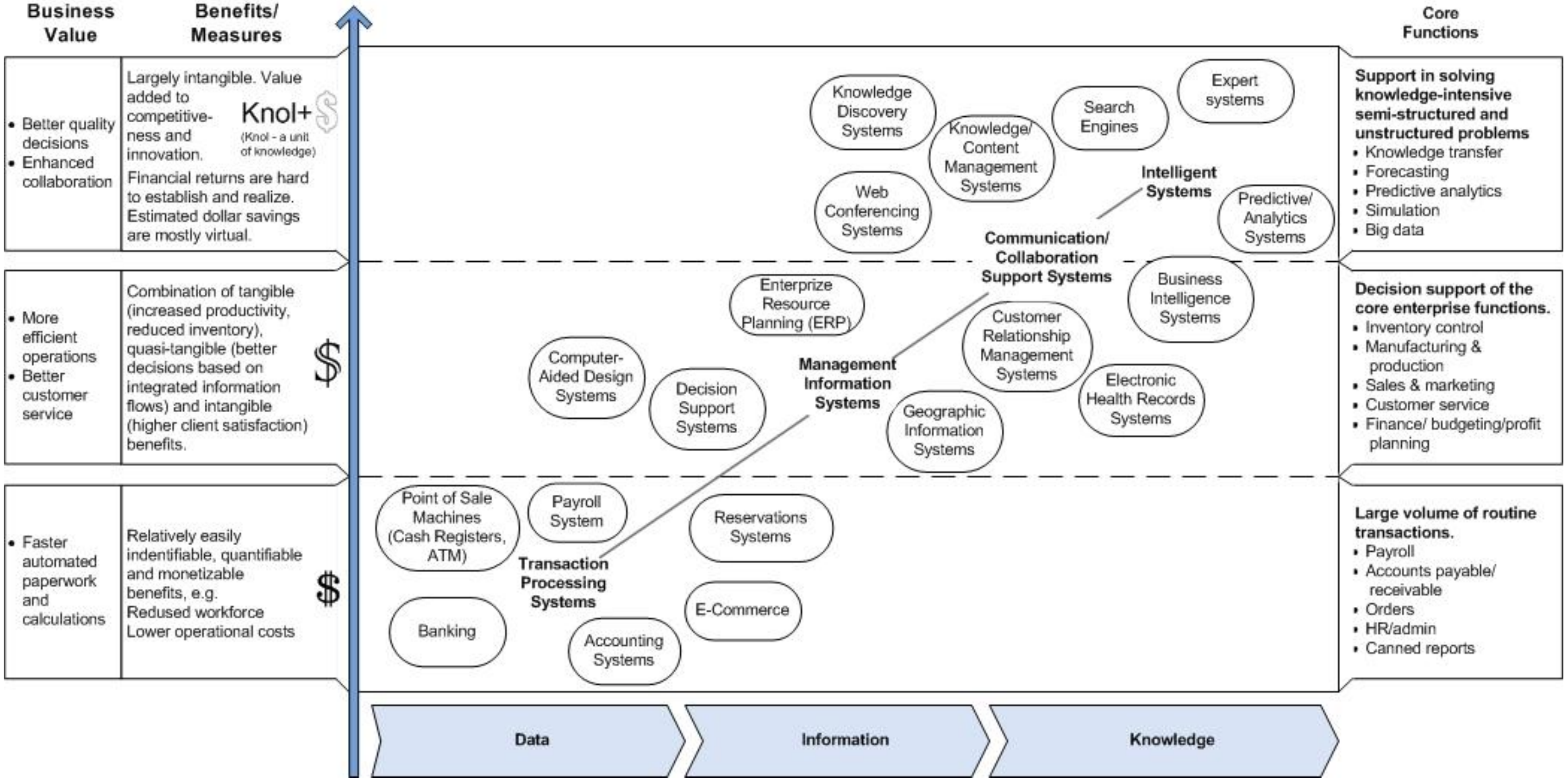


*Figure 2*. Evolution of the Information Systems and their Benefits

## 3.0 Analytical Estimation of the ROI Accuracy

The ROI is defined as:

$$R_{est} = \frac{B_{est} - C_{est}}{C_{est}} \quad (1)$$

where $C_{est}$ is an estimate of the cost to implement a project (predicted cost);

$B_{est}$ is an estimate of the benefit (financial return) from the project implementation (predicted benefit);

$R_{est}$ is the value of the ROI calculated based on the estimated costs and benefits (predicted ROI).

Equation (1) represents a complex non-linear function.

Due to the uncertainties of the estimation process, actual costs $(C_{act})$ and actual benefits $(B_{act})$, realized after the project is completed, will be different from the estimated ones. Because of multiple impacting uncertainties the absolute estimating errors could be considered random and expressed as follows:

$$\delta C = C_{act} - C_{est}; \quad \delta B = B_{act} - B_{est}$$

Hence, the actual ROI will also be different from the estimated one. The error of estimating ROI can be written as:

$$\delta R = R_{act} - R_{est}$$

The problem is to define an analytical expression for the ROI estimation error as a function of the uncertainties measuring costs and benefits:

$$\delta R = f[\delta C, \delta B, R(C, B)]$$

or for the relative ROI error:

$$\frac{\delta R}{R} = F\left(\frac{\delta C}{C}, \frac{\delta B}{B}, R(C, B)\right)$$

Similar problem is well-known in the physical sciences and engineering, and studied in the measurements theory and error analysis (Taylor 1997; Hughes & Hase 2010). In measurements, involving readings from two or more physical devices/meters, there is a need to assess the error of the experimental result when the readings are combined in an equation, e.g. three sides of a block are measured with a tape measure and then the volume of the block is calculated by multiplying these readings and the volume of the block is determined. Uncertainties that occurred in measuring the sides will propagate through the equation/formulae and affect the uncertainty of the calculated result. Usually, this area of studies is called error propagation or propagation of uncertainties and it is based on the mathematics of stochastic processes and, specifically, on algebra of stochastic variables. Measurement theory developed certain methods of calculating output errors depending on the type of the equations/formulae used: whether the measured parameters are added, deducted, multiplied, etc. This research follows the considerations accepted in the measurements theory. However, it should be noted that some assumptions and subsequent mathematical approximations common for the measurement field (e.g. the absolute error of the measurement is much smaller than the value of the measured quantity) may not be valid for all ROI evaluation scenarios. So, error analysis mathematics should be applied with caution.

**Maximum probable error – worst-case scenario.** Let's determine the maximum probable error for ROI. In the equation (1), a variable $(B_{est})$ is used more than once. That may lead to an effect of errors cancelling themselves (i.e. compensating errors) (Taylor 1997, p. 74). We can re-arrange equation (1) to avoid using a variable more than once

$$R_{est} = \frac{B_{est}}{C_{est}} - 1 \tag{2}$$

According to (Taylor 1997 p. 66), any problem for propagation error can be subdivided into sequence of steps, each of them based on the elementary mathematical operation. The second term in equation (2) doesn't include error component and could be neglected in the further error analysis. The first term is a quotient of two variables and error propagation for such a function is well-known (Taylor 1997; Hughes & Hase 2010). The maximum value of the ROI in equation (2) will occur when the numerator will be maximum and denominator will be minimum:

$$R_{est} + \delta R = \frac{B_{est} + \delta B}{C_{est} - \delta C} \tag{3}$$

Minimum value can be expressed as

$$R_{est} - \delta R = \frac{B_{est} - \delta B}{C_{est} + \delta C} \tag{4}$$

Following (Lindberg 2000; Physics Laboratory Companion), we can rewrite equation (3)

$$B_{est} + \delta B = (R_{est} + \delta R)(C_{est} - \delta C) = R_{est}C_{est} - R_{est}\delta C + C_{est}\delta R - \delta R\delta C$$

Assuming the errors are small, the last term $(\delta R\delta C)$can be neglected, and absolute ROI error can be written as

$$\delta R \approx (B_{est} + \delta B - R_{est}C_{est} + R_{est}\delta C)/C_{est} \tag{5}$$

Taking into account that $R_{est} = B_{est}/C_{est}$and substituting into equation (5), the expression for the maximum probable absolute error will be:

$$\delta R \approx \frac{C_{est}\delta B + B_{est}\delta C}{C_{est}^2} \tag{6}$$

or, multiplying both numerator and denominator by $B_{est}$, and rearranging

$$\delta R \approx \frac{B_{est}}{C_{est}}\left(\frac{\delta B}{B_{est}} + \frac{\delta C}{C_{est}}\right) \tag{7}$$

As it is observed in (Taylor 1997; Hughes & Hase 2010; Physics Laboratory Companion), error for a quotient is better expressed in terms of the relative error. Dividing both parts of equation (7) by $R_{est}$, we get the following formula

$$\frac{\delta R}{|R_{est}|} \approx \frac{\delta B}{B_{est}} + \frac{\delta C}{C_{est}} \qquad (8)$$

We arrived at a formula that is commonly used in the error propagation assessments for quotients (Taylor 1997; Hughes & Hase 2010; Physics Laboratory Companion).

Another approach to calculate maximum probable error is as follows. Equation (2) may be rewritten to show maximum and minimum levels of the ROI

Maximum

$$R_{est} + \delta R = \frac{B_{est} + \delta B}{C_{est} - \delta C} - 1 \qquad (9)$$

Minimum

$$R_{est} - \delta R = \frac{B_{est} - \delta B}{C_{est} + \delta C} - 1 \qquad (10)$$

Following a method used in (Taylor 1997 pp. 51; Palmer n.d.), equation (10) can be rewritten as

$$R_{est} + \delta R = \frac{B_{est}}{C_{est}}\left(\frac{1 + \delta B/B_{est}}{1 - \delta C/C_{est}}\right) - 1 \qquad (11)$$

Assuming the errors are small and using a binomial theorem, a component of (11) can be simplified (approximated by a Taylor series)

$$\frac{1}{1 - \delta C/C_{est}} \approx 1 + \delta C/C_{est} + (\delta C/C_{est})^2 + \cdots$$

Using only the first two terms of the approximation, equation (11) can be rewritten as

$$R_{est} + \delta R \approx \frac{B_{est}}{C_{est}}\left(1 + \frac{\delta B}{B_{est}}\right)\left(1 + \frac{\delta C}{C_{est}}\right) - 1 \qquad (12)$$

Rearranging equation (12), the error can be expressed as

$$\delta R \approx \frac{B_{est}}{C_{est}}\left(1 + \frac{\delta B}{B_{est}} + \frac{\delta C}{C_{est}} + \frac{\delta B}{B_{est}}\frac{\delta C}{C_{est}}\right) - 1 - R_{est}$$

Assuming again that the relative errors are small, so the last term in the brackets can be neglected and substituting $R_{est} = (B_{est}/C_{est}) - 1$

$$\delta R \approx \frac{B_{est}}{C_{est}}\left(1 + \frac{\delta B}{B_{est}} + \frac{\delta C}{C_{est}}\right) - 1 - \frac{B_{est}}{C_{est}} + 1 =$$

$$\frac{B_{est}}{C_{est}}\left(\frac{\delta B}{B_{est}} + \frac{\delta C}{C_{est}}\right) \tag{13}$$

Similar results can be gained if we use a generalized formula for a maximum probable error which for our case could be expressed through the total differential of a function (Taylor 1997 pp. 51; Palmer n.d.)

$$dR = \left(\frac{\partial R}{\partial B}\right) dB + \left(\frac{\partial R}{\partial C}\right) dC$$

Assuming $dR = \delta R$, and likewise for the other differentials, and that the variables $C$ and $B$are independent, the result for errors

$$\delta R \approx \left|\frac{\partial R}{\partial B}\right| \delta B + \left|\frac{\partial R}{\partial C}\right| \delta C \tag{14}$$

Formula (14) neglects higher order derivatives of the function which is considered a good approximation when the errors are small.

Substituting equation (2) into (14) and taking partial derivatives of the ROI function with respect of B and C

$$\delta R \approx \left|\frac{\partial}{\partial B}\left(\frac{B_{est}}{C_{est}} - 1\right)\right| \delta B + \left|\frac{\partial}{\partial C}\left(\frac{B_{est}}{C_{est}} - 1\right)\right| \delta C = \tag{15}$$

$$\left|\left(\frac{1}{C_{est}}\delta B\right)\right| + \left|B_{est}\frac{\partial}{\partial C}\left(\frac{1}{C_{est}}\right)\right|\delta C =$$

$$\left|\frac{1}{C_{est}}\delta B\right| + \left|B_{est}\left(-\frac{1}{C_{est}^2}\delta C\right)\right| =$$

$$\frac{C_{est}\delta B + B_{est}\delta C}{C_{est}^2} =$$

$$\frac{B_{est}}{C_{est}}\left(\frac{\delta B}{B_{est}} + \frac{\delta C}{C_{est}}\right)$$

We can observe that equations (7), (13, (15) provide the same result. ROI maximum probable error approximately equals benefits-costs ratio multiplied by the sum of benefits and costs relative errors.

**Probable error.** Maximum probable error, presented in a previous subsection, dealt with a worst-case scenario: the errors assume largest possible values and in a most "unpleasant" way, i.e. benefits are overestimated and costs are underestimated, or vice versa. Although important and conceivable, this scenario will not occur often. In a more likely scenario, when errors are random and independent, errors of estimating benefits and costs will have different signs and may be partially compensating each other. This scenario also needs to be assessed.

A generalized formula for a probable error for a two-variable function R has been derived in (Taylor 1997 pp. 62, 141; Hughes & Hase 2010):

$$\delta R \approx \sqrt{\left(\frac{\partial R}{\partial B}\delta B\right)^2 + \left(\frac{\partial R}{\partial C}\delta C\right)^2} \tag{16}$$

Substituting equation (2) into (16) and taking partial derivatives of the ROI function with respect of B and C, equation (16) can be transformed

$$\delta R \approx \sqrt{\left(\frac{\partial R}{\partial B}\delta B\right)^2 + \left(\frac{\partial R}{\partial C}\delta C\right)^2} =$$

$$\sqrt{\left[\frac{\partial}{\partial B}\left(\frac{B_{est}}{C_{est}} - 1\right)\delta B\right]^2 + \left[\frac{\partial}{\partial C}\left(\frac{B_{est}}{C_{est}} - 1\right)\delta C\right]^2} =$$

$$\sqrt{\left(\frac{1}{C_{est}}\delta B\right)^2 + \left[B_{est}\frac{\partial}{\partial C}\left(\frac{1}{C_{est}}\right)\delta C\right]^2} =$$

$$\sqrt{\left(\frac{1}{C_{est}}\delta B\right)^2 + \left[B_{est}\left(-\frac{1}{C_{est}^2}\delta C\right)\right]^2} = \qquad (17)$$

$$\sqrt{\frac{C_{est}^2\delta B^2 + B_{est}^2\delta C^2}{C_{est}^4} * \frac{B_{est}^2}{B_{est}^2}} =$$

$$\frac{B_{est}}{C_{est}}\sqrt{\left(\frac{\delta B}{B_{est}}\right)^2 + \left(\frac{\delta C}{C_{est}}\right)^2}$$

ROI probable error approximately equals benefits-costs ratio multiplied by the square root of the sum of squared benefits and costs relative errors.

**Breakdown of benefits and costs.** So far in this section to simplify the layout of the mathematical formulae, it was assumed that the value of the benefits (financial returns) is given by a single number $B_{est}$. For example, the project has a single type of benefits: cost savings due to downsizing, e.g. salaries and wages of the full time employees saved due to the system implementation. In general, there could be a variety of the benefits types: e.g. increased revenues due to increased sales, or sales margins; revenue enhancement, e.g. additional revenues were gained due to better targeted marketed and advertising; revenue protection, e.g. imminent fine was avoided (due to demonstrated compliance with regulatory requirements).The same refers to the costs. Common types of the costs include:

Cost of software development or customization/configuration.

Cost of IT infrastructure, e.g. Software/licenses - initial and annual maintenance; Hardware - if IS run in-house (e.g. purchasing and installation of new servers); Hosting - if information system provided as Software as a Service by a third party.

Cost of labour, e.g. direct operating expenses (DOE). Salaries and wages plus benefits for full time equivalent positions; Consultant services of installation, configuration, software customization, integration that requires skills not available within the I&IT Department.

Cost of training, e.g. IT personnel training by a third party; Program area end-user training by a third party.

So for a generic project, benefits $B_{est}$ and costs $C_{est}$ will be represented by summations of individual benefits and costs

$$B_{est} = \sum_i B_i ; \qquad\qquad C_{est} = \sum_j C_j$$

where $B_i$-$i$-th component of the financial return; and $C_j$-$j$-th component of the system cost.

Most likely, each of these benefits and costs types will be estimated separately using different tools/methods, and have their own (specific) estimation error values, i.e. $\delta B_i$ and $\delta C_j$. As it is derived in (Taylor 1997; Hughes & Hase 2010), uncertainty propagation for the operation of summation can be estimated using the following formulae:

Maximum probable error

$$\delta B \approx \sum_i \delta B_i \qquad \delta C \approx \sum_j \delta C_j \tag{18}$$

Probable error (sum in quadrature)

$$\delta B \approx \sqrt{\sum_i (\delta B_i)^2} \qquad \delta C \approx \sqrt{\sum_j \left(\delta C_j\right)^2} \tag{19}$$

General procedure for estimating ROI errors will be to calculate overall errors of benefits and costs using equations (18) or (19) and then substitute the results in equations (15) or (17).

## 4.0 Estimating ROI Accuracy with Monte Carlo Simulation

Monte Carlo simulation offers itself as a flexible technique for estimating ROI accuracy. It provides much more comprehensive insights into dependences of the costs and benefits uncertainties and ROI errors. Spreadsheet software packages have been widely used for Monte Carlo simulations due to their availability and simplicity (Chew and Walczyk 2012, Farrance and Frenkel 2014). In this study, the simulation was implemented on Microsoft Excel 2010 spreadsheets using Visual Basic for Applications (VBA). Earlier versions (1998, 2000, 2003 and 2007) of Excel were strongly criticized by the statistical community for their accuracy flaws (McCullough and Wilson 2005, McCullough and Heiser 2008). Recent research provides evidence that Excel 2010 demonstrates certain improvements, although still not perfect (Keeling and Pavur 2011, Mélard 2014, Kallner 2015). Known Excel limitations (specifically, relatively short cycle length and low numerical accuracy) are not critical for this application. The number of simulation trials and generated random numbers in the study is significantly smaller than the Excel cycle length – 2^24 (over 16 million). Also, there are no very small numbers or numbers that would differ in the fifth or sixth decimal place – issues that make Excel unsuitable in certain physical or mathematical sciences (Farrance and Frenkel 2014).

The Monte Carlo simulation process flowchart used in the study is shown in Figure 3.

As a first step of setting a new case, a project cost value (used as an actual cost) was randomly selected from one of the three project ranges: small (100K-500K), medium (501K-900K) or large (901K-1,300K). Using the cost value, benefit amount was calculated at a certain benefit-cost ratio. Actual ROI was calculated using a standard formula:

$$R_{act} = \frac{B_{act} - C_{act}}{C_{act}}$$

Estimated ROI will differ from the actual value due to the uncertainties in estimating benefits and costs. These uncertainties were generated through a range of relative errors of benefits and costs $\delta B/B_{act}, \delta C/C_{act}$.

Upper and lower levels of the estimated benefits were calculated as follows

$$B_{estU} = B_{act} + B_{act}(\delta B/B_{act})$$

$$B_{estL} = B_{act} - B_{act}(\delta B/B_{act})$$

Then, estimated value of benefits $\beta_i$was generated as a random number within the lower and upper bounds $\beta_i \in [B_{estL}, B_{estU}]$. Microsoft Excel VBA RND function was used to generate random numbers uniformly distributed within the specified interval.

Estimates of costs $\zeta_i$were generated using the same approach $\zeta_i \in [C_{estL}, C_{estU}]$.

Estimated ROI values were calculated as

$$R_{esti} = \frac{\beta_i - \zeta_i}{\zeta_i}$$

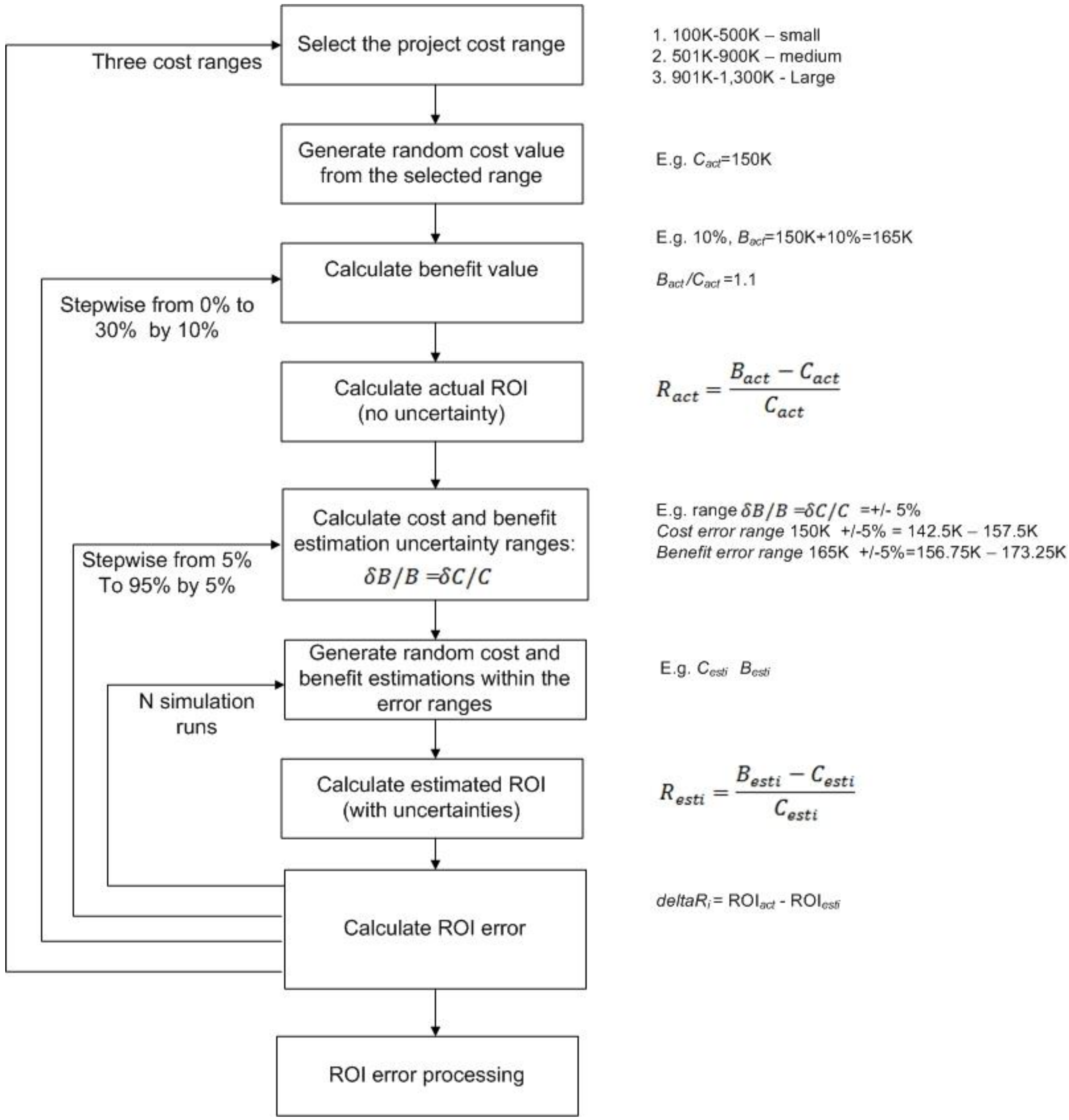


*Figure 3.* Simulation Process Flowchart

Finally, ROI error $\delta R$ (mean absolute error), after N Monte Carlo iterations, was calculated as

$$\delta R = \frac{1}{N}\sum_{i}^{N} |R_{act} - R_{esti}|$$

Several cases were run to determine the required number of iterations (similar to the approach of Farrance and Frenkel 2014). The results demonstrated that the amount of the ROI error converges to the first or second decimal of a percent when the number of iterations reaches 15,000 to 20,000. As the runtime was not an issue (under 10 sec for a single point) due to a relatively simple model, the number of iterations was set to 30,000.

Results of the simulation are shown on the Figures 4-5. Fig. 4 shows dependences of the ROI error $\delta R$ with the increase of the relative errors of benefits and costs estimates $\delta B/B_{act}$, $\delta C/C_{act}$ for the errors in the range from 0 to 45%. Fig. 5 shows similar data for the larger errors: 40% to 95%.

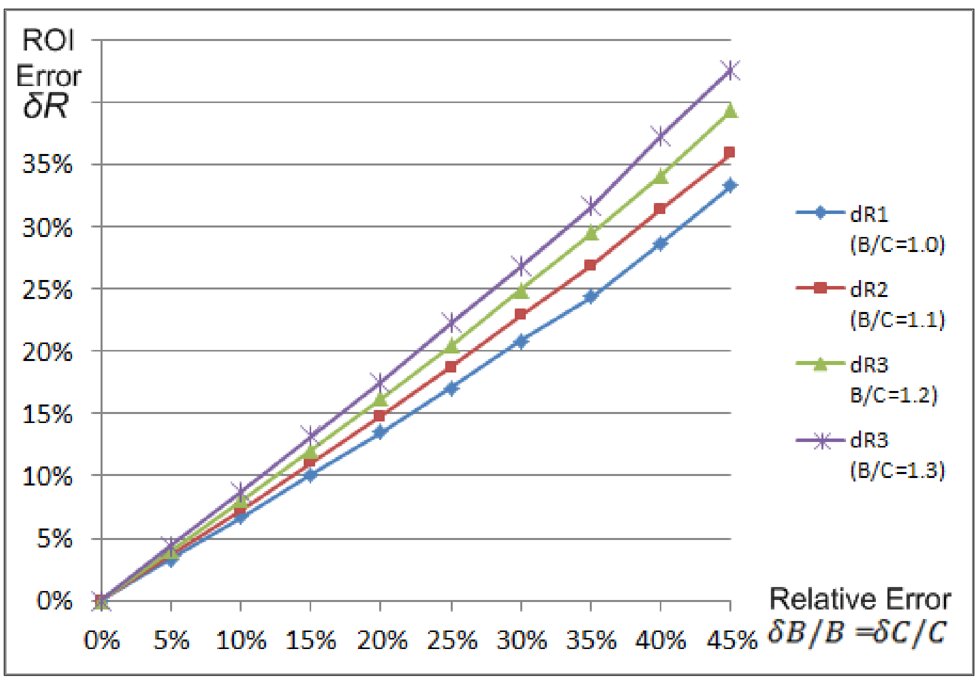


*Figure 4.* ROI error for the low lower-level benefits and costs relative errors

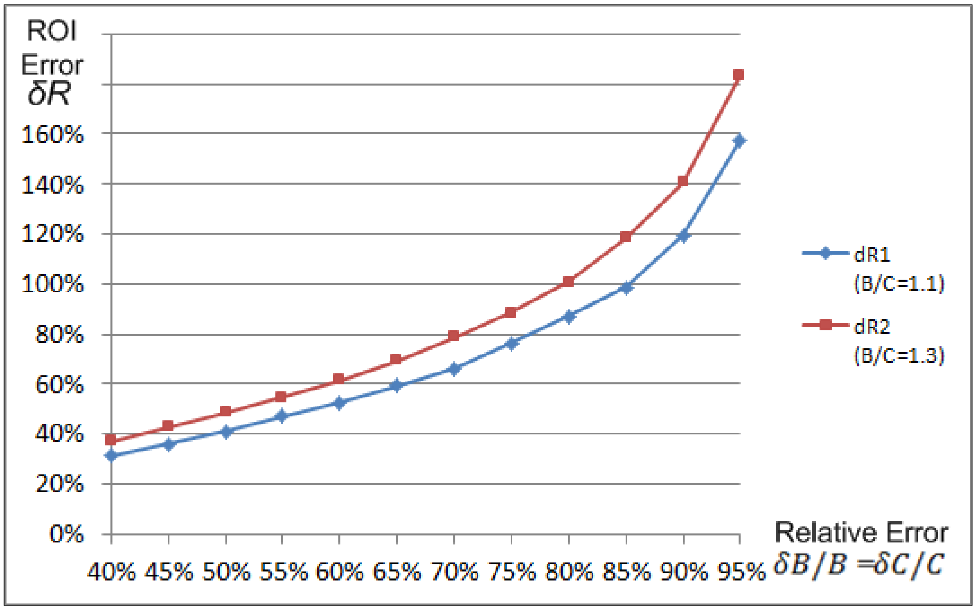


*Figure 5*. ROI error for the low higher-level benefits and costs relative errors

## 5.0 Discussion

Analytical expressions for the ROI errors derived in Section 3.0 are based on certain assumptions and simplifications. The prime one is that benefits and costs estimating errors are small and Taylor series expansion can be used. The validity of the resulting formulae needs to be checked to verify applicability of the approximations.

Equation (13) for the ROI maximum probable error was derived using the first two items in the Taylor series expansion:

$$\frac{1}{1 - \delta C/C_{est}} \approx 1 + \delta C/C_{est} \tag{20}$$

Fig 6 demonstrates the graphs for the left (exact) and right (approximated) parts of the equation (20) for a range of the cost relative errors $\delta C/C_{est}$. M is a numeric value of the approximated term.

The graph shows that variance between the exact and approximated solutions increases rapidly when value of the relative error exceeds 15-20%. This data suggests that approximated expressions for the ROI errors are best used for relative errors under 15-20%. It should be noted that the approximated line goes below the exact line. As a result approximated errors may underestimate real ROI errors.

Analytical expressions (with better ROI accuracy) for the cases with larger errors of costs and benefits are difficult to derive. There are studies in this area, e.g. (Seiler 1987), but the complexity of the solutions precludes them from being recommended for practitioners.

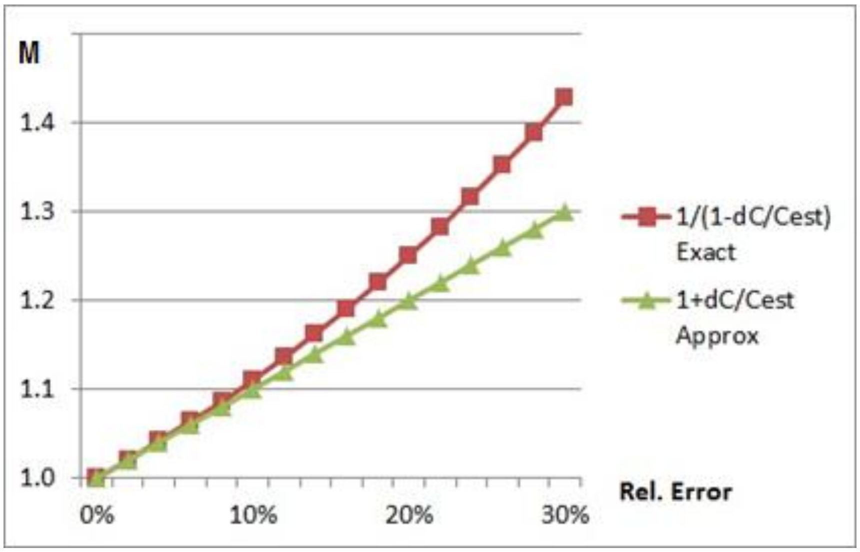


*Figure 6.* Taylor series expansion

A simple example can illustrate the levels of error using direct ROI calculations with a standard formula (not equations derived for the ROI errors). For this example, information system is being implemented with estimated benefits $B_{est}$ = \$120,000 and costs $C_{est}$= \$100,000. Using equation (1), the value of the ROI for this case $R_{est}$ =0.2 or 20%. Now, assuming there are uncertainties in benefits and costs estimations, we can calculate maximum and minimum values of the ROI using equations (9) and (10) for different values of errors. The results of these calculations are presented in Fig 7. Note that ROI is a two-variable function and generally requires 3-dimensional representation. In order for better 2-dimensional visualization, each point of the graph was calculated for equal values of errors $\delta B/B = \delta C/C$ (horizontal axis). The graph also shows maximum and minimum absolute ROI errors (variations from the calculated ROI which equals to

20%). Actual ROI value will be within the upper (maximum) and lower (minimum) boundaries shown on the graph. It should be noted that the “funnel” of ROI errors is not symmetrical regarding the expected value of 20% with zero errors. Overestimating ROI is more likely than underestimating.  Also, with no surprise, it should be noted that even with modest levels of the benefits and costs estimation errors ROI errors tend to be rather high (e.g. for $\delta B/B = \delta C/C =$ +/- 10%, absolute ROI errors could be up to +27% ÷ -22%).

Fig 8 shows ROI errors calculated according to different formulae using the same sample case: D-RIO – direct calculation using (9) and (10); MP-ROI – maximum probable ROI using (15) and summing in quadrature using (17).

Results of the simulation presented in the Section 4 show how the ROI absolute mean error is changing with the relative errors of benefits and costs. The behaviour of the graphs is different for the lower and higher levels of the relative errors. For better visual perception they are demonstrated separately. The graph for the lower error levels (see Fig. 5) shows almost linear relationship between the ROI absolute error and relative errors of benefits and costs (especially when relative errors are under 30%). The graph for the higher error levels (over 40%) shows

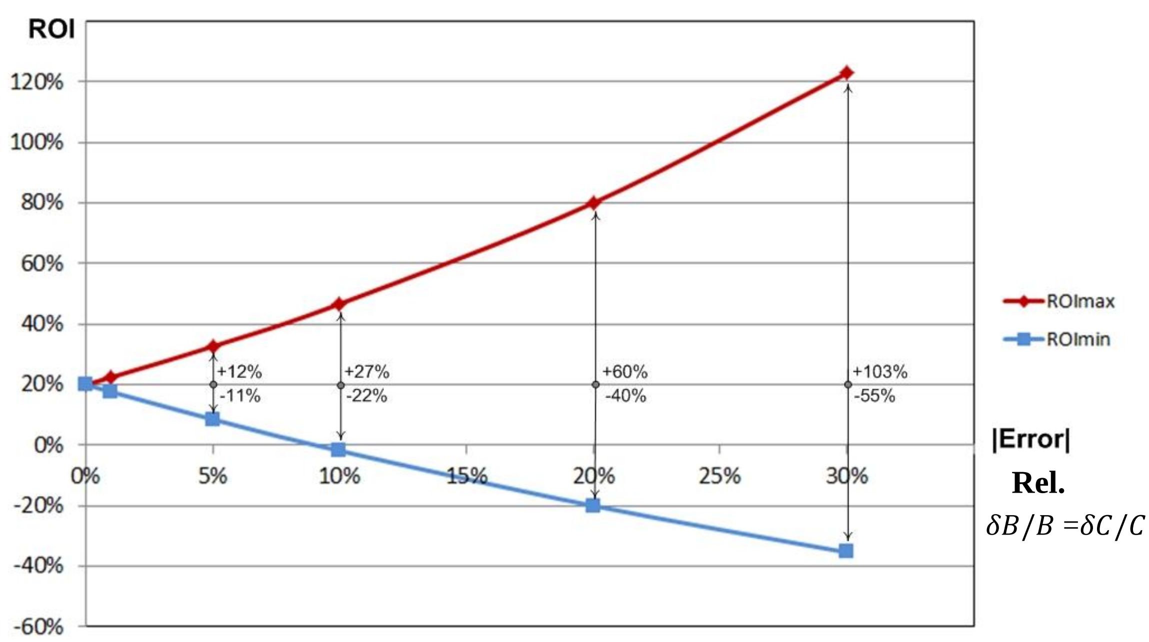


*Figure 7.* Maximum and minimum ROI levels and ROI errors for a sample case

exponential growth (see Fig. 5). As it might be expected, simulation has shown no difference for the ROI error behaviour for the projects of different sizes. The results show that ROI errors for the small and large projects (for the same relative errors of benefits and costs) are identical.

It should be noted that analytically derived formulas are approximations based on the Taylor series expansion. They should be applied with caution, especially when the relative errors of benefits and costs are over 15-20%.

Simulation results include the assumption that the relative errors of benefits and costs are equal (to ensure better visual presentation). Also, the distribution of the relative errors of benefits and costs was set to be uniform.

To round off the Discussion section, it is important to note that as any project is a unique endeavour (by definition), the same characteristic applies to the value of ROI errors in each project. It means that there are no any standard or expected ROI error amounts. Everything depends on how accurate were the financial assessments of the project benefits and costs. Project manager or analyst has to make ROI error estimations in specific conditions of the project. The results of this study provide a foundation for such estimations.

When assessing the ROI uncertainty, it is also noteworthy to take into account the ultimate financial implications not the intermediate parameters. For example, a company is developing a new software solution. The workload has been estimated with uncertainty of +/-50%. It seems at this point that expected ROI error will also be very large. And it is true, if the project would be developed in-house and workload will be directly translated into costs with the similar errors. However, if the software development would be outsourced through a fixed-price contract – the financial/cost uncertainty for the company will be close to zero, and so will be ROI error.

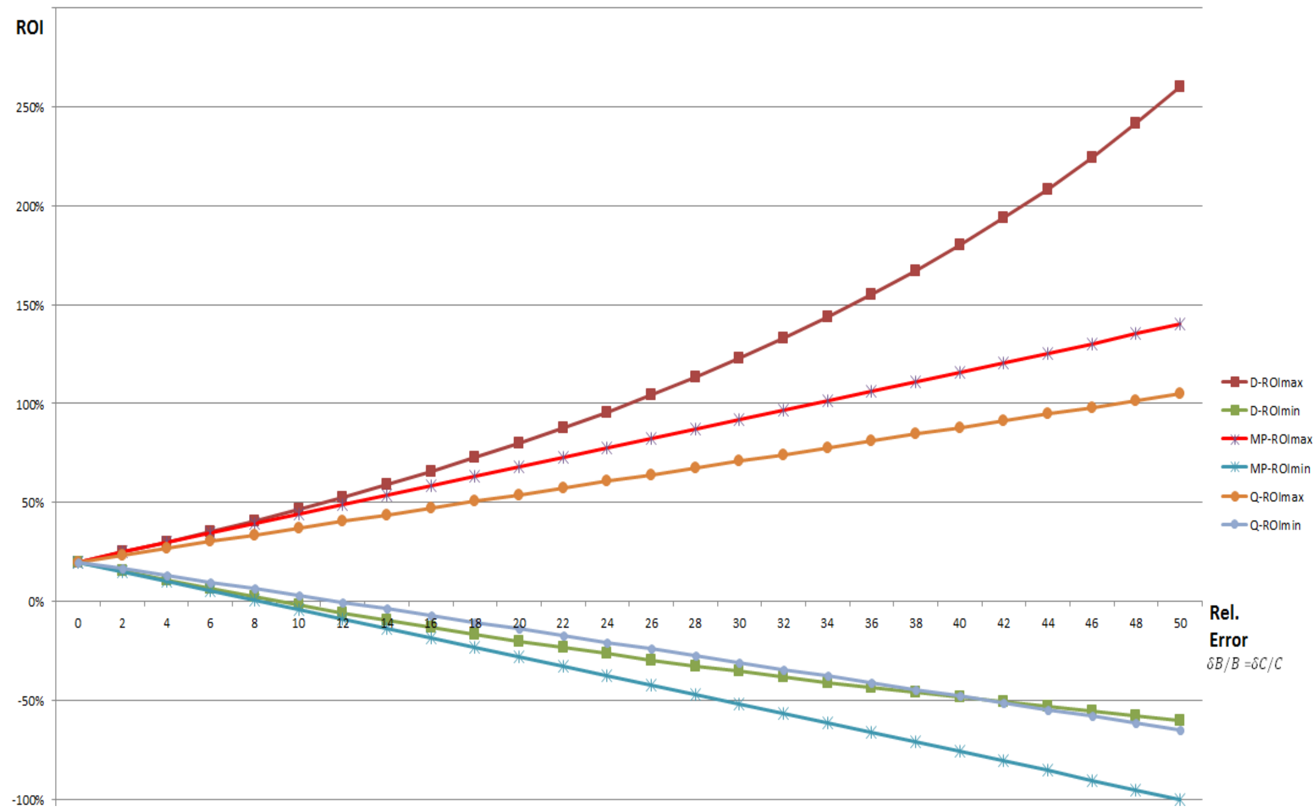


*Figure 8*. Sample case comparison of errors calculated with different methods

## 6.0 Concluding Remarks

Estimating accuracy of the ROI evaluations should become a part of the ROI assessments' best practices in order to avoid erroneous investment decisions. This study provided the first (to the best knowledge of the author) systematic research (both analytical and using simulation) of the accuracy of the ROI evaluations in the context of the information systems implementations and laid foundation for further theoretical and practical works in this area.

**Future research** may be focused on developing a framework of assessing and presenting benefits accuracy in a more standardized way. Also, research can be conducted into mathematical aspects of estimating ROI accuracy in the cases when estimating errors of benefits and costs are large, and have various probability distribution functions.

**Reference List**


Albrecht, A.J. (1979). Measuring Application Development Productivity, *Joint SHARE/ GUIDE/IBM Application Development Symposium*, 83-92.

Alves, R., Valente, P., & Nunes, N. J. (2013). Improving Software Effort Estimation with Human-Centric Models: a comparison of UCP and iUCP accuracy. *In Proceedings of the 5th ACM SIGCHI symposium on Engineering interactive computing systems*, EICS '13, 33-42, New York , NY, USA. ACM. DOI=10.1145/2480296.2480300 http://doi.acm.org/10.1145/2480296.2480300

Andolsen, A.A. (2004). Investing Wisely for the Future, *The Information Management Journal*, *8*(5), 47-54.

Andru, P., Botchkarev, A. (2011). The Use of Return on Investment ROI) in the Performance Measurement and Evaluation of Information Systems. *Information Management, Access and Privacy Symposium.* Presentation available at E-prints in Library and Information Science E-LIS), http://hdl.handle.net/10760/15503

Attarzadeh, I., & Ow, S. H. (2010). Improving the accuracy of software cost estimation model based on a new fuzzy logic model. *World Appl. Sci. J*, *82*(2), 177-184. http://www.idosi.org/wasj/wasj8210/7.pdf

Azzeh, M., Neagu, D., & Cowling, P. I. (2010). Fuzzy grey relational analysis for software effort estimation. *Empirical Software Engineering*, *15*(1), 60-90. https://pure.york.ac.uk/portal/files/13014230/fuzzylogic_fugrreansoefes.pdf

Basha, S., & Ponnurangam, D. (2010). Analysis of empirical software effort estimation models. *International Journal of Computer Science and Information Security*, *7*(3), 68-77. arXiv preprint arXiv:1004.1239

Boehm, B., Abts, C., Brown, A.W., Chulani, Clark, B.K., Horowitz, E., Madachy, R., Reifer, D.J., and Steece, B. (2000). Software Cost Estimation with COCOMO II with CD-ROM). Englewood Cliffs, NJ: Prentice-Hall. ISBN 0-13-026692-2

Boehm, B. W., & Sullivan, K. J. (2000). Software economics: a roadmap. *In Proceedings of the conference on The future of Software engineering* 319-343. ACM.

Bojanc, R., & Jerman-Blažic, B. (2012). Quantitative Model for Economic Analyses of Information Security Investment in an Enterprise Information System. *Organizacija 45*(6), 276-288. DOI: 10.2478/v10051-012-0027-z

Botchkarev, A., & Andru, P. (2011). A return on investment as a metric for evaluating information systems: Taxonomy and application. *Interdisciplinary Journal of Information, Knowledge, and Management, 6*, 245-269. http://www.ijikm.org/Volume6/IJIKMv6p245-269Botchkarev566.pdf

Buglione, L., & Ebert, C. (2011). Estimation Tools and Techniques. *IEEE Software, 28*(3), 15-18.

Burke, D. E., & Menachemi, N. (2004). Opening the black box: Measuring hospital information technology capability. *Health Care Management Review, 29*(3), 207-217.

Carroll, E. R. (2005). Estimating software based on use case points. In *Companion to the 20th annual ACM SIGPLAN conference on Object-oriented programming, systems, languages, and applications*. 257-265. ACM.

Chemuturi, M. (2009). Software Estimation Best Practices, Tools and Techniques for Software Project Estimators, J. Ross Publishing, 84-87.

Chew, G., & Walczyk, T. (2012). A Monte Carlo approach for estimating measurement uncertainty using standard spreadsheet software. *Analytical and bioanalytical chemistry, 402*(7), 2463-2469.

COCOMOII Constructive Cost Model II http://sunset.usc.edu/csse/research/COCOMOII/cocomo_main.html

Daneva, M., and Wieringa, R. (2008). "Cost estimation for cross-organizational ERP projects: research perspectives" *Software Quality Journal, 16*(3), 459-481. doi:10.1007/s11219-008-9045-8

Driessen, J., Cioffi, M., Alide, N., Landis-Lewis, Z., Gamadzi, G., Gadabu, O. J., & Douglas, G. (2013). Modeling return on investment for an electronic medical record system in

Lilongwe, Malawi. *Journal of the American Medical Informatics Association*, 743-748 doi:10.1136/amiajnl-2012-001242

Du, W. L., Capretz, L. F., Nassif, A. B., & Ho, D. (2013). A Hybrid Intelligent Model for Software Cost Estimation. *Journal of Computer Science*, *9*(11), 1506. http://thescipub.com/abstract/10.3844/jcssp.2013.1506.1513

Eckartz, S. M. (2009). Costs, Benefits and Value Distribution - Ingredients for Successful Cross-Organizational ES Business Cases. In: *32nd Information Systems Research Seminar in Scandinavia, IRIS 32, Inclusive Design*, 9-12 August 2009, Molde, Norway.

Erdogmus, H., Favaro, J. and Strigel, W. (2004). Guest Editors' Introduction: Return on Investment. *IEEE Software, 21*(3), 18-22. doi:10.1109/MS.2004.1293068 http://www.computer.org/csdl/mags/so/2004/03/s3018.html

Farrance, I., & Frenkel, R. (2014). Uncertainty in Measurement: A Review of Monte Carlo Simulation Using Microsoft Excel for the Calculation of Uncertainties Through Functional Relationships, Including Uncertainties in Empirically Derived Constants. *The Clinical Biochemist Reviews*, *35*(1), 37. http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3961998/pdf/cbr-35-37.pdf

Goh, Y. M., Newnes, L. B., Mileham, A. R., McMahon, C. A., & Saravi, M. E. (2010). Uncertainty in Through-Life Costing--Review and Perspectives. *Engineering Management, IEEE Transactions on*, *57*(4), 689-701.

Huang, S. J., & Chiu, N. H. (2009). Applying fuzzy neural network to estimate software development effort. *Applied Intelligence*, *30*(2), 73-83.

Hughes, I., & Hase, T. (2010). Measurements and their uncertainties: a practical guide to modern error analysis. Oxford University Press.

Irani, Z., & Love, P. (Eds.) (2013). Evaluating Information Systems. Routledge.

Jørgensen, M. (2007). A critique of how we measure and interpret the accuracy of software development effort estimation. In *First International Workshop on Software Productivity Analysis and Cost Estimation. Information Processing Society of Japan, Nagoya.* http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.110.4666&rep=rep1&type=pdf#page=25

Jorgensen, M., Shepperd, M. (2007). A Systematic Review of Software Development Cost Estimation Studies. *IEEE Transactions on Software Engineering*, *33*(1), 33 -53. doi:10.1109/TSE.2007.256943

Kallner, A. (2015). Microsoft EXCEL 2010 offers an improved random number generator allowing efficient simulation in chemical laboratory studies. *Clinica chimica acta; international journal of clinical chemistry*, *438*, 210-211.

Kaur D.S.S. (2013). Bayesian Regularization Based Neural Network Tool for Software Effort Estimation. Global Journal of Computer Science and Technology, *13*(2). http://computerresearch.org/stpr/index.php/gjcst/article/viewPDFInterstitial/1381/1247

Keeling, K.B and Pavur, R.J. (2011). Statistical Accuracy of Spreadsheet Software, *The American Statistician*, *65*(4), 265-273, http://dx.doi.org/10.1198/tas.2011.09076 http://www.tandfonline.com/doi/pdf/10.1198/tas.2011.09076

Keung, J., Kocaguneli, E., Menzies, T. (2011). A ranking stability indicator for selecting the best effort estimator in software cost estimation. *Automated Software Engineering (submitted)* Available on-line at http://menzies.us/pdf/11draftranking.pdf

Khatibi, V., & Jawawi, D. N. (2011). Software cost estimation methods: A review. *Journal of emerging trends in computing and information sciences*, *2*(1), 21-29.

Kitchenham, B. A., Pickard, L. M., MacDonell, S. G., & Shepperd, M. J. (2001). What accuracy statistics really measure software estimation. *In Software, IEE Proceedings. 148*(3), 81-85. IET. http://v-cheiner.brunel.ac.uk/bitstream/2438/1855/1/00942859.pdf

Kumari, S., & Pushkar, S. (2013). Performance Analysis of the Software Cost Estimation Methods: A Review. *International Journal of Advanced Research in Computer Science and Software Engineering, 3*(7), 229-238. http://www.ijarcsse.com/docs/papers/Volume_3/7_July2013/V3I7-0247.pdf

Laudon, K.C., Laudon, J.P., (2005). *Essentials of Management Information Systems: Managing the Digital Firm*, Sixth ed., Prentice Hall, Upper Saddle River, New Jersey,.

Lavazza, L., & Morasca, S. (2013). Measuring the Functional Size of Real-Time and Embedded Software: a Comparison of Function Point Analysis and COSMIC. In ICSEA 2013, *The Eighth International Conference on Software Engineering Advances,* 465-470.

Lewis, R. A., & Rao, J. M. (2013). On the Near Impossibility of Measuring the Returns to Advertising. ftp://ftp.zew.de/pub/zew-docs/veranstaltungen/ICT2013/Papers/ICT2013_Rao.pdf

Lim, J. Y., Kim, M. J., Park, C. G., & Kim, J. Y. (2011). Comparison of Benefit Estimation Models in Cost-Benefit Analysis: A Case of Chronic Hypertension Management Programs. *Journal of Korean Academy of Nursing*, *41*(6), 750-757. http://synapse.koreamed.org/search.php?where=aview&id=10.4040/jkan.2011.41.6.750&code=0006JKAN&vmode=FULL

Lin, J. C., Lin, Y. T., Tzeng, H. Y., & Wang, Y. C. (2013). Using Computing Intelligence Techniques to Estimate Software Effort. *International Journal of Software Engineering & Applications IJSEA*, *4*(1), 43-53. http://www.airccse.org/journal/ijsea/papers/4113ijsea04.pdf

Lindberg, V. (2000). *Uncertainties and error propagation. Manual on Uncertainties, Graphing and the Vernier Caliper,* Part I. Rochester Institute of Technology, New York, USA. http://www.rit.edu/~w-uphysi/uncertainties/Uncertaintiespart2.html

Matthews, J., (2013). Adding value: getting to the heart of the matter. *Performance Measurement and Metrics, 14*(3),162 - 174. DOI: 10.1108/PMM-08-2013-0024

McConnell, S. (2009). *Software Estimation: Demystifying the Black Art: Demystifying the Black Art*. O'Reilly Media, Inc.

McCullough, B. D., & Heiser, D. A. (2008). On the accuracy of statistical procedures in Microsoft Excel 2007. *Computational Statistics & Data Analysis*, *52*(10), 4570-4578.

McCullough, B. D., & Wilson, B. (2005). On the accuracy of statistical procedures in Microsoft Excel 2003. *Computational Statistics & Data Analysis*, *49*(4), 1244-1252.

Mélard, G. (2014). On the accuracy of statistical procedures in Microsoft Excel 2010. *Computational Statistics*, 1-34.

Menachemi, N., Burkhardt, J., Shewchuk, R., Burke, D., & Brooks, R. G. (2006). Hospital information technology and positive financial performance: a different approach to finding an ROI. *Journal of Healthcare Management*. American College of Healthcare Executives, *51*(1), 40-58.

Mogollon, M. and Raisinghani, M. (2003). Measuring ROI in E-Business: A Practical Approach. *Information Systems Management*, *20*(2), 63 - 81.

Molokken, K., & Jorgensen, M. (2003). A review of software surveys on software effort estimation. In *Empirical Software Engineering*, ISESE 2003. Proceedings. International Symposium on, 223-230. IEEE.

Morgenshtern, O., Raz, T., & Dvir, D. (2007). Factors affecting duration and effort estimation errors in software development projects. *Information and Software Technology*, *49*(8), 827-837.

Muzaffar, Z., & Ahmed, M. A. (2010). Software development effort prediction: A study on the factors impacting the accuracy of fuzzy logic systems. *Information and Software Technology*, *52*(1), 92-109.

Nassif, A. B., Ho, D., & Capretz, L. F. (2013). Towards an early software estimation using log-linear regression and a multilayer perceptron model. *Journal of Systems and Software*. *86*(1), 144- 160. http://dx.doi.org/10.1016/j.jss.2012.07.050

Palmer, M. *Propagation of uncertainty through mathematical operations*. Massachusetts Institute of Technology. http://web.mit.edu/fluids-modules/www/exper_techniques

*Physics Laboratory Companion*, Chapter 3 Error Propagation. University of Regina. http://uregina.ca/~szymanss/uglabs/companion/Ch3_Error_Prop.pdf

Reddy, P. V. G. D., Sudha, K. R., Sree, P. R., & Ramesh, S. N. S. V. S. C. (2010). Software effort estimation using radial basis and generalized regression neural networks. Journal of Computing, *2*(5), ISSN 2151-9617 arXiv preprint arXiv:1005.4021. http://arxiv.org/ftp/arxiv/papers/1005/1005.4021.pdf

Return On Investment (ROI), Glossary. Centers for Disease Control and Prevention. http://www.cdc.gov/leanworks/resources/glossary.html accessed June 27, 2014

Seiler, F. A. (1987). Error propagation for large errors. *Risk Analysis*, *7*(4), 509-518.

Shepperd, M., & MacDonell, S. (2012). Evaluating prediction systems in software project estimation. *Information and Software Technology*, *54*(8), 820-827. http://v-scheiner.brunel.ac.uk/bitstream/2438/6473/4/IST_Invited_2011_v7.pdf

Sidorov, J. (2006). It ain't necessarily so: the electronic health record and the unlikely prospect of reducing health care costs. *Health Affairs*, *25*(4), 1079-1085. doi: 10.1377/hlthaff.25.4.1079 http://content.healthaffairs.org/content/25/4/1079.full.html

Smith, Roderick; Burnett, Simon. (2003). Beyond first-generation KM. *Inside Knowledge 6*(7). http://www.ikmagazine.com/xq/asp/sid.0/articleid.DE0ADE1B-4060-49F7-B0F1-CC75898E101C/eTitle.Beyond_firstgeneration_KM/qx/display.htm

Song, Q., & Shepperd, M. (2011). Predicting software project effort: A grey relational analysis based method. *Expert Systems with Applications*. *38*(6), 7302-7316. http://zema.gr.xjtu.edu.cn/LiferayFCKeditor/UserFiles/File/QSong/Paper/ESWA5582.pdf

Stamelos, I., Angelis L. (2001). Managing uncertainty in project portfolio cost estimation. *Information and Software Technology*. *43*(13), 759-768. doi:10.1016/S0950-5849(01)00183-5

Taylor, J. R. (1997). *An introduction to error analysis: the study of uncertainties in physical measurements*. University science books. 2nd edition.

Toka, D., Turetken, O. (2013). Accuracy of Contemporary Parametric Software Estimation Models: A Comparative Analysis. *39th Euromicro Conference on Software Engineering and Advanced Applications SEAA 2013* IEEE, 313-316. http://dx.doi.org/10.1109/SEAA.2013.49

Uzoka, F. M. E. (2009). Fuzzy-Expert system for cost Benefit Analysis of Enterprise information systems, A Frame work. *International Journal on Computer Science and Engineering*, *1*(3), 254-262.

Vogel, L. H. (2002). Finding value from IT investments: exploring the elusive ROI in healthcare. *Journal of healthcare information management. JHIM*, *17*(4), 20-28.

Wagner, S., Xie, S., Rübel-Otterbach, M., & Sell, B. (2007). Profitability estimation of software projects: A combined framework. In *The First International Workshop on Software Productivity Analysis and Cost Estimation (SPACE'07)*. 37-43.

Yang, D., Wang, Q., Li, M., Yang, Y., Ye, K., & Du, J. (2008). A survey on software cost estimation in the chinese software industry. In Proceedings of the *Second ACM-IEEE* *international symposium on Empirical software engineering and measurement*. 253-262. ACM.

Zapata, A. H., & Chaudron, M. R. (2012). An analysis of accuracy and learning in software project estimating. In *Software Engineering and Advanced Applications (SEAA), 2012 38th EUROMICRO Conference on*, 414-421 IEEE.

Zapata, A. H., & Chaudron, M. R. V. (2013). An Empirical Study into the Accuracy of IT Estimations and its Influencing Factors. *International Journal of Software Engineering and Knowledge Engineering. 23*(4), 409-432.

## APPENDIX 1

TABLE 1

Sample Estimating Errors by Method/Technique

| No. | Estimation Method/Model | Estimated Project Parameter | Error Measure | Error/ Accuracy | Reference |
|---|---|---|---|---|---|
| 1 | UCP | Cost | MMRE | 34.3% | Alves, Valente & Nunes 2013 |
| 2 | iUCP | Cost | MMRE | 69.6% | Alves, Valente & Nunes 2013 |
| 3 | UCP | Cost | MMRE for 95% of the projects | 9% | Carroll 2005 |
| 4 | N/A | Duration | MMRE | 22% | Morgenshtern, Raz, & Dvir 2007 |
| 5 | N/A | Effort | MMRE | 24% | Morgenshtern, Raz, & Dvir 2007 |
| 6 | Initermediate COCOMO | Effort | MMRE | 18.6% | Reddy et al 2010 |
| 7 | Radial Basis Neural Network | Effort | MMRE | 17.3% | Reddy et al 2010 |
| 8 | Generalized Regression Neural Network | Effort | MMRE | 34.6% | Reddy et al 2010 |
| 9 | COCOMO | Effort | MMRE | 52% | Kaur & Salaria 2013 |
| 10 | Levenberg-Marquardt Based Neural Network | Effort | MMRE | 123% | Kaur & Salaria 2013 |

| No. | Estimation Method/Model | Estimated Project Parameter | Error Measure | Error/ Accuracy | Reference |
|---|---|---|---|---|---|
| 11 | Back Propagation Based Neural Network | Effort | MMRE | 1,218% | Kaur & Salaria 2013 |
| 12 | Bayesian Regularization Based Neural Network | Effort | MMRE | 48% | Kaur & Salaria 2013 |
| 13 | SEER-SEM | Effort | MMRE | 57% | Du et al 2013 |
| 14 | SEER-SEM Enhanced with Nuero-Fuzzy Model | Effort | MMRE | 39% | Du et al 2013 |
| 15 | COCOMO Enhanced with Computing Intelligence Techniques | Effort | MMRE | 23% | Lin et al 2013 |
| 16 | COCOMO | Effort | MMRE | 26% | Lin et al 2013 |
| 17 | Fuzzy Neural Network | Effort | MMRE | 22% | Huang & Chiu 2009 |
| 18 | COCOMOII | Effort | MMRE | 74% | Toka & Turetken 2013 |
| 19 | COCOMOII | Duration | MMRE | 91% | Toka & Turetken 2013 |
| 20 | SEER-SEM | Effort | MMRE | 36% | Toka & Turetken 2013 |
| 21 | SEER-SEM | Duration | MMRE | 81% | Toka & Turetken 2013 |
| 22 | SLIM by QSM | Effort | MMRE | 41% | Toka & Turetken 2013 |
| 23 | SLIM by QSM | Duration | MMRE | 84% | Toka & Turetken 2013 |
| 24 | TruePlanning by Price Systems | Effort | MMRE | 34% | Toka & Turetken 2013 |
| 25 | TruePlanning by Price Systems | Duration | MMRE | 99% | Toka & Turetken 2013 |
| 26 | UCP with log-linear regression model | Effort | MMERE | 39.2% | Nassif, Ho & Capretz 2013 |
| 27 | UCP with Multilayer Perceptron (MLP) | Effort | MMERE | 40% | Nassif, Ho & Capretz 2013 |
| 28 | UCP | Effort | MMERE | 46.7% | Nassif, Ho & Capretz 2013 |
| 29 | N/A | Cost | MMRE | 26% | Zapata & Chaudron 2012 |
| 30 | N/A | Effort | MMRE | 103% | Zapata & Chaudron 2012 |
| 31 | COCOMO II | Effort | MMRE | 41% | Attarzadeh & Ow 2010 |
| 32 | COCOMO II enhanced with Fuzzy Model | Effort | MMRE | 37% | Attarzadeh & Ow 2010 |
| 33 | Grey Relational Model | Effort | MMRE | 41.4% | Song & Shepperd 2011 |
| 34 | Stepwise Regression Model | Effort | MMRE | 46.5% | Song & Shepperd 2011 |

| No. | Estimation Method/Model | Estimated Project Parameter | Error Measure | Error/ Accuracy | Reference |
|---|---|---|---|---|---|
| 35 | Estimation by Analogy enhanced with fuzzy grey relational analysis | Effort | MMRE | 30.6% | Azzeh, Neagu & Cowling 2010 |
| 36 | Case-Based Reasoning | Effort | MMRE | 38.2% | Azzeh, Neagu & Cowling 2010 |
| 37 | Multiple Linear Regression | Effort | MMRE | 39.9% | Azzeh, Neagu & Cowling 2010 |
| 38 | Artificial Neural Networks | Effort | MMRE | 61.2% | Azzeh, Neagu & Cowling 2010 |
| 39 | Intermediate COCOMO | Effort | MMRE | 64% | Kumari & Pushkar 2013 |
| 40 | COCOMO II | Effort | MMRE | 45% | Kumari & Pushkar 2013 |
| 41 | MOPSO Model | Effort | MMRE | 58% | Kumari & Pushkar 2013 |
| 42 | Support Vector Regression (SVR) Model | Effort | MMRE | 46% | Kumari & Pushkar 2013 |
| 43 | Software Engineering Laboratory (SEL) Model | Effort | MMRE | 81% | Kumari & Pushkar 2013 |
| 44 | Walton-Felix Model | Effort | MMRE | 52% | Kumari & Pushkar 2013 |
| 45 | Bailey-Basil Model | Effort | MMRE | 84% | Kumari & Pushkar 2013 |
| 46 | Halsted Model | Effort | MMRE | 43% | Kumari & Pushkar 2013 |
| 47 | Doty Model | Effort | MMRE | 49% | Kumari & Pushkar 2013 |

Abbreviations used in the table:

MMRE - Mean Magnitude of Relative Error

MMERE - Mean Magnitude of Error Relative to the Estimate