ANALYTICAL METHOD FOR QUANTIFICATION OF ECONOMIC RISKS DURING FEASIBILITY ANALYSIS FOR LARGE ENGINEERING PROJECTS

By

KULATILAKA ARTHANAYAKE MALIK KUMAR RANASINGHE

B. Sc. (Engineering) Honors, University of Moratuwa, Sri Lanka.

M. A. Sc., The University of British Columbia, Canada.

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF

THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

in

THE FACULTY OF GRADUATE STUDIES

DEPARTMENT OF CIVIL ENGINEERING

We accept this thesis as conforming

to the required standard

THE UNIVERSITY OF BRITISH COLUMBIA

August 1990

5

Abstract

The objectives of this thesis are to develop an analytical method for economic risk quantification during feasibility analysis for large engineering projects and to computerize the method to explore its behavior, to validate it and to test its practicality for the measurement of uncertainty of decision variables such as project duration, cost, revenue, net present value and internal rate of return. Based on the probability of project success the method can be utilized to assist on strategic feasibility analysis issues such as contingency provision, "go-no go" decisions and adopting phased or fast track construction.

The method is developed by applying a risk measurement framework to the project economic structure. The risk measurement framework is developed for any function $Y = g(\mathbf{X})$, between a derived variable and its correlated primary variables. Using a variable transformation, it transforms the correlated primary variables and the function to the uncorrelated space. Then utilizing the truncated Taylor series expansion of the transformed function and the first four moments of the transformed uncorrelated variables it approximates the first four moments of the derived variable. Using these first four moments and the Pearson family of distributions the uncertainty of the derived variable is quantified as a cumulative distribution function. The first four moments for the primary variables are evaluated from the Pearson family of distributions using accurate, calibrated and coherent subjective percentile estimates elicited from experts. The correlations between the primary variables are elicited as positive definite correlation matrices. The project economic structure describes an engineering project in three hierarchical levels, namely, work package/revenue stream, project

performance and project decision. Each of these levels can be described by $Y = g(\mathbf{X})$, with the derived variables of the lower levels as the primary variables for the upper level. Therefore, the input as expert judgements is only at the work package/revenue stream level.

Project duration is estimated by combining the generalized PNET algorithm to the project economic structure. This permits the evaluation of the multiple paths in the project network. Also, the limiting values of the PNET transitional correlation (0,1) permits the estimation of bounds on all of the derived variables. Project cost and revenue are evaluated in terms of current, total and discounted dollars, thereby emphasizing the economic effects of time, inflation and interest on net present value and internal rate of return. The internal rate of return is evaluated from a variation of Hillier's method.

The analytical method is validated using Monte Carlo simulation. The validations show that the analytical method is a comprehensive and extremely economical alternative to Monte Carlo simulation for economic risk quantification of large engineering projects. In addition, they highlight the ability of the analytical method to go beyond the capabilities of simulation in the treatment of correlation, which are seen to be significant in the application problems. From these applications a technique to provide contingencies based on the probability of project success and to distribute the contingency to individual work packages is developed.

Table of Contents

List of Tables

List of Figures

Acknowledgement

I wish to express my most sincere gratitude to Dr. Alan Russell, my teacher and supervisor, who has had a profound and positive impact on my academic and professional attitudes. I greatly appreciate his advice, guidance and support throughout my graduate studies. This thesis would not exist without his patient efforts and valuable suggestions.

My special thanks to Dr. Bill Caselton and Dr. Ricardo Foschi for the numerous stimulating discussions, especially on the areas of correlations and simulations. Their efforts in reviewing this thesis are greatly appreciated. I thank Dr Frank Navin and Dr. Karl Bury for serving on my supervisory committee.

Acknowledgement is most gratefully extended to the Canadian Commonwealth Scholarship and Fellowship Plan who provided the scholarship which enabled me to pursue graduate studies in Canada. My special thanks to Miss. Deirdre Roeser of the above plan for making my stay in Canada an enjoyable experience.

To Ashley Herath, Arif Rahemtulla, Gerard Canisius, Damika Wickremesinghe, Chanaka Edirisinghe, Ron Yaworsky, Ibrahim Al-Hammad, fellow colleagues and friends, many thanks for your moral support and encouragement. You made the bad times more tolerable and the good times more enjoyable.

Finally, to Deepthi, your patience, support and encouragement is most gratefully acknowledged.

To my parents and to my grandmother

for their support and encouragement

in all my endeavors.

Chapter 1

Introduction

"Far better an approximate answer to the right question,
which is often vague, than an exact answer to
the wrong question, which can always be made precise."
John W. Tukey,
Annals of Mathematical Statistics, vol. 33, 1962, p.13.

1.1 General

This thesis describes the development, validation and application of an analytical method for time and economic risk quantification during feasibility analysis for large engineering projects. The method has the ability to quantify the uncertainty in and estimate the bounds on decision variables of a project implemented in traditional, fast track or phased construction.

The pragmatic convention of treating risk and uncertainty as synonyms is adopted in this thesis. The precision in the computations presented is to facilitate comparisons with other risk quantification techniques. This precision however, belies the accuracy of estimations which can be achieved for real life projects.

This chapter describes the background for the research problem, the economic structure adopted to represent an engineering project, the research objectives, a brief state-of-the-art and an overview of the thesis.

1.2 Background for the Research

Large, complex engineering projects will continue to be undertaken both in the developed and developing worlds to meet the increase in demand for infrastructure, energy, raw materials and employment. Typically these projects have long durations, high costs, multiple investors and are undertaken in uncertain environments. The generation of benefits at the earliest possible date to pay back or justify the large investments required for such projects has necessitated the adoption of concepts such as fast track and phased construction. The very nature of these concepts coupled with the increasing size and complexity of such projects necessitates explicit treatment of risk and uncertainty, especially in the early stages.

The World Bank reports that about 20% of the projects evaluated between 1974 and 1986 were determined to be unsatisfactory (see figure 1.1). The "satisfactory" projects between 1974 - 1984 were based on the achievement of at least a 10% economic rate of return, or other significant benefits if the economic rate of return was lower, or an evaluator's qualitative judgement about the performance if no economic rate of return was calculated. The classifications for 1985 and 1986 were based on achievement of one of the three states: 1. wholly satisfactory : project achieves or exceeds all its major objectives, achieves substantial results in almost all respects; 2. satisfactory : project achieves most of its objectives and has satisfactory results with no major shortcomings; 3. marginally satisfactory : project reveals major shortcomings in meeting objectives and/or achievements but is still considered worthwhile. (Project Performance Results for 1986 (1988)). Figure (1.2) depicts the average economic rates of return at appraisal and average re-estimated economic rates of return calculated shortly after final disbursement of Bank funds. Both rates are based on future flows of economic benefits. The first is calculated from project costs and economic events predicted in the appraisal phase, while the second is based on the actual project cost, relative price changes and current economic events. The figures clearly display the risks and uncertainty associated with the predictions that are made during feasibility analysis.

The critical reasons for project failure, besides an adverse economic environment, were deficiencies in project design. These include the lack of: clarity and acceptance of objectives in terms of technical, economic and administrative criteria; and/or the thoroughness with which the project design is prepared and appraised. Over one third of the projects reviewed by the World Bank in 1985 were judged to have been adversely affected by deficiencies in preparation or appraisal (The Twelfth Annual Review of Project Performance Results, 1987).

A profile of the project completion time overruns/underruns for 1513 projects reviewed by World Bank between 1974 to 1986 is shown in figure (1.3). Time overrun/underrun refers to the difference between actual and appraised project execution time. The execution time is from the signing date of the loan/credit to actual completion date. The average project execution time for those reviewed in 1986 was 6.1 years. The principal reasons for completion delays were inadequate project preparation, changes in project scope, administrative constraints within the country and the unfamiliarity of the borrower with Bank procurement procedures, delays in the appointment of staff or consultants, and lack of financial support for the project by the borrower (The Twelfth Annual Review of Project Performance Results, 1987). The average cost overruns for 1269 projects reviewed by the World Bank between 1974 to 1986 are depicted in figure (1.4). The average cost overrun is the unweighted mean of the percent cost overrun for individual projects.

The World Bank states that while Bank forecasting methods deserve continual scrutiny to enhance their effectiveness in identifying development opportunities, the

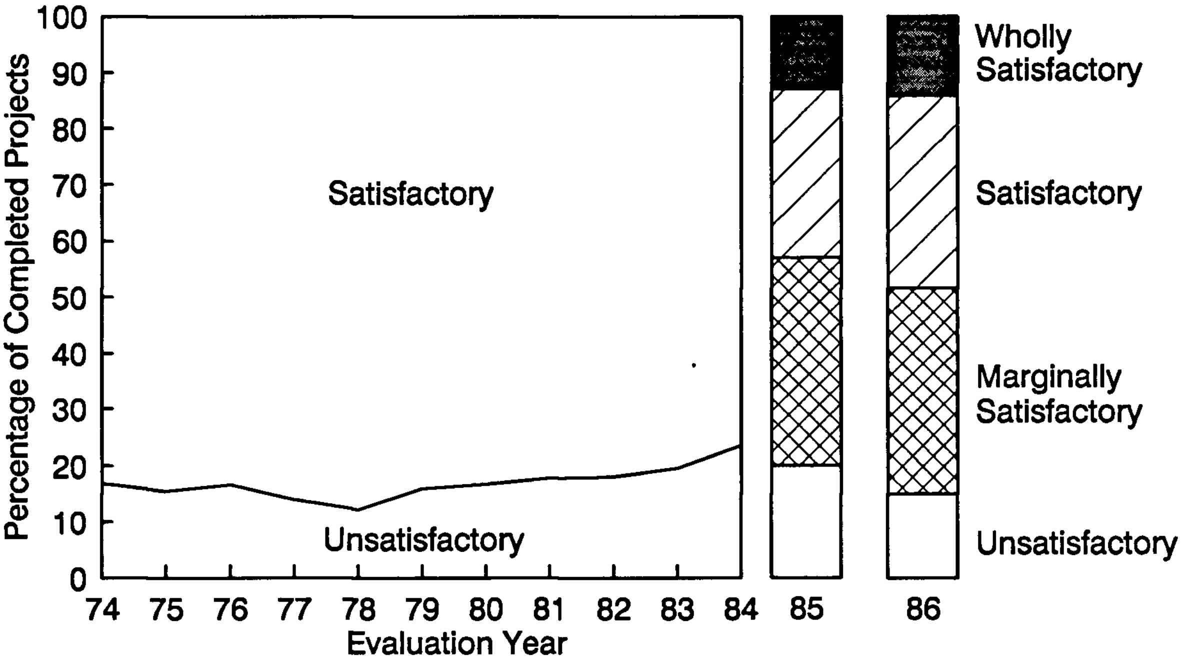

Figure 1.1: Overall Assessment of Project Results

Source : Appendix Tables 8 and 9. - Project Performance Results for 1986 (1988), Appendix Table 1.14 - 12th Annual Review of Project Performance Results, (1987).

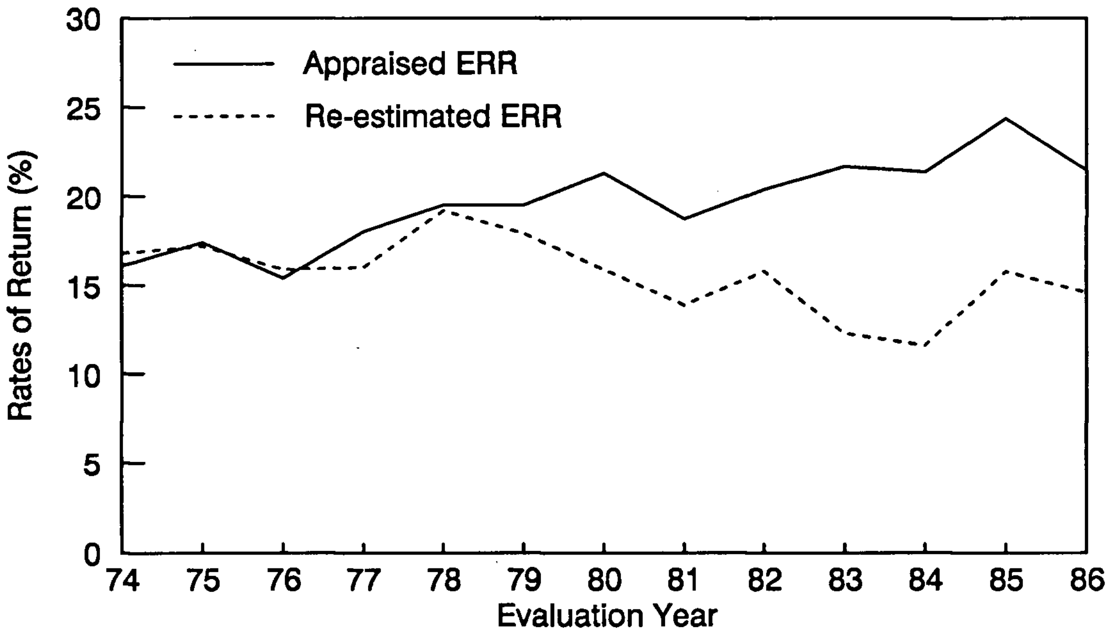

Figure 1.2: Average Economic Rates of Return for Evaluated Projects

Source : Appendix Table 10 - Project Performance Results for 1986 (1988).

resulting investments will continue to face considerable risk and uncertainty. The array of difficulties now confronting borrowers, such as foreign debt, domestic inflation, exchange rates and the continued volatility of external factors, implies that risk will remain an important issue, calling for broader risk analysis and more deliberate efforts at risk management. It is suggested that the way to address risks directly at the feasibility stage is to present the probability of project success (Project Performance Results for 1986 (1988)).

After an extensive study on risk management in engineering construction, Hayes et al. (1986) concluded that: all too often, risks are either ignored, or dealt with in a completely arbitrary way (simply adding 10% contingency onto the estimated cost of a project is typical); and the greatest uncertainty is present in the earliest stages in the life of a project, which is also when decisions of greatest impact are made. Risks must be treated at this phase; and since all parties involved in construction projects and contracts would benefit from reduction in uncertainty prior to financial commitment, more effort should be devoted to risk management. While risk and uncertainty are distinguished in the context of decision analysis (Siddall, 1972), Perry and Hayes (1985b) state that risk and uncertainty are inherently present in all construction projects and in the practice of construction risk management such distinctions are unnecessary and may even be unhelpful.

The objective of the feasibility analysis is to develop and evaluate alternatives so that the most desirable ones are selected and implemented. Generally speaking, the selected alternatives should be, in the decision maker's view, the best in terms of technical, economic and socio-political feasibility. However, in practice technical feasibility is considered as the dominant criterion (Jaafari, 1988a, 1988b). Youker (1989) states that the economic analysis should be treated as the decision criterion and done before, rather than after detailed engineering design.

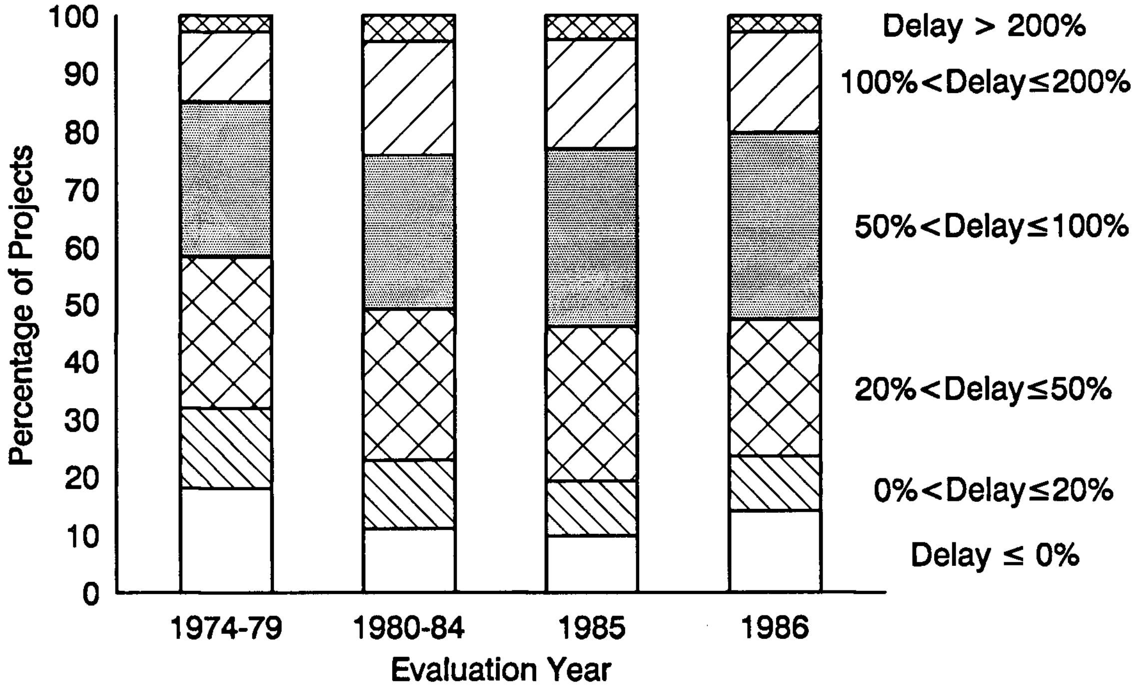

Figure 1.3: Project Completion Time Overruns/Underruns

Source : Appendix Table 18 - Project Performance Results for 1986 (1988).

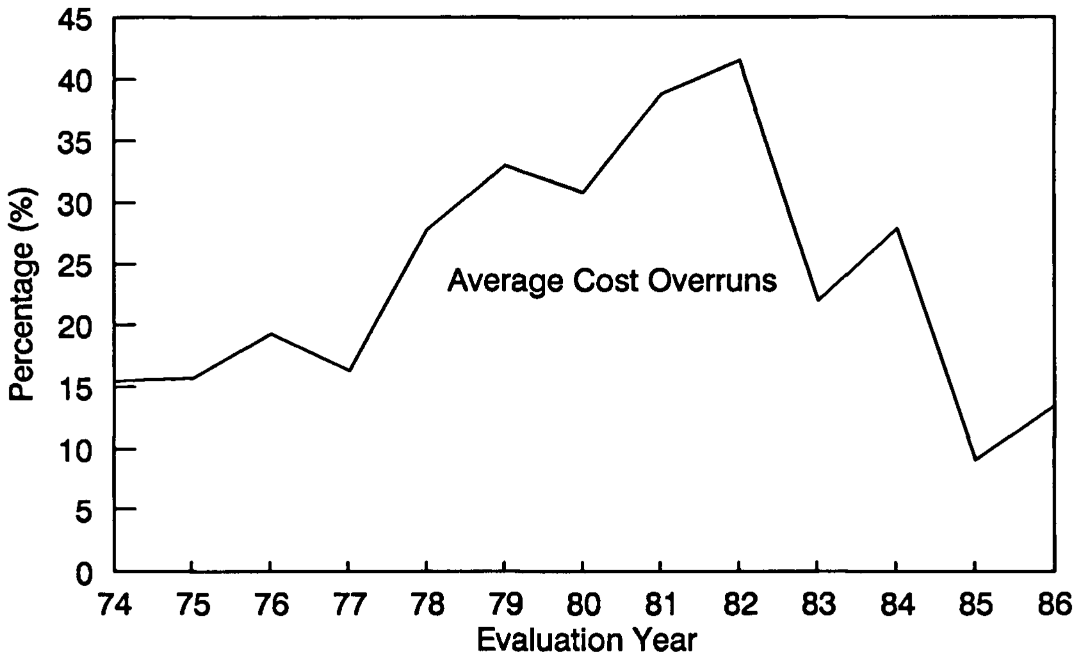

Figure 1.4: Average Project Cost Overruns

Source : Appendix Table 17 - Project Performance Results for 1986 (1988).

A project is economically feasible if the net present value of the benefits generated from it exceeds the net present value of its cost at the minimum attractive rate of return (marr). This thesis treats net present value of a project at marr and its internal (economic) rate of return as the two basic measures that guide the decisions on the economic feasibility of an engineering project (Au, 1988; Bonini, 1975; Cooper and Chapman, 1987; Thompson and Wilmer, 1985). Other measures such as: ratio of net present value over total initial capital investment (Jaafari, 1988b) to complement total life cycle cost (Jaafari, 1988a, 1988c) and risk adjusted discount rate (Farid et al., 1989) have been suggested for construction projects. Taylor (1988) argues that the most reliable approach for appraising projects is using the criterion of net present value alone, and not net present value divided by the initial cost.

The greatest degree of uncertainty about the future estimates is encountered at the feasibility stage. Consequently, decisions taken during this stage of a project can have a large impact on its final cost and its duration. However, it is in this stage that decision makers have the greatest leeway to make changes in the scope of the project, restructure a marginally unfeasible project into a feasible one, or even to cancel the project with minimum loss (Youker, 1989). The limited information available at this stage increases the uncertainty of such decisions. The ability to identify, measure and respond to potential risks and uncertainties will significantly improve the quality of decisions made during feasibility analysis. This process of risk identification, risk quantification and risk response is considered as the most suitable approach for risk management in engineering projects (Flanagan et al., 1987; Perry and Hayes, 1985b).

More comprehensive discussions on risk management in engineering projects are found in Ashley (1980a, 1980b); Ashley and Bonner (1987), Chicken and Hayns (1989); Cooper and Chapman (1987); Hayes et al. (1986); Jaafari (1986, 1987, 1988b); Perry (1986); Perry and Hayes (1985a, 1985b); Thompson (1981); Youker (1989).

1.3 Problem Statement and Structure

Economic risk quantification is a vital step for risk management in large engineering projects because it develops the basis for the decision maker to respond to identified risks. While economic risk quantification techniques for engineering projects are available, in their current form many lack the ability to model large engineering projects realistically for a comprehensive feasibility analysis. Some of the considerations for realistic modeling are: limitation of data and the need for judgements; interaction of time with cost and revenue; correlation among variables; existence of multiple paths to complete a project; the number of variables that can be used in the analysis; and most importantly the need to evaluate a range of alternatives economically to select the best strategy to develop a project. These issues are dealt with explicitly in this thesis in the formulation of the analytical method for risk quantification.

Central to this method is the description of the project economic structure as a hierarchy containing all of the derived time and economic variables of an engineering project. The one presented is an extension of the structure developed by Ranasinghe (1987) to represent an engineering project. In this thesis, three levels of description, namely, project decision, project performance and work package/revenue stream, describe the project economic structure.

Work packages and revenue streams of a project are linked together by way of a precedence network (see figure 1.5). The work package/revenue stream level is considered as the lowest level at which meaningful information can be obtained during feasibility analysis. However, if necessary, a work package can be further decomposed to a sub-network of activities, with each activity having a duration and cost function. Definitions relevant to this structure are as follows.

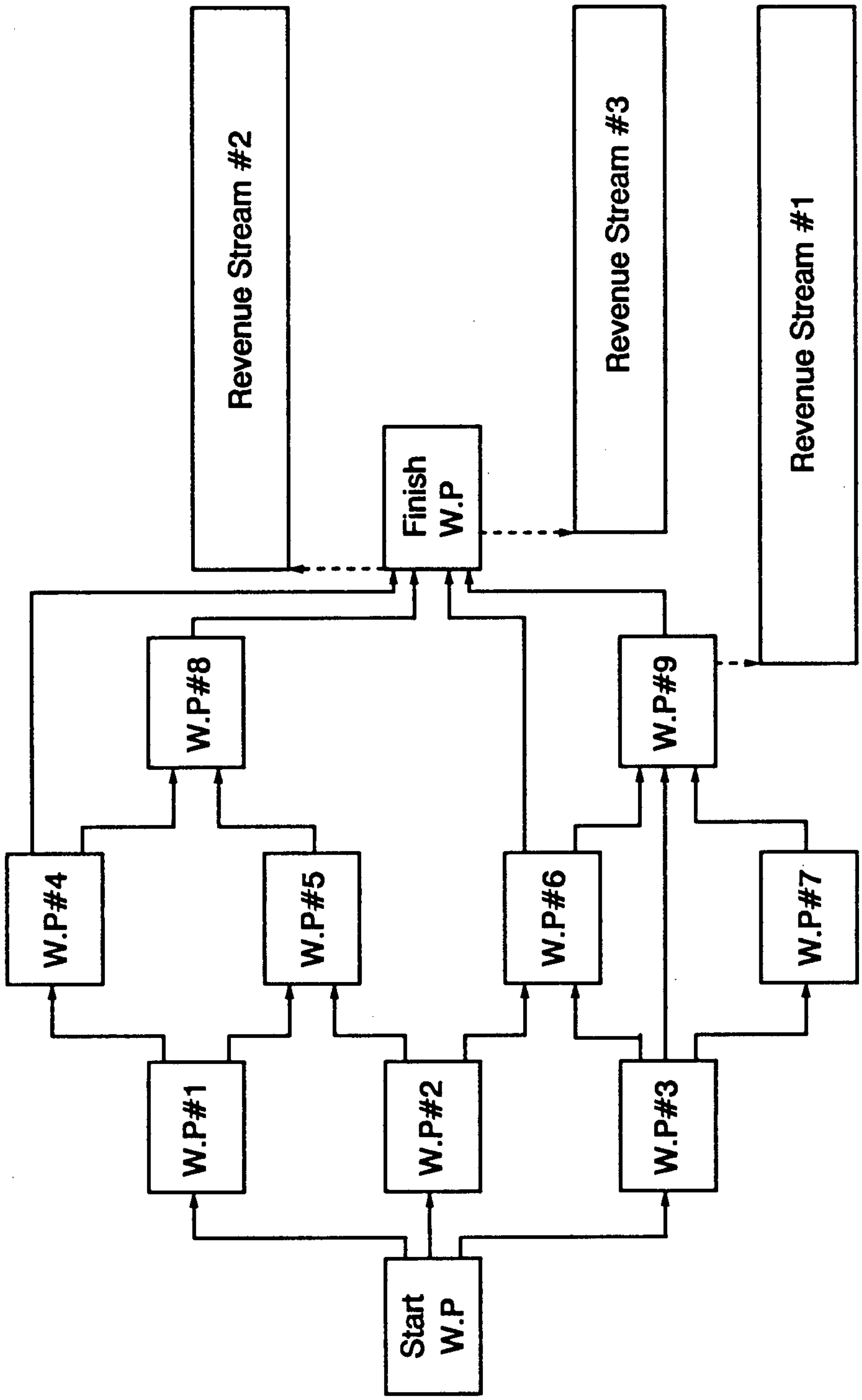

Figure 1.5: Precedence Network for an Engineering Project

1.3.1 Work Package/Revenue Stream Level

The variables at the work package/revenue stream level are,

Work Package Duration : Work package duration can be estimated directly by the analyst (holistic value) or derived using a functional relationship which treats work scope, anticipated job conditions, likely construction methods, productivity and resource levels or a sub-network of activities.

Work Package Cost : A generalized expression for work package cost which can be used to estimate constant, current, total dollar cost and discounted value is as follows:

$$\begin{aligned} WPC_i &= f\ e^{(\theta_{C_i}-y)T_{S_{Ci}}} \int_0^{T_{Ci}} C_{0i}(\tau)\ e^{(\theta_{C_i}-y)\tau}\ d\tau \\ &\quad +(1-f)e^{(r-y)T_p} e^{\theta_{C_i} T_{S_{Ci}}} \int_0^{T_{Ci}} C_{0i}(\tau) e^{(\theta_{C_i}-r)\tau} d\tau \end{aligned} \tag{1.1}$$

where WPC_i is the discounted i^{th} work package cost, $C_{0i}(\tau)$ is the function for constant dollar cash flow for the i^{th} work package, $T_{S_{C_i}}$ and T_{Ci} are work package start time and duration, T_p is the time at which the repayment of interim financing is due for all work packages, f is the equity fraction, θ_{C_i}, r and y are inflation, interest and discount rates respectively. The time τ is measured from the start of the i^{th} work package. $C_{0i}(\tau)$ can be either holistic or a decomposed function of work scope, resources applied, and productivity.

The work package cost is expressed in discounted dollars for generality. When required, the work package cost can be expressed: in total dollars (constant + inflation + financing) by setting the discount rate to zero (WPC_{TDi}); in current dollars by setting the discount rate to zero and equity fraction to one (WPC_{CDi}); or in constant dollars by setting discount and inflation rates to zero and equity fraction to one (WPC_{C0i}).

Net Revenue Stream : The present value of a net revenue stream can be expressed as follows:

$$NRS_i = \int_{T_{S_{Ri}}}^{T_{S_{Ri}}+T_{Ri}} \left[R_{0i}(t)e^{\theta_{R_i}(t-T_{S_{Ri}})} - M_{0i}(t)e^{\theta_{M_i}t}\right] e^{-yt}dt \tag{1.2}$$

where NRS_i is the discounted i^{th} net revenue stream, $R_{0i}(t)$ and $M_{0i}(t)$ are the functions for constant dollar cash flow for i^{th} gross revenue and operation and maintenance cost, $T_{S_{Ri}}$ and T_{Ri} are early start time and duration of the revenue stream, $\theta_{R_i}, \theta_{M_i}$ and y are inflation and discount rates respectively. $R_{0i}(t)$ and $M_{0i}(t)$ can be either holistic or decomposed functional forms.

1.3.2 Project Performance Level

The variables at the project performance level are as follows.

Project Duration

$$T_j = \sum_{i=1}^{n} WPD_{ij} \tag{1.3}$$

where T_j is the duration of the j^{th} path and WPD_{ij} is the duration of the i^{th} work package on the j^{th} path. For the deterministic case, project duration is given by,

$$T = max\ \forall_j\ (T_j) \qquad j = 1, ..., n \tag{1.4}$$

When time is uncertain, the probability of completing the project in time t, denoted by $p(t)$, (Ang et al., 1975), is given by,

$$\begin{aligned} p(t) &= 1 - [P(T_1 > t) + P(T_1 \leq t, T_2 > t) \\ &\quad + ... + P(T_1 \leq t, T_2 \leq t, ..., T_{n-1} \leq t, T_n > t)] \end{aligned} \tag{1.5}$$

where $T_1, T_2, ..., T_n$ are durations of the possible paths to complete the project. Times and probabilities for intermediate milestones can be determined in a similar manner.

Project Cost

$$\begin{aligned}
Discounted\ Project\ Cost &= \sum_{i=1}^{n} WPC_i && (1.6)\\
Project\ Cost\ in\ Total\ Dollars &= \sum_{i=1}^{n} WPC_{TDi} && (1.7)\\
Project\ Cost\ in\ Current\ Dollars &= \sum_{i=1}^{n} WPC_{CDi} && (1.8)\\
Project\ Cost\ in\ Constant\ Dollars &= \sum_{i=1}^{n} WPC_{COi} && (1.9)
\end{aligned}$$

Project Revenue

$$Discounted\ Project\ Revenue = \sum_{i=1}^{r} NRS_i \qquad (1.10)$$

1.3.3 Project Decision Level

The variables at the project decision level are as follows.

Net Present Value

$$NPV = Discounted\ Project\ Revenue - Discounted\ Project\ Cost \qquad (1.11)$$

Internal Rate of Return

$$IRR = Discount\ Rate\ when\ NPV = 0 \qquad (1.12)$$

1.3.4 Observation

The variables at every level of the project economic structure can be described by $Y = g(\mathbf{X})$, where Y is defined as the derived variable and $\mathbf{X}$ is the vector of its primary variables. The derived variables of the lower levels of the project economic structure are the primary variables for the higher levels. At the work package/revenue stream level the derived variables are work package duration, start time, cost and net revenue streams. At the project performance level the derived variables are the project duration, cost, revenue and cash flow profile while the primary variables are the derived variables at the work package/revenue stream level. At the project decision level the derived variables are project net present value and internal rate of return, while the primary variables are discounted project cost and revenue.

At the work package/revenue stream level, time, cost and revenue may be predicted using a variety of functional forms - growth and decay functions for revenue, production functions for time and cost through to network models. These production functions are generally multiplicative and/or additive in nature. The functions for derived variables at the project performance and decision levels are always predetermined and linear.

1.4 Objectives of the Research

The primary objectives of this research are:

1. to develop an analytical method for economic risk quantification during feasibility analysis for large engineering projects,
2. to computerize the method to explore its behavior, to validate it and to test its practicality in the measurement of uncertainty of performance and decision variables.

The secondary objective of this research is to lay the foundation for obtaining the input data necessary to make the analytical method a practical tool for the construction industry. The input data are in the form of subjective probabilities and correlation matrices for primary variables.

The desired features of the analytical method are: model the interaction of time, cost and revenue throughout the life cycle of the project; provide the freedom to model a project to any level of detail using any number of variables; recognize the constraints that exist during feasibility stage, such as data limitations and the need for subjective probabilities; quantify the uncertainty in and estimate bounds on performance variables such as duration, cost, revenue, net present value and internal rate of return of a project; perform sensitivity and probabilistic analysis; consider multiple paths (shorter paths with higher variance or skewness) when evaluating project duration; treat correlations at all levels of the project; estimate individual contributions to overall uncertainty; provide intermediate milestone information to set realistic targets for performance; and have the flexibility to model and evaluate a range of alternatives economically to select the best strategy to develop a project.

1.5 Previous Research and Motivation

A review of the literature shows that estimates for project decision and performance variables are still treated deterministically by most authors. A number of authors have recognized the random nature of estimates and adopted probabilistic concepts in developing their methods. These methods are classified below depending on their individual applications.

Probabilistic Time Methods : those which evaluate the duration of activities and the project as the decision criterion (Ahuja and Nandakumar, 1985; Ang et al., 1975; Carr et al., 1974; Crandall, 1976, 1977; Crandall and Woolery, 1982; Elmaghraby, 1977; Hall, 1986; Jaafari, 1984; Kennedy and Thrall, 1976; King and Wilson, 1967; King et al., 1967; King and Lukas, 1973; McGough, 1982; Mirchandani, 1976; Pritsker and Happ, 1966; Pritsker and Whitehouse, 1966; Woolery and Crandall, 1983).

Probabilistic Cost Methods : those which evaluate project cost as the decision criterion (Bjornsson, 1977; Deshmukh, 1976; Flanagan and Norman, 1980; Flanagan et al., 1987; Hemphill, 1968; Reinschmidt and Frank, 1976; Shafer, 1974; Smith and Thoem, 1976; Spooner, 1974; Vergara and Boyer, 1974; Wallace, 1977).

Probabilistic Time/Cost Methods : those which treat cost as time dependent when evaluating project cost as the decision criterion (Ahuja and Arunachalam, 1984; Baker, 1986; Borcherding, 1977; Chapman, 1979; Chapman and Cooper, 1983; Chapman et al., 1985; Cooper et al., 1985; DeCoster, 1976; Diekmann, 1983; Jaafari, 1988a; 1988c; Moeller, 1972; Thompson and Whitman, 1973; Van Tetterode, 1971).

Probabilistic Present Value Methods : those which evaluate project net present value and internal rate of return as decision criteria (Cooper and Chapman, 1987; Hillier, 1963, 1969; Hull, 1980; Pouliquen, 1970; Reutlinger, 1970; Thompson, 1981; Thompson and Whitman, 1974; Thompson and Wilmer, 1985; Wagle, 1967; Zinn et al., 1977).

From this classification, only those methods which evaluate project net present value and internal rate of return are suitable for economic feasibility studies because they represent the family of criteria used to evaluate a project. Of these, CASPAR (Computer Aided Simulation for Project Appraisal and Review) developed by Thompson and Wilmer (1985) is the widely applied model for large engineering projects - Severn Tidal Power (Thompson et al., 1980), Mersey Barrage (Perry et al, 1983).

CASPAR (Thompson and Wilmer, 1985) is a project management tool designed to model the interaction of time, cost and revenue throughout the entire life of a project. It differs from the normal economic appraisal model as it is network based and is designed to simulate the realistic interaction of time and money. CASPAR models a project in four stages. The first is a definitive model of the project constructed from a network of inter-related activities using a precedence diagram, to which costs and revenues are attached. The second stage identifies and investigates major uncertainties by performing a sensitivity analysis. During the third stage the definitive model is reviewed in light of the sensitivity analysis. At the fourth stage a probabilistic risk analysis is performed using the revised definitive model in a Monte Carlo simulation. A suitable probability distribution is assumed for the uncertain variables - a generalized triangular distribution has been assumed for variables in the applications. The decision criteria are the project net present value and internal rate of return. PROJECT (Thompson and Whitman, 1974) is an older version of CASPAR.

When CASPAR and other simulation based methods (Bjornsson, 1977; Flanagan et al., 1987; Hull, 1980; Jaafari, 1988a; 1988c; Moeller, 1972; Pouliquen, 1970; Thompson and Whitman, 1974; Van Tetterode, 1971) are considered in the context of the desired features of the analytical method, issues such as modelling interaction of time, cost and revenue throughout the life cycle of the project; quantifying uncertainty of decision variables by developing cumulative distribution functions; performing sensitivity and probabilistic analysis; treating correlations at the level of variable input; the effect of multiple paths in the project network when evaluating project duration; and obtaining milestone information are resolved.

However, when the number of variables in the analysis is large and the variables are correlated, Monte Carlo simulation can be both time consuming and computationally expensive, precluding exploration of a wide range of alternative strategies. Hence, the motivation for an analytical method that can handle a realistic formulation of the problem, a large number of correlated variables in the analysis and yet is computationally economical, thereby permitting the evaluation of several strategies.

1.6 Structure of the Thesis

Chapter two develops the risk measurement framework which is the foundation for the analytical method. The framework, based on four assumptions, quantifies the uncertainty of a derived variable that is functionally related to a set of primary variables ($Y = g(\mathbf{X})$). The uncertainty of the derived variable is quantified by developing a cumulative distribution function for it. This development is based on the first four moments of a derived variable obtained from moment analysis using the truncated Taylor series expansion of the transformed function for $g(\mathbf{X})$, and the first four moments of transformed variables. The first four moments are the expected value and second to fourth central moments. The correlations between primary variables are treated by using a variable transformation approach. A numerical example is used to demonstrate the framework, while the stochastic breakeven problem is used for comparison with some published results (Kottas and Lau, 1978).

Chapter three develops an approach to elicit accurate and calibrated subjective probabilities as percentile estimates of an expert's subjective prior probability distribution for a primary variable. The analysis and verification method ensures that the measured belief is coherent and useful for the quantification of uncertainty of a derived variable.

Chapter four discusses the correlations between variables. The discussion highlights the positive definite correlation matrix. A method to elicit a positive definite correlation matrix for primary variables and a method to obtain a positive definite correlation matrix for the derived variables when only linear correlations between primary variables are available are developed. Numerical comparisons under general conditions and multicollinearity are performed to show that the variable transformation approach is more robust than the standard method used to treat correlations in moment analysis.

Chapter five describes a study on the decomposition of a derived variable that is sometimes estimated holistically in the elicitation of subjective probabilities. The study contains hypotheses, an experiment and test statistics to compare the two estimation approaches used in engineering risk quantification. The duration of an activity is used as the example for the derived variable to compare holistic versus decomposed methods of estimation.

Chapter six combines all of the developments and studies done in chapters two to five with the project economic structure to develop the analytical method for time and economic risk quantification for large engineering projects. The method computes the moments for derived variables at the work package/revenue stream level (work package duration, cost, and net revenue), project performance level (project duration, cost and revenue) and project decision level (net present value and internal rate of return) using the moments and correlation matrices for primary variables in their functional forms. The shape characteristics of the derived variables are used to approximate Pearson type distributions to quantify their uncertainty. The computed moments for derived variables at project decision and performance levels are exact. The approximations for moments are only for the derived variables at the work package/revenue stream level. The expected value, standard deviation and cumulative distribution function for project duration are obtained from the modified PNET approach (Ang et al., 1975), while those for project internal rate of return are derived from a variation of Hillier's method (Hillier, 1963).

Chapter seven describes the validations and applications of the analytical method. The validations are done using Monte Carlo simulation. The modified PNET algorithm is validated by solving two numerical examples that were presented by Ang et al. (1975). The Monte Carlo simulation process is first validated using two limiting cases. The first limiting case is a parallel network while the second is a single dominant path in a highly interrelated network. Six simulations of two engineering projects are used to validate the analytical method. The data for the first example is obtained from an actual deterministic feasibility analysis. The second is a hypothetical engineering project developed to demonstrate the full potential of the analytical method. The types of sensitivity analyses that can be performed by the analytical method are explored. A detailed method to distribute the contingency for a derived variable to its primary variables using one of the sensitivity analyses is presented.

Chapter eight contains the conclusions and recommendations for future work.

Appendices A, B, C and E contain proofs and derivations used for the developments described by this thesis. Appendix D describes the two computer programs, "ELICIT" and "TIERA", developed to obtain input data and perform time and economic risk quantification. Appendix F contains the input values used for the numerical examples presented in chapter seven.

Chapter 2

Risk Measurement Framework

2.1 General

The framework to quantify the uncertainty of a derived variable is developed in this chapter. The inspiration for this development is the "unified statistical framework for probabilistic planning models" suggested by Kottas and Lau (1980), (1982). Their framework is a computational alternative to simulation for models involving additive and multiplicative functions of random variables. The proposed framework is for any arbitrary function, $g(\mathbf{X})$, between the derived variable and its primary variables.

This development is based on four assumptions which are explicitly identified and discussed, moment analysis and a function $g(\mathbf{X})$. The use of a truncated Taylor series expansion of the function, $g(\mathbf{X})$, for moment analysis generalizes the type of functional relationship between the derived variable and its primary variables. In addition to $g(\mathbf{X})$, moments of the primary variables are required to evaluate the moments of the derived variable. The Pearson family of distributions and subjective percentile estimates are used to approximate the moments of the primary variables. The cumulative distribution function for the derived variable is approximated from the Pearson family of distributions using its first four moments.

The next section describes the Pearson family of distributions. In the third section an iterative process for approximating the first four moments of a primary variable is developed. The fourth section describes the approach to approximate the first four moments of the derived variable. The approximation of the cumulative distribution function for the derived variable is described in the fifth section. The application of the risk measurement framework to three examples: duration of a construction activity; the breakeven analysis problem; and a linear function is presented in the sixth section. The seventh section highlights the contributions of this chapter.

2.2 The Pearson Family of Distributions

The Pearson family of distributions are obtained as solutions of the differential equation which, when the origin of x is at the mean has the form,

$$\frac{dy}{dx} = \frac{-\,y\,(x\,+\,b)}{a\,+\,b\,x\,+\,c\,x^2} \qquad L_1 \,<\, x \,<\, L_2 \tag{2.1}$$

where the coefficients a, b, and c are functions of the moment ratios $(\sqrt{\beta_1}, \beta_2)$, and may be expressed as (Amos and Daniel, 1971),

$$a = \frac{\mu_2\,(4\,\beta_2\,-\,3\,\beta_1)}{10\,\beta_2\,-\,12\,\beta_1\,-\,18} \tag{2.2}$$

$$b = \frac{\sqrt{\beta_1}\,\mu_2\,(\beta_2\,+\,3)}{10\,\beta_2\,-\,12\,\beta_1\,-\,18} \tag{2.3}$$

$$c = \frac{2\,\beta_2\,-\,3\,\beta_1\,-\,6}{10\,\beta_2\,-\,12\,\beta_1\,-\,18} \tag{2.4}$$

where $\beta_1 = \frac{\mu_3^2}{\mu_2^3}$ and $\beta_2 = \frac{\mu_4}{\mu_2^2}$. μ_2, μ_3, and μ_4 are the second, third, and fourth central moments of the random variable x.

For each pair of $(\sqrt{\beta_1}, \beta_2)$ the solution of equation (2.1) defines a density function with mean zero on an interval $L_1 < x < L_2$. The solutions assume a variety of different mathematical forms according to the values of the moment ratios (Johnson et al., 1963). These forms or "types of distributions" may be associated with different regions in a plane having rectangular co-ordinate axes $\sqrt{\beta_1}$ and β_2 (see figure 2.1).

Since the shapes of the distributions change continuously across the boundaries of the regions, Johnson et al. (1963) compiled tables for the Pearson family of distributions by treating the problem as a single unity. These tables, tabulate the standardized deviate for fifteen percentage points based on the values of $\sqrt{\beta_1}$ and β_2. The fifteen percentage points are namely the median, upper and lower 0.25, 0.5, 1.0, 2.5, 5.0, 10.0, and 25.0 percentage points. Amos and Daniel (1971) extended the Pearson tables to cover a much larger area of the plane $(\sqrt{\beta_1}, \beta_2)$.

Assumption 2.1 : The derived and the primary variables are continuous and their probability distributions are approximated by the Pearson family of distributions.

The variables of the project economic structure such as time, cost, revenue, inflation and interest rates are all continuous in nature. The continuous random variable is a convenience for probabilistic applications. Most of the probability distributions used for applications in engineering such as - Normal, Beta (*TypeI*), Gamma (*TypeIII*), Exponential (*TypeX*), Uniform, Lognormal (*TypeV*), Student's t (*TypeVII*), Chi-square (*TypeIII*), F (*TypeVI*), are members of the Pearson family of distributions (Harr, 1987; Ord, 1985).

While there is no guarantee that a Pearson type distribution will always provide a good fit for a variable, the theoretical developments of the Pearson family (Kendall and Stuart, 1969; Ord, 1972) and the widespread applications of the Pearson system indicate that it will provide a good fit to most "real life" distributions (Kottas

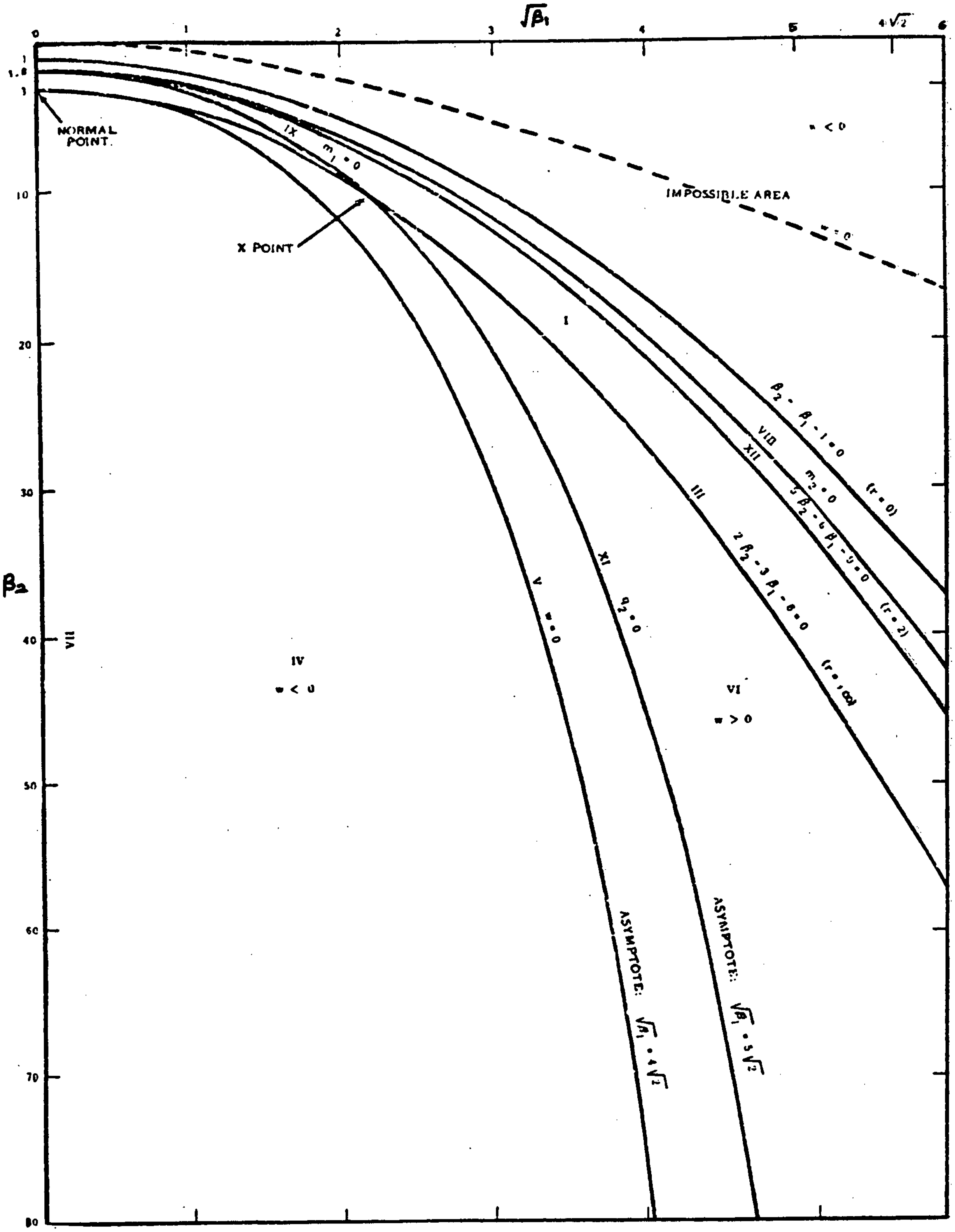

Figure 2.1: Moment Ratio Plane Showing Pearson *Types I-XII*

Source : Amos and Daniel, (1971)

and Lau, 1982). However, it must be noted that there are theoretical examples for which Pearson type provides a poor fit for a variable with the same first four moments (Pearson, 1963). The assumption permits the development of the framework as a "distribution free" method because of the high flexibility of the Pearson system (Siddall, 1972).

2.3 The Moments of a Primary Variable

A continuous random primary variable approximated by a Pearson type distribution can be expressed by its first four moments (Kendall and Stuart, 1969; Ord, 1972). The first four moments of a primary variable are its expected value and second to fourth central moments.

Assumption 2.2 : An expert can provide estimates for percentiles of his subjective prior probability distribution for a primary variable at the input level.

The use of subjective probabilities to quantify the uncertainty about the primary variables stems from the observation that actuarial or relative frequency based data are unavailable or not meaningful for direct input as probability forecasts for estimating future events (Wright and Ayton, 1987). To use subjective probabilities as input to risk analyses, they have to be accurate, calibrated and coherent (Lindley et al., 1979). In chapter three a method to elicit accurate, calibrated and coherent subjective probabilities as percentile estimates of an expert's subjective prior probability distributions for primary variables is developed.

A step by step iterative process for approximating the first four moments of a primary variable (see figure 2.2) is set out in this section. The sole purpose of approximating third and fourth central moments of primary variables is to approximate third and fourth central moments of the derived variable. This information is required to fit a Pearson distribution and to make probabilistic statements about the derived variable.

The starting point follows from the first two assumptions. The first assumption permits the use of the table for percentage points of standardized Pearson distributions (Amos and Daniel, 1971; Johnson et al., 1963). From the second assumption estimates for the 5, 25, 50, 75, and 95 percentiles of an expert's subjective prior probability distribution for a primary variable are elicited. The process of approximating the first four moments of a primary variable stops when **either** of the following conditions are met.

Condition 1 : When $\sigma^*_{0.05}$ (equation 2.8) is greater than $\sigma^*_{0.025}$ (equation 2.9),

Condition 2 : When "best fit" distribution is the same as that of the previous cycle.

The step by step process for generating the first four moments for a primary variable is as follows.

Step 1 : Subjective Estimates

Obtain the estimates for the 5, 25, 50, 75, and 95 percentiles of the expert's subjective prior probability distribution for the primary variable (assumption 2.2).

Step 2 : Expected Value and Standard Deviation

Since the time Malcolm et al. (1959) suggested the well known approximations for PERT, a number of different studies have been done on the approximations for the expected value and standard deviation of a continuous random variable (Britney, 1976; Davidson and Cooper, 1976; Hull, 1978; Keefer and Bodily, 1983; Moder and Rodgers, 1968; Pearson and Tukey, 1965; Perry and Greig, 1975). From an extensive empirical study, Pearson and Tukey (1965) developed approximations to the expected value and standard deviation for the Pearson family of distributions. This development was based on the constancy of the ratio of the distances between suitable symmetrical percentage points to the standard deviation.

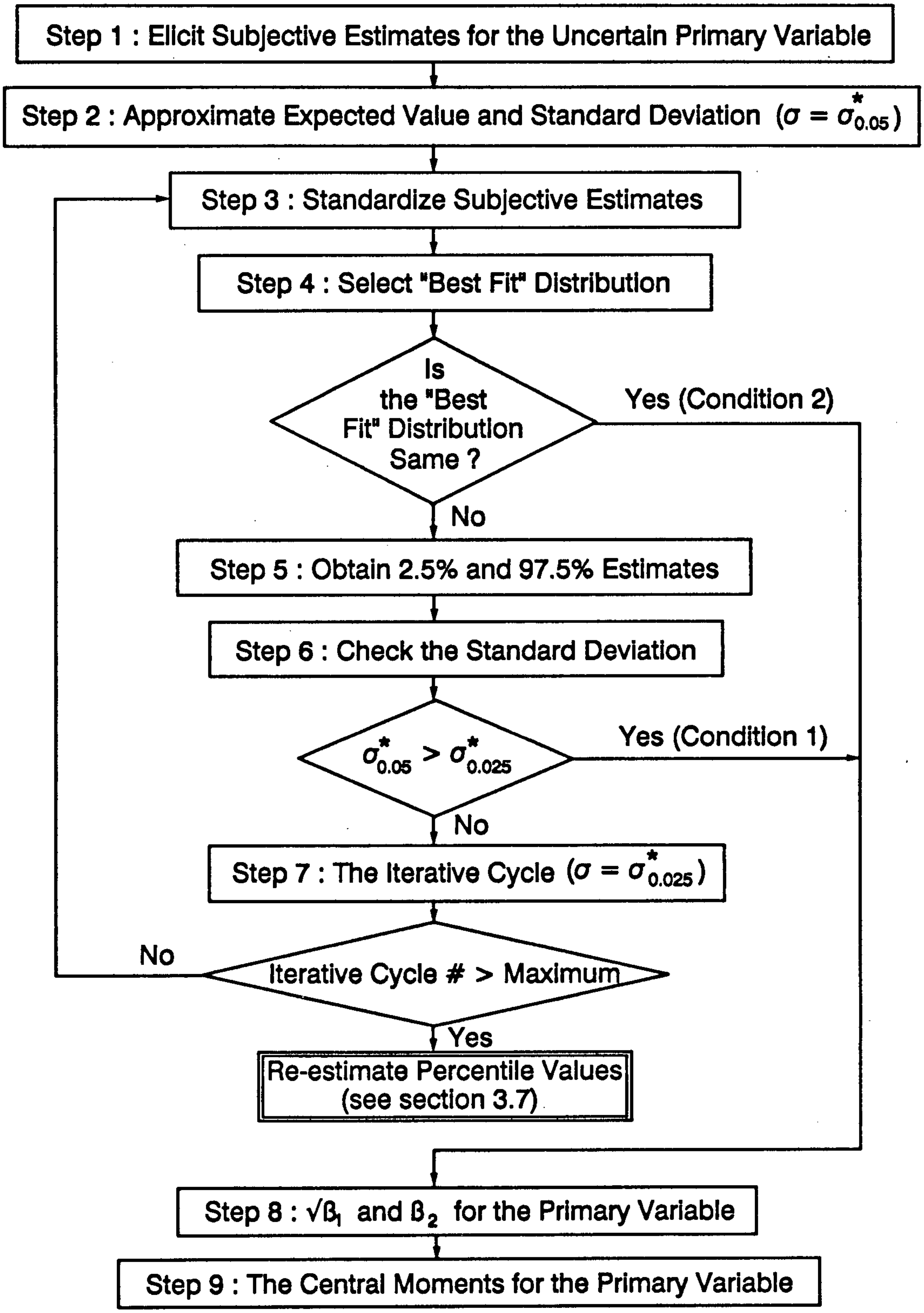

Figure 2.2: The Steps of the Iterative Process

The approximation for the expected value from percentile values ([P%]) is,

$$E[X] \;=\; [50\%] \;+\; 0.185\;\Delta \tag{2.5}$$

where

$$\Delta \;=\; [95\%] \;+\; [5\%] \;-\; 2\;[50\%]. \tag{2.6}$$

The approximation for the standard deviation using percentile values and the iteration scheme suggested by Pearson and Tukey (1965) is,

$$\sigma \;=\; max\;\{\sigma^{*}_{0.05}\;,\;\sigma^{*}_{0.025}\} \tag{2.7}$$

where

$$\sigma^{*}_{0.05} \;=\; \frac{[95\%] \;-\; [5\%]}{max\left\{3.29 \;-\; 0.1\left[\dfrac{\Delta}{\sigma_{0.05}}\right]^2, 3.08\right\}} \tag{2.8}$$

$$\sigma^{*}_{0.025} \;=\; \frac{[97.5\%] \;-\; [2.5\%]}{max\left\{3.98 \;-\; 0.1\left[\dfrac{\Delta}{\sigma_{0.025}}\right]^2, 3.66\right\}} \tag{2.9}$$

$\sigma_{0.05}$ and $\sigma_{0.025}$ are the approximations for the standard deviation from the previous iteration. For the first iteration $\sigma_{0.05}$ and $\sigma_{0.025}$ are defined on the basis of figure (3) of Pearson and Tukey (1965) as,

$$\sigma_{0.05} \;=\; \frac{[95\%] \;-\; [5\%]}{3.25} \tag{2.10}$$

$$\sigma_{0.025} \;=\; \frac{[97.5\%] \;-\; [2.5\%]}{3.92} \tag{2.11}$$

Pearson and Tukey (1965) state that the error in the approximation of the expected value is not more than 0.1% for a large area of the $(\sqrt{\beta_1}, \beta_2)$ plane and not more than

0.5% for the rest. The error for the standard deviation is less than 0.5% for a very large area of the $(\sqrt{\beta_1}, \beta_2)$ plane.

After comparing most of the approximations available to estimate expected value and standard deviation of continuous random variables from judgmental (subjective) estimates, Keefer and Bodily (1983) concluded that the approximations suggested by Pearson and Tukey (1965) are more accurate, often by a wide margin, than their competitors. For their study Keefer and Bodily (1983) used only the approximation given by equation (2.8) for the standard deviation because of the difficulty of assessing the 2.5 and 97.5 percentiles subjectively. However, the standard deviation approximated from equation (2.8) alone is an underestimation for a large part of the Pearson family (Pearson and Tukey, 1965). In developing the framework both approximations for the standard deviation (equations 2.8 and 2.9) are included in the iterative approach, thereby ensuring that the approximated standard deviation for the primary variable is the maximum. The 2.5 and 97.5 percentiles for equations (2.9) and (2.11) are obtained as described in steps 3 through 7.

The five subjective estimates from step 1 are used in equations (2.5), (2.6), (2.8) and (2.10) to approximate the expected value and the standard deviation for the primary variable. The process of determining the standard deviation using equation (2.7) starts with σ equal to $\sigma^*_{0.05}$.

Step 3 : Standardize the Subjective Estimates

Using the approximations for the expected value and the standard deviation of the primary variable from step 2, the five subjective estimates from step 1 are standardized by,

$$X_p = \frac{x_p - E[X]}{\sigma} \tag{2.12}$$

where x_p is a subjective percentile estimate and X_p is its standardized value.

Step 4 : The "Best Fit" Distribution

The standardized estimates from step 3 are then compared with the 5.0, 25.0, 50.0, 75.0, and 95.0 percentage points for the standardized Pearson variable tabulated by Amos and Daniel (1971), by minimizing the sum of squared deviations as suggested by Ord (1972) to approximate the "best fit" distribution. The acceptable error of the approximation (square root of the sum of squared deviations) for the "best fit" distribution should be specified by the user. A maximum cumulative error of 10% of the standard deviation is used as a default value in the computer program.

Step 5 : 2.5% and 97.5% Estimates

For the "best fit" distribution from step 4 obtain the standardized Pearson variable values for 2.5% and 97.5% points (see figure 2.3). From these standardized values generate the actual values for the two percentiles from,

$$x_p = X_p \, \sigma + E[X] \tag{2.13}$$

Step 6 : Check for the Standard Deviation

From the generated values for 2.5% and 97.5% in step 5 and equations (2.9) and (2.11) evaluate $\sigma^*_{0.025}$.

If $\sigma^*_{0.05} > \sigma^*_{0.025}$: go to step 8 as **Condition 1** is satisfied.

If $\sigma^*_{0.05} < \sigma^*_{0.025}$: go to step 7 for the iterative cycle. The standard deviation for the primary variable σ is now equal to $\sigma^*_{0.025}$.

Step 7 : The Iterative Cycle

Go back to step 3 to start the iterative cycle. If the "best fit" distribution from step 4 is same as for the previous cycle then go to step 8 as **Condition 2** is satisfied. If not continue till either of the conditions are met for a specified number of iterative cycles.

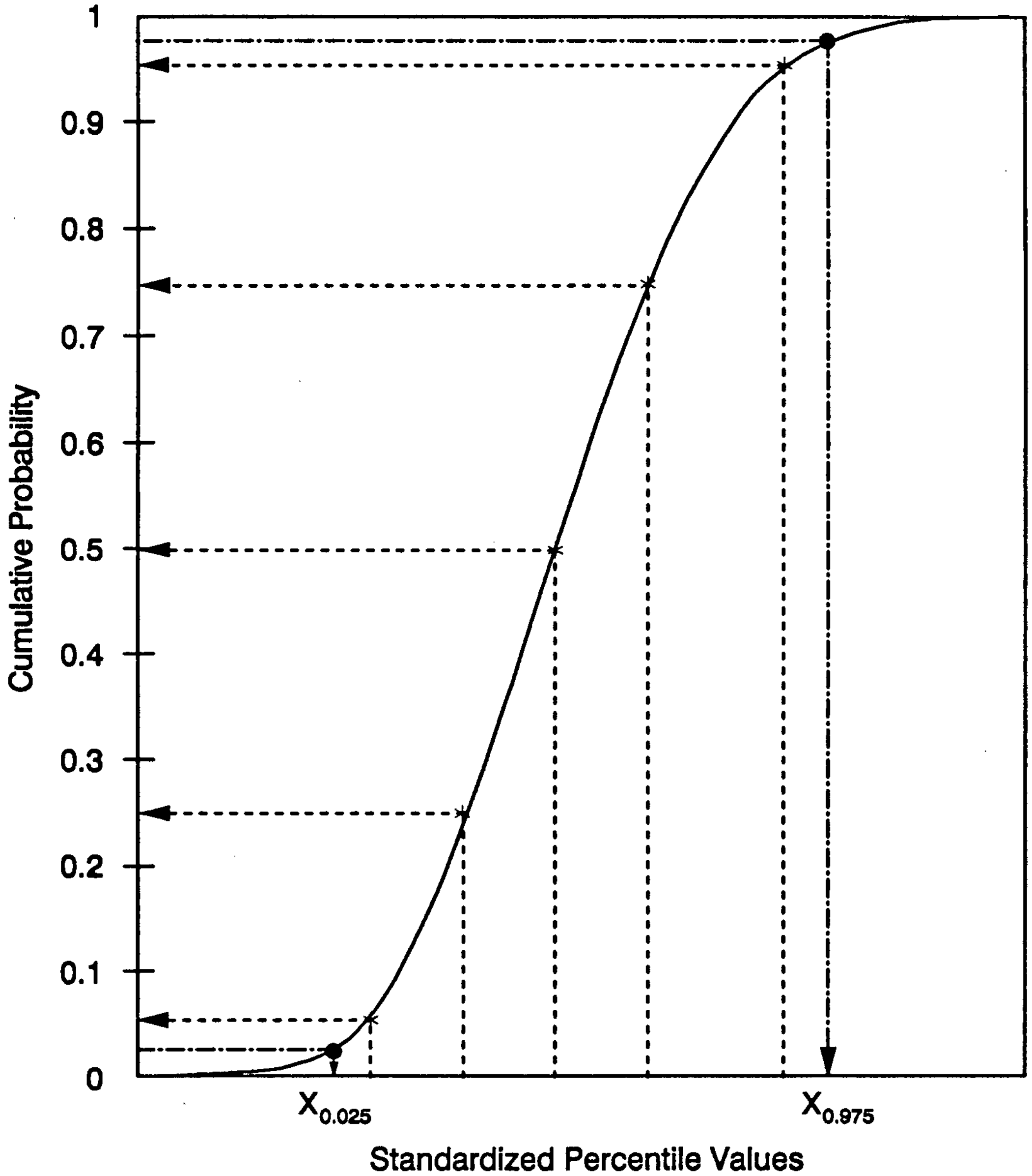

Figure 2.3: The "Best Fit" Distribution

Step 8 : $\sqrt{\beta_1}$ and β_2 for the Primary Variable

Obtain $\sqrt{\beta_1}$ and β_2 from the Pearson table (Amos and Daniel, 1971) for the selected "best fit" distribution in step 4. When **Condition 1** is satisfied, from equation (2.7) the standard deviation for the variable is $\sigma^*_{0.05}$ (the condition used by Keefer and Bodily, 1983). When **Condition 2** is satisfied the standard deviation for the variable is $\sigma^*_{0.025}$. Then, the requirement specified by Pearson and Tukey (1965) in equation (2.7) is fulfilled.

Step 9 : The Central Moments

From the standard deviation approximated at step 8 and the $\sqrt{\beta_1}$ and β_2 for the "best fit" distribution, the second, third and fourth central moments of the primary variable are evaluated from,

$$\mu_2(X) = \sigma^2 \tag{2.14}$$

$$\mu_3(X) = \sqrt{\beta_1}\,\sigma^3 \tag{2.15}$$

$$\mu_4(X) = \beta_2\,\sigma^4 \tag{2.16}$$

2.4 Moments of the Derived Variable

The method to approximate the moments of the derived variable is based on moment analysis. The moment analysis use the moments of the transformed variables and a truncated Taylor series expansion of the transformed function for $g(\mathbf{X})$ to approximate the first four moments of the derived variable.

Assumption 2.3 : A derived variable can be more accurately estimated from a set of primary variables that are functionally related to it than by direct estimation.

When a functional form between a derived variable and primary variables is used in stochastic applications, it is based on the premise that it is more accurate to estimate the primary variables individually than to estimate the derived variable directly (Kottas and Lau, 1982). It reflects the engineering penchant to seek more detail as a way of seeking greater precision. The analytical method developed in chapter six does not require this assumption at all levels but allows for elaboration of time and cost estimating relationships to achieve more precision. However, when variables which are sometimes assessed holistically are used in decomposed estimation (duration, productivity) the assumption becomes debatable. Chapter five examines the validity of assumption (2.3) for such variables.

2.4.1 Truncated Taylor Series

For a system of n primary variables described by the function, $Y = g(\mathbf{X})$, which has continuous partial derivatives, the Taylor series expansion of the function $g(\mathbf{X})$ about the mean values $\bar{\mathbf{X}}$ is given by,

$$\begin{aligned} g(\mathbf{X}) &= g(\bar{\mathbf{X}}) + \sum_{i=1}^{n} (X_i - \bar{X}_i) \frac{\partial g}{\partial X_i} \\ &\quad + \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} (X_i - \bar{X}_i)(X_j - \bar{X}_j) \frac{\partial^2 g}{\partial X_i \partial X_j} + \ldots\ldots. \end{aligned} \tag{2.17}$$

The Taylor series is then truncated at the second order such that the truncation error of the approximation is,

$$|R_2| = \frac{1}{3!} \sum_{i=1}^{n} \sum_{j=1}^{n} \sum_{k=1}^{n} (X_i - \bar{X}_i)(X_j - \bar{X}_j)(X_k - \bar{X}_k) \frac{\partial^3 g}{\partial X_i \partial X_j \partial X_k} \tag{2.18}$$

The truncated second order approximation of the expansion is,

$$
\begin{aligned}
g(\mathbf{X}) \approx\ & g(\bar{\mathbf{X}}) + \sum_{i=1}^{n} (X_i - \bar{X}_i) \frac{\partial g}{\partial X_i} \\
& + \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} (X_i - \bar{X}_i)(X_j - \bar{X}_j) \frac{\partial^2 g}{\partial X_i \partial X_j}
\end{aligned}
\tag{2.19}
$$

where the partial derivatives are evaluated at $\bar{\mathbf{X}}$. The partial derivatives constitute sensitivity coefficients and either increase or decrease the contribution of each term, depending on the importance of each variable to the derived variable, thereby, acting as an in-built sensitivity analysis.

The second order approximation provides reasonable mathematical ease for moment analysis. A third or higher order approximation would give more accurate results (Tukey, 1954), but mathematical complexities that are involved when treating statistical dependencies prohibit their use. The moments of a derived variable can be evaluated using the truncated Taylor series expansion with the definition of moments (Siddall, 1972). Then, the first four moments of the derived variable are,

$$E[Y] = E[g(\mathbf{X})] \tag{2.20}$$

$$\mu_2(Y) = E\left[(Y - E[Y])^2\right] \tag{2.21}$$

$$\mu_3(Y) = E\left[(Y - E[Y])^3\right] \tag{2.22}$$

$$\mu_4(Y) = E\left[(Y - E[Y])^4\right] \tag{2.23}$$

To evaluate accurate first four moments, correlations between primary variables have to be treated. The standard approach to treat correlations in moment analysis is by expanding the above equations (Ang and Tang, 1975). This approach can include the linear correlations easily only in the approximation for the first two moments. A variable transformation approach that can include the linear correlations in the higher order moments of the derived variable is used in the development of this framework.

Assumption 2.4 : The correlations between primary variables are linear.

Generally, when the correlations among primary variables are treated it is the linear correlations. If all the variables in the system are normally distributed then the linear correlations between variables are the true correlations. In general, the primary variables which describe a work package are not normally distributed. Consequently, one is faced with the prospect of non-linear correlations. Obtaining non-linear correlations or treating non-linear correlation in a multivariate situation are still complex and largely unresolved theoretical issues. Most four moment methods (Jackson 1982; Siddall, 1972) avoid the treatment of correlations; their treatment is important, however, if one wishes to establish an accurate measurement of risk (Perry and Hayes, 1985b; Cooper and Chapman, 1987) and a realistic estimate of bounds.

2.4.2 Variable Transformation Method

A set of correlated variables are transformed to a set of uncorrelated variables having mean values and unit variances by,

$$\mathbf{Z} = \mathbf{L}^{-1}\,\mathbf{D}^{-1}\,\mathbf{X} \tag{2.24}$$

where $\mathbf{X}$ is the vector of correlated random variables, $\mathbf{X} = [X_1, ..., X_n]^T$; $\mathbf{Z}$ is the vector of transformed variables with unit covariance matrix; $\mathbf{L}^{-1}$ is the inverse of the lower triangular matrix obtained from the Cholesky decomposition of the correlation matrix $\mathbf{R}$ $(= \mathbf{LL^T})$; and $\mathbf{D}^{-1}$ is the inverse of the diagonal matrix of standard deviations of the $\mathbf{X}$ vector $(\mathbf{D} = \text{diag}\,[\sigma_i])$.

Proof of the Transformation

Let $\mathbf{X}$ be a vector of correlated random variables with covariance matrix $\mathbf{C_x}$ and correlation matrix $\mathbf{R}$. Let $\mathbf{Z}$ be the vector of transformed variables from equation (2.24) with covariance matrix $\mathbf{C_z}$. Then,

$$\mathbf{Var[Z]} = \mathbf{Var[L^{-1}\,D^{-1}\,X]} \tag{2.25}$$

$$\mathbf{C_z} = \mathbf{L^{-1}\,D^{-1}\,C_x}\,\left[\mathbf{L^{-1}\,D^{-1}}\right]^{\mathbf{T}} \tag{2.26}$$

Using the relationship $\mathbf{R} = \mathbf{D^{-1}\,C_x\,D^{-1}}$ and the Cholesky decomposition of the correlation matrix $\mathbf{R} = \mathbf{LL^T}$,

$$\mathbf{L^{-1}\,D^{-1}\,C_x\,D^{-1}} = \mathbf{L^{-1}\,L\,L^T} = \mathbf{L^T} \tag{2.27}$$

Similarly,

$$\mathbf{L^{-1}\,D^{-1}\,C_x\,D^{-1}}\,\left[\mathbf{L^T}\right]^{-1} = \mathbf{L^T}\,\left[\mathbf{L^T}\right]^{-1} = \mathbf{I} \tag{2.28}$$

Since $\left[\mathbf{L^T}\right]^{-1} = \left[\mathbf{L}^{-1}\right]^{\mathbf{T}}$ and $\mathbf{D}^{-1} \equiv \left[\mathbf{D}^{-1}\right]^{\mathbf{T}}$ because $\mathbf{D}^{-1}$ is symmetric,

$$\mathbf{L^{-1}\,D^{-1}\,C_x}\,\left[\mathbf{L^{-1}\,D^{-1}}\right]^{\mathbf{T}} = \mathbf{I} \tag{2.29}$$

From equations (2.26) and (2.29),

$$\mathbf{C_z} = \mathbf{I} \tag{2.30}$$

Therefore, the transformed variables are uncorrelated with unit variances.

Even with assumption (2.4) it is not possible to prove that the variable transformation precludes the existence of non-linear correlations amongst the transformed variables. This has implications for the terms treated in approximating the fourth central moment (see section 2.4.5, chapter four and Appendix A).

A similar transformation to obtain a set of standard variates with zero means and unit covariance matrix from a set of correlated variables was used by Der Kiureghian and Liu (1986) for applications in structural reliability.

2.4.3 Moments of the Uncorrelated Variables

Since the transformation given by equation (2.24) is linear the first four moments of the transformed uncorrelated variables can be evaluated directly from the moments of the correlated primary variables. Then, the first four moments of a transformed uncorrelated primary variable are,

$$E[Z_i] = \sum_{j=1}^{n} A_{ij}\ E[X_j] \tag{2.31}$$

$$\begin{aligned}\mu_2(Z_i) &= \sum_{j=1}^{n} A_{ij}^2\ \mu_2(X_j) \\ &\quad + 2 \sum_{j=1}^{n} \sum_{k=j+1}^{n} A_{ij}\ A_{ik}\ cov(X_j, X_k) = 1\end{aligned} \tag{2.32}$$

$$\mu_3(Z_i) \approx \sum_{j=1}^{n} A_{ij}^3\ \mu_3(X_j) \tag{2.33}$$

$$\mu_4(Z_i) \approx \sum_{j=1}^{n} A_{ij}^4\ \mu_4(X_j) \tag{2.34}$$

where $\mathbf{A} = \mathbf{L}^{-1}\mathbf{D}^{-1}$ and $E[X_j]$, $\mu_2(X_j)$, $\mu_3(X_j)$, $\mu_4(X_j)$ are the first four moments of the j^{th} correlated primary variable in the $\mathbf{X}$ vector.

2.4.4 The Function

To use moments of the transformed uncorrelated primary variables to evaluate the first four moments of the derived variable Y, the function $g(\mathbf{X})$ has to be transformed to the uncorrelated space. This transformation is done from,

$$\mathbf{X} = \mathbf{D}\,\mathbf{L}\,\mathbf{Z} \tag{2.35}$$

Then the transformed function is $Y = G(\mathbf{Z})$. If the original function $g(\mathbf{X})$ was complicated, this transformation increases the complexity as each variable in the $\mathbf{X}$ vector is replaced by a linear combination of variables from the $\mathbf{Z}$ vector. However, since this transformation is linear and in practice the replacement will be done by the computer, the increased complexity of the transformed function is not apparent to the user.

2.4.5 The First Four Moments

The first four moments of the derived variable are now evaluated using the transformed function, $G(\mathbf{Z})$, as the function for the derived variable. The moment analysis considers the terms involving up to the fourth order because moment information is available up to the fourth order. The cross moment terms for which information is not available are neglected (see Appendix A for derivations). The approximation for the expected value is,

$$E[Y] \approx G(\bar{\mathbf{Z}}) + \frac{1}{2}\sum_{i=1}^{n} \frac{\partial^2 G}{\partial Z_i^2}\,\mu_2(Z_i) \tag{2.36}$$

the approximation for the second central moment is,

$$\begin{aligned}\mu_2(Y) \approx & \sum_{i=1}^{n}\left[\frac{\partial G}{\partial Z_i}\right]^2 \mu_2(Z_i) \\ & + \sum_{i=1}^{n}\frac{\partial G}{\partial Z_i}\frac{\partial^2 G}{\partial Z_i^2}\mu_3(Z_i) \\ & + \frac{1}{4}\sum_{i=1}^{n}\left[\frac{\partial^2 G}{\partial Z_i^2}\right]^2\left[\mu_4(Z_i) - [\mu_2(Z_i)]^2\right]\end{aligned} \tag{2.37}$$

the approximation for the third central moment is,

$$\begin{aligned}\mu_3(Y) \approx & \sum_{i=1}^{n}\left[\frac{\partial G}{\partial Z_i}\right]^3 \mu_3(Z_i) \\ & + \frac{3}{2}\sum_{i=1}^{n}\left[\frac{\partial G}{\partial Z_i}\right]^2\frac{\partial^2 G}{\partial Z_i^2}\left[\mu_4(Z_i) - [\mu_2(Z_i)]^2\right]\end{aligned} \tag{2.38}$$

and the approximation for the fourth central moment is,

$$\mu_4(Y) \approx \sum_{i=1}^{n}\left[\frac{\partial G}{\partial Z_i}\right]^4 \mu_4(Z_i) \tag{2.39}$$

where $\mathbf{Z}$ is the vector of transformed uncorrelated variables and $G(\mathbf{Z})$ is the transformed function for the derived variable. The first $(\frac{\partial G}{\partial Z_i})$ and second $(\frac{\partial^2 G}{\partial Z_i^2})$ partial derivatives with respect to the transformed variables are evaluated numerically (Howard, 1971).

The sole purpose of approximating the third and fourth central moments of the derived variable is to approximate a cumulative distribution function for it. In the approximation for the fourth central moment it is evident that only the first term of the expansion is used (see Appendix A.5). A second fourth order term which cannot be evaluated except for the case of statistical independence and the special case when there are no non-linear correlations between transformed variables, has been ignored. The underestimation of the fourth central moment can create a problem however, because one may not be able to fit a valid Pearson distribution to the derived variable. A valid distribution requires the relation $\beta_2 - \beta_1 - 1 \geq 0$ be satisfied (Johnson et al., 1963).

2.5 Cumulative Distribution Function

The approximated central moments are then used to evaluate the shape characteristics, skewness ($\sqrt{\beta_1}$) and kurtosis (β_2) for the derived variable from,

$$\sqrt{\beta_1} = \frac{\mu_3(Y)}{\mu_2(Y)^{1.5}} \tag{2.40}$$

$$\beta_2 = \frac{\mu_4(Y)}{\mu_2(Y)^2} \tag{2.41}$$

where $\mu_2(Y)$, $\mu_3(Y)$ and $\mu_4(Y)$ are the approximated central moments for the derived variable. A cumulative distribution function for the derived variable is approximated from the Pearson family of distributions (assumption 2.1) using the method suggested by Johnson et al. (1963). The approximated Pearson distribution is the one which corresponds most closely to the shape characteristics of the derived variable. The cumulative distribution function is the quantification of the uncertainty associated with the derived variable.

2.6 Application of the Framework

Three examples are presented to demonstrate the application of the framework. The first is a numerical example for a real construction activity. For the second example, results for the breakeven analysis problem from the framework are compared to those reported by Kottas and Lau (1978). The third is a linear function of the primary variables in the breakeven problem. It is used to highlight some of the reasons for the differences between exact moments and those approximated by the framework.

2.6.1 Example 1 : Activity Duration

The duration to fly form a typical slab of 3000 ft^2 in a single suite per floor high-rise is considered as the derived variable for the numerical example to demonstrate the risk measurement framework. The duration can be estimated from the decomposed relationship given by,

$$T = A + \frac{Q}{P\,L} \tag{2.42}$$

where T is the duration to fly form a typical slab in *days*, Q is the estimated quantity in ft^2, P is the estimated labour productivity in $ft^2/manhour$ once the fly forms are placed, L is the estimated labour usage in $manhours/day$ and A is the time required to fly the forms in *days*. Other authors (Jaafari, 1984; Hendrickson, 1987) have used decomposed relations to compute the activity duration. While the quantity is deterministic, the other three variables are considered as random. Then, equation (2.42) can be re-written as,

$$T = X_1 + \frac{3000}{X_2\,X_3} = g(\mathbf{X}) \tag{2.43}$$

The subjective percentile estimates for the random variables and the positive definite correlation matrix ($\mathbf{R}$) elicited from an experienced engineer are given below.

Table 2.1: Subjective Percentile Estimates for A, P and L

Variable	5%	25%	50%	75%	95%
$A(days)$	0.25	0.33	0.375	0.42	0.5
$P(ft^2/mh)$	17.0	19.0	20.0	21.0	22.0
$L(mh/day)$	75.0	83.0	88.0	92.0	96.0

$$\mathbf{R} = \begin{bmatrix} 1.0 & -0.5 & 0 \\ -0.5 & 1.0 & -0.4 \\ 0 & -0.4 & 1.0 \end{bmatrix}$$

The negative correlation between X_1 and X_2 suggests the greater the productivity, probably the greater the efficiency of flying the forms and vice versa. The negative correlation between X_2 and X_3 implies that the smaller the crew the greater the productivity (minimum congestion, all crew members visible and not able to hide).

The expected values, standard deviations, and shape characteristics of the approximated Pearson type distributions for the random variables from equations (2.5) to (2.16) are given in Table 2.2.

Table 2.2: Statistics for the Random Variables

Variable	Expected Value	Standard Deviation	$\sqrt{\beta_1}$	β_2
A (X_1)	0.375	0.08	0.0	9.0
P (X_2)	19.815	1.54	-0.8	4.1
L (X_3)	87.075	6.5	-0.7	3.3

The diagonal matrix of standard deviations ($\mathbf{D}$) and lower triangular matrix from the Cholesky decomposition of the correlation matrix ($\mathbf{R} = \mathbf{LL^T}$) are,

$$\mathbf{D} = \begin{bmatrix} 0.08 & 0.0 & 0.0 \\ 0.0 & 1.54 & 0.0 \\ 0.0 & 0.0 & 6.5 \end{bmatrix} \quad ; \quad \mathbf{L} = \begin{bmatrix} 1.0 & 0.0 & 0.0 \\ -0.5 & 0.866 & 0.0 \\ 0.0 & -0.46188 & 0.88694 \end{bmatrix}$$

The transformed function $G(\mathbf{Z})$, for the duration to fly form a typical slab is obtained using the above in equation (2.35). $\mathbf{Z}$ is the vector of transformed uncorrelated variables. The first four moments of the transformed uncorrelated variables, and the first ($\frac{\partial G}{\partial Z_i}$) and second ($\frac{\partial^2 G}{\partial Z_i^2}$) partial derivatives with respect to the transformed variables which are evaluated numerically (Howard, 1971) are given in Table 2.3.

Table 2.3: First Four Moments and Partial Derivatives of Transformed Variables

Variable	$E[Z_i]$	$\mu_2(Z_i)$	$\mu_3(Z_i)$	$\mu_4(Z_i)$	$\frac{\partial G}{\partial Z_i}$	$\frac{\partial^2 G}{\partial Z_i^2}$
Z_1	4.68177	1.0	0.0	9.0	0.14764	0.00525
Z_2	17.56525	1.0	-1.23168	8.28889	-0.05705	0.01181
Z_3	24.25121	1.0	-1.17720	5.94210	-0.11515	0.01525

The expected value, standard deviation, skewness and kurtosis respectively for duration to fly form a typical slab from equations (2.36) to (2.41) are 2.13 *days*, 0.2045 *days*, 0.622 and 3.095.

2.6.2 Example 2 : Stochastic Breakeven Analysis

The problem of breakeven (or cost-volume-profit) analysis under uncertainty has had considerable discussion in the management literature (Jaedicke and Robichek, 1964; Hilliard and Leitch, 1975; Starr and Tapiero, 1975; Kottas and Lau, 1978; Cooper and Chapman, 1987). Kottas and Lau (1978) used the breakeven analysis problem reported by Starr and Tapiero (1975) to present an "exact" four moment solution.

Their solution to the breakeven equation given by,

$$P(x) \;=\; (p-c)\,x \;-\; K \tag{2.44}$$

where p is the unit sale price; c is unit variable cost; x is sales volume; K is fixed cost; and $P(x)$ is profit realized; was shown to be superior to that given by Starr and Tapiero (1975) using Chebyshev's Inequality. The framework is applied to the same numerical example used by Kottas and Lau (1978). In the numerical example p, c, x and K were assumed to be normally distributed with expected values, standard deviations and correlation coefficients as shown below.

$$\begin{array}{llll}
\mu_p = 1000; & \mu_c = 600; & \mu_x = 1000; & \mu_K = 250000; \\
\sigma_p = 100; & \sigma_c = 60; & \sigma_x = 200; & \sigma_K = 20000; \\
\rho_{pc} = 0.3; & \rho_{px} = -0.4; & \rho_{cx} = -0.2; & \rho_{Kx} = 0.2;
\end{array}$$

Since all the primary variables are normally distributed, there are no non-linear correlations between the transformed variables (see equations A.9, A.15 and A.20). Table 2.4 shows the comparison of the moments for $P(x)$ approximated by the framework with those computed by Kottas and Lau (1978). Figure (2.4) shows the Pearson distributions approximated by Kottas and Lau (1978) and by the framework.

Table 2.4: Comparison of Moments and Shape Characteristics

	Kottas and Lau	Framework	Difference
$E[P(x)]$	$144,400$	$144,400.68$	0%
$\mu_2(P)$	$1.2335 * 10^{10}$	$1.2111 * 10^{10}$	-1.81%
$\mu_3(P)$	$4.3964 * 10^{14}$	$4.5454 * 10^{14}$	3.39%
$\mu_4(P)$	$4.5895 * 10^{20}$	$4.5135 * 10^{20}$	-1.66%
σ_P	$111,063$	$110,048$	-0.91%
$\sqrt{\beta_1}$	0.3209	0.3411	6.29%
β_2	3.0164	3.0775	2.02%

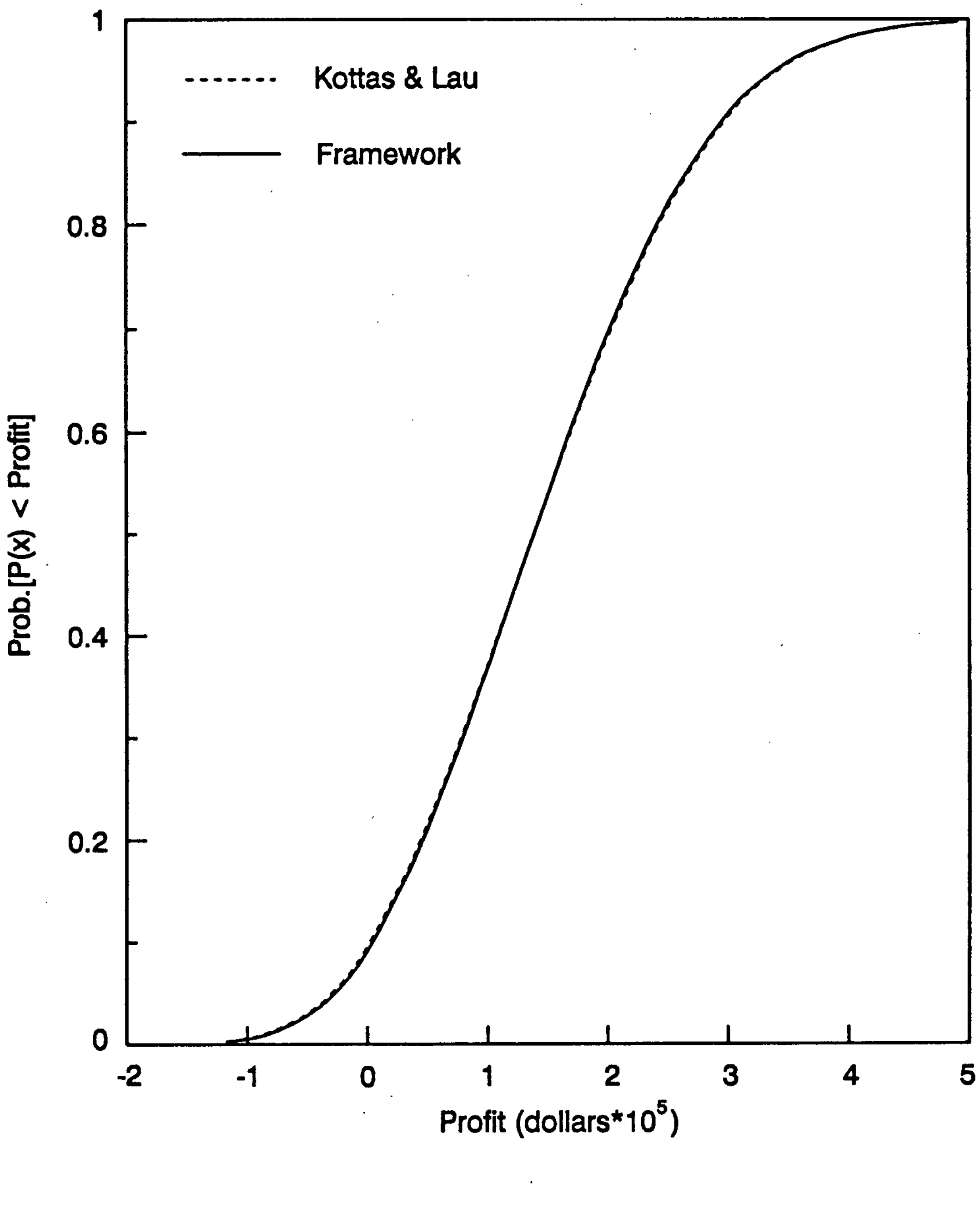

Figure 2.4: Approximated Pearson Type Distributions for $P(x)$

Kottas and Lau (1978) compared estimates from their approach to (1) what is the probability of at least breaking even ? and (2) what is the probability of realizing more than the expected profit of $ 144,400 ? with those from Chebyshev's Inequality and a simulation with a sample size of 50,000. Table 2.5 shows the comparison of those results with that from the framework.

Table 2.5: Comparison of Estimation Approaches

	Starr and Tapiero	Kottas and Lau	Simulation $n = 50,000$	Framework
Prob.$[P(x) > 0]$	$> 41\%$	90.9%	91.2%	91.2%
Prob.$[P(x) > 144,400]$	$> 0\%$	47.9%	47.7%	47.9%

The comparison of the approximated moments to the exact moments, and of the estimation approaches show that the proposed framework is robust. The next example is presented to highlight some of the reasons for the underestimation of the moments.

2.6.3 Example 3 : Linear Function

Assume a linear functional form given by,

$$Y = p + c + x + K \tag{2.45}$$

where p, c, x and K are the same variables as in the previous example. In addition, assume that all of the variables are uncorrelated. Since all of the primary variables are normally distributed, they are now statistically independent. Hence, Y is also normally distributed and its exact moments can be computed. Table 2.6 shows the exact moments of Y and those approximated by the framework.

Table 2.6: Comparison of Moments and Shape Characteristics

	Exact	Framework	Difference
$E[Y]$	$252,600$	$252,600$	0%
$\mu_2(Y)$	400053600	400004963	-0.012%
$\mu_3(Y)$	0	0	0%
$\mu_4(Y)$	$4.8012 * 10^{17}$	$4.7988 * 10^{17}$	-0.050%
$\sqrt{\beta_1}$	0.0	0.0	0%
β_2	3.0	2.9992	-0.027%

The variance of Y is underestimated due to numerical differentiation. The first and second partial derivatives in equations (2.36) to (2.39) are computed numerically to provide for generality of the function. Table 2.7 gives the first partial derivatives of Y with respect to the transformed variables. The second partial derivatives of Y with respect to the transformed variables are zero because the transformed functional form is also linear.

Table 2.7: Comparison of the First Partial Derivatives of Y

	Exact	Framework	Difference
$\frac{\partial Y}{\partial Z_p}$	100	99.99392	0.006%
$\frac{\partial Y}{\partial Z_c}$	60	59.99635	0.006%
$\frac{\partial Y}{\partial Z_r}$	200	199.98784	0.006%
$\frac{\partial Y}{\partial Z_K}$	20000	19998.79419	0.006%

Hence, from equation (2.37),

$$\begin{aligned} \mu_2(Y)_{ex} &= 100^2 * 1 + 60^2 * 1 + 200^2 * 1 + 20000^2 * 1 \\ &= 400053600 \end{aligned}$$

$$\begin{aligned} \mu_2(Y)_{fw} &= 99.99392^2 * 1 + 59.99635^2 * 1 + 199.98784^2 * 1 + 19998.79419^2 * 1 \\ &= 400004963 \end{aligned}$$

When primary variables are statistically independent equation (2.34) should be,

$$\mu_4(Z_i) = \sum_{j=1}^{n} A_{ij}^4\, \mu_4(X_j) + 6 \sum_{j=1}^{n} \sum_{k=j+1}^{n} A_{ij}^2 A_{ik}^2\, \mu_2(X_j)\, \mu_2(X_k) \tag{2.46}$$

and equation (2.39) should be,

$$\mu_4(Y) = \sum_{i=1}^{n} \left[\frac{\partial Y}{\partial Z_i}\right]^4 \mu_4(Z_i) + 6 \sum_{i=1}^{n} \sum_{j=i+1}^{n} \left[\frac{\partial Y}{\partial Z_i}\right]^2 \left[\frac{\partial Y}{\partial Z_j}\right]^2 \mu_2(Z_i)\mu_2(Z_j) \tag{2.47}$$

For generality, the approximation for $\mu_4(Y)$ is based on the assumption that transformed variables will only be uncorrelated. This assumption is reasonable because statistical independence will occur only when all the primary variables are normally distributed. Hence, it is evident that the fourth central moment for the derived variable from the framework will always be an approximation.

Table 2.8 gives the first four moments and shape characteristics for Y when exact, those approximated by the framework in general and when corrected for this example by using equations (2.46) and (2.47) instead of (2.34) and (2.39).

Table 2.8: Comparison of Moments and Shape Characteristics

	Exact	Framework	Corrected
$E[Y]$	252,600	252,600	252,600
$\mu_2(Y)$	400053600	400004963	400004963
$\mu_3(Y)$	0	0	0
$\mu_4(Y)$	$4.8012 * 10^{17}$	$4.7988 * 10^{17}$	$4.8001 * 10^{17}$
$\sqrt{\beta_1}$	0.0	0.0	0.0
β_2	3.0	2.9992	3.0

2.7 Summary

The proposed framework requires: a functional relationship, $g(\mathbf{X})$, between the derived variable and its primary variables; approximation of the first four moments of a primary variable from subjective estimates; approximation of the first four moments of the derived variable from moment analysis using a truncated second order Taylor series expansion of the transformed function and moments of the transformed variables; evaluation of shape characteristics of the derived variable; and approximation of the derived variable to a Pearson type distribution using its shape characteristics. The framework is suitable for systems where pre-determined functions are available, data limitations exist and the decisions are not based on extreme probabilities. The results from the application to the stochastic breakeven problem show that the framework is accurate.

The use of a truncated Taylor series expansion of the system function for moment analysis (Ang and Tang, 1975; Benjamin and Cornell, 1970; Jackson, 1982; Siddall, 1972; Smith, 1971) or the four moment approach for the quantification of uncertainty of a derived variable (Kottas and Lau, 1980, 1982; Siddall, 1972; Jackson, 1982) are not unique. The method to approximate the first four moments of a primary variable from subjective probabilities and the variable transformation method to treat correlations between primary variables in the approximation of the first four moments of the derived variable are unique for this framework. The use of subjective probabilities recognizes the lack of input data for most risk analyses performed during the feasibility stage. The variable transformation method permits the inclusion of correlation information in the approximations for higher order moments of the derived variable which is neglected by the standard approach for moment analysis (see section 4.2 and Appendix A).

In the context of time and economic feasibility of an engineering project, all of the decision and performance parameters have well defined functional forms (even though the functions for derived variables at the work package/revenue stream level can change from analyst to analyst) and significant data limitations exist. In addition, strategic decisions such as contingencies and tolerances for those parameters rarely require probability values beyond the 90^{th} percentile. Therefore, the framework becomes the foundation for the proposed method.

The practical advantages of the framework are the rigor it imparts on the analysis process and the formalized procedure it imparts upon the participants. The analysts and the experts are forced to consider that the inputs are random and to structure their thinking in terms of range estimates. Hence, it quickly becomes apparent what primary variables are the major contributors to the uncertainty of the derived variable.

Chapter 3

Elicitation of Subjective Probabilities

3.1 General

The framework developed in the previous chapter to quantify the uncertainty of a derived variable is based on the assumption that experts can provide estimates for percentile values of their subjective prior probability distributions for primary variables in construction estimation. This is the measurement of the experts' belief about the uncertainty of primary variables. For the measured belief to be useful in the quantification of uncertainty of the derived variable it has to be accurate, calibrated, coherent and also be converted to moments.

While the work described in this chapter is not conclusive, it provides a foundation for obtaining input data necessary to make the analytical method a practical tool for engineering construction. Also, it should be seen as a vital step towards standardizing and computerizing, to the extent possible, elicitation of expert input dealing with uncertainty. Consequently, this chapter achieves the secondary research objective identified in chapter one of this thesis.

Developed in this chapter are an approach to elicit the desired percentile values of an expert's subjective prior probability distributions for variables in engineering construction and a method to ensure the elicited subjective probabilities are coherent and useful in the quantification of uncertainty of the derived variable.

3.2 Subjective Probabilities

After the detailed work of DeFinetti (1970) and Savage (1954), the use of subjective probabilities - the *degree of belief* in the occurrence of an event attributed by a given person at a given instant and a given set of information, is considered a quantification of uncertainty, because it represents the extent to which the person believes a statement is true, based on the information available to him at that time (Hampton et al., 1973).

Subjective probabilities are generally elicited for use in Bayesian decision analysis. Lindley et al., (1979) state that to use subjective assessments in decision analysis they have to be accurate, calibrated and coherent. A person is calibrated if for all events assigned a probability, q, the proportion that actually occur is in fact equal (or close) to q (Budescu and Wallsten, 1987). A set of subjective probabilities are coherent if they are compatible with the probability axioms. Coherence is essential if the assessments are to be manipulated according to probabilistic laws (Lindley et al., 1979).

Wallsten and Budescu (1983) argue that it is not necessary for encodings to obey axioms of additive probability theory in order to be valid measures of belief. Such conformity is necessary only if the user of the judgements wants to treat them as additive probability measures. Wright and Ayton (1987) were surprised by the lack of significant relations between coherence and forecasting performance (*i.e* calibration), because the two ways of assessing the adequacy of a forecaster are logically interrelated. They state that if a forecaster is incoherent he cannot be well calibrated, but it does not follow that coherence necessarily produces good calibration.

A review of the subjective probability literature show that it can be classified into three broad categories, namely, theoretical, review and empirical. The theoretical literature can be further divided into axioms on subjective probabilities (DeFinetti, 1970; DeGroot, 1970, 1975, 1979; French, 1980, 1982; Lindley, 1982; Lindley et al., 1979; Pratt et al., 1964; Savage, 1954, 1971; Suppes, 1975), assessment and consensus of subjective probabilities (Ashton and Ashton, 1985; Bacharach, 1975; Bordley, 1982; Bordley and Wolff, 1981; Diaconis and Ylvisaker, 1985; Dickey, 1979; Dickey and Chen, 1985; Dickey et al., 1986; French, 1985; Holt, 1986; Press, 1979; Winkler, 1986b) and expert resolution (Ashton, 1986; Clemen, 1986; Einhorn, 1972; French, 1980, 1986; Lindley, 1986; Lock, 1987; Morris, 1974, 1977, 1983, 1986; Schervish, 1986; Winkler, 1981, 1986a).

Some of the review literature on subjective probabilities are (Beach, 1975; Beach et al., 1987; Budescu and Wallsten, 1987; Bunn, 1979a, 1979b; Chesley, 1975; Christensen-Szalanski and Beach, 1984; Cooper and Chapman, 1987; Green, 1967; Hampton et al., 1973; Hogarth 1975; Huber, 1974; Ludke et al., 1977; Moore, 1977; Morrison, 1967; Phillips, 1987; Wallsten and Budescu, 1983; Winkler, 1983; Wright and Ayton, 1987), while the empirical studies are (Bunn, 1975; Gustafson et al., 1973; Hull, 1978; Milkovich et al., 1972; Murphy and Winkler, 1971a, 1971b, 1975, 1984; Murphy and Daan, 1984; Murphy et al., 1985; Pratt and Schlaifer, 1985; Press, 1985; Seaver, 1977; Seaver et al., 1978; Smith, 1967; Spetzler and Stael von Holstein, 1975; Stael von Holstein, 1971, 1972; Tversky and Kahneman, 1984; Winkler, 1967a, 1967b, 1968, 1971).

The literature review shows that theoretical investigations on the topic of subjective probabilities have currently far outstripped the empirical studies. Christensen-Szalanski and Beach (1984), after reviewing over 3500 abstracts of articles on probability judgements and decision making found only 84 (2.4%) empirical studies. This is unfortunate because available guidance for the elicitation of subjective probabilities is not on a par with the theoretical analyses. Nevertheless, proven techniques from other fields are used for the development of the elicitation approach described herein.

3.3 Definitions and Assumptions

In this thesis, the terms analyst and expert are used throughout. This section will define these terms and state the assumptions that are central to the development of the elicitation approach.

Analyst

Analyst refers to the individual (or group of individuals) within the firm responsible for conducting economic and financial feasibility, scheduling and cost analyses. He is the key person in the elicitation approach because he must elicit from the expert his belief about the uncertainty of variables as subjective probabilities. To achieve this, the analyst must know the problem, concepts in subjective probabilities and be able to build a rapport with the expert.

Expert

Expert refers to individuals both within and external to the firm who provide key input dealing with economic, revenue, cost, financial, productivity and schedule information and is a person who in his area has some degree of training, experience and/or knowledge significantly greater than that in the general population (Wallsten and Budescu, 1983). These experts are drawn from the fields of economics, finance, design, construction and so forth, when the analyst believes that they possess the most relevant knowledge and information regarding the uncertainty of a primary variable. In general, they are substantive experts, who in a given domain, assess events in their field of expertise (Wallsten and Budescu, 1983).

Based on the literature reviewed, three assumptions that are central for the elicitation of subjective probabilities are stated.

Assumption 3.1 : The experts involved with engineering projects are calibrated.

Budescu and Wallsten (1987) state that calibration is the most important criterion for an expert because it directly compares his performance with empirical reality, and while experienced experts are highly calibrated, calibration can be further improved with training. The calibration curve as shown in figure (3.1) is a bivariate plot of the proportion of events occurring versus the expert's probability assigned to the events. It is linear with unit slope and zero intercept for a "perfectly" calibrated expert (Murphy and Winkler, 1984). Phillips (1987) states that calibration of assessments are usually better for future events made by experts in a group when training and feedback are available. Past studies show that when experts are required to encode subjective probabilities within their area of competence, they can be exceedingly well calibrated (Wallsten and Budescu, 1983).

Assumption 3.2 : Interaction between the analyst and the expert is an essential part of the process.

The main reason for the interaction between the analyst and the expert is to avoid serious misunderstandings and biases. Spetzler and Stael von Holstein (1975) state that even subjects who are well trained in probability or statistics, when having to assign a probability distribution without the help of an analyst often provide poor assignments. Past studies show that interaction is useful (Chesley, 1975; Cooper and Chapman, 1987; Huber, 1974; Hull, 1980; Spetzler and Stael von Holstein, 1975) especially when experts lack experience in providing subjective probabilities.

However, the interaction hinders the practicality of the framework. Firstly, it makes the implicit assumption of an additional person. Secondly, it discourages self-elicitation. Thirdly, every problem may not justify the time and cost associated with the interaction during the elicitation. Spetzler and Stael von Holstein (1975) state

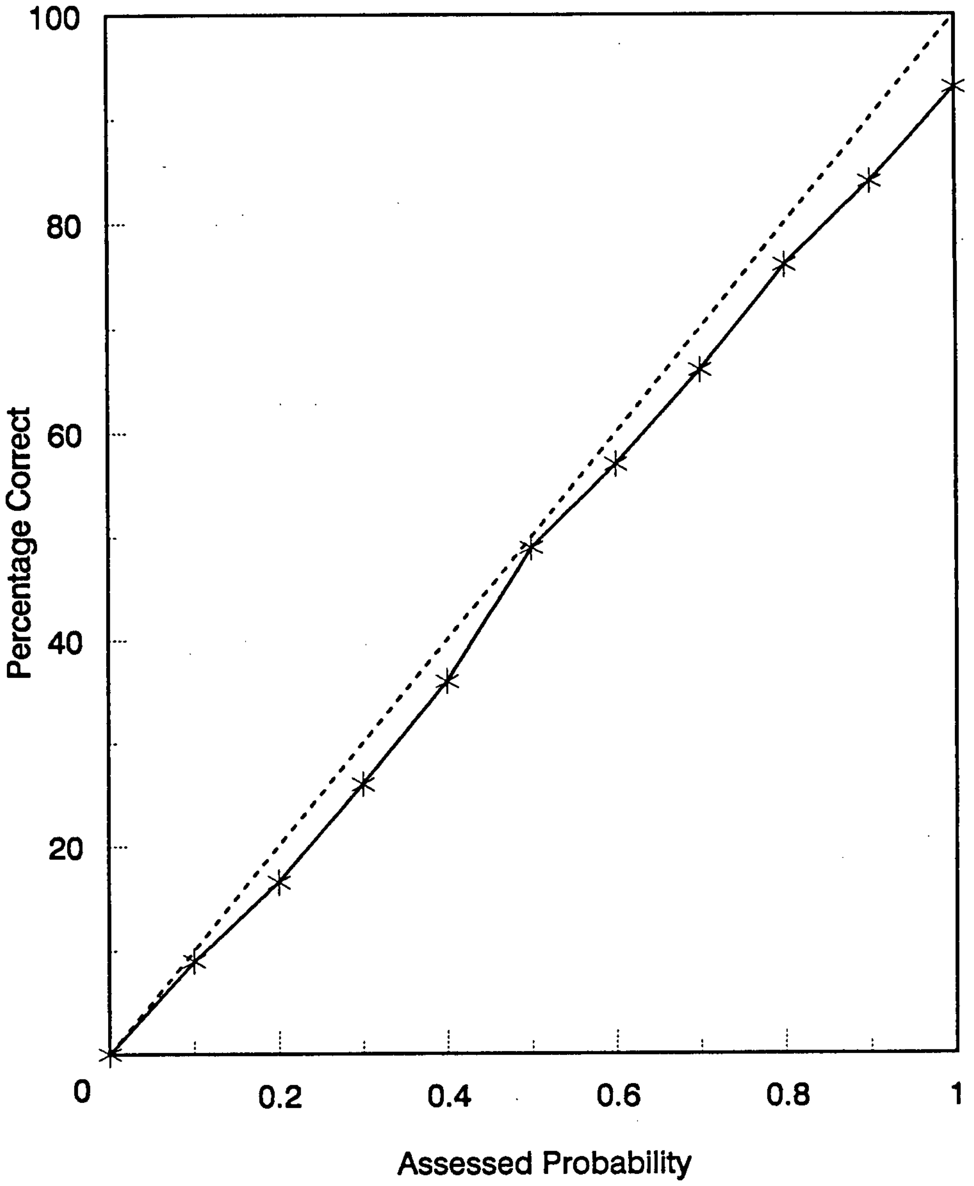

Figure 3.1: Calibration Curve

that in such situations or when the firm uses probabilities regularly to communicate about uncertainty, interactive computer interviews might be valuable. Some real applications, such as probabilistic weather forecasting, rarely used interaction for the elicitation of subjective probabilities (Murphy and Winkler, 1975).

Due to the need to obtain assessments for a large number of variables required for engineering risk analysis, it is necessary to standardize and computerize the elicitation approach, to the extent possible. While Spetzler and Stael von Holstein (1975) and Chapman and Cooper (1987) assert that the role of the computer should be minimized in the elicitation process, Wallsten and Budescu (1983) recommend the study of unaided judgements, because of their applied interest, and because only through studying unaided judgements can the benefits of interaction be determined. Those stages of the approach that can be standardized and computerized to increase the efficiency of the process are explicitly identified in this chapter.

Assumption 3.3 : Questions based on those from previous non-construction related applications and studies will elicit accurate subjective percentiles for the construction context.

The developments in this research are restricted to the measurement of an expert's belief as percentiles of subjective prior probability distributions. In developing the elicitation approach, many proven techniques are used to compensate for the lack of experience in formal elicitation of subjective probabilities in engineering construction. The elicitation of accurate and calibrated subjective probabilities involves three phases - pre-elicitation, elicitation and feedback.

3.4 Pre-Elicitation Stage

The objective of pre-elicitation is for the analyst to train the expert in the task of quantifying his belief as subjective probabilities. It is done in the three phase approach of motivating, structuring and conditioning developed by Spetzler and Stael von Holstein (1975).

3.4.1 Motivating Phase

The motivating phase has two purposes. The first is to build a rapport with the expert by introducing him to the elicitation task. The second is to explore whether any motivational biases might operate (Spetzler and Stael von Holstein, 1975).

Introduce the expert to the elicitation task

The analyst attempts to build a rapport with the expert by giving an explanation on the importance and purpose of probability encoding. This is useful in motivating the expert to become fully involved in the elicitation task (Spetzler and Stael von Holstein, 1975). The need for subjective probabilities in engineering construction, because the variables represent predictions of future events, is emphasized. As most engineers prefer to make deterministic predictions (and then add safety factors for the uncertainty), the difference between deterministic and probabilistic prediction is explained. This discussion is helpful when the expert is asked to respond to probabilistic questions during the elicitation stage.

Explore whether any motivational biases are operating

Motivational biases are defined as either conscious or subconscious adjustments in expert's responses motivated by his perceived system of personal rewards for various responses (Spetzler and Stael von Holstein, 1975). The analyst points out to the expert that there is no commitment (firm projection) inherent in a probability assessment and that the only aim is to elicit a probability distribution that represents belief of the expert about the uncertain variable.

3.4.2 Structuring Phase

The structuring phase concerns the uncertain variable. It also has two purposes.

Define the uncertain variable

The uncertain variable is clearly defined in terms of the structure of the problem. The definition includes relevant units for the variable. The importance of the variable to the decision problem is explained to demonstrate the relevance of the elicitation process to gain the expert's full cooperation. Such cooperation is essential for a successful elicitation (Cooper and Chapman, 1987; Huber, 1974; Hull, 1980; Spetzler and Stael von Holstein, 1975).

Expert is asked to think the variable through

Having defined the uncertain variable the expert is then asked to think the variable through carefully. This enables the analyst to find out what background information is relevant to the elicitation process. If relevant historical data are available it is used in the discussion. The meanings of any descriptive terms (such as highest and lowest or shortest and longest) used in the questionnaire are explained. Winkler (1967a) observed that when subjects are asked for a shape of their subjective distribution many try to associate it to a normal distribution. The author's experience is similar, some experts believing that percentiles should be symmetric to be consistent. The expert is made aware of this common mistake, so that features like skewness will be considered during the elicitation.

3.4.3 Conditioning Phase

The aim of this phase is to condition the expert to think fundamentally about his judgements and to avoid cognitive biases. Cognitive biases are defined as either conscious or subconscious adjustments in the expert's responses that are systematically introduced by the way the expert intellectually processes his perceptions (Spetzler and Stael von Holstein, 1975). For example, a response may be biased towards the most recent piece of information simply because the information is the easiest to recall. Spetzler and Stael von Holstein (1975) state that cognitive biases depend on the expert's "modes of judgement".

Find out how the expert makes probability assignments

The analyst tries to discover what "mode of judgement" the expert might be using to make probability assessments and then adapts the interview to minimize possible biases. Spetzler and Stael von Holstein (1975) define five "modes of judgement" and how each might operate in producing bias, based on the work by Tversky and Kahneman (1984).

1. Availability: Probability is based on the ease with which relevant information is recalled or visualized. This occurs when recent information or information that made a strong impression at the time it was first presented is given more weight than old information. While availability as a mode of judgement can produce biases due to retrievability of instances or imaginability, it can also be introduced deliberately by the analyst to help compensate for an expert's bias. If the analyst believes that the expert has a central bias, he asks the expert to make up scenarios for extreme outcomes, which become more available and help counteract the central bias.

2. Adjustment and Anchoring : The initial response in an interview often serves as a basis for later responses, especially when the first question concerns a likely value for the uncertain variable. Most often experts' adjustment from such a basis is insufficient. Thus, anchoring occurs from a failure to process information about other points on the distribution independently from the point under consideration.

3. Representativeness : The probability of an event is evaluated according to the degree to which it is considered representative of, or similar to, some specific major characteristics of the process from which it originated (*i.e* probability judgements are reduced to judgements of similarity). When this mode is operating there is a strong tendency to place more confidence in a single piece of information that is considered representative than in a larger body of more generalized information.

4. Unstated Assumptions : Expert's responses are conditional on various unstated assumptions. Since the expert cannot be held responsible for taking into account all possible eventualities that may effect the variable, the analyst states the assumptions he is making about the uncertain variable. Once identified the experts can assign their probabilities.

5. Coherence : People tend to assign probabilities based on the ease with which they can fabricate a plausible scenario that would lead to an outcome. Therefore any discussion of scenarios leading to possible outcomes for an uncertain variable should be well balanced, since the relative coherence of various arguments can have an effect on the probability assignment.

Be alert for biases symptomatic of modes of judgements

Asking the expert to specify the most important bases for his judgement, and what information he is taking into account in making his estimates, will indicate possible biases symptomatic of the modes. The first often acts as anchor and possibly leads to central bias while the second will indicate what information is easily available. These observations are also used as checks when obtaining responses for subjective estimates.

3.5 Elicitation Stage

With the completion of the pre-elicitation stage the expert is ready to quantify subjectively his belief about the uncertain variable. The elicitation session is based on a questionnaire that would elicit the desired percentiles of the subjective prior probability distribution for each uncertain variable. This section develops the questionnaire and describes how the elicitation session is conducted.

In developing the questionnaire, central bias (Bunn, 1975, 1979; Chesley, 1975; Hampton et al., 1973; Huber, 1974; Hull, 1978, 1980; Seaver et al., 1978; Spetzler and Stael von Holstein, 1975; Tversky and Kahneman, 1984; Wallsten and Budescu, 1983; Winkler, 1967a) and its effect on the elicitation of tail probabilities (5^{th} and 95^{th} percentiles) must be treated. Therefore, the questionnaire begins by establishing the extremes of the distribution (Budescu and Wallsten, 1987; Cooper and Chapman, 1987; Hull, 1980; Selvidge, 1975; Spetzler and Stael von Holstein, 1975). This prepares the expert to respond to questions on tail probabilities. The deliberate use of scenarios for extreme outcomes counteracts the effect of central bias that is otherwise likely to occur. Also, this has the overall effect of increasing the range of the assigned distribution for the uncertain variable (Hull, 1980). Estimation of the time required to construct a floor slab is used as the example for an uncertain variable to demonstrate a sample questionnaire. Each question is followed with an explanatory comment. Duration assignments for different percentiles are depicted in figure (3.2).

Question 1 : What in your opinion is the shortest possible duration to construct the floor slab for which the probability is so small as to equal zero for practical purposes ? (say the value is A)

Comment : The pre-elicitation stage would have clarified the terms used in the question and explained the range of scenarios the experts should consider in their quantification of judgements.

Question 2 : So, A is in your opinion the shortest possible duration, is that correct ?
Comment : A check to clarify the expert's thinking about the lower tail value of the distribution.

Question 3 : If A in your opinion has a zero probability of not exceeding the actual duration, what is the duration which would not exceed a probability of 0.05 ? (Say the value is C)
Comment : Having established the point for zero probability the expert should be able to give a value for the 5^{th} percentile. This value would be anchored to that of zero probability. However, the anchoring is the result of forcing the expert to think of extreme outcomes to counteract central bias.

Question 4 : So, you associate a 1 in 20 chance that the actual duration will be less than C. Is that correct ?
Comment : Here, odds are used to check the consistency of the elicited 5^{th} percentile. This is helpful to verify the expert's thinking. If the expert confirms his estimate, go to Question 6, if not, ask Question 5.

Question 5 : If not, what is the value for the actual duration that you consider to have a 1 in 20 chance of not being exceeded ?
Comment : A follow up question to the consistency check attempted in Question 4.

Question 6 : What in your opinion is the longest possible duration to construct the floor slab for which the probability is so large as to be equal to one for practical purposes ? (Say the value is Z)

Comment : Going from one extreme to the other increases the range and would reduce even more the possible effects of the central bias that may occur when the 25^{th} and 75^{th} percentiles are elicited after the median value.

Question 7 : So, Z is in your opinion the longest possible duration, is that correct ?
Comment : A check to clarify the expert's thinking about the upper value of the distribution.

Question 8 : If Z in your opinion has a unit probability of not exceeding the actual duration, what is the duration which would not exceed a probability of 0.95 ? (Say the value is X)
Comment : Same as for Question 3

Question 9 : So, you associate a 1 in 20 chance that the actual duration will be above X. Is that correct ?
Comment : Again, odds are used to check the consistency of the elicited 95^{th} percentile. If the expert confirms his estimate, go to Question 11, if not, ask Question 10.

Question 10 : If not, what is the value for the actual duration that you consider to have a 1 in 20 chance of being exceeded ?
Comment : A follow up question to 9.

Question 11 : What in your opinion is the value for actual duration such that it is equally likely to be above as it is to be below ? (Say the value is M)
Comment : This question would elicit the median value of the expert's subjective probability distribution for duration to construct a floor slab.

Question 12 : So, you are willing to bet equal odds that the actual duration is either above or below M , is that correct ?

Comment : A check to clarify the expert's response to the median.

Question 13 : What is the value for duration that you feel will divide the region below M equally, thus it is just as likely that duration will fall below this value as it will be between this value and M ? (Say the value is L)

Comment : The expert is asked to bisect the area below the median to give an estimate for his 25^{th} percentile value.

Question 14 : So, you associate a 1 in 4 chance that the actual duration will be below L, is that correct ?

Comment : A consistency check to clarify that the expert is thinking about the 25^{th} percentile with the bisected value. If the expert confirms his estimate, go to Question 16, if not ask Question 15.

Question 15 : If not, what is the value for the actual duration that you consider to have a 1 in 4 chance of not being exceeded ?

Comment : A follow up question to 14.

Question 16 : Now, concentrate on the case where the duration could be above M, which you felt would be 50% of the time. What is the value that you feel will divide the region above M equally, thus it is just as likely that duration will be above this value as it will be between this value and M ? (Say the value is N)

Comment : The expert is asked to bisect the area above the median to give an estimate for his 75^{th} percentile value. In addition the expert is reminded of his estimate for the median. This gives him a further opportunity to change or confirm his estimate for the median, now that he has given an estimate for the 25^{th} percentile.

Question 17 : So, you associate a 1 in 4 chance that the actual duration will be above N, is that correct ?

Comment : A check to clarify that the expert is thinking about the 75^{th} percentile with the bisected value. If the expert confirms his estimate, stop the interview, if not ask Question 18.

Question 18 : If not, what is the value for the actual duration that you consider to have a 1 in 4 chance of being exceeded ?

Comment : A follow up question to 17.

The questionnaire combines direct probability responses and chance responses to provide cross checking for consistency. The basis for direct probability responses is the variable interval method (Huber, 1974), because it elicits the percentiles required by the framework. Hull (1978) and Seaver et al. (1978) have reported that the fixed interval method performed better than the variable interval method because the variable interval method gave distributions that were "too-tight". It must be noted that both studies assessed the median first, giving rise to possible central bias. Murphy and Winkler (1975) studying experienced weather forecasters conclude that the variable interval method performed better than the fixed interval method in probabilistic weather forecasting. Since the questionnaire starts with the tails of the distribution, the elicited percentiles should overcome the effects of central bias and display sufficient spread.

The elicitation session is based on the questionnaire and the analyst is expected to follow the general format of the questionnaire when conducting the session. However, at his discretion the analyst can adapt the interview to suit different situations. Since the questionnaire is based primarily on the variable interval technique, it is easily standardized and automated to use for those variables selected for interactive computer interviews.

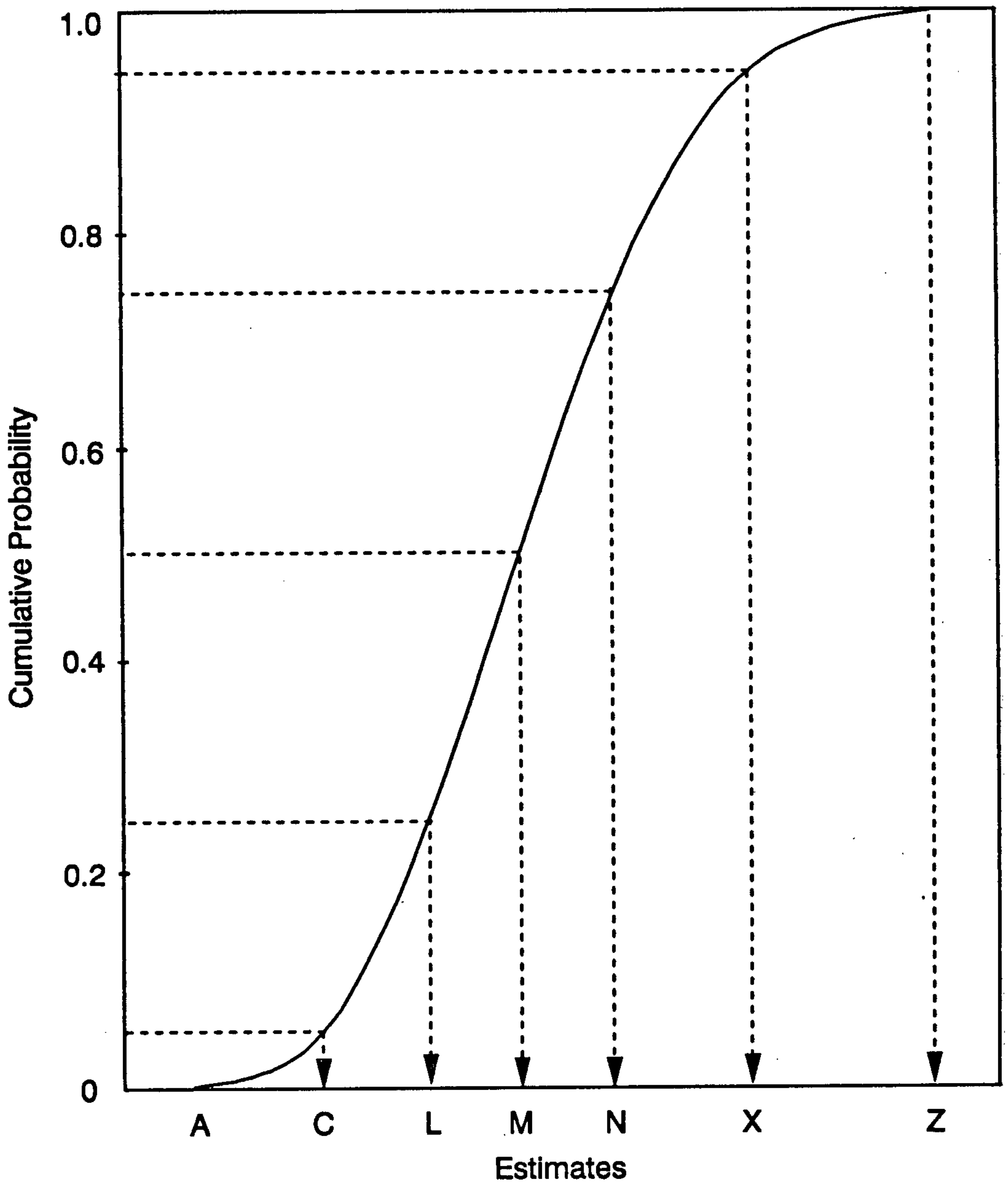

Figure 3.2: Subjective Percentile Estimates

3.6 Feedback and Consensus Estimates

Whenever possible a group of experts are used for the elicitation because it has been found that consensus judgements from a group to be better than the individual judgements (Ashton and Ashton, 1985; Ashton, 1986; Bacharach, 1975; Beach, 1975; Bordley, 1982; Bordley and Wolff, 1981; Cooper and Chapman, 1987; French, 1985; Hampton et al., 1973; Huber, 1974; Stael von Holstein, 1971; Winkler, 1968, 1971). Huber (1974) states that the aggregation of responses from several experts improve subjective judgements because aggregation from a statistical viewpoint tends to reduce random error as well as reduce the impact of biases. This view is confirmed by Beach (1975) who in addition states that combining the opinions of several experts would aid in eliminating conservatism and/or extremism and promote more nearly optimal decisions.

Once the initial subjective estimates are made by the expert, he is provided feedback on his assessments in the form of a discussion (graphically if necessary) between the analyst and expert, and expert and expert. The expert can revise his prior judgements after the discussion. This process is based on the nominal group technique (Delbecq et al., 1975). While there is no consensus in the literature as to which is the best method to provide feedback, there is agreement that feedback improves the original estimates (Beach, 1975; Chesley, 1975; Gustafson et al., 1973; Lock, 1987; Stael von Holstein, 1971; Winkler, 1971). However, Gustafson et al. (1973) have shown that assessments from the nominal group technique (estimate-talk-estimate) to be more accurate than those from Delphi technique (estimate-feedback-estimate), conventional group technique or individual estimates. Lock (1987) in proposing a general approach to group judgmental forecasting concludes that there are benefits to communication and discussion between group members, so long as these are structured as in nominal group approaches.

Winkler (1968) used several mathematical and behavioral approaches for arriving at consensus subjective probability distributions. The mathematical approaches were either using a weighted average or Bayes' theorem. The behavioral approaches, Delphi and nominal group led the group to arrive at the final probability distribution. He could not determine which method was most accurate, because there was no "correct" opinion, but he did find that different methods produced different results. Makridakis and Winkler (1983) used ten different forecasting methods to combine forecasts. Performance was compared in terms of the mean average percentage error (MAPE). They found that the accuracy increased when additional methods were added to the forecast. However, the gains tailed off after about four or five were combined.

Ashton and Ashton (1985) compared equal weighting with four differential weighting methods to examine the impact of aggregation in forecasting of annual advertising sales at TIME magazine. They found: aggregates of subjective forecasts to be more accurate than the individual forecasts that comprised the aggregates; incremental accuracy of differential weighting methods over equal weighting was small; regardless of the weighting method, accuracy attributable to aggregation was achieved by combining a small number of individual forecasts. They concluded that equal weighting appears to be the solution to the problem of choosing a weighting method for subjective forecasting. Lock (1987) states that for most purposes linear models are adequate for aggregation and differential weighting do not offer any real advantages in practical terms.

Since the experts are provided feedback, their subjective estimates would be close to consensus. Also, because of the difficulty in measuring the variability in expert accuracy, consensus subjective estimates are obtained by assigning equal weights to all the experts. The routine aspects of feedback such as exchanging individual estimates, obtaining revised estimates and the task of combining subjectives estimates can be readily automated.

3.7 Analysis Stage

The requirement for eliciting coherent subjective probabilities has been discussed previously. An automated approach to ensure the coherence of subjective probabilities has been developed as a part of this research effort. It is documented in the form of an interactive computer program called "ELICIT" (see Appendix D). This program, based on the method to convert subjective estimates to moments (section 2.3.2), enables the analyst to approximate the subjective estimates for a variable to a Pearson type distribution. The high flexibility of the Pearson family (Amos and Daniel, 1971) approximates most of the subjective estimates to Pearson type distributions.

However, in some instances subjective estimates may not approximate to a Pearson distribution for the specified maximum cumulative error. In these situations the expert is made aware of the necessity for subjective probabilities to be coherent and is asked to modify the 25^{th} and 75^{th} percentile estimates. These two estimates are elicited only to approximate a Pearson type distribution. The expected value and standard deviation for the uncertain variable are derived from the 5^{th}, 50^{th} and 95^{th} percentile values and are initially independent of the approximated distribution. In addition, the conversion of subjective estimates to moments ensures that the measured belief is useful in the quantification of uncertainty of the derived variable.

For example, assume that the five percentile values elicited for a variable are 1.0, 2.5, 5.0, 7.5 and 9.0. The expected value and standard deviation from step 2 of section (2.3.2) is 5.0 and 2.432. However, the five subjective estimates do not approximate to a Pearson type distribution. It is obvious from the 5^{th} and 95^{th} estimates that the expert is thinking of a symmetric distribution. If a value for the 25^{th} percentile between 2.9 and 3.5 with a symmetrical value for the 75^{th} percentile between 6.5 and 7.1 is acceptable to the expert, a Pearson distribution with $E[X] = 5.0$, $\sigma = 2.432$, $\sqrt{\beta_1} = 0$ and β_2 between 2.0 and 4.8 can be approximated. In the author's limited experience in eliciting subjective estimates, the consensus among experts is that if necessary they would be willing to change within reason the 25^{th} and 75^{th} percentile estimates because those two are the least important to them.

Ideally, the interactive program should give guidance for the change as in the case of correlation coefficients for a positive definite correlation matrix. However, at present it is the responsibility of the analyst to guide the expert. The analyst can recognize the shape of the distribution (symmetric, positively or negatively skewed) by observing the 5^{th}, 50^{th} and 95^{th} percentile estimates and guide the expert to acceptable estimates for the 25^{th} and 75^{th} percentiles. It is planned that this facility be added to "ELICIT" as a future improvement.

3.8 Verification

As the final stage of the elicitation, the subjective prior probability distribution is verified to see if the expert is in total agreement with it (*i.e* it reflects his belief). Cooper and Chapman (1987) state that verification can be conducted by: using cross checking for consistency between values; using different elicitation methods especially when indirect methods have been used; and having the expert examine and confirm the final result.

Since the questionnaire has performed cross checking for consistency of all the percentile values, as verification, the computer program "ELICIT" informs the user of the expected value, standard deviation and shape characteristics of the approximated

Pearson type distribution. While the skewness and the kurtosis give an indication of the shape of the distribution, a better verification method is to provide a graphical display of the approximated density function. The next step in the development of "ELICIT" is to display the probability density function of the approximated Pearson type to be viewed by the expert. Incorporating such a verification process to the elicitation technique would provide the analyst and expert with greater confidence that the approximated distribution represents the expert's belief. In addition, it would eliminate approximating variables with bell shaped probability distributions to Pearson distributions that are U or J shaped.

3.9 Summary

An approach to elicit an expert's belief of uncertainty as subjective probabilities has been described in this chapter. The approach combines the theoretical requirements of subjective probabilities with a practical process. The process is developed by transforming proven techniques from other fields of study to the requirements of risk quantification in engineering construction. The role of the computer and the use of a standard approach to expedite the process is identified at every stage.

The pre-elicitation stage based on the developments by Spetzler and Stael von Holstein (1975) trains and prepares the expert to quantify his belief as subjective probabilities. This stage requires a high level of person to person interaction. Hence, there is little use of the computer during pre-elicitation. The elicitation stage elicits the percentile values of an expert's subjective prior probability distributions for uncertain variables using the developed questionnaire. These subjective probabilities are accurate and calibrated. Since the questionnaire is based primarily on the variable interval technique, the elicitation stage can be standardized and automated for interactive computer interviews. If more than one expert participates in the elicitation, consensus subjective estimates are obtained by assigning equal weights to all the experts. The routine aspects of feedback and obtaining consensus subjective estimates can be done by the computer. The coherence of subjective probabilities is ensured by the interactive computer program "ELICIT" (see Appendix D) , using the Pearson family of distributions. The moments for the uncertain primary variable are used to verify whether the shape of the approximated distribution is similar to that which the expert has in mind.

Distinct roles for computerization and standardization exist to expedite the process of elicitation and verification of an expert's belief about uncertainty. At present, the experience in using these approaches in field applications is limited. The next stage of this research will concentrate on building up experience from field applications and on refining and validating the elicitation approach. This is essential for the proposed method to become a practical tool in risk quantification for large engineering projects.

Chapter 4

Correlations Between Variables

4.1 General

The risk measurement framework developed in chapter two was based on the assumption that the correlations between variables were linear. From that assumption a variable transformation method was developed to treat linear correlations among the primary variables when evaluating the moments of the derived variable. This transformation was based on the correlation matrix for the primary variables.

A correlation matrix is defined by Graybill (1983) as follows. Let $\mathbf{X}$ be an nx1 random vector with positive definite covariance matrix denoted by $\mathbf{C} = [v_{ij}]$. The correlation matrix of $\mathbf{X}$ is $\mathbf{R} = [\rho_{ij}]$ where ρ_{ij} is defined by,

$$\rho_{ij} = \frac{v_{ij}}{\sqrt{v_{ii}\ v_{jj}}} \tag{4.1}$$

for all i and j.

This chapter addresses some of the issues that arise in treating analytically the linear correlations between variables (these are equally relevant when treating correlations for Monte Carlo simulation). In the next section the correlations between primary variables are discussed. The discussion highlights an often ignored theoretical requirement of the correlation matrix and thereby develops a method to elicit a positive definite correlation matrix for primary variables.

Developed in the third section is a method to obtain a positive definite correlation matrix for derived variables. The method is developed by extending the approximation for the covariance between two functions suggested by Kendall and Stuart (1969) to the multivariate case. The fourth section address the issue of multicollinearity in the correlation matrix and suggests a mathematical manipulation to overcome the effect of multicollinearity for practical applications.

A numerical study is presented in the fifth section. The first part of the study compares the variable transformation method to the standard approach used in moment analysis to treat correlation among primary variables (Ang and Tang, 1975; Benjamin and Cornell, 1970) under general conditions. The second part explores the behavior of the two methods in the presence of multicollinearity. This study demonstrates that while the variable transformation method is stable in the presence of multicollinearity, the standard approach could fail. The intention of the third part is to study the susceptibility of the transformation to the effect of multicollinearity.

4.2 Correlation between Primary Variables

The correlation information between primary variables, required by the framework, will have to be obtained subjectively from experts because of data limitations. A number of authors in both simulation and approximate applications have recognized this necessity (Eilon and Fowkes, 1973; Inyang, 1983; Howard, 1971; Hull, 1977, 1980; Kadane et al., 1980; Keefer and Bodily, 1983; Kryzanowski et al., 1972; Wagle 1967). Other than for Kadane et al., (1980), who develop an approach to elicit a positive definite correlation matrix, all of the others obtain only the correlation coefficients between variables.

4.2.1 Positive Definite Correlation Matrix

A positive definite correlation matrix ensures theoretical consistency of a system. A correlation matrix is positive definite if there are no linear dependencies among the primary variables. If an elicited correlation matrix is not positive definite, then it has to be positive semi-definite because the variance of a vector of random variables is always greater than or equal to zero.

Proof that a Correlation Matrix is Positive Definite

Let $\mathbf{X}$ be the vector of n random variables with covariance matrix $\mathbf{C_x}$ and correlation matrix $\mathbf{R}$. Let $\mathbf{a}$ be a vector of n scalars. From definition,

$$\mathbf{Var}\left[\mathbf{a}^{\mathrm{T}}\ \mathbf{X}\right] \geq 0 \tag{4.2}$$

$$\mathbf{a}^{\mathrm{T}}\ \mathbf{C_x}\ \mathbf{a} \geq 0 \tag{4.3}$$

Therefore, covariance matrix $\mathbf{C_x}$ is always positive definite ($i.e > 0$) or positive semi definite ($i.e = 0$). Rewriting equation (4.3) as,

$$\mathbf{a}^{\mathrm{T}}\ \mathbf{C_x}\ \mathbf{a} \quad = \quad \mathbf{a}^{\mathrm{T}}\ (\mathbf{X} - \bar{\mathbf{X}})\ (\mathbf{X} - \bar{\mathbf{X}})^{\mathrm{T}}\ \mathbf{a} \tag{4.4}$$

Let $b \ = \ (\mathbf{X} - \bar{\mathbf{X}})^{\mathrm{T}}\ \mathbf{a}$, where b is a number and $b \ = \ b^T$. When $b = 0$ (positive semi definite condition),

$$(\mathbf{X} - \bar{\mathbf{X}})^{\mathrm{T}}\ \mathbf{a} \ = \ 0 \tag{4.5}$$

$$(X_1 - \bar{X}_1)\, a_1 \ + \ (X_2 - \bar{X}_2)\, a_2 \ + \ \ldots\ldots. \ + \ (X_n - \bar{X}_n)\, a_n \ = \ 0 \tag{4.6}$$

variable X_n is a linear combination of the others.

If there are no linear dependencies (combinations) then the covariance matrix is always positive definite.

For the correlation matrix, starting from the relationship $\mathbf{R} \ = \ \mathbf{D}^{-1}\ \mathbf{C_x}\ \mathbf{D}^{-1}$, where $\mathbf{D}^{-1}$ is the inverse of the diagonal matrix of standard deviations of the $\mathbf{X}$

vector, it follows that,

$$\mathbf{a}^{\mathbf{T}}\ \mathbf{R}\ \mathbf{a}\ =\ \mathbf{a}^{\mathbf{T}}\ \mathbf{D}^{-1}\ \mathbf{C}_{\mathbf{x}}\ \mathbf{D}^{-1}\ \mathbf{a} \tag{4.7}$$

Since $\mathbf{D}^{-1} \equiv \left[\mathbf{D}^{-1}\right]^{\mathbf{T}}$ because $\mathbf{D}^{-1}$ is symmetric,

$$\mathbf{a}^{\mathbf{T}}\ \left[\mathbf{D}^{-1}\right]^{\mathbf{T}}\ \mathbf{C}_{\mathbf{x}}\ \mathbf{D}^{-1}\ \mathbf{a}\ =\ \mathbf{b}^{\mathbf{T}}\ \mathbf{C}_{\mathbf{x}}\ \mathbf{b} \tag{4.8}$$

where $\mathbf{b}\ =\ \mathbf{D}^{-1}\ \mathbf{a}$.

Since $\mathbf{D}^{-1}$ is non-singular and symmetric, when $\mathbf{C}_{\mathbf{x}}$ is positive definite,

$$\mathbf{b}^{\mathbf{T}}\ \mathbf{C}_{\mathbf{x}}\ \mathbf{b}\ >\ 0 \tag{4.9}$$

If the covariance matrix is positive definite then the correlation matrix is always positive definite.

A correlation matrix could be positive semi-definite even when all the variables are not perfectly correlated. For example, consider the correlation matrix for a three variable system given by $\mathbf{R_0}$,

$$\mathbf{R_0} = \begin{bmatrix} 1.0 & 0.5 & 0.5 \\ 0.5 & 1.0 & -0.5 \\ 0.5 & -0.5 & 1.0 \end{bmatrix} \tag{4.10}$$

At first glance, the correlation coefficients between the variables seem reasonable. However, the determinant of matrix $\mathbf{R_0}$ is equal to zero (*i.e* positive semi-definite). A further investigation shows that a linear combination of variables 2 and 3 is perfectly correlated with variable 1 (see Appendix B).

4.2.2 Elicitation of a Correlation Matrix

The proposed method of elicitation is a combination of a two stage process. The first is the elicitation of the linear correlation coefficients between the primary variables, while the second is ensuring the positive definiteness of the correlation matrix.

Linear Correlation Coefficients

The linear correlation coefficient between two primary variables X_i and X_j can be approximated from the conditional expected value of $X_j|X_i = Q$. The conditional expected value of $X_j|X_i = Q$ from Bury (1975) is,

$$E[X_j|X_i = Q] \;=\; E[X_j] \;+\; \rho_{ij}\,\frac{\sigma_j}{\sigma_i}\,(Q - E[X_i]) \qquad (4.11)$$

Hence,

$$\rho_{ij} \;=\; \frac{(E[X_j|X_i = Q] \;-\; E[X_j])\,\sigma_i}{(Q \;-\; E[X_i])\,\sigma_j} \qquad (4.12)$$

where $E[X_i]$ and $E[X_j]$ are the expected values and σ_i and σ_j are the standard deviations for X_i and X_j; Q is the conditional value for X_i; and ρ_{ij} is the linear correlation coefficient between primary variables X_i and X_j.

Then, similarly to the method suggested by Hull (1977) for risk simulation, the correlation coefficient between X_i and X_j is approximated by averaging three or four values for ρ_{ij}. The values for ρ_{ij} are evaluated from equation (4.12). The conditional expected value of X_j, $E[X_j|X_i = Q]$, is elicited by asking the question "**What is the expected value for $\mathbf{X_j}$, when $\mathbf{X_i = Q}$?**" from the experts. Different percentile values of X_i can be used for the conditional value Q.

The Elicitation Procedure

Let $\mathbf{R_n}$ be a subjectively elicited nxn correlation matrix partitioned as,

$$\mathbf{R_n} = \begin{bmatrix} \mathbf{R_{n-1}} & \mathbf{b} \\ \mathbf{b^T} & 1 \end{bmatrix} \qquad (4.13)$$

where $\mathbf{R_{n-1}}$ is a $(n-1)$x$(n-1)$ correlation matrix for $n = 2, 3, ...$ and $\mathbf{b^T} = [\rho_{1n}\ \rho_{2n}\ ....\ \rho_{n-1n}]$.

Then $\mathbf{R_n}$ is positive definite if $\mathbf{R_{n-1}}$ is positive definite and,

$$\mathbf{b}^{\mathbf{T}} \ \mathbf{R}_{\mathbf{n-1}}^{-1} \ \mathbf{b} < 1 \tag{4.14}$$

(Kadane et al., 1980; for proof see Appendix C)

First, the primary variables in the function for the derived variable are ordered according to the expert's confidence in them and their relationship with the other variables. The variable that is selected as the "best" is numbered one and the "worst" numbered n (when there are n variables in the functional form). Then, ρ_{12} is elicited as suggested in the previous section. This value is assumed to be consistent with the expert's belief because the 2x2 matrix is always positive definite.

Thereafter, ρ_{13} and ρ_{23} are elicited. If the condition given by equation (4.14) is satisfied, the correlation values are accepted because the 3x3 matrix is positive definite. If the condition is violated, the expert is made aware of the inconsistency and given the option to change one of the correlation coefficient values in the $\mathbf{b}$ vector. When a value is selected (say ρ_{23}) the expert is informed of the real bounds for ρ_{23} in which the 3x3 matrix will be positive definite. The bounds, Γ_1 and Γ_2, (if they exist - see figure 4.1) are the real roots of the quadratic equation (see Appendix C for the derivation),

$$S_{jj}\, \Gamma^2 + [C_{1j} + C_{2j} + \sum_{i=1}^{j-1} S_{ji}\, B_{1i} + \sum_{i=j+1}^{n-1} S_{ji}\, B_{2i}]\, \Gamma \tag{4.15}$$

$$+ \sum_{i=1}^{j-1} (C_{1i} + C_{2i})\, B_{1i} + \sum_{i=j+1}^{n-1} (C_{1i} + C_{2i})\, B_{2i} - 1 < 0$$

$$\text{where } \mathbf{R}_{\mathbf{n}-\mathbf{1}}^{-1} = \begin{bmatrix} \mathbf{S_1} \\ \cdots \\ \mathbf{S_j} \\ \cdots \\ \mathbf{S_2} \end{bmatrix} ; \qquad \mathbf{b}^{\mathbf{T}} = \begin{bmatrix} \mathbf{B_1^T} & \vdots & \Gamma & \vdots & \mathbf{B_2^T} \end{bmatrix} ;$$

$\mathbf{C_1} = \mathbf{B_1^T}\ \mathbf{S_1}$; and $\mathbf{C_2} = \mathbf{B_2^T}\ \mathbf{S_2}$.

Γ is the correlation coefficient (ρ_{jn}) for which bounds are required (for ρ_{23}, $j = 2$ and $n = 3$),

$\mathbf{S_1}$ is a $(j-1)$x$(n-1)$ matrix and $\mathbf{S_2}$ is a $(n-1-j)$x$(n-1)$ matrix,

$\mathbf{B_1^T}$ and $\mathbf{B_2^T}$ are 1x$(j-1)$ and 1x$(n-1-j)$ row matrices, and

$\mathbf{S_j}$, $\mathbf{C_1}$ and $\mathbf{C_2}$ are 1x$(n-1)$ row matrices.

This procedure, of introducing the next ordered variable, eliciting correlation coefficients between that and the previous variables and ensuring that the correlation matrix is positive definite is continued until the $\mathbf{R_n}$ is positive definite.

Once accepted, the elicitation procedure does not permit the positive definite $\mathbf{R_{n-1}}$ to be changed. If at any stage the expert refuses to change a value from the $\mathbf{b}$ vector when $\mathbf{R_n}$ is not positive definite, he is implying that the function for the derived variable is not consistent with his belief and it should be changed by removing one or more of the already used ordered variables from the function.

4.3 Correlation between Derived Variables

Assumption 4.1 : Correlation between two derived variables arise only from common (shared) variables in their functional forms.

The common (shared) primary variables are defined as those of the same type having

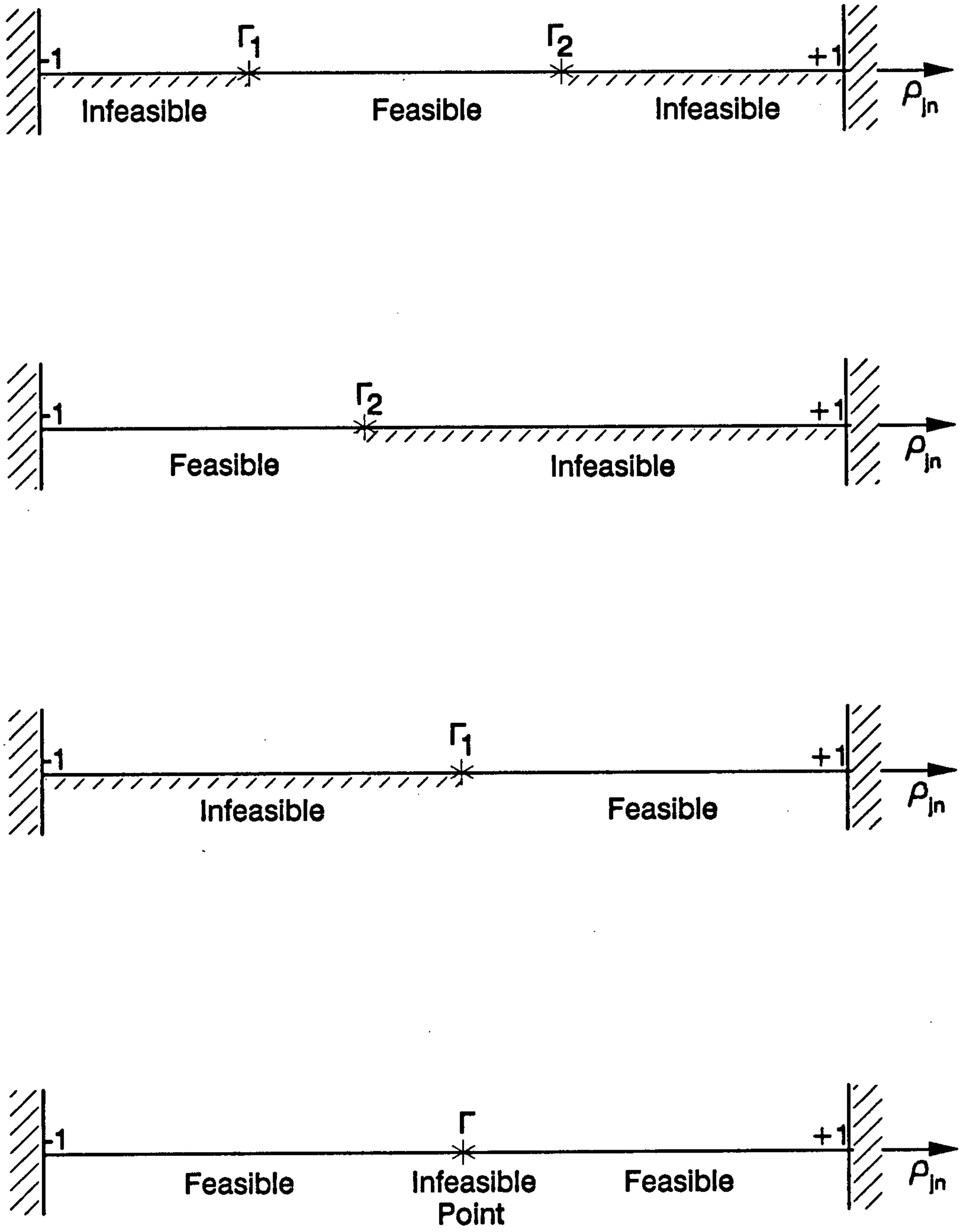

Figure 4.1: Feasible Regions for Γ for $\mathbf{R_n}$ to be Positive Definite.

the same first four moments in the functional forms for two or more derived variables (see figure 4.2). Hence, correlation between two derived variables arise only from those primary variables that are quantified for the functions. Correlation arising due to unquantifiable variables in construction such as management, methods, or weather are ignored by assumption (4.1).

The correlation coefficient between two derived variables Y_a and Y_b can be evaluated from,

$$\rho_{ab} = \frac{cov(Y_a, Y_b)}{\sqrt{\mu_2(Y_a)\ \mu_2(Y_b)}} \tag{4.16}$$

where $cov(Y_a, Y_b)$ is the covariance between Y_a and Y_b; $\mu_2(Y_a)$ and $\mu_2(Y_b)$ are the variance of Y_a and Y_b; and ρ_{ab} is the correlation coefficient between Y_a and Y_b. The covariance between two derived variables can be approximated from the approximation given by Kendall and Stuart (1969), using only the linear correlation information between primary variables. The approximation is,

$$\begin{aligned} cov(Y_a, Y_b) \quad \approx \quad & \sum_{i=1}^{l} \sum_{j=1}^{l} \frac{\partial g_a}{\partial X_i} \frac{\partial g_b}{\partial X_j}\ cov(X_i, X_j) \\ & + \sum_{i=1}^{l} \sum_{j=l+1}^{n} \frac{\partial g_a}{\partial X_i} \frac{\partial g_b}{\partial X_j}\ cov(X_i, X_j) \\ & + \sum_{i=l+1}^{m} \sum_{j=1}^{l} \frac{\partial g_a}{\partial X_i} \frac{\partial g_b}{\partial X_j}\ cov(X_i, X_j) \end{aligned} \tag{4.17}$$

where $Y_a = g_a(\mathbf{X})$ has m random variables; $Y_b = g_b(\mathbf{X})$ has n random variables; l is the number of common (shared) primary variables in the functions $g_a(\mathbf{X})$ and $g_b(\mathbf{X})$, (*i.e* $X_1, X_2, ..., X_l$); and $cov(X_i, X_j)$ is the covariance between two primary variables X_i and X_j. First order Taylor series expansions of the functions for derived variables are used in equations (4.16) and (4.17). For a vector of derived variables $\mathbf{Y}$ in a system, the correlation matrix must also be positive definite.

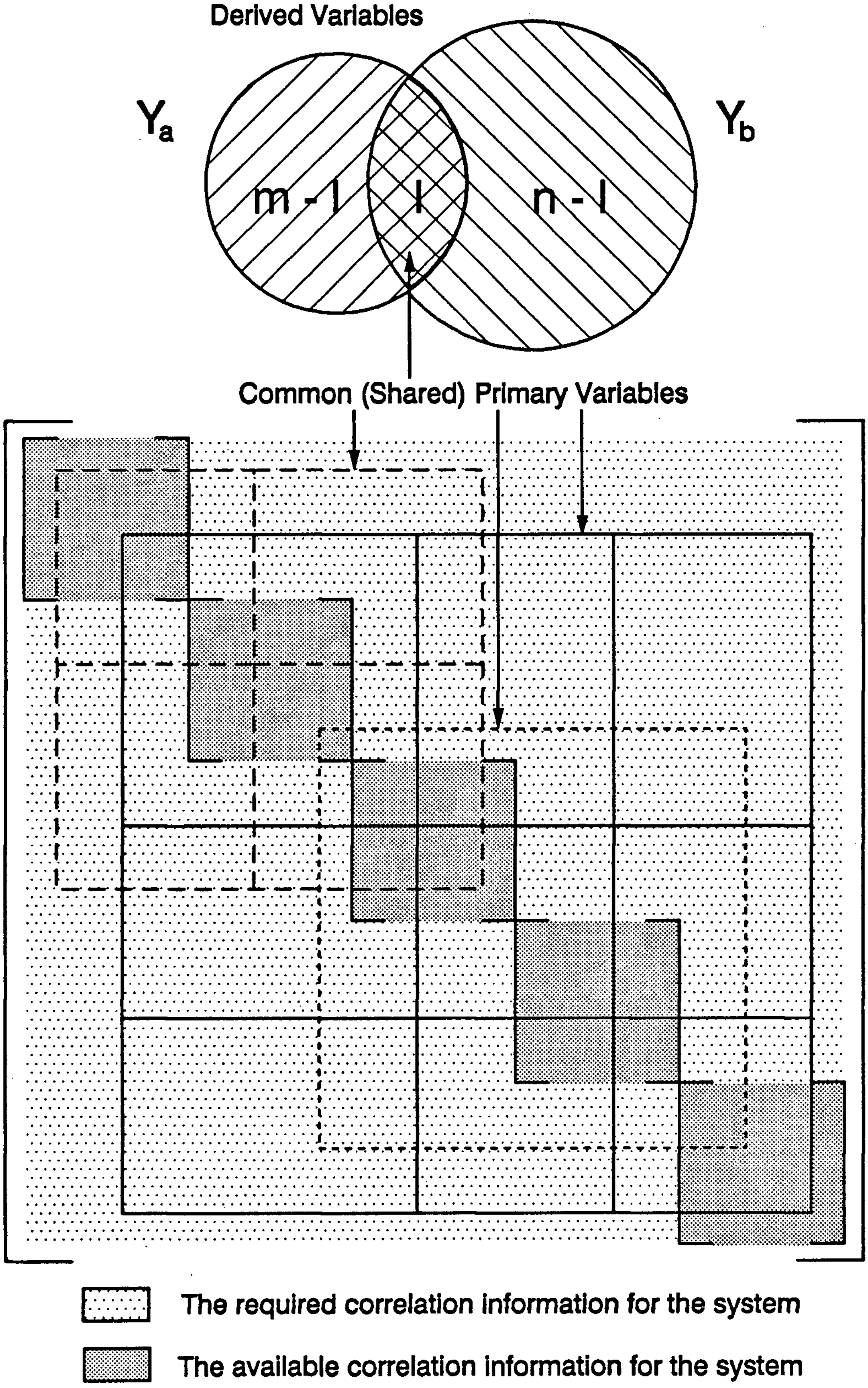

Figure 4.2: Correlation from Common (Shared) Primary Variables

Proof that Correlation Matrix for Y is Positive Definite

Let $\mathbf{Y}$ be a vector of derived variables where, $\mathbf{Y} = [Y_1....Y_a\ Y_b....Y_z]^{\mathbf{T}}$ and $Y_1 = g_1(\mathbf{X})$, ..., $Y_a = g_a(\mathbf{X})$, $Y_b = g_b(\mathbf{X})$, ..., $Y_z = g_z(\mathbf{X})$. Let $\mathbf{C_y}$ be the covariance matrix, $\mathbf{R_y}$ be the correlation matrix and $\mathbf{D_y}$ be the diagonal matrix of standard deviations of the vector $\mathbf{Y}$.

The covariance matrix of vector $\mathbf{Y}$ is positive definite if there are no linear dependencies among the derived variables. The linear dependencies can occur if the functional form for two or more derived variables are identical and all the primary variables are shared. However, since the models are not perfect and unquantifiable variables exist in all systems, the true models are,

$Y_1^* = g_1(\mathbf{X}) + \epsilon_1, ..., Y_a^* = g_a(\mathbf{X}) + \epsilon_a, Y_b^* = g_b(\mathbf{X}) + \epsilon_b, ..., Y_z^* = g_z(\mathbf{X}) + \epsilon_z.$

where ϵ is a vector of independent error variables to represent the unquantifiable variables in the systems. For simplicity assume all error variables have the same variance σ^2. Then,

$$\mathbf{Y}^* = \mathbf{Y} + \epsilon \tag{4.18}$$

Let $\mathbf{a}$ be a vector of scalars. From the definition for variance,

$$\mathbf{Var}\left[\mathbf{a}^{\mathbf{T}}\ \mathbf{Y}^*\right] \geq 0 \tag{4.19}$$

$$\mathbf{a}^{\mathbf{T}}\ \left(\mathbf{Y}^* - \bar{\mathbf{Y}}^*\right)\left(\mathbf{Y}^* - \bar{\mathbf{Y}}^*\right)^{\mathbf{T}}\ \mathbf{a} \geq 0 \tag{4.20}$$

Since $E[\epsilon] = 0$, $\left(\mathbf{Y}^* - \bar{\mathbf{Y}}^*\right) = \left(\mathbf{Y} - \bar{\mathbf{Y}} + \epsilon\right)$. Then,

$$\mathbf{a}^{\mathbf{T}}\ \left(\mathbf{Y} - \bar{\mathbf{Y}} + \epsilon\right)\left(\mathbf{Y} - \bar{\mathbf{Y}} + \epsilon\right)^{\mathbf{T}}\ \mathbf{a} \geq 0 \tag{4.21}$$

Since error variables are statistically independent, from mathematical induction,

$$\left(\mathbf{Y} - \bar{\mathbf{Y}} + \epsilon\right)\left(\mathbf{Y} - \bar{\mathbf{Y}} + \epsilon\right)^{\mathbf{T}} = \left(\mathbf{Y} - \bar{\mathbf{Y}}\right)\left(\mathbf{Y} - \bar{\mathbf{Y}}\right)^{\mathbf{T}} + \sigma^2\ \mathbf{I} \tag{4.22}$$

Substituting (4.22) in (4.21),

$$\mathbf{a}^{\mathbf{T}}\ \left[\left(\mathbf{Y} - \bar{\mathbf{Y}}\right)\left(\mathbf{Y} - \bar{\mathbf{Y}}\right)^{\mathbf{T}} + \sigma^2\ \mathbf{I}\right]\mathbf{a} \geq 0 \tag{4.23}$$

Expanding equation (4.23),

$$\mathbf{a}^{\mathbf{T}} \left(\mathbf{Y} - \bar{\mathbf{Y}}\right)\left(\mathbf{Y} - \bar{\mathbf{Y}}\right)^{\mathbf{T}} \mathbf{a} + \mathbf{a}^{\mathbf{T}} \sigma^2 \mathbf{I}\, \mathbf{a} \geq 0 \tag{4.24}$$

For the covariance matrix to be positive semi-definite both terms in (4.24) have to be equal to zero at the same time. But, $\mathbf{a}^{\mathbf{T}} \sigma^2 \mathbf{I}\, \mathbf{a} > 0$. Therefore,

$$\mathbf{a}^{\mathbf{T}} \left(\mathbf{Y}^{*} - \bar{\mathbf{Y}}^{*}\right)\left(\mathbf{Y}^{*} - \bar{\mathbf{Y}}^{*}\right)^{\mathbf{T}} \mathbf{a} > 0 \tag{4.25}$$

Since no linear dependencies exist in the vector of derived variables,

$$\mathbf{a}^{\mathbf{T}} \left(\mathbf{Y} - \bar{\mathbf{Y}}\right)\left(\mathbf{Y} - \bar{\mathbf{Y}}\right)^{\mathbf{T}} \mathbf{a} > 0 \tag{4.26}$$

Hence,

$$\mathbf{a}^{\mathbf{T}} \mathbf{C}_{\mathbf{y}}\, \mathbf{a} > 0 \tag{4.27}$$

The covariance matrix for derived variables is always positive definite.

In a correlation matrix the following requirements must hold: (Graybill, 1983) (1) $\rho_{ii} = 1$; $i = 1, ..., a, b, ..., z$; (2) $-1 < \rho_{ij} < 1$; for all $i \neq j$

Since only the first order Taylor series expansions of the vector $\mathbf{Y}$ are used to evaluate correlation and covariance, the first requirement is obtained by evaluating the i^{th} diagonal element of $\mathbf{D}_{\mathbf{y}}^{-1}\, \mathbf{C}_{\mathbf{y}}\, \mathbf{D}_{\mathbf{y}}^{-1}$ ($= \mathbf{R}_{\mathbf{y}}$). The second requirement is obtained by setting the i^{th} element of the vector of scalars $\mathbf{b}$ equal to $+1$, the j^{th} element equal to $+1$, and the other elements equal to zero. From $\mathbf{b}^{\mathbf{T}}\, \mathbf{R}_{\mathbf{y}}\, \mathbf{b} > 0$, $\rho_{ii} + \rho_{ij} + \rho_{ji} + \rho_{jj} > 0$ or $\rho_{ij} > -1$. Similarly, by changing the j^{th} element of $\mathbf{b}$ to -1, $\rho_{ii} - \rho_{ij} - \rho_{ji} + \rho_{jj} > 0$ or $\rho_{ij} < 1$.

Since the covariance matrix is positive definite and the requirements hold, the correlation matrix for derived variables is always positive definite.

4.4 Multicollinearity

For a successful transformation, the elicited correlation matrix, in addition to being positive definite, should also be stable because of matrix inversion. The instability can occur when the determinant of the correlation matrix is close to zero. This problem is called multicollinearity . The term multicollinearity defines itself, multi implying many and collinear implying linear dependencies (Myers, 1986).

Multicollinearity occurs when there are near linear dependencies among the columns of a correlation matrix. That is, there is a vector of constants $\mathbf{c}$ (not all zero) for which,

$$\sum_{j=1}^{n} \mathbf{c}^{\mathbf{T}} x_j \cong 0 \tag{4.28}$$

where x_j are the columns of the n x n correlation matrix. If the right hand side of equation (4.28) is identically zero then the correlation matrix is positive semi-definite. Thus, the linear dependencies are exact and the inverse of the correlation matrix and hence $\mathbf{L}^{-1}$ does not exist.

Myers (1986) states that if multicollinearity is present, then there exists at least one $\lambda_i \cong 0$, where λ_i are the eigenvalues of the correlation matrix. While the nearness to zero of the smallest eigenvalue is a measure of the strength of a linear dependency, the ratio,

$$\phi = \frac{\lambda_{max}}{\lambda_{min}} \tag{4.29}$$

which is called the condition number of the correlation matrix is the true measure of multicollinearity. As a rule of thumb, a correlation matrix with $\phi \leq 100$ is considered to be stable. However, when ϕ exceeds 1000 then one should be concerned about the effect of multicollinearity (*i.e* instability in the correlation matrix) (Myers, 1986).

The instability in the correlation matrix can hamper a successful transformation. It is suggested that for practical applications the concept called the "k value" be utilized. The k value is used in ridge regression as a mathematical manipulation to stabilize an unstable correlation matrix (Myers, 1986). The stability is achieved by replacing the correlation matrix $\mathbf{R}$ by $(\mathbf{R} + k\,\mathbf{I})$, where k is a small positive quantity. Similarly, an elicited unstable correlation matrix can be stabilized by introducing a $k\,\mathbf{I}$ matrix to the correlation matrix. The k value would be the smallest value that would make the correlation matrix stable.

Stability is defined in terms of the desired stabilizing condition number ϕ_s given by,

$$\phi_s = \frac{\lambda_{max} + k}{\lambda_{min} + k} \tag{4.30}$$

Therefore,

$$k = \frac{\lambda_{max} - \phi_s\,\lambda_{min}}{\phi_s - 1} \tag{4.31}$$

would stabilize the correlation matrix to the desired condition number ϕ_s.

An upper bound on the stabilizing k value can be established in terms of the number of the variables in the functional form (n) and the desired stabilizing condition number (ϕ_s), from the fact that the largest eigenvalue of a correlation matrix is always less than n (Graybill, 1983). Hence, k is always less than $\left[\frac{n}{\phi_s - 1}\right]$. For example, if a condition number $\phi_s = 100$ is desired (the empirical limit at which regression analysis considers a correlation matrix to be stable) the k value for the function described by equation (4.33) is less than 0.030303.

Therefore, the k value that stabilizes a correlation matrix to a desired ϕ_s can be bound as,

$$\left[\frac{\lambda_{max} - \phi_s\,\lambda_{min}}{\phi_s - 1}\right] \leq k_s < \left[\frac{n}{\phi_s - 1}\right] \tag{4.32}$$

4.5 Numerical Study

The first part of the numerical study compares the variable transformation method to the standard approach used in moment analysis under general conditions (*i.e* stable correlation matrices). The second part explores the behavior of the two methods in the presence of multicollinearity. The intention of the third part is to study the susceptibility of the transformation to the effect of multicollinearity.

The duration of a work package in a construction project is used as the derived variable for the study. The duration of a work package (T) can be evaluated from the simple relationship given by,

$$T = \frac{Q}{P_L \; L} = g(\mathbf{X}) \tag{4.33}$$

where Q is the quantity descriptor, P_L is the labour productivity rate and L is the labour usage, the primary variables of the work package duration model.

4.5.1 Variable Transformation Method

For the work package duration model described by equation (4.33),

$$\mathbf{X}^{\mathbf{T}} = [Q \; P_L \; L] = [X_1 \; X_2 \; X_3] \qquad \text{and} \qquad \mathbf{D} = \mathrm{diag}[\sigma_i].$$

If the correlation matrix for the work package duration model is $\mathbf{R}$, where

$$\mathbf{R} = \begin{bmatrix} 1.0 & \rho_{12} & \rho_{13} \\ \rho_{12} & 1.0 & \rho_{23} \\ \rho_{13} & \rho_{23} & 1.0 \end{bmatrix} \tag{4.34}$$

then the lower triangular matrix obtained from the Cholesky decomposition of the correlation matrix ($\mathbf{R} = \mathbf{L}\mathbf{L}^{\mathbf{T}}$) is,

$$\mathbf{L} = \begin{bmatrix} L_{11} & 0.0 & 0.0 \\ L_{12} & L_{22} & 0.0 \\ L_{13} & L_{23} & L_{33} \end{bmatrix} \tag{4.35}$$

where $L_{11} = 1.0$; $L_{12} = \rho_{12}$; $L_{13} = \rho_{13}$; $L_{22} = \sqrt{1 - \rho_{12}^2}$; $L_{23} = \dfrac{\rho_{23} - \rho_{12}\,\rho_{13}}{\sqrt{1 - \rho_{12}^2}}$; and $L_{33} = \sqrt{1 - L_{13}^2 - L_{23}^2}$.

The transformation for the uncorrelated variables is $\mathbf{Z} = \mathbf{L}^{-1}\,\mathbf{D}^{-1}\,\mathbf{X}$ and for the functional form is $\mathbf{X} = \mathbf{D}\;\mathbf{L}\;\mathbf{Z}$.

Hence, the first four moments of the work package duration are approximated from equations (2.36) to (2.39). The expected value of the work package duration is,

$$E[T] \approx G(\bar{\mathbf{Z}}) + \frac{1}{2}\sum_{i=1}^{3}\frac{\partial^2 G}{\partial Z_i^2}\,\mu_2(Z_i) \tag{4.36}$$

the second central moment is,

$$\begin{aligned} \mu_2(T) \approx & \sum_{i=1}^{3}\left[\frac{\partial G}{\partial Z_i}\right]^2 \mu_2(Z_i) \\ & + \sum_{i=1}^{3}\frac{\partial G}{\partial Z_i}\,\frac{\partial^2 G}{\partial Z_i^2}\,\mu_3(Z_i) \\ & + \frac{1}{4}\sum_{i=1}^{3}\left[\frac{\partial^2 G}{\partial Z_i^2}\right]^2 \left[\mu_4(Z_i) - [\mu_2(Z_i)]^2\right] \end{aligned} \tag{4.37}$$

the third central moment is,

$$\begin{aligned}\mu_3(T) \approx & \sum_{i=1}^{3}\left[\frac{\partial G}{\partial Z_i}\right]^3 \mu_3(Z_i) \\ & +\frac{3}{2}\sum_{i=1}^{3}\left[\frac{\partial G}{\partial Z_i}\right]^2 \frac{\partial^2 G}{\partial Z_i^2}\left[\mu_4(Z_i) - [\mu_2(Z_i)]^2\right]\end{aligned} \tag{4.38}$$

and the fourth central moment is,

$$\mu_4(T) \approx \sum_{i=1}^{3}\left[\frac{\partial G}{\partial Z_i}\right]^4 \mu_4(Z_i) \tag{4.39}$$

where Z_1, Z_2, Z_3 are the transformed uncorrelated variables of Q, P_L, L and $G(\mathbf{Z})$ is the transformed function for work package duration.

4.5.2 The Standard Approach

The first four moments of the work package duration from the standard approach are derived by expanding equations (2.20) to (2.23) (see Appendix A for the general derivation). The boxed terms are those due to the linear correlations between primary variables. Then the expected value of the work package duration is,

$$\begin{aligned}E[T] \approx & \frac{\bar{X}_1}{\bar{X}_2\bar{X}_3} + \frac{1}{2}\sum_{i=1}^{3}\frac{\partial^2 g}{\partial X_i^2}\mu_2(X_i) \\ & \boxed{+\sum_{i=1}^{3}\sum_{j=i+1}^{3}\frac{\partial^2 g}{\partial X_i \partial X_j} cov(X_i, X_j)}\end{aligned} \tag{4.40}$$

the second central moment is,

$$
\begin{aligned}
\mu_2(T) \approx\ & \sum_{i=1}^{3} \left[\frac{\partial g}{\partial X_i}\right]^2 \mu_2(X_i) \\
& + \sum_{i=1}^{3} \frac{\partial g}{\partial X_i} \frac{\partial^2 g}{\partial X_i^2} \mu_3(X_i) \\
& + \frac{1}{4} \sum_{i=1}^{3} \left[\frac{\partial^2 g}{\partial X_i^2}\right]^2 \left[\mu_4(Z_i) - [\mu_2(Z_i)]^2\right] \\
& \boxed{+ 2 \sum_{i=1}^{3} \sum_{j=i+1}^{3} \frac{\partial g}{\partial X_i} \frac{\partial g}{\partial X_j} cov(X_i, X_j)} \\
& \boxed{- \sum_{i=1}^{3} \sum_{j=i+1}^{3} \left[\frac{\partial^2 g}{\partial X_i \partial X_j}\right]^2 [cov(X_i, X_j)]^2}
\end{aligned}
\tag{4.41}
$$

the third central moment is,

$$
\begin{aligned}
\mu_3(T) \approx\ & \sum_{i=1}^{3} \left[\frac{\partial g}{\partial X_i}\right]^3 \mu_3(X_i) \\
& + \frac{3}{2} \sum_{i=1}^{3} \left[\frac{\partial g}{\partial X_i}\right]^2 \frac{\partial^2 g}{\partial X_i^2} \left[\mu_4(Z_i) - [\mu_2(Z_i)]^2\right] \\
& \boxed{- 6 \sum_{i=1}^{3} \sum_{j=i+1}^{3} \frac{\partial g}{\partial X_i} \frac{\partial g}{\partial X_j} \frac{\partial^2 g}{\partial X_i \partial X_j} [cov(X_i, X_j)]^2}
\end{aligned}
\tag{4.42}
$$

and the fourth central moment is,

$$
\mu_4(T) \approx \sum_{i=1}^{3} \left[\frac{\partial g}{\partial X_i}\right]^4 \mu_4(X_i) \tag{4.43}
$$

where X_1 is Q, X_2 is P_L and X_3 is L.

Table 4.1: Quantity Descriptors (Q) (ft^3)

W.P	E[Q]	σ_Q	$\sqrt{\beta_1}$	β_2
01	38397.3	12186.1	0.5	3.3
02	60555.0	8829.3	0.9	9.0
03	76850.0	24440.5	0.5	3.2
04	16185.0	3527.4	0.8	7.8
05	32429.2	7030.8	0.8	7.8
06	38397.3	12186.1	0.5	3.3
07	21998.0	2621.4	0.2	2.4
08	76850.0	24440.5	0.5	3.2
09	20413.0	5782.4	0.7	8.5
10	76850.0	24440.5	0.5	3.2

4.5.3 The Comparison

The moment analyses for both approaches consider terms up to the fourth order. While both methods treat the same correlations, the variable transformation method simplifies the approximations by the transformation (see the boxed terms in equations 4.40, 4.41 and 4.42). The two approaches are compared for ten hypothetical work package durations. The values for the primary variables and correlation coefficients used for the numerical study are given in Tables 4.1 to 4.4. Table 4.5 shows the moments for work package durations evaluated from the two approaches when the correlation matrices are stable. The time unit is in years.

When the primary variables are assumed to be uncorrelated, both methods give identical moments indicating that they are comparable (see Table 4.5). When there is correlation between the primary variables the expected values from both methods are identical. The second and third central moments from the variable transformation method are larger for all work packages. The fourth central moments from the standard approach are same as when uncorrelated or highly correlated (see Table 4.6) because there are no covariance terms in equation (4.43). If the moment analysis

Table 4.2: Labour Productivity Rates, P_L; $(ft^3/m.d)$

W.P	$E[P_L]$	σ_{P_L}	$\sqrt{\beta_1}$	β_2
01	9.0	1.25	0.0	5.6
02	9.0	1.25	0.0	5.6
03	9.0	1.25	0.0	5.6
04	10.1	2.28	0.1	2.2
05	8.4	1.28	0.1	8.8
06	9.0	1.25	0.0	5.6
07	10.1	2.28	0.1	2.2
08	9.0	1.25	0.0	5.6
09	9.9	2.22	0.9	9.0
10	10.2	2.23	0.8	8.0

Table 4.3: Labour Usage, L; $(m.d/year)$

W.P	$E[L]$	σ_L	$\sqrt{\beta_1}$	β_2
01	6833.2	692.7	0.4	2.4
02	15185.0	1539.5	0.4	2.3
03	15185.0	1539.5	0.4	2.3
04	6074.0	615.8	0.4	2.4
05	7777.5	2339.8	1.1	5.7
06	9055.5	832.9	0.4	4.3
07	6074.0	615.8	0.4	2.4
08	15092.5	1388.1	0.4	4.3
09	3850.8	393.4	0.4	2.3
10	15092.5	1388.1	0.4	4.3

Table 4.4: Condition Number (ϕ) and Correlation Coefficients

W.P	ϕ	ρ_{QP_L}	ρ_{QL}	ρ_{P_LL}
01	6.77	-0.48	0.42	-0.69
02	9.05	-0.55	0.62	-0.74
03	8.60	-0.53	0.56	-0.74
04	6.77	-0.48	0.42	-0.69
05	8.60	-0.53	0.56	-0.74
06	8.22	-0.48	0.62	-0.70
07	6.77	-0.48	0.42	-0.69
08	11.4	-0.53	0.56	-0.80
09	7.94	-0.48	0.62	-0.69
10	8.60	-0.53	0.56	-0.74

considered only the terms up to the third order (Bury, 1975; Siddall, 1972), there will be no covariance term in equation (4.42). Then the third central moments from the standard approach will also be same as when uncorrelated or highly correlated.

4.5.4 Transformation under Multicollinearity

Two numerical studies are done to demonstrate the behavior of the transformation in the presence of multicollinearity. The first, compares the variable transformation and the standard method using the same correlation matrix for all the work package durations. The correlation matrix used is,

$$\mathbf{R_m} = \begin{bmatrix} 1.0 & -0.999 & 0.999 \\ -0.999 & 1.0 & -0.999 \\ 0.999 & -0.999 & 1.0 \end{bmatrix} \tag{4.44}$$

which has a condition number ϕ equal to 2998.04.

Table (4.6) shows the moments from the two methods. Again, the expected values are identical, but some of the central moments are different. While some of the variances were comparable, the others showed considerable differences. The most

WP	$E[T]$			$\mu_2(T)$			$\mu_3(T)$			$\mu_4(T)$		
	Uncor	Trans	Stdrd	Uncor	Trans	Stdrd	Uncor	Trans	Stdrd	Uncor	Trans	Stdrd
01	.64294	.64168	.64168	.05127	.05109	.04942	.00517	.00750	.00410	.00545	.00692	.00545
02	.45628	.45254	.45254	.01022	.00727	.00684	.00069	.00088	.00053	.00025	.00025	.00025
03	.57906	.57625	.57625	.04172	.03951	.03825	.00379	.00500	.00273	.00351	.00409	.00351
04	.28024	.27986	.27986	.00762	.00842	.00736	.00031	.00116	.00022	.00012	.00026	.00012
05	.55286	.52657	.52657	.03521	.01447	.00842	.01296	.01721	.01092	.00417	.00361	.00417
06	.48430	.48156	.48156	.02886	.02669	.02609	.00226	.00266	.00172	.00177	.00190	.00177
07	.38089	.37804	.37804	.00979	.00801	.00769	.00043	.00043	.00031	.00011	.00008	.00011
08	.58158	.57981	.57981	.04175	.04091	.03946	.00393	.00556	.00293	.00360	.00447	.00360
09	.56646	.56472	.56472	.03997	.04224	.03768	.01011	.01584	.00874	.00632	.00866	.00632
10	.52807	.53086	.53086	.03906	.04559	.04139	.00701	.00974	.00562	.00320	.00465	.00321

Table 4.5 : First Four Moments of the Work Package Durations

Table 4.6: Moments of the Duration with an Unstable Correlation Matrix

	$E[T]$		$\mu_2(T)$		$\mu_3(T)$		$\mu_4(T)$	
WP	Trans	Stdrd	Trans	Stdrd	Trans	Stdrd	Trans	Stdrd
01	.6417	.6417	.05392	.04845	.01217	.00051	.00814	.00545
02	.4525	.4525	.00885	.00669	.00244	.00026	.00045	.00025
03	.5779	.5779	.04386	.03945	.00874	.00037	.00524	.00351
04	.2814	.2814	.01304	.00798	.00413	−.00006	.00057	.00012
05	.5142	.5142	.00642	−.00505	.01732	.00751	.00336	.00417
06	.4854	.4854	.03258	.02921	.00585	.00039	.00293	.00177
07	.3780	.3780	.00929	.00754	.00154	.00005	.00022	.00011
08	.5829	.5829	.04711	.04227	.00995	.00069	.00596	.00360
09	.5727	.5727	.07155	.04508	.05648	.00519	.01994	.00632
10	.5381	.5381	.06099	.04761	.02216	.00245	.00845	.00320

startling observation is the negative variance for the fifth work package duration, indicating that in the presence of multicollinearity the standard approach could fail.

When the moments from the variable transformation method in Tables (4.5) and (4.6) are compared, the expected values compare well while the central moments are reasonably close, considering the fact that they have different correlation values. This study indicates that the transformation is not too susceptible to the instability in the correlation matrix for this example.

The intention of the second study is to see how susceptible the transformation is to the effect of multicollinearity. The "k value" concept discussed in section (4.4) is used to study the percentage change in moments of a work package duration. If the percentage change from the base value moment (*i.e* at $k = 0$), for stable (small ϕ) and unstable (large ϕ) correlation matrices, are similar with increasing k values (*i.e* as the matrices got more and more stable), then the transformation is not susceptible to instability in the correlation matrix.

Figures (4.3) to (4.6) show the absolute percentage changes in the first four moments from the base values, for increasing k values. The percentage changes in the moments are similar when the condition number (ϕ) for the correlation matrices vary from 50 to 2998, indicating that the transformation is not susceptible to the instability ($i.e$ effect of multicollinearity) in the correlation matrix for this study.

However, it must be noted that in another situation it is possible for multicollinearity to effect the transformation. It is suggested that in practical applications of the variable transformation method (or the standard method) the condition number (ϕ) of the correlation matrix be checked for multicollinearity. If unstable correlation matrices have been elicited, they can be stabilized using a small k value at the discretion of the analyst. This check is equally valid for the treatment of correlations in Monte Carlo simulation.

4.6 Summary

The correlations between the primary variables and between the derived variables is addressed in this chapter.

The second section highlighted the often ignored requirement for the correlation matrix to be positive definite and developed a subjective elicitation method to obtain a positive definite correlation matrix for primary variables. A positive definite correlation matrix recognizes the existence of multivariates in a system. The third section suggested a method to obtain a positive definite correlation matrix for derived variables when only the linear correlations between the primary variables are available. The theoretical developments in these two sections are the basis for the part in the computer program "ELICIT" (see Appendix D) to obtain interactively the correlations between variables. The fourth section highlighted the concept of multicollinearity and its possible effects on the variable transformation. A mathematical manipulation that could provide stability to the correlation matrix for practical applications was suggested.

The final section utilizing the example of the work package duration showed numerically that the variable transformation method is comparable to the standard approach in treating correlation between primary variables under general conditions (stable correlation matrices). Also, that the transformation simplifies the approximations for the first four moments and treats the linear correlations consistently. The other two numerical studies explored the behavior of the transformation in the presence of multicollinearity. The first showed that the standard approach can fail in the presence of multicollinearity while the variable transformation method was more stable. The second study showed that the transformation was not susceptible to instabilities in the correlation matrix.

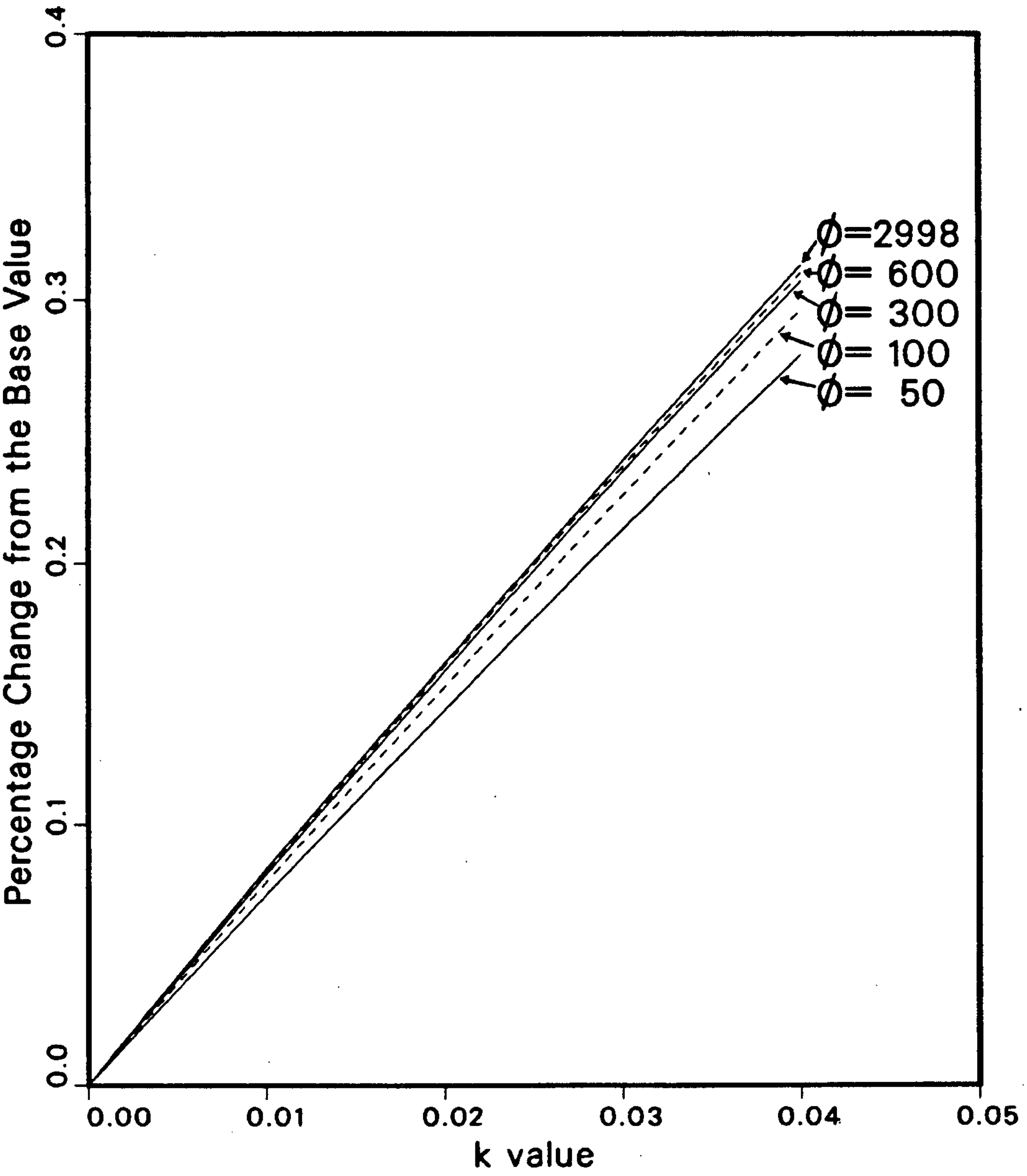

Figure 4.3: Expected Values

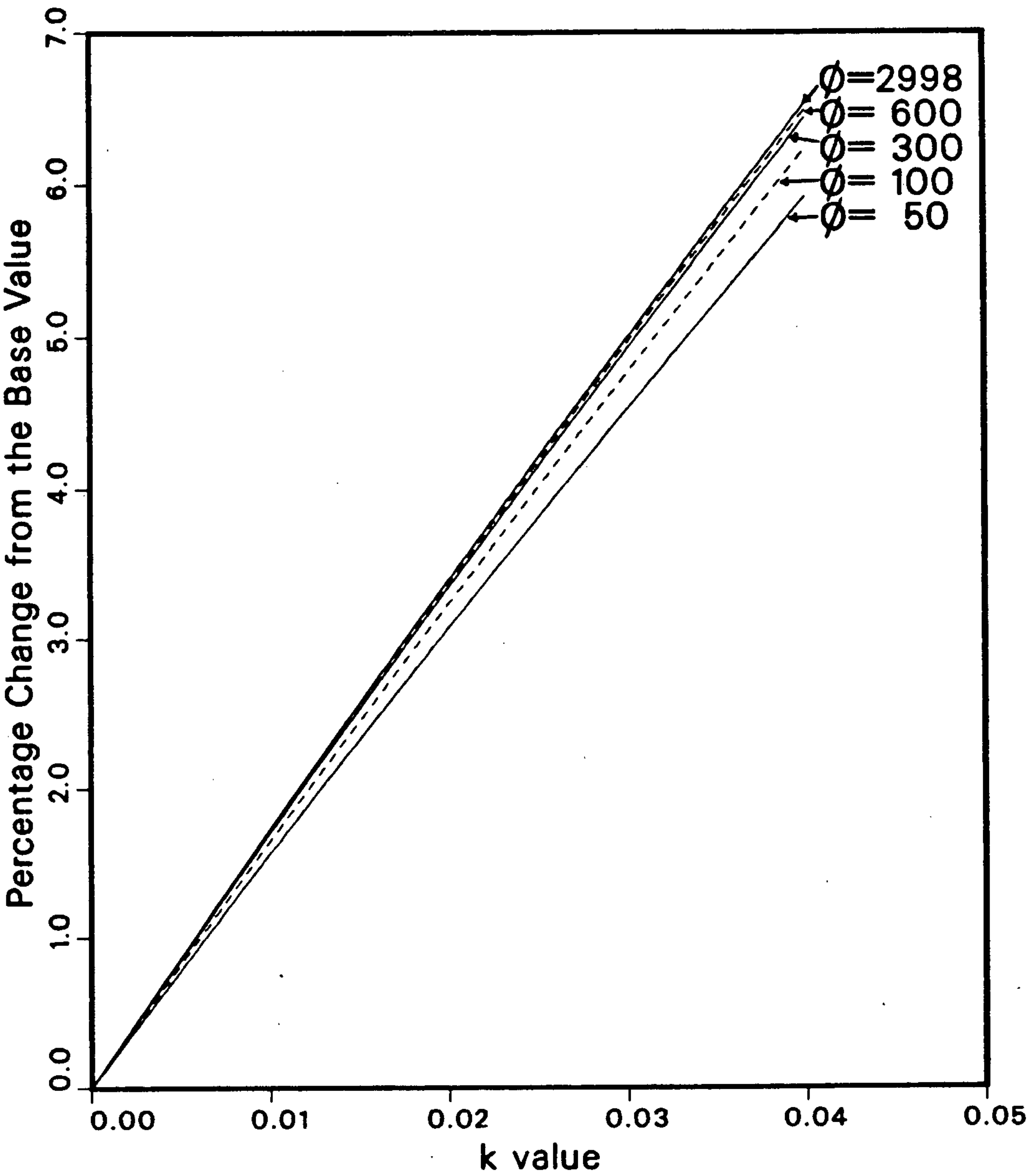

Figure 4.4: Second Central Moment

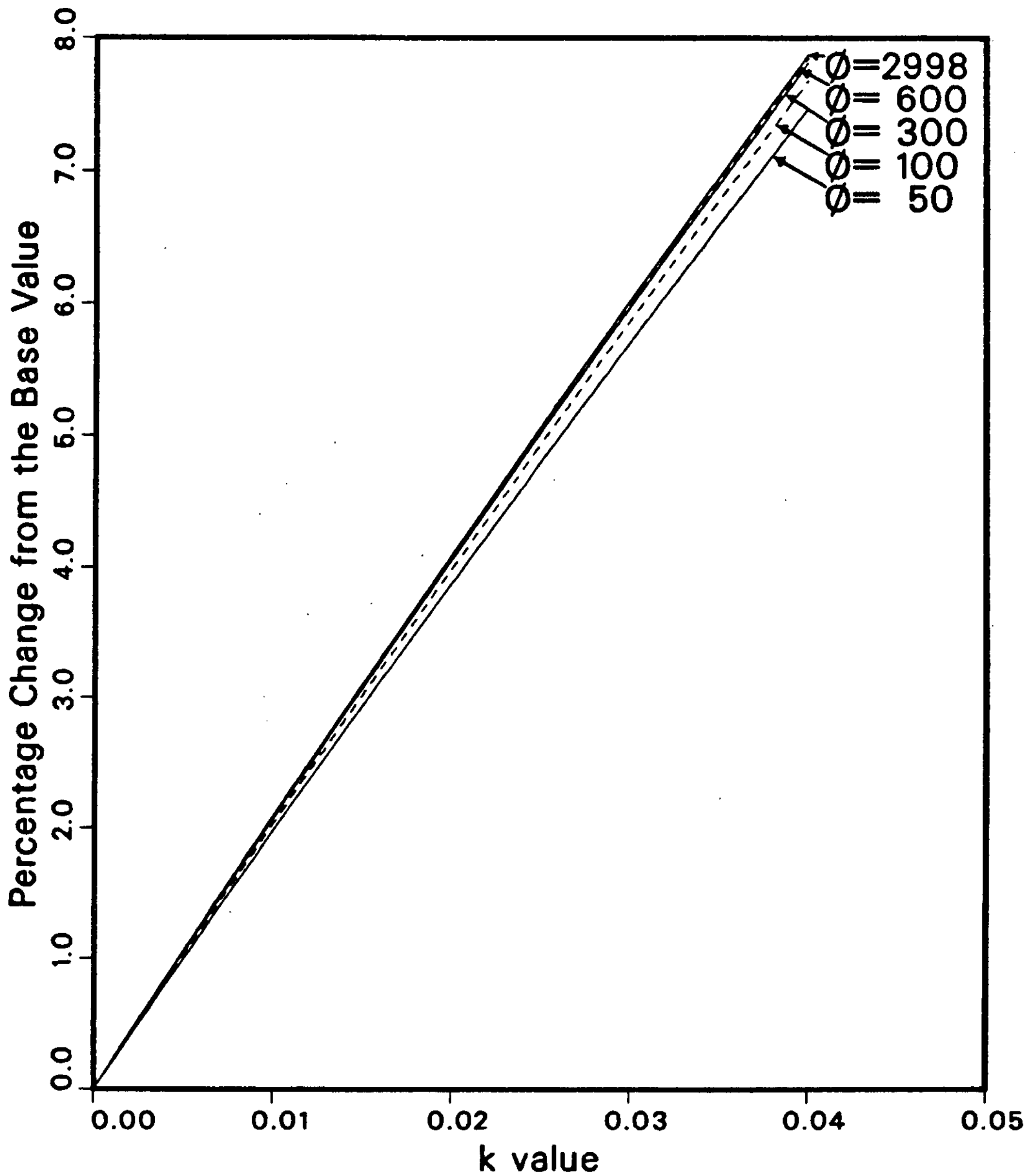

Figure 4.5: Third Central Moment

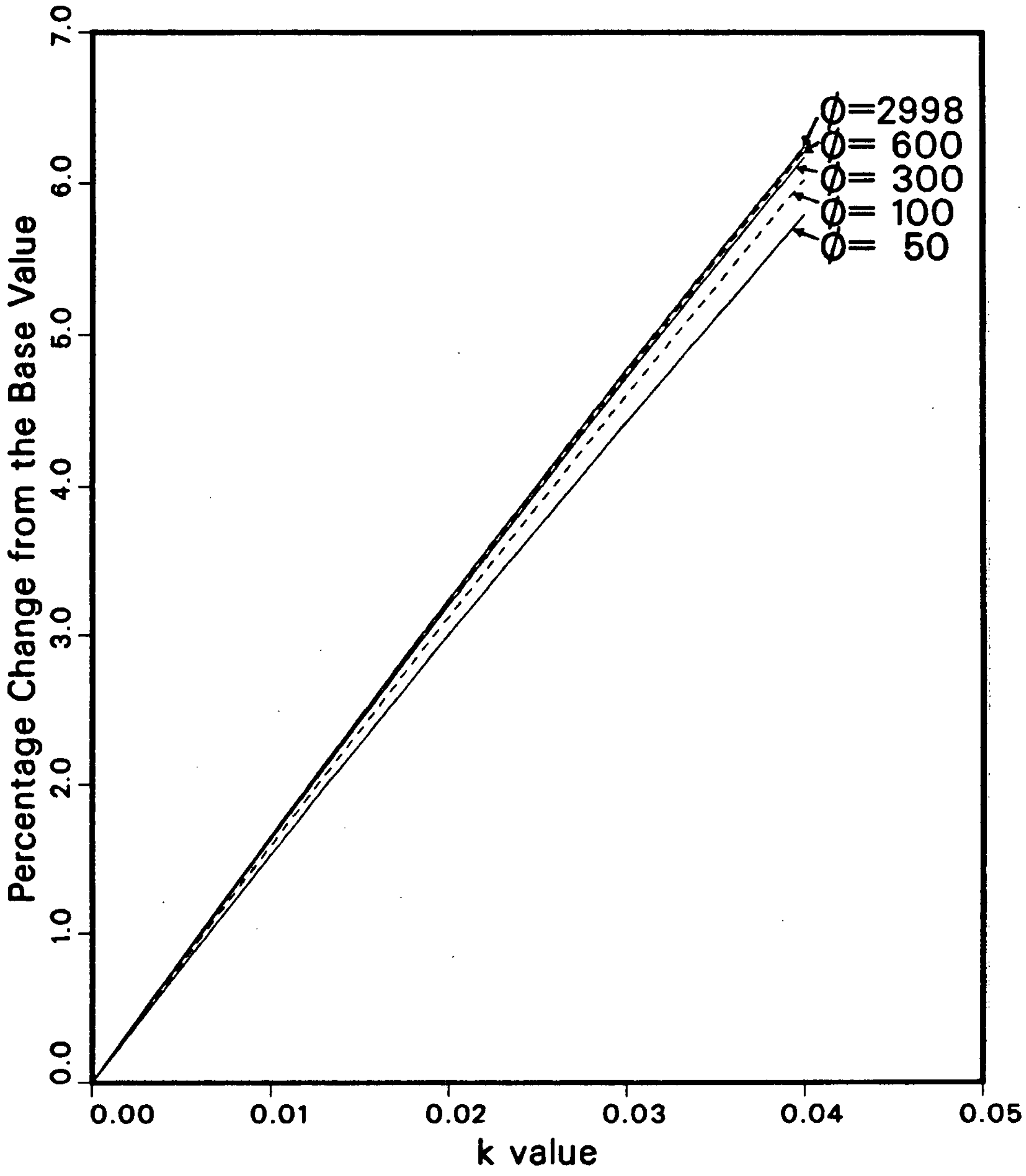

Figure 4.6: Fourth Central Moment

Chapter 5

Decomposition of a Derived Variable

5.1 General

In developing the risk measurement framework to quantify the uncertainty of a derived variable it was assumed that a derived variable can be more accurately estimated from a set of primary variables that are functionally related to it than by direct estimation (assumption 2.3). This reflects the engineering penchant to seek more detail as a way of seeking greater precision. For most derived variables in engineering construction, assumption (2.3) is reasonable.

However, this assumption becomes debatable when variables which are sometimes estimated holistically in the elicitation of subjective judgments (probabilities) - (*eg.* duration, productivity) are decomposed. This chapter describes a study on the decomposition of such a derived variable. The duration of an activity is used as the example for the derived variable to compare holistic versus decomposed methods of estimation.

The objective of this chapter is to make a small step towards exploring an issue that is largely ignored in the estimation literature and to provide the motivation for a more extensive study on the elicitation of subjective probabilities for continuous random primary variables.

5.2 Decomposition

Ravinder et al. (1988) state that decomposition is often regarded as a useful technique for reducing the complexity of difficult judgment problems. They studied the application of decomposition to the elicitation of subjective probabilities for discrete events. A target event for which probability judgments were required was decomposed into background events in the form of conditional probabilities of the target event. Once the individual conditional distributions for background events were elicited, the law of total probability was used for the aggregation.

The probability of the target event $Pr(A)$ was defined as,

$$Pr(A) = \sum_{i=1}^{n} Pr(A|B_i)\ Pr(B_i) \tag{5.1}$$

where the background events denoted $B_1, \ldots.., B_n$ form a mutually exclusive and exhaustive partition of the relevant event space (*i.e* $\sum_{i=1}^{n} Pr(B_i) = 1$).

They concluded: if the component probabilities can be assessed with no greater precision than holistic assessment, decomposition reduces random errors associated with probability encoding; but as the number of events increases, error reduction will only occur up to a point (*i.e* a limit for decomposition exists). While it is not possible to generalize their conclusions to the decomposition of a derived variable to a functionally related set of primary variables, they are used as guidance for this study.

In the context of this research, the main reason for decomposing work package variables to their primary variables is to develop a link between cost and time of the work package for economic analysis. It is incorrect to assume that cost is independent of time, because when a work package duration is either reduced (more resources) or increased (less resources) the net result is a change in the cost. The link between cost and time permits the use of net present value and internal rate of return as decision variables.

The second reason is the basis for assumption (2.3), to reduce the complexity of holistic estimation because experts in construction (engineers) find it easier to quantify the decomposed primary variables. The same reasoning has been used by other authors for decomposing the activity duration into its primary variables (Jaafari, 1984 ; Hendrickson et al., 1987). This raises another question; won't it be more accurate if the primary variables are further decomposed.

The main disadvantage of decomposition is the loss of the mental awareness of interdependencies between primary variables that exists when estimating from a holistic approach. While it may be possible to relate the primary variables functionally to the derived variable, it is also difficult to model all the interdependencies (Inyang, 1983).

Secondly, even if accurate estimates for primary variables are obtained, as the decomposition is continued a model which can link them to provide a reliable estimate of the derived variable may be lacking. As Ravinder et al. (1988) have shown a definite limit exists for decomposition. Thirdly, unless decomposition improves the system significantly, it would be hard to convince an expert that decomposition is necessary for the elicitation of subjective probabilities. It was stated in chapter three that convincing experts about the relevance of the primary variables was essential to gain their full cooperation during the elicitation (Cooper and Chapman, 1987; Huber, 1974; Hull, 1980; Spetzler and Stael von Holstein, 1975).

The next four sections propose hypotheses, test statistics, an experiment and the analysis to study a derived variable that is sometimes estimated holistically. In addition, some of the beliefs that exist in engineering construction regarding decomposed versus holistic estimation of judgments are explored.

5.3 Hypotheses

Nine hypotheses are suggested to study decomposed versus holistic estimation of an activity duration. The first hypothesis is about the precision of assessments for an activity duration from the two approaches. The other eight are for assessed expected values and standard deviations for an activity duration to compare holistic versus decomposed estimation.

Hypothesis 5.1

The precision of assessments for an activity duration from holistic or decomposed estimation are similar.

The precision of assessments from the two approaches are measured using the coefficient of variation for duration of an activity, a non-dimensional measure of variation.

$$H_0 \; : \; E\left[V_i\right] \; = \; 0 \quad ; \quad H_1 \; : \; E\left[V_i\right] \; \neq \; 0$$

where $V_i = V_{D_i} - V_{W_i}$, the difference between a pair of coefficients of variation of the assessed activity duration from decomposed (V_{D_i}) and holistic (V_{W_i}) estimation.

Hypothesis 5.2

When experts are asked to assess the expected value for duration of an activity from the holistic approach that assessment will be the true value.

$$H_0 \; : \; E[T] \; = \; E\left[T_W\right] \quad ; \quad H_1 \; : \; E[T] \; \neq \; E\left[T_W\right]$$

Hypothesis 5.3

When experts are asked to assess the standard deviation for duration of an activity from the holistic approach, that assessment will be the true value.

$$H_0 : \sigma_T = \sigma_{T_W} \quad ; \quad H_1 : \sigma_T \neq \sigma_{T_W}$$

Hypothesis 5.4

When experts are asked to assess the expected value for duration of an activity from the decomposed approach that assessment will be the true value.

$$H_0 : E[T] = E[T_D] \quad ; \quad H_1 : E[T] \neq E[T_D]$$

Hypothesis 5.5

When experts are asked to assess the standard deviation for duration of an activity from the decomposed approach, that assessment will be the true value.

$$H_0 : \sigma_T = \sigma_{T_D} \quad ; \quad H_1 : \sigma_T \neq \sigma_{T_D}$$

Hypothesis 5.6

When experts are asked to assess the expected value for duration of an activity from the holistic approach, that assessment will be an underestimation of the true value.

$$H_0 : E[T] = E[T_W] \quad ; \quad H_1 : E[T] > E[T_W]$$

Hypothesis 5.7

When experts are asked to assess the standard deviation for duration of an activity from the holistic approach, that assessment will be an underestimation of the true value.

$$H_0 : \sigma_T = \sigma_{T_W} \quad ; \quad H_1 : \sigma_T > \sigma_{T_W}$$

Hypothesis 5.8

When experts are asked to assess the expected value for duration of an activity from the decomposed approach, that assessment will be an underestimation of the true value.

$$H_0 \;:\; E[T] \;=\; E\left[T_D\right] \qquad ; \qquad H_1 \;:\; E[T] \;>\; E\left[T_D\right]$$

Hypothesis 5.9

When experts are asked to assess the standard deviation for duration of an activity from the decomposed approach, that assessment will be an underestimation of the true value.

$$H_0 \;:\; \sigma_T \;=\; \sigma_{T_D} \qquad ; \qquad H_1 \;:\; \sigma_T \;>\; \sigma_{T_D}$$

where $E\left[T_W\right]$ and $E\left[T_D\right]$ are the expected values and σ_{T_W} and σ_{T_D} are the standard deviations assessed for the activity duration from holistic and decomposed subjective estimation.

While a hypothesis test is done for the first hypothesis, only significance tests (*i.e* a hypothesis can only be rejected) are done for the next eight because there is only one sample to test all of the hypotheses. Hypothesis (5.1) tests whether there is a difference between the precision of assessments (*i.e* coefficients of variation) from the two approaches. The next four hypotheses (5.2 to 5.5) provide the basis to compare the two approaches. By calculating the percentages of the number of times an individual hypothesis is rejected, given the group of experts and the amount of information available during the elicitation, the two approaches are compared. Hypotheses (5.6) to (5.9) are included because of the traditional belief in engineering construction that holistic approach underestimates duration more regularly than the decomposed approach (*i.e* holistic will be rejected more times than the decomposed).

It must be stressed that the objective of this study is not to select the "better" method, but to compare the two methods available for estimating duration when experts participate in subjective elicitation. The "better" method is a consensus approach after estimating from both approaches. However, it is not practical as a subjective elicitation technique.

5.4 Test Statistics

Assumption 5.1 : A sample of durations to complete an activity constitutes a random sample from a normal distribution with both μ and σ unknown.

Since the sample of the measured durations are for the same activity it is reasonable to expect the measurements to be symmetric around the mean value. Then the test statistic for the difference between paired coefficients of variation of the assessed activity duration is (Devore, 1982),

$$t_{paired} = \frac{\bar{V}}{S_v/\sqrt{n}} \tag{5.2}$$

where $\bar{V}$ and S_v are sample mean and standard deviation, respectively for V_i's and n is the sample size.

The rejection regions for level α tests are (see figure 5.1),

Hypothesis	Rejection Region
5.1	$t_{paired} \geq t_{\frac{\alpha}{2},n-1}$ or $t_{paired} \leq -t_{\frac{\alpha}{2},n-1}$

Test statistic for the assessed expected value for activity duration is (Devore, 1982),

$$t = \frac{\bar{T} - E[T_0]}{S/\sqrt{n}} \tag{5.3}$$

where $\bar{T}$ and S are the sample mean and standard deviation, $E[T_0]$ is either the assessed $E[T_W]$ or $E[T_D]$.

The rejection regions for level α tests are (see figures 5.1 and 5.2),

Hypothesis	Rejection Region
5.2 and 5.4	$t \geq t_{\frac{\alpha}{2},n-1}$ or $t \leq -t_{\frac{\alpha}{2},n-1}$
5.6 and 5.8	$t \geq t_{\alpha,n-1}$

Test statistic for the assessed standard deviation for activity duration is (Devore, 1982),

$$\chi^2 = \frac{(n-1)\,S^2}{\sigma^2_{T_0}} \tag{5.4}$$

where S is the sample standard deviation and σ_{T_0} is the assessed σ_{T_W} or σ_{T_D}.

The rejection regions for level α tests are (see figures 5.3 and 5.4),

Hypothesis	Rejection Region
5.3 and 5.5	$\chi^2 \geq \chi^2_{\frac{\alpha}{2},n-1}$ or $\chi^2 \leq \chi^2_{1-\frac{\alpha}{2},n-1}$
5.7 and 5.9	$\chi^2 \geq \chi^2_{\alpha,n-1}$

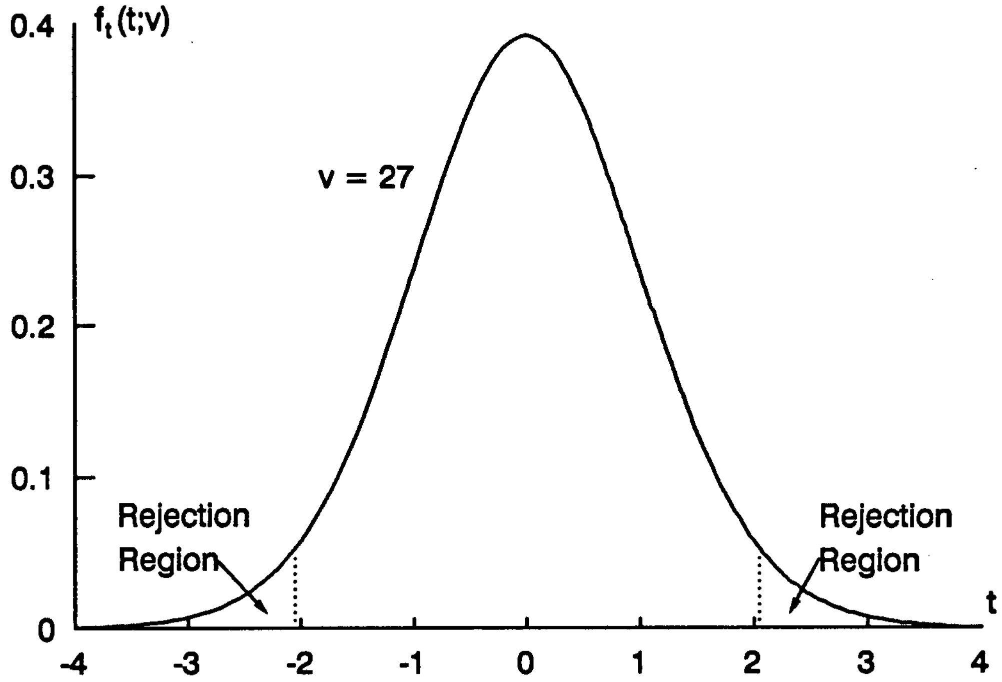

Figure 5.1: t Distribution for Two Tailed Test

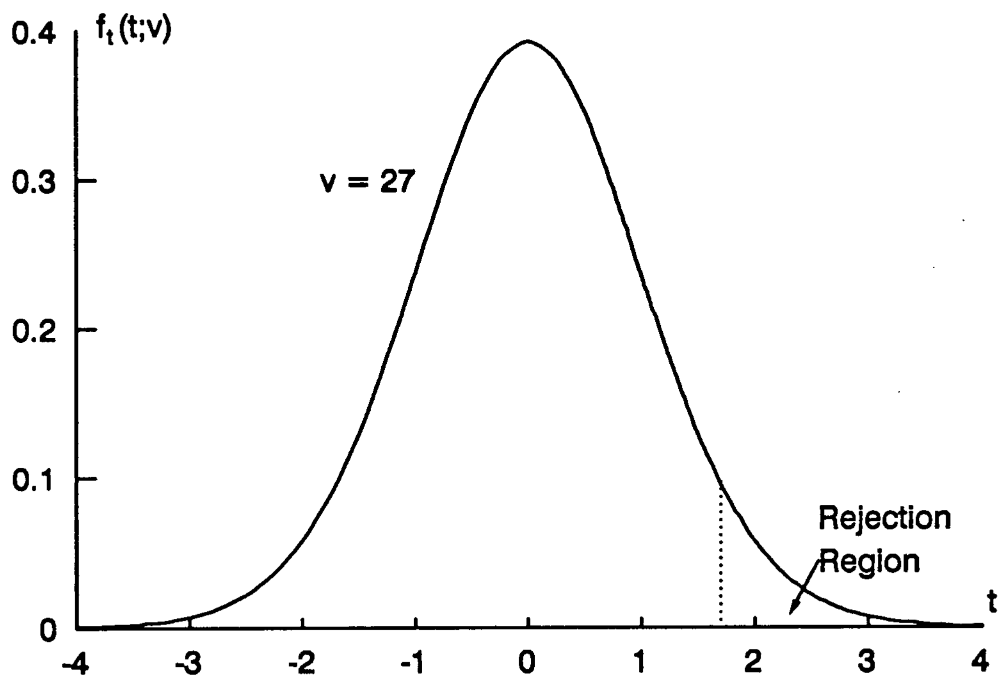

Figure 5.2: t Distribution for Upper Tailed Test

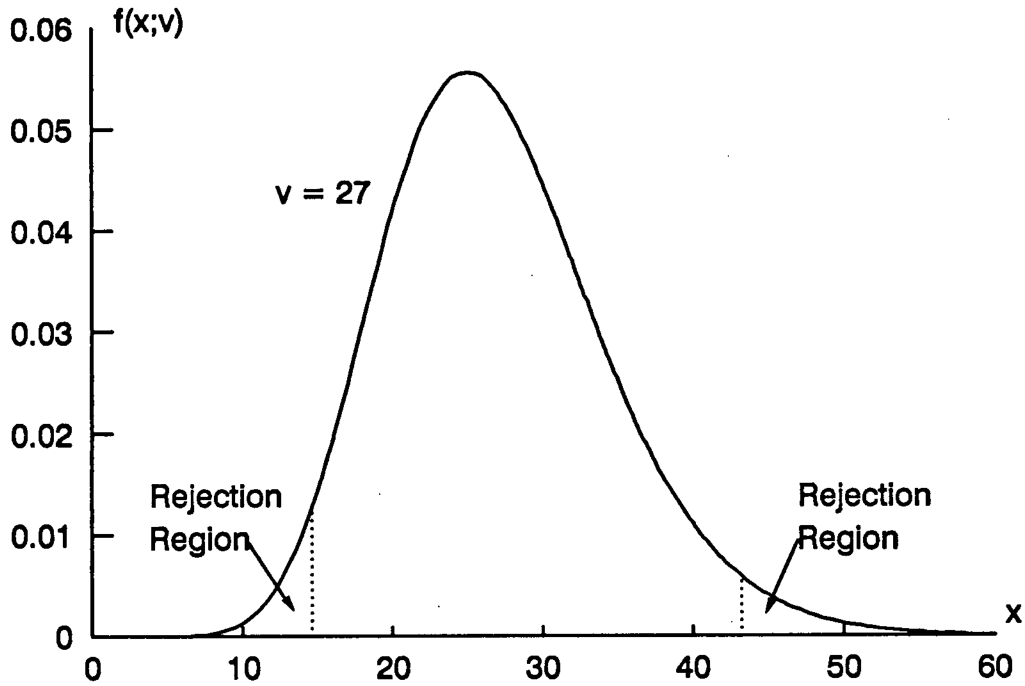

Figure 5.3: χ^2 Distribution for Two Tailed Test

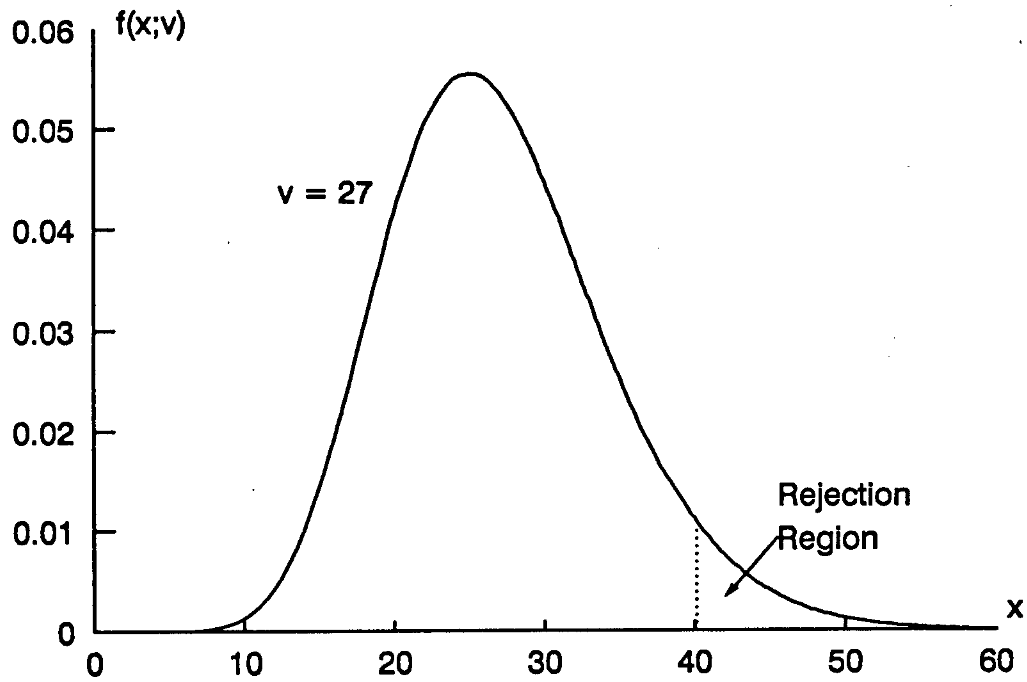

Figure 5.4: χ^2 Distribution for Upper Tailed Test

5.5 Experiment

5.5.1 The Activity

The activity to obtain a sample of durations to test the hypotheses should: permit the assessment of duration from holistic and decomposed subjective estimation; permit the measurement of actual duration; be repetitive; utilize the expertise of construction engineers - read, interpret and visualize construction drawings.

While an activity such as the repetitive construction of a column, a beam or a footing is ideal, the inherent difficulties of field experiments such as: free access to a construction site; measurement of actual duration; and time constraints; makes the selection infeasible. Instead, the assembly of a LEGOLAND wheel loader (model #6658) was selected as the activity for the experiment. In addition to satisfying the requirements of an activity for the experiment, the LEGOLAND model permitted the experiment to be conducted in a laboratory setting.

5.5.2 Procedure

First, the objectives of the experiment were explained to the participants. This explanation was based on the procedure of pre-elicitation discussed in section (3.3). Thereafter, using two questionnaires the desired subjective percentile values for the activity duration were elicited from all the participants. The first questionnaire elicited the duration in minutes to assemble the complete model in accordance with the drawings (*i.e* holistic estimation). The second elicited the duration in seconds to identify and attach one component to the model in accordance with the drawings (*i.e* decomposed estimation). Finally, the actual duration to assemble the model by each participant was measured (see Table 5.1).

The participants were graduate students and final year undergraduates in civil engineering who had followed courses in engineering economics and risk analysis. The subjective elicitation based on the drawings for the LEGOLAND model and its assembly in accordance with drawings utilized their expertise as civil engineers.

Each participant is considered as an independent source for hypotheses testing. That is, evaluated expected values, standard deviations and coefficients of variation from the two approaches for each participant are the basis for a set of hypotheses. The measured actual durations constitute the sample to obtain statistics to test each set of hypotheses.

5.6 Analysis

The expected values, standard deviations and coefficients of variation for duration (holistic - duration to assemble the complete model; decomposed - duration to identify and attach one component to the model) are evaluated from equations (2.5) to (2.11) using the elicited subjective percentile values. However, for hypotheses on decomposed estimation, the moments for duration to assemble the complete model have to be evaluated.

5.6.1 Moments from Decomposition

For a LEGOLAND model consisting of l components, the duration to assemble the complete model is,

$$T_D = \sum_{i=1}^{l} t \tag{5.5}$$

where t is the duration to identify and attach one component to the model. It is assumed that t is identical for all the components.

The expected value for duration to assemble the complete model from decomposed estimation is,

$$E\left[T_D\right] \;=\; l\; E[t] \tag{5.6}$$

where $E[t]$ is the evaluated expected value for duration to identify and attach one component to the model and assumed to be identical for all the components.

Assumption 5.2 : The estimated coefficients of variation for duration for an activity from holistic and decomposed estimation are similar (*i.e* $V_D \;\equiv\; V_W \equiv\; V$).

Assumption (5.2) is tested by hypothesis (5.1). Therefore, from the definition for coefficient of variation,

$$\frac{\sigma_{T_W}}{E\left[T_W\right]} \;\equiv\; \frac{\sigma_{T_D}}{E\left[T_D\right]} \;\equiv\; \frac{\sigma_t}{E[t]} \;\equiv\; V. \tag{5.7}$$

The standard deviation for duration to assemble the complete model depends on the assumption regarding the correlation between duration to assemble individual components. The variance for duration to assemble the complete model is,

$$\mu_2(T_D) \;=\; \sum_{i=1}^{l} \sum_{j=1}^{l} cov(t_i, t_j) \tag{5.8}$$

From definition, $\mu_2(T_D) \;\geq\; 0.$ Hence,

$$\mu_2(T_D) \;=\; l\; \sigma_t^2 \;+\; \rho l(l-1)\; \sigma_t^2 \;\;\geq\;\; 0 \tag{5.9}$$

where σ_t is the evaluated standard deviation for duration to identify and attach one component to the model and ρ is the correlation coefficient between two component durations. Rewriting equation (5.9),

$$\mu_2(T_D) \;=\; \sigma_t^2 \left[l \;+\; \rho(l^2 - l)\right] \;\geq\; 0 \tag{5.10}$$

Since $\sigma_t^2 \geq 0$, for $\mu_2(T_D)$ to exist

$$\left[l \;+\; \rho(l^2 - l)\right] \;>\; 0 \tag{5.11}$$

Therefore,

$$\rho \;>\; \frac{-l}{l^2 - l} \tag{5.12}$$

Hence, $\lim_{l \to \infty} \rho \;=\; 0.$ From definition $\rho \;<\; 1.$

At the extremes, component durations are either uncorrelated or perfect positive correlated.

When the component durations are assumed to be uncorrelated, the variance for duration to assemble the complete model from equation (5.9) is,

$$\mu_2(T_D) \;=\; l\;\sigma_t^2 \tag{5.13}$$

Hence, the standard deviation is,

$$\sigma_{T_D} \;=\; \sqrt{l}\;\sigma_t \tag{5.14}$$

When the component durations are assumed to be perfect positive correlated, the variance for duration to assemble the complete model from equation (5.9) is,

$$\mu_2(T_D) \;=\; l^2\;\sigma_t^2 \tag{5.15}$$

Hence, the standard deviation is,

$$\sigma_{T_D} = l \, \sigma_t \tag{5.16}$$

where σ_t is the evaluated standard deviation for duration to identify and attach one component to the model and assumed to be identical for all the components.

Comparing equations (5.6), (5.7), (5.14) and (5.16) it is evident that relationship given by equation (5.16) evaluates the standard deviation for duration to assemble the complete model from decomposed estimation.

5.6.2 Experimental Results

The actual duration to assemble the LEGOLAND model and the expected values, standard deviations and coefficients of variation for duration from holistic and decomposed estimation are given in Table 5.1. All of the participant are given an identification #. The subjective estimates of participant # 15 were rejected.

The sample mean and standard deviation for the sample of actual measured durations to assemble the LEGOLAND model in minutes are, $\bar{T} = 15.95$ and $S = 7.99$. The sample mean and standard deviation for V_i's, the difference between paired coefficients of variation are, $\bar{V} = 0.0518$ and $S_v = 0.1714$. The test statistic t_{paired} for 27 participants from equation (5.2) is equal to 1.5698.

The t values for assessed expected values and χ^2 values for assessed standard deviations for holistic and decomposed estimation evaluated from equations (5.3) and (5.4) are given in Table 5.2. Table 5.1, Table 5.2, sample means and standard deviations are obtained from a computer program called "LEGO".

Table 5.1: Actual and Estimated Statistics for the Activity Duration (minutes)

	Actual	Holistic Estimation			Decomposed Estimation		
#	Duration	$E[T_W]$	σ_{T_W}	V_W	$E[T_D]$	σ_{T_D}	V_D
01	10.0	12.00	4.02	0.335	15.01	6.32	0.421
02	13.0	3.18	1.41	0.442	10.67	3.75	0.352
03	16.0	28.70	13.94	0.485	24.29	15.84	0.652
04	16.0	15.55	6.33	0.407	24.49	15.65	0.639
05	24.0	17.44	7.03	0.403	5.29	3.03	0.572
06	18.0	30.00	12.56	0.418	16.00	7.50	0.469
07	21.0	11.81	5.91	0.500	11.26	5.57	0.495
08	7.75	5.37	2.57	0.478	6.99	3.58	0.512
09	16.83	29.07	8.89	0.306	52.43	34.00	0.636
10	13.75	20.92	8.89	0.425	11.54	4.03	0.349
11	25.5	6.81	2.53	0.371	15.80	5.37	0.340
12	13.42	6.28	2.05	0.326	5.23	0.95	0.181
13	47.0	30.00	10.05	0.335	16.00	8.43	0.527
14	24.5	14.18	3.96	0.279	10.47	2.96	0.283
15	9.25	–	–	–	–	–	–
16	9.5	10.18	1.78	0.175	39.86	1.96	0.049
17	18.92	10.00	3.77	0.377	10.86	7.91	0.728
18	16.5	25.00	4.27	0.171	32.99	5.54	0.168
19	9.92	19.07	9.56	0.502	5.48	0.96	0.175
20	10.05	10.55	3.12	0.295	6.30	0.68	0.109
21	10.82	3.00	0.75	0.251	4.27	1.07	0.251
22	21.3	30.00	7.54	0.251	32.00	8.04	0.251
23	9.08	9.18	3.53	0.384	7.57	3.75	0.496
24	16.88	18.52	6.14	0.332	9.88	2.90	0.293
25	17.0	5.99	3.65	0.610	13.13	8.31	0.632
26	11.9	12.74	3.19	0.250	26.86	18.01	0.671
27	10.78	10.00	3.02	0.302	5.33	1.61	0.302
28	8.09	9.63	2.81	0.292	3.40	2.03	0.597

Table 5.2: Test Statistics for Expected Values and Standard Deviations

#	Holistic Estimation		Decomposed Estimation	
	t	χ^2	t	χ^2
01	2.6204	106.55	0.6238	43.11
02	8.4610	870.43	3.5038	122.32
03	−8.4479	8.86	−5.5249	6.87
04	0.2649	42.97	−5.6556	7.03
05	−0.9873	34.85	7.0641	187.63
06	−9.3060	10.91	−0.0299	30.58
07	2.7430	49.29	3.1116	55.55
08	7.0133	261.42	5.9386	134.12
09	−8.6931	21.78	−24.8280	1.49
10	−3.2931	21.78	2.9278	106.04
11	6.0559	269.82	0.1008	59.77
12	6.4120	410.52	7.1030	1914.45
13	−9.3060	17.05	−0.0299	24.23
14	1.1727	109.87	3.6346	196.36
16	3.8230	544.55	−15.8399	449.77
17	3.9456	121.24	3.3731	27.51
18	−5.9931	94.39	−11.2849	56.14
19	−2.0640	18.83	6.9395	1864.32
20	3.5778	177.44	6.3962	3676.43
21	8.5836	3030.90	7.7443	1498.43
22	−9.3060	30.31	−10.6311	26.64
23	4.4856	138.40	5.5587	122.15
24	−1.6996	45.67	4.0268	205.07
25	6.6033	129.07	1.8695	24.96
26	2.1301	169.33	−7.2246	5.31
27	3.9456	189.43	7.0376	665.97
28	4.1907	217.61	8.3203	418.09

5.6.3 Hypotheses Testing

All of the hypotheses are tested at confidence level $\alpha = 95\%$. Then, $t_{\frac{\alpha}{2},n-1} = 2.052$, $t_{\alpha,n-1} = 1.703$, $\chi^2_{\frac{\alpha}{2},n-1} = 43.194$, $\chi^2_{\frac{1-\alpha}{2},n-1} = 14.573$, and $\chi^2_{\alpha,n-1} = 40.113$ for $n = 28$. Since t_{paired} is within the acceptance region, hypothesis (5.1) is accepted at 95% confidence level. Hence, the assumption that coefficients of variation for activity duration from holistic and decomposed estimation are similar is verified.

The results of the significance tests for hypotheses (5.2) to (5.9), total and percentages of the number of times an individual hypothesis is rejected at 95% confidence level are given in Table 5.3. 'R' indicates that the hypothesis is rejected, while 'NR' indicates that it is not rejected at 95% confidence level. The significance tests show high rejection rates and similar percentage values for both methods of estimation.

The classical approaches for hypotheses testing and the generally accepted confidence levels used in this study may not be the most suitable to test human ability to predict future events because of the high variability in predictions from individual to individual. While broader confidence levels reduce the rejection rates they may not be acceptable from a statistical view point. This highlights the inherent difficulties in developing experiments to measure human ability to predict future events.

Similar rejection percentages of individual hypothesis confirm the view that neither is the "better" method. Those for hypotheses (5.6) to (5.9) contradict the traditional belief that holistic estimation underestimates duration more regularly than decomposed estimation. If decomposition is not critical to the decision problem - when only work package and project duration estimates are desired; the approach preferred by the analyst and experts can be used for subjective elicitation. However, if decomposition is important to the decision problem - Ravinder et al., (1988); decomposed estimation alone can be used with confidence that the precision of assessments are similar to those from holistic estimation and that it can reduce random

Table 5.3: Significance Tests for Hypotheses (5.2) to (5.9) at 95% confidence level

#	Hypotheses							
	5.2	5.3	5.4	5.5	5.6	5.7	5.8	5.9
01	R	R	NR	NR	R	R	NR	R
02	R	R	R	R	R	R	R	R
03	R	R	R	R	NR	NR	NR	NR
04	NR	NR	R	R	NR	R	NR	NR
05	NR	NR	R	R	NR	NR	R	R
06	R	R	NR	NR	NR	NR	NR	NR
07	R	R	R	R	R	R	R	R
08	R	R	R	R	R	R	R	R
09	R	NR	R	R	NR	NR	NR	NR
10	R	NR	R	R	NR	NR	R	R
11	R	R	NR	R	R	R	NR	R
12	R	R	R	R	R	R	R	R
13	R	NR	NR	NR	NR	NR	NR	NR
14	NR	R	R	R	NR	R	R	R
16	R	R	R	R	R	R	NR	R
17	R	R	R	NR	R	R	R	NR
18	R	R	R	R	NR	R	NR	R
19	R	NR	R	R	NR	NR	R	R
20	R	R	R	R	R	R	R	R
21	R	R	R	R	R	R	R	R
22	R	NR	R	NR	NR	NR	NR	NR
23	R	R	R	R	R	R	R	R
24	R	R	R	R	R	R	R	R
25	R	R	NR	NR	R	R	R	NR
26	R	R	R	R	R	R	NR	NR
27	R	R	R	R	R	R	R	R
28	R	R	R	R	R	R	R	R
Total	24R	20R	22R	21R	16R	19R	16R	18R
	88.89%	74.07%	81.48%	77.77%	59.26%	70.37%	59.26%	66.66%

errors (Ravinder et al., 1988).

5.7 Summary

This chapter described a study on the decomposition of a derived variable that is sometimes estimated holistically. The study consists of a set of hypotheses, test statistics, an experiment and analysis to compare holistic versus decomposed methods for estimating duration when experts participate in subjective elicitation.

While classical approaches and confidence levels used in this study may be too restrictive to test the human ability to predict future events, they provide a statistically accepted framework. The interpretation of the results as percentages of rejection rates reduce some of the restrictions. The first hypothesis verified the assumption that coefficients of variation in subjective assessments for duration from holistic and decomposed estimation are similar. The next four support the view that neither is the "better" estimation approach to elicit subjective assessments for duration. The last four while contradicting the traditional belief in construction about holistic estimation of duration confirm the view regarding the "better" approach. It must be stressed that these are observations based on this study and in no way can they be generalized to the decomposition of derived variables that are sometimes estimated holistically.

The recognition that some of the implicit assumptions and beliefs in engineering construction (assumption 2.3; holistic estimation underestimates duration more regularly) should be explored when dealing with the human ability to predict future events and the inherent difficulties in developing experiments and methods to test such beliefs are some of the benefits of this study. It is recommended that this topic be explored further.

Chapter 6

The Analytical Method

6.1 General

The generation of economic benefits is one of the fundamental objectives of an investment in a project. Hence, the initial decision to invest is governed by the ability of the project to generate a return that would justify the investment. Figure (6.1) shows the generalized cash flow diagram for an engineering project. However, a more simplified cash flow diagram as shown in figure (6.2) is used for the development of the analytical method. In this simplified scenario, the expenditure for design and construction comes from a combination of equity and interim financing. It is assumed that repayment of interim financing is due at the end of the construction period. No attempt is made to include permanent financing in the analytical method because of the numerous financing alternatives available in the market.

The analytical method described herein is developed by applying the framework to quantify the uncertainty of a derived variable to the three levels of the project economic structure as shown in figure (6.3). At the work package/revenue stream level the derived variables are work package duration, start time, cost and net revenue streams. The primary variables at the work package/revenue stream are those variables in the functions specified by the analyst. At the project performance level the derived variables are the project duration, cost and revenue while the primary

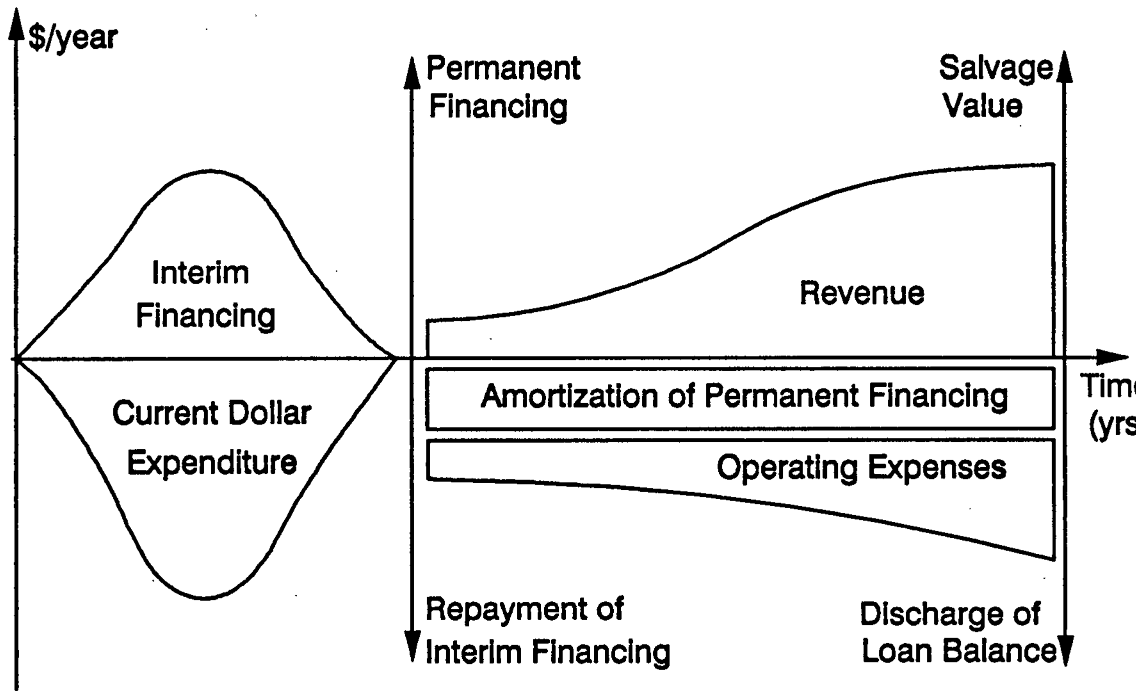

Figure 6.1: Generalized Cash Flow Diagram for an Engineering Project

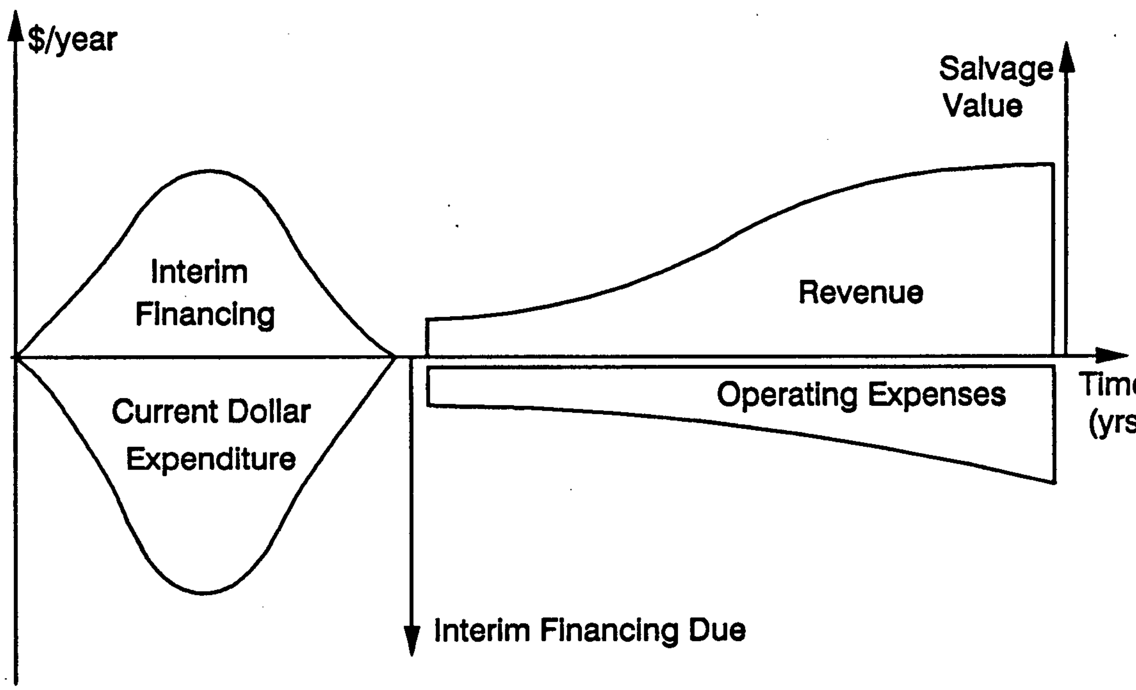

Figure 6.2: Cash Flow Diagram for the Analytical Method

variables are the derived variables at the work package/revenue stream level. At the project decision level the derived variables are project net present value and internal rate of return while the primary variables are discounted project cost and revenue.

This application combines all of the developments and studies from chapters two to five. For generality, the analytical method treats cost and revenue as continuous cash flows under continuous discounting (Tanchoco et al., 1981; Buck, 1989).

6.2 Work Package/Revenue Stream Level

The work package/revenue stream is the first level of application. At this level, the framework is applied as developed, permitting the analyst to use general functional forms for work package durations, costs and revenue streams.

6.2.1 Work Package Duration

Work package duration can be estimated directly as a holistic value or derived using a functional relationship which treats work scope, anticipated job conditions, likely construction methods, productivity and resource levels or a sub-network of activities.

When the estimation is holistic, the analyst/experts provide percentile values for their subjective prior probability distributions and the correlation matrix for work package durations. The first four moments for a work package duration are evaluated from the method described in section (2.3.2) using these percentile values.

When the estimation is decomposed, the analyst must specify the functional forms for work package durations. The analyst/experts provide percentile values for their subjective prior probability distributions and the correlation coefficients for primary variables in the functions for work package durations and identify common (shared)

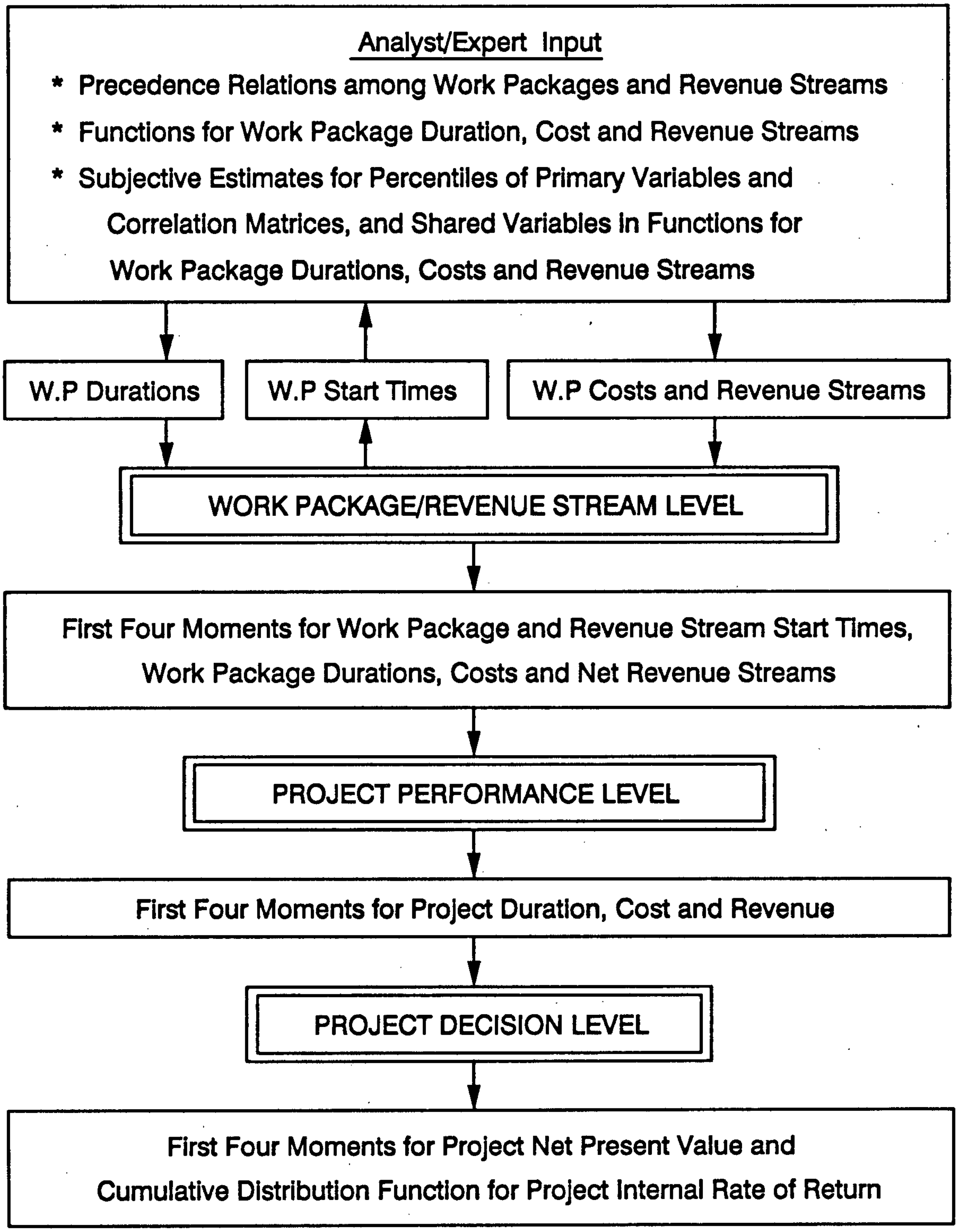

Figure 6.3: Flowchart for the Analytical Method

primary variables among the functions. The correlation matrix for work package durations is evaluated from this information. The specified function for a work package duration is treated as $g(\mathbf{X})$ in equation (2.19). The first four moments for a work package duration are evaluated from equations (2.36) to (2.39) using the elicited positive definite correlation matrix for the variable transformation.

6.2.2 Work Package Start Time

Since start time positions the work package with respect to time it is the variable that links time and cost. Consequently, it is important to have accurate estimates of the moments for start times. In most analytical methods, the start time of a work package is determined by the longest path to that work package.

Let T_i^E be the start time of the i^{th} work package, T_h^E be the start time and T_h be the duration of the preceding h^{th} work package. The start time of the i^{th} work package from the longest path is defined as,

$$T_i^E = max\forall \left[T_h^E + T_h\right] \tag{6.1}$$

where $max\forall$ implies that the maximization is to be over all the links "*h to i*" terminating at the i^{th} work package. While equation (6.1) gives the maximum expected value for the i^{th} work package start time, it does not necessarily evaluate the maximum uncertainty because it ignores shorter but more uncertain (higher variance or skewed) paths. This is the main drawback in using the longest path approach in stochastic network analysis. In theory, an accurate estimate of the moments for start times would involve the analysis of all paths leading to the work packages.

Ang et al., (1975), proposed an analytical technique called "Probabilistic Network Evaluation Technique (PNET)" to evaluate the completion time probability of project duration by considering multiple paths to complete the project. Since project duration is the start time of the finish work package of the precedence network the developments in PNET can be generalized to the work package start time.

PNET

For a project network with a specified number of activities and a set of n possible paths from the start node to the end node, Ang et al., (1975) state the probability of completing the project in time t, denoted $p(t)$ is,

$$\begin{aligned} p(t) &= 1 - [P(T_1 > t) + P(T_1 \leq t, T_2 > t) \\ &\quad + \ldots.. + P(T_1 \leq t, T_2 \leq t, \ldots.., T_{n-1} \leq t, T_n > t)] \end{aligned} \tag{6.2}$$

where $T_1, T_2, \ldots.., T_n$ are the durations of the respective n paths. The bounds on the completion time probability $p(t)$ are (Ang et al., 1975),

$$\prod_{i=1}^{n} P(T_i \leq t) \leq p(t) \leq \min_i P(T_i \leq t) \tag{6.3}$$

When all the paths are assumed to be statistically independent, (*i.e* all possible paths to a node or work package are used to evaluate $p(t)$), the value for $p(t)$ is the most pessimistic (lower bound) and when all the paths are assumed to be perfectly correlated (so that one path is representative of all paths), the value for $p(t)$ is the most optimistic (upper bound). The lower bound of $p(t)$ is the upper bound for duration (see figure 6.5). When all the paths are perfectly correlated, duration is represented by the longest path.

The longest path duration always gives an optimistic estimate for completion time probability (Ang et al., 1975). In other words, the longest path always yields the most optimistic mean duration for work package start time. Since work package cost and revenue stream calculations are linked to start time, longest path based analytical solutions do not adequately estimate the statistics of the derived variables. This is the rationale for a "better" solution from Monte Carlo simulation. On the other hand, if the work package start times are based on the lower bound of $p(t)$, it yields the most pessimistic mean duration. Therefore, when an alternative is evaluated at the bounds of equation (6.3), the resulting solutions are the bounds for the derived variables in the project economic structure.

The start times on which the derived variables should be estimated can be obtained from equation (6.2) if the joint probabilities between the path durations are evaluated. However, the evaluation of joint probabilities for equation (6.2) is complex (Ang et al., 1975). Instead, PNET works around this problem by considering all major paths for estimating $p(t)$ while avoiding the evaluation of joint probabilities. PNET assumes that the activity durations are statistically independent. Also, it is limited to treating $finish\ to\ start = 0$ relationships between activities to evaluate the expected values and variances of individual path durations. Although individual activities are considered to be statistically independent, two different paths are considered to be correlated as a result of common activities. Then, the correlation between two paths i and j having m common activities is defined as (Ang et al., 1975),

$$\rho_{ij} = \frac{\sum_{k=1}^{m} \sigma_{ijk}^2}{\sigma_i \ \sigma_j} \tag{6.4}$$

where σ_{ijk}^2 is the variance of the k^{th} common activity on paths i and j, σ_i and σ_j are the standard deviations for duration of paths i and j and ρ_{ij} is the correlation coefficient between paths i and j.

An approximation for computing $p(t)$ was derived by Ang et al., (1975) from the following observations: (1) paths with long mean durations and high coefficients of variations have the greatest impact on $p(t)$ (defined as major paths); (2) if several paths are each highly correlated with a major path, then those paths can be represented by that major path (upper bound of $p(t)$); (3) if representative paths have low correlations, $p(t)$ can be approximated by the product of the respective path probabilities (lower bound). Consequently, PNET approximates the project completion time probability, $p(t)$ by,

$$p(t) \approx P(T_1 \leq t)\, P(T_2 \leq t) \,.....\, P(T_r \leq t) \tag{6.5}$$

where $P(T_1 \leq t), P(T_2 \leq t), ..., P(T_r \leq t)$ are the probabilities of each representative path completing the project in time t, for r representative paths.

Those paths with $\rho_{ij} \geq \rho$ are represented by path i (the longer path because it has a lower $p(t)$) from the assumption that ρ represents the transition between high and low correlation. When $\rho = 1$, the estimate for $p(t)$ is the lowest (upper bound on duration), whereas when $\rho = 0$, $p(t)$ is the highest (lower bound on duration). If all the major paths are correlated with the longest path, PNET reduces to PERT. In applying PNET, Ang et al. (1975) estimate $p(t)$ from equation (6.5) using a transitional correlation value of $\rho = 0.5$ and assuming the representative path durations to be normally distributed.

Some of the shortcomings of PNET are: (1) by assuming the individual activities (or work packages) to be statistically independent it ignores the correlation brought about by the use of shared resources such as manpower, equipment, management, etc; (2) $p(t)$ is dependent upon the level of interdependence between various paths, *i.e* the selection of the most suitable transitional correlation ρ (Crandall, 1977); (3) a representative path duration may not be normally distributed if a few skewed work package durations dominate the path to a work package or if the work packages appear early in the network.

Modified PNET

The PNET algorithm developed by Ang et al., (1975), is modified to overcome some of the shortcomings in applying it to work package start time. The modifications are: (1) include the linear correlations between work package durations in evaluating the first four moments of path durations; (2) include the shape characteristics (skewness and kurtosis) of representative paths in evaluating the first four moments of the work package start time.

The modified PNET approach to compute the first four moments of a work package start time are as follows. First, the first four moments for duration of each path to a work package are evaluated using equations (6.11) to (6.14), thereby including the linear correlations between work package durations. To facilitate the treatment of correlations between work package durations on a path, only *finish to start* $= 0$ relationships are permitted. Then, in order of decreasing mean path durations all of the individual paths are sequentially ordered. Second, representative paths to a work package are identified as in PNET. Similar to PNET, the transitional correlation ρ must be specified by the analyst. Third, the first four moments of the representative path durations are used to approximate cumulative distribution functions from the Pearson family of distributions. This ensures that shape characteristics of a representative path are not ignored. However, as discussed in section (2.4.5), it may not always be possible to approximate a Pearson type distribution. In such a situation, modified PNET defaults to PNET. Fourth, the cumulative distribution function for start time of a work package is developed by evaluating $p(t)$ from equation (6.5) for a range of durations. The starting duration for the distribution range is obtained from,

$$t_{start} \;=\; E[T_i]_{max} \;-\; 3\;\sigma_{i_{max}} \tag{6.6}$$

where $E[T_i]_{max}$ is the largest expected value from all the path durations (*i.e* the expected value of the longest path) and $\sigma_{i_{max}}$ is the largest standard deviation for all the path durations. If $p(t) > 0$ for the starting duration, then t_{start} in equation (6.6) is reduced until the starting $p(t) = 0$. The duration range for the cumulative distribution function is complete when $p(t) = 1$ is obtained. Finally, given the tableau of values for $p(t)$ versus t, the first four moments for work package start times are evaluated similar to section (2.3.2). In the author's experience, the developed cumulative distribution functions for start times have always approximated to Pearson type distributions. However, the default is the PNET algorithm.

The improvements to the work package start time by applying modified PNET instead of PNET are: (1) since the work package start time is always a primary variable in the functional form for work package cost, the treatment of correlation at two levels, - between work package durations on an individual path and between paths due to common work packages makes the evaluation of first four moments for work package start times, costs and their bounds more precise; (2) considering skewness and kurtosis of the individual paths makes the first four moments for start time of work packages at the beginning of a project more precise, because the number of predecessor work packages on an individual path are too few to invoke the central limit theorem. When there are sufficient predecessor work packages on a path to invoke the central limit theorem (as done by PNET), the approximation of the path durations to the Pearson family of distributions will reflect it because the normal distribution is a member of the Pearson family.

The drawbacks of modified PNET are: (1) it is also dependent upon the level of interdependence between various paths (Crandall, 1977). However, the estimation of upper and lower bounds provides a sensitivity analysis on the transitional correlation specified by the analyst; (2) allows only single (*finish to start* = *0*) logic relationships to sequence work packages. Ability to sequence work packages in overlapping and/or compound relationships will enhance the practicality of the application. However, the treatment of correlation between work packages in overlapping and/or compound relationships on a path or between paths are still theoretically complex. Harris (1978) has shown that overlapping relationships can be transformed to single relationships (*finish to start* = *0*) using one or two additional work packages, and compound relationships using two additional work packages with time-discontinuous assumption.

6.2.3 Work Package Cost

The estimate for expenditure to design and construct a work package is defined as the work package cost. External economic variables have a strong influence on the work package cost estimate. Escalation primarily due to inflation and interest payments for the construction loan (interim financing) are a significant portion of the cost estimate. In estimating the escalation during construction in work package cost, the analytical method allows different rates for different categories of cost.

For the simplification of the derivation, it is assumed that the inflation rates and interest rate for financing of work package cost are constant over the construction period. However, if necessary both of these quantities can be expressed as functions of time. The generalized discounted work package cost is represented by (see figure 6.4),

$$
\begin{aligned}
WPC_i \quad = \quad & f \; e^{(\theta_{C_i} - y) T_{S_{Ci}}} \int_0^{T_{Ci}} C_{0i}(\tau) \; e^{(\theta_{C_i} - y)\tau} \; d\tau \qquad (6.7) \\
& + (1-f) \; e^{(r-y)T_p} \; e^{\theta_{C_i} T_{S_{Ci}}} \int_0^{T_{Ci}} C_{0i}(\tau) \; e^{(\theta_{C_i} - r)\tau} \; d\tau
\end{aligned}
$$

where WPC_i is the discounted i^{th} work package cost, $C_{0i}(\tau)$ is the function for constant dollar cash flow for the i^{th} work package, $T_{S_{Ci}}$ and T_{Ci} are work package start time and duration, T_p is the time at which the repayment of interim financing is due for all work packages, f is the equity fraction, θ_{C_i}, r and y are inflation, interest and discount rates respectively. The time τ is measured from the start of the i^{th} work package. $C_{0i}(\tau)$ can be either holistic or a decomposed function of work scope, resources applied, and productivity.

The estimation of discounted work package cost is always decomposed. The analyst specifies the functional form $C_{0i}(t)$ for equation (6.7). The analyst/experts provide percentile values for their subjective prior probability distributions and the correlation coefficients for primary variables in the functions for discounted work package costs and identify common primary variables among the functions. The correlation matrix for work package costs is evaluated from this information.

The system function $g(\mathbf{X})$ to approximate the first four moments for a discounted work package cost is equation (6.7). The first four moments for a discounted work package cost are evaluated from equations (2.36) to (2.39) using the elicited positive definite correlation matrix for the variable transformation. The bounds for work package costs are obtained when the transitional correlation $\rho = 1$ and $\rho = 0$.

6.2.4 Net Revenue Stream

The possibility of generating a number of revenue streams at different points in time is typical of large engineering projects. Therefore, the ability to study the economic effects of projected revenue with respect to time is essential. The start time of a revenue stream is its link to the precedence network describing the development and operation phases. The analyst must specify the work package and the fraction of that

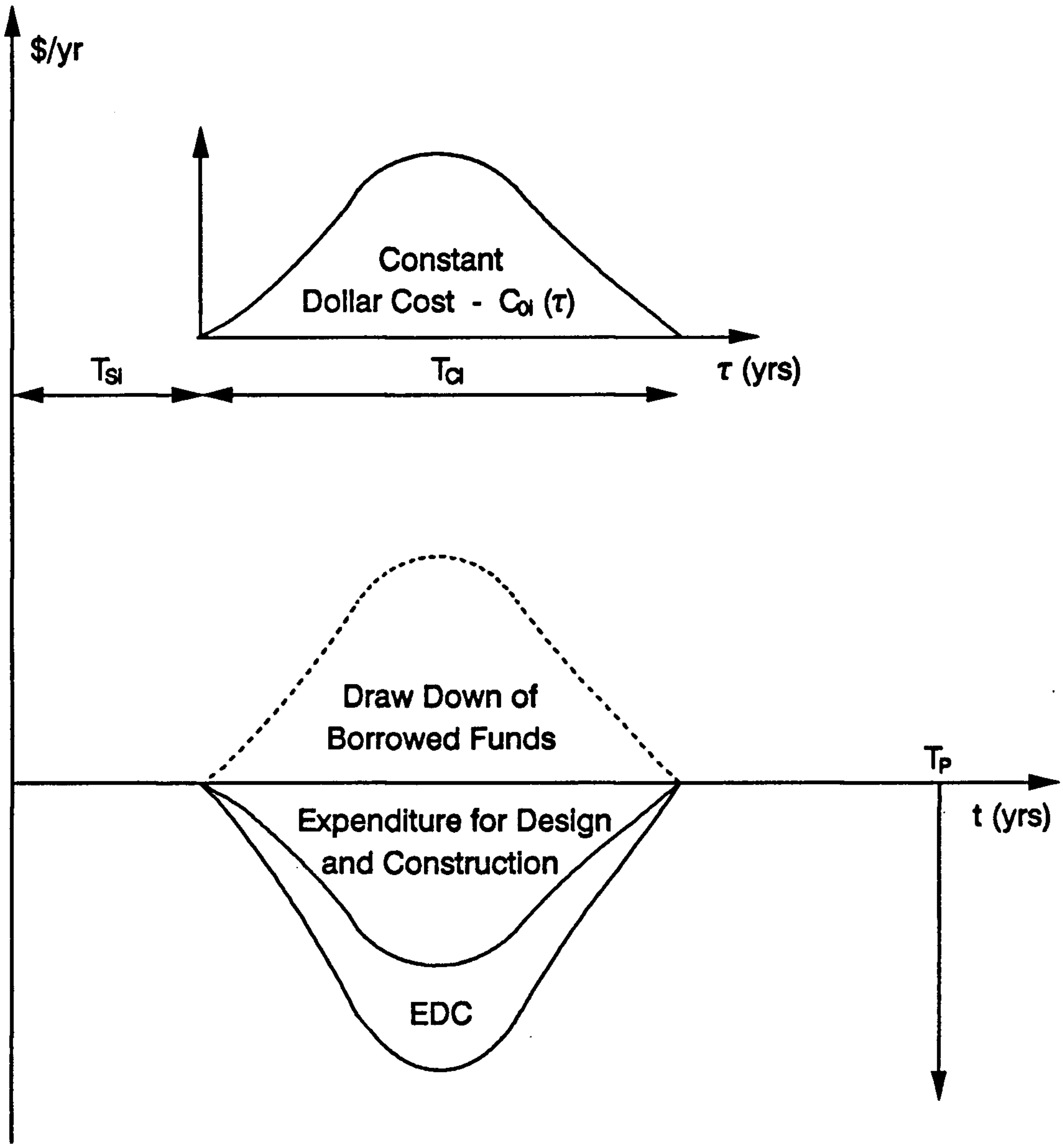

Figure 6.4: Generalized Discounted Work Package Cost

work package duration after which the revenue stream is projected to begin. The start time of the revenue stream is then evaluated from network analysis. To link revenue streams beginning after construction, the operation period is specified as the duration for finish work package. The duration for an individual revenue stream is a primary variable of the function for discounted net revenue.

The net revenue stream is defined as the difference between gross revenue and its operation and maintenance cost. Both, the gross revenue and the operation and maintenance cost are inflated with different rates, and revenues are assumed to inflate once operation starts. The discounted net revenue stream is represented by,

$$NRS_i = \int_{T_{S_{Ri}}}^{T_{S_{Ri}}+T_{Ri}} \left[R_{0i}(t)e^{\theta_{R_i}(t-T_{S_{Ri}})} - M_{0i}(t)e^{\theta_{M_i}t}\right] e^{-yt}dt \qquad (6.8)$$

where NRS_i is the discounted i^{th} net revenue stream, $R_{0i}(t)$ and $M_{0i}(t)$ are the functions for constant dollar cash flow for i^{th} gross revenue and operation and maintenance cost, $T_{S_{Ri}}$ and T_{Ri} are early start time and duration of the revenue stream, $\theta_{R_i}, \theta_{M_i}$ and y are inflation and discount rates respectively.

The estimation for discounted net revenue stream is also decomposed. The analyst specifies $R_{0i}(t)$ and $M_{0i}(t)$ for equation (6.8) as functional forms or holistic constant dollar values. The analyst/experts provide percentile values for their subjective prior probability distributions and the correlation coefficients for primary variables in the functions for discounted net revenue streams and identify common primary variables among the functions. The correlation matrix for net revenue streams is evaluated from this information.

The system function $g(\mathbf{X})$ to approximate the first four moments for a discounted net revenue stream is equation (6.8). The first four moments for a discounted net revenue stream are evaluated from equations (2.36) to (2.39) using the elicited positive definite correlation matrix for the variable transformation. The bounds for net revenue streams are obtained when the transitional correlation $\rho = 1$ and $\rho = 0$.

6.3 Project Performance Level

The functions for all the derived variables at the project performance level are linear additive. The derived variables at this level are project duration, project cost and project revenue, while the primary variables are the derived variables at the work package/revenue stream level.

Assumption 6.1 : There are no non-linear correlations between the transformed variables at the project performance level.

Let Y be a derived variable at the project performance level. Then,

$$Y = g(\mathbf{X}) = \sum_{i=1}^{n} X_i \tag{6.9}$$

where $\mathbf{X}$ is the vector of derived variables from the work package/revenue stream level. Let $\mathbf{Z}$ be the vector of transformed variables at project performance level (from equation 2.24). Since g($\mathbf{X}$) is always linear, the transformed functional form $G(\mathbf{Z})$ at the project performance level from equation (2.35) is,

$$Y = G(\mathbf{Z}) = \sum_{i=1}^{n} \left[\sum_{j=1}^{n} B_{ji} \right] Z_i \tag{6.10}$$

where $\mathbf{B} = \mathbf{D}\ \mathbf{L}$. The expected value of the derived variable Y is,

$$E[Y] = \sum_{i=1}^{n} \left[\sum_{j=1}^{n} B_{ji} \right] E[Z_i] \tag{6.11}$$

the second central moment of Y is,

$$\mu_2(Y) = \sum_{i=1}^{n} \left[\frac{\partial G}{\partial Z_i}\right]^2 \mu_2(Z_i) \tag{6.12}$$

the third central moment of Y is,

$$\mu_3(Y) = \sum_{i=1}^{n} \left[\frac{\partial G}{\partial Z_i}\right]^3 \mu_3(Z_i) \tag{6.13}$$

the fourth central moment of Y is,

$$\begin{aligned} \mu_4(Y) &= \sum_{i=1}^{n} \left[\frac{\partial G}{\partial Z_i}\right]^4 \mu_4(Z_i) \\ &+ 6 \sum_{i=1}^{n} \sum_{j=i+1}^{n} \left[\frac{\partial G}{\partial Z_i}\right]^2 \left[\frac{\partial G}{\partial Z_j}\right]^2 \mu_2(Z_i)\, \mu_2(Z_j) \end{aligned} \tag{6.14}$$

where $\left[\frac{\partial G}{\partial Z_i}\right] = \sum_{j=1}^{n} B_{ji}$; and $E[Z_i]$, $\mu_2(Z_i)$, $\mu_3(Z_i)$, $\mu_4(Z_i)$ are the first four moments of the i^{th} transformed uncorrelated variable.

The first two moments of the derived variable are exact with or without assumption (6.1) because the transformed function $G(\mathbf{Z})$ is linear. With assumption (6.1), the third and fourth moments are also exact. The correlations between primary variables at the project performance level are linear because correlations for derived variables approximated from section (4.3) are always linear. These correlations are included in the moments for $\mathbf{Z}$, and therefore in the first four moments for the derived variable.

Even if there are no non-linear correlations among the primary variables, it is not possible to conclude that the transformed variables are free of non-linear correlations.

Hence, third and fourth moments will be in error only if non-linear correlations develop between the transformed variables. Since the measurement and treatment of non-linear correlations are still theoretically complex this assumption is reasonable. In addition, it permits the computation of exact first four moments for a derived variable at the project performance level.

6.3.1 Project Duration

The project duration is the start time of the finish work package of the precedence network. The first four moments of project duration are obtained from the modified PNET algorithm. The upper bound for project duration is computed when the transitional correlation $\rho = 1$ while the lower bound is when $\rho = 0$ (see figure 6.5).

6.3.2 Project Cost

The project cost is the summation of all the work package costs. When there are n work packages in the construction project, the discounted project cost is given by,

$$DPC = \sum_{i=1}^{n} WPC_i \tag{6.15}$$

where WPC_i is the discounted i^{th} work package cost from equation (6.7). The function $g(\mathbf{X})$ for discounted project cost is equation (6.15). The first four moments for discounted project cost are computed from equations (6.11) to (6.14).

The project cost is expressed in discounted dollars for generality. When required the project cost can be expressed in total, current or constant dollars. The bounds for project cost in total, current or constant dollars are obtained when the transitional correlation $\rho = 1$ and $\rho = 0$. A typical example for upper and lower bounds of project cost in current dollars is depicted in figure (6.6).

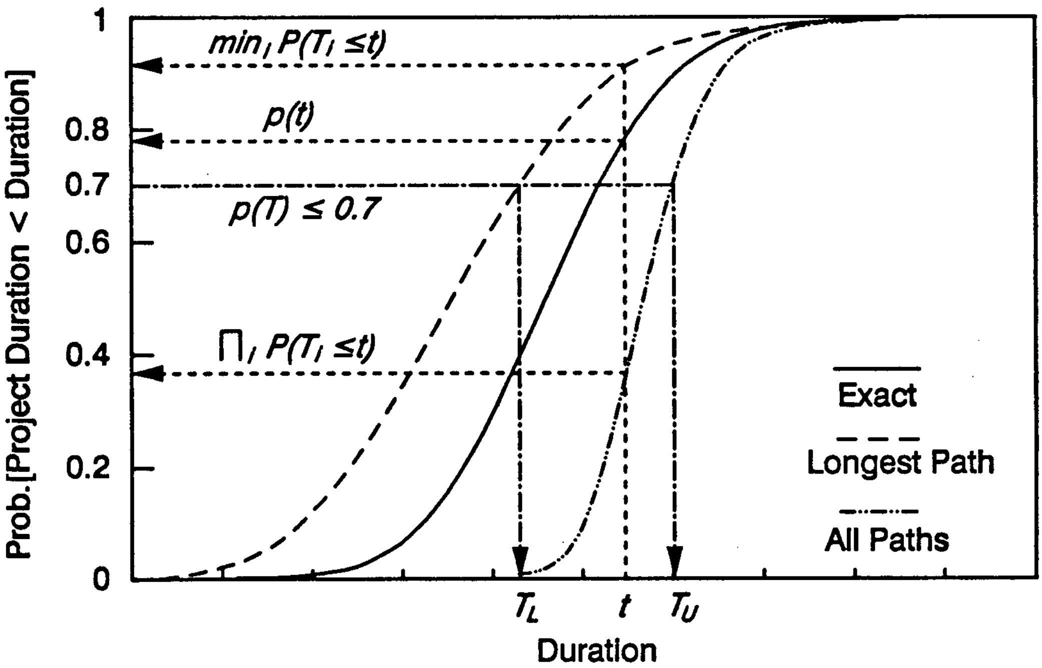

Figure 6.5: Upper and Lower Bounds for Project Duration

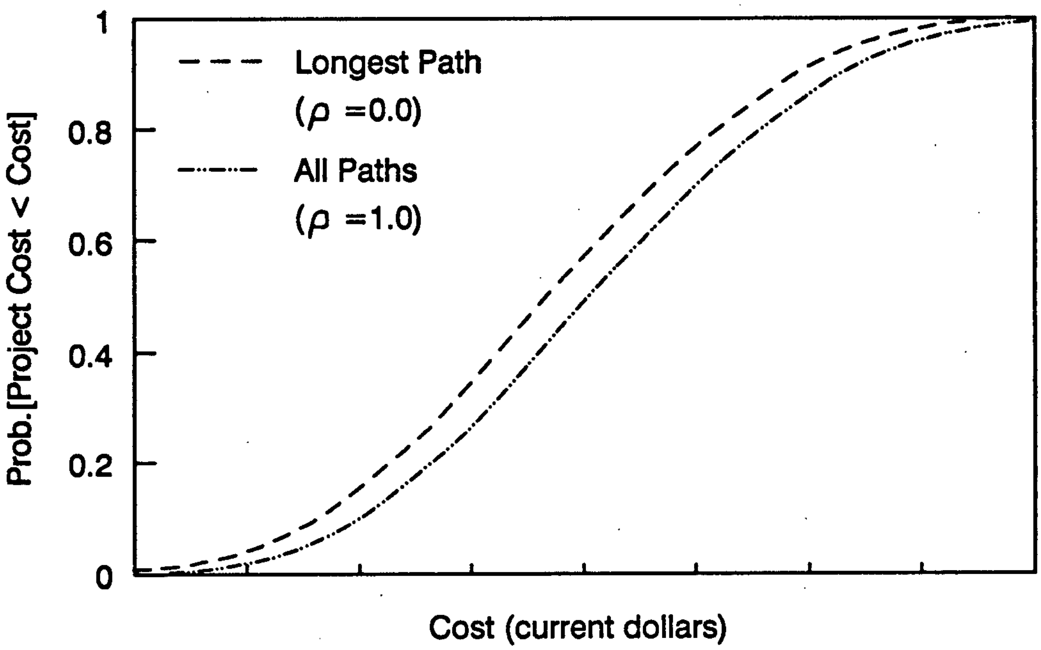

Figure 6.6: Upper and Lower Bounds for Project Cost

6.3.3 Project Revenue

The project revenue is the summation of all the revenue streams. When there are m revenue streams in the construction project, the discounted project revenue is given by,

$$DPR = \sum_{i=1}^{m} NRS_i \tag{6.16}$$

where NRS_i is the discounted i^{th} net revenue stream from equation (6.8). The function $g(\mathbf{X})$ for discounted project revenue is equation (6.16). The first four moments for discounted project revenue are computed from equations (6.11) to (6.14).

6.4 Project Decision Level

The project decision level is the top of the hierarchy of the project economic structure. The derived variables at this level, project net present value and internal rate of return are the decision criteria for an investment. To quantify their uncertainty, the analytical method exploits the fact that the functions for these derived variables are the same for all engineering projects.

6.4.1 Project Net Present Value

The net present value of a project is the difference between the project revenue and the project cost discounted at minimum attractive rate of return. The first four moments for project net present value are computed by assuming discounted project cost and discounted project revenue to be independent. Then the first four moments of project net present value are,

$$E[NPV] = E[DPR] - E[DPC] \tag{6.17}$$

$$\mu_2(NPV) = \mu_2(DPR) + \mu_2(DPC) \tag{6.18}$$

$$\mu_3(NPV) = \mu_3(DPR) - \mu_3(DPC) \tag{6.19}$$

$$\mu_4(NPV) = \mu_4(DPR) + \mu_4(DPC) + 6\,\mu_2(DPR)\,\mu_2(DPC) \tag{6.20}$$

where DPR and DPC are discounted project revenue and discounted project cost respectively.

6.4.2 Project Internal Rate of Return

The internal rate of return of a project is the discount rate at which the discounted project revenue is equal to the discounted project cost. In other words, the discount rate at which the project net present value is zero. The internal rate of return is an implicit function of the net present value and therefore does not provide a direct functional form to apply the framework.

Hillier (1963) proposed a method to develop the cumulative distribution function for internal rate of return utilizing its definition. A number of authors have since discussed this method for applications (Bonini, 1975; Davidson and Cooper, 1976; Wagle, 1967; Zinn et al., 1977). The analytical method develops the expected value, standard deviation and cumulative distribution function for internal rate of return by using a variation of the method suggested by Hillier (1963).

Initially, first four moments for net present value at a discount rate, $r = 0.01$, are evaluated. Using these first four moments a Pearson type distribution is approximated for net present value. (The author's experience is that it is always possible to approximate a Pearson distribution for net present value because the first four moments for net present value, discounted project revenue and cost are exact. However, the default is Hillier's approach.) From this distribution the probability for $NPV \leq 0|r$ is obtained. This is the probability that $IRR \leq r$. Summarizing in equation form (equation 9 from Hillier, 1963),

$$P(IRR \leq r) = P(NPV \leq 0|r) \tag{6.21}$$

The cumulative distribution function for internal rate of return is developed, by repeating the above process while incrementing the discount rate by 0.01, until the range $0 < P(IRR \leq r) < 1$ is obtained from equation (6.21). Then using the 2.5%, 5%, 50%, 95% and 97.5% values of the developed cumulative distribution function, the expected value and standard deviation for internal rate of return are computed from equations (2.5) to (2.11).

Hillier (1963) approximated the cumulative distribution functions for net present value to the normal distribution to develop the cumulative distribution function for internal rate of return. The cumulative distribution function for internal rate of return was also approximated to the normal distribution to obtain the expected value and the standard deviation for internal rate of return. Inyang (1983) showed that the assumption of normality made by Hillier (1963),(1969) and Wagle (1967) is in error because skewness develops for situations where input variables are skewed; response of the decision criterion to changes in input variables are non-linear; input variables are insufficient; discontinuity in cash flow occurs (staged construction).

The analytical method utilizes the first four moments for net present values at different discount rates to approximate Pearson type distributions in developing the cumulative distribution function for internal rate of return, thus allowing for the treatment of skewness. Also, since equations (2.5) to (2.11) are used to compute the expected value and standard deviation for internal rate of return, there is no necessity to approximate the developed cumulative distribution function to a normal distribution.

The upper and lower bounds for the project net present value at minimum attractive rate of return and the project internal rate of return are computed when the transitional correlation $\rho = 1$ and $\rho = 0$ respectively. Typical examples of bounds for the project net present value and the project internal rate of return are depicted in figures (6.7) and (6.8).

6.5 Discussion

Cooper and Chapman (1987) state that four moment methods (those using the first four moments of primary variables to calculate the first four moments of a derived variable) for risk analysis achieve a large increase in generality over two moment methods (mean and variance) and are more versatile because: they allow primary variable distributions to be quite general; the moments are related to the distributions shape characteristics; computational requirements are modest. However, they question the computational accuracy of the four moment methods because: calculations for the first four moments of a derived variable require the central moments of primary variables higher than the fourth order, which can be numerically significant; and the restrictions generally imposed on the possible forms of interdependence relationships between primary variables.

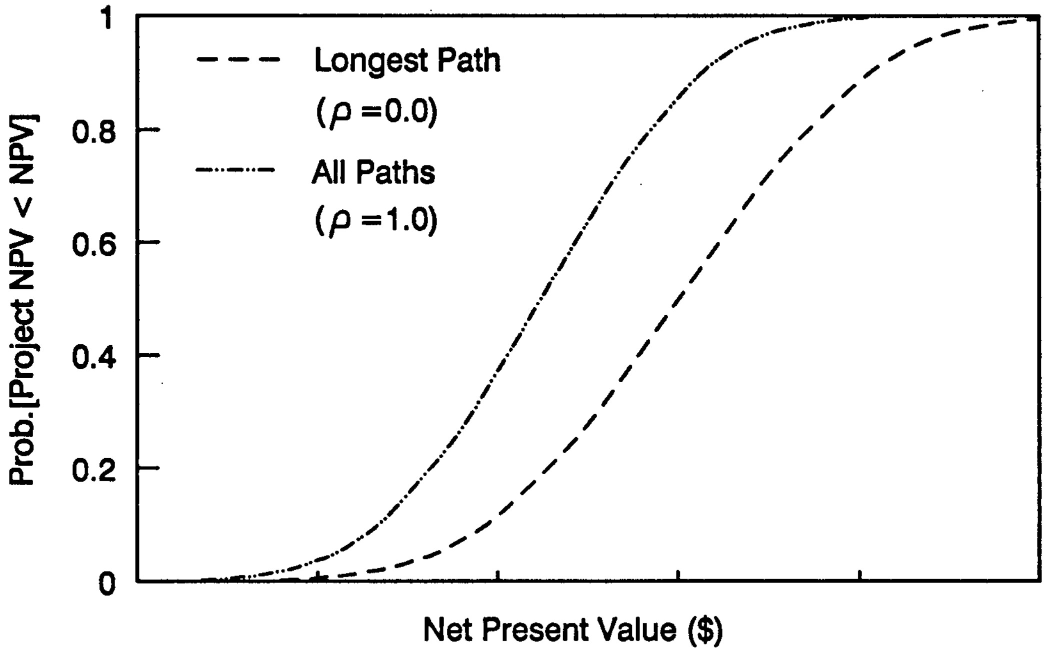

Figure 6.7: Bounds for the Project Net Present Value

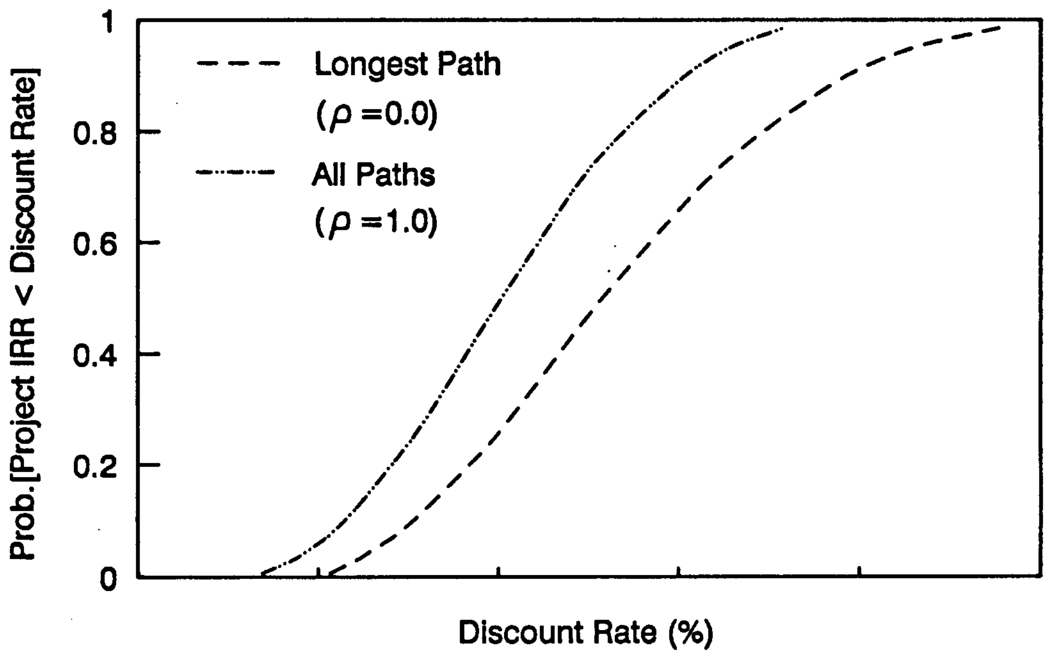

Figure 6.8: Bounds for the Project Internal Rate of Return

This section will discuss how some of the issues raised by Cooper and Chapman (1987) affect the analytical method, and what can be done to increase computational accuracy where possible. In addition, this analytical solution is compared to that which can be obtained from the currently available moment analysis approach (standard approach) to show the improvement of the derivation.

6.5.1 Computational Accuracy

The first four moments for derived variables at project performance and decision levels computed by the analytical method are exact because of the linear functional forms. Therefore, at these two levels only the first four moments of primary variables are required. However, at work package/revenue stream level the issue raised by Cooper and Chapman (1987) regarding higher order moments are valid because general functional forms are permitted for derived variables.

The second, third and fourth central moments for the derived variables require up to fourth, sixth and eighth order moments of primary variables. Since the framework considers moments up to the fourth order, the approximation for the second central moment has considered the necessary central moments of primary variables. As all the primary variables are approximated to Pearson type distributions it is possible to generate moments up to the eighth order from the recurrence property of the Pearson family (Kendall and Stuart, 1969 - see Appendix A.6). Then the approximations for third and fourth central moments for a derived variable can consider the necessary central moments of primary variables. However, until more practical experience is gained in the elicitation of subjective probabilities from experts, it is prudent to use only the first four moments for primary variables. With experience, higher order moments of primary variables can be included in the approximations for third and fourth central moments of a derived variable.

The question whether the fifth and higher order central moments of primary variables are numerically significant in the approximations for the third and fourth central moments is neither proved nor disproved in the literature, possibly because of the difficulty of the exercise. After a rigorous theoretical study, Tukey (1954) concluded that the approximations for first four moments of a derived variable are much better than seems to be usually realized. His study used terms up to the fifth order.

When generalized four moment methods are suggested for risk analysis, primary variables are assumed to be statistically independent (Siddall, 1972; Jackson, 1982). The variable transformation approach used by the analytical method treats linear correlations at all levels of the project economic structure in a consistent manner. The concern raised by Cooper and Chapman (1987) regarding treating interdependencies between primary variables is overcome to the extent that the analytical method treats the correlation information that is generally available during feasibility analysis.

6.5.2 Standard Approach

In the fourth chapter, the variable transformation method was compared numerically to the standard approach to show that it treats correlation information more consistently at the work package/revenue stream level. Similarly, when the solution for derived variables at the project performance level using standard approach is compared, it is evident that the analytical method using variable transformation treats linear correlations accurately and consistently.

The correlations between primary variables at project performance level (*i.e* derived variables at work package/revenue stream) are restricted to linear correlations from section (4.3). Assuming that there are no non-linear correlations between primary variables at project performance level and using equation (6.9) as $g(\mathbf{X})$, the first four moments for a derived variable from the standard approach are,

$$E[Y] = \sum_{i=1}^{n} E[X_i] \tag{6.22}$$

$$\mu_2(Y) = \sum_{i=1}^{n} \mu_2(X_i) + 2 \sum_{i=1}^{n} \sum_{j=i+1}^{n} cov(X_i, X_j) \tag{6.23}$$

$$\mu_3(Y) \approx \sum_{i=1}^{n} \mu_3(X_i) \tag{6.24}$$

$$\mu_4(Y) \approx \sum_{i=1}^{n} \mu_4(X_i) + 6 \sum_{i=1}^{n} \sum_{j=i+1}^{n} \mu_2(X_i)\, \mu_2(X_j) \tag{6.25}$$

where $E[X_i]$, $\mu_2(X_i)$, $\mu_3(X_i)$, $\mu_4(X_i)$ are the first four moments of the i^{th} primary variable at project performance level.

Consider an engineering project consisting of five work packages, with expected values, standard deviations and shape characteristics for work package costs as shown in Table 6.1. The correlation matrix for work package costs is $\mathbf{R_{wpc}}$. Table 6.2 shows the first four moments and shape characteristics for project cost computed by the analytical method (equations 6.11 to 6.14), standard approach (equations 6.22 to 6.25), and when work package costs are assumed to be statistically independent (Siddall, 1972).

Table 6.1: Statistics for Work Package Costs

W.P #	E[WPC]	σ_{wpc}	$\sqrt{\beta_1}$	β_2
01	107.40	43.67	0.5	2.2
02	194.82	22.92	−0.8	2.8
03	305.55	28.32	0.6	2.4
04	411.10	50.78	0.7	2.5
05	492.60	37.76	−0.6	2.4

$$\mathbf{R_{wpc}} = \begin{bmatrix} 1.00 & 0.41 & 0.58 & 0.67 & 0.51 \\ 0.41 & 1.00 & 0.28 & 0.48 & 0.39 \\ 0.58 & 0.28 & 1.00 & 0.61 & 0.60 \\ 0.67 & 0.48 & 0.61 & 1.00 & 0.48 \\ 0.51 & 0.39 & 0.60 & 0.48 & 1.00 \end{bmatrix}$$

Table 6.2: First Four Moments and Shape Characteristics for Project Cost

Moments	Analytical Method	Standard Approach	Statistically Independent
$E[PC]$	1511.47	1511.47	1511.47
$\mu_2(PC)$	21184.46	21184.46	7239.68
$\mu_3(PC)$	1018621.	105012.	105012.
$\mu_4(PC)$	1202791440.	149342432.	149342432.
$\sqrt{\beta_1}$	0.3303	0.0341	0.1705
β_2	2.6801	0.3328	2.8493

The expected value (equations 6.11 and 6.22) and second central moment (equations 6.12 and 6.23) for Y are identical, indicating that linear correlation is treated accurately by the analytical method because equations (6.22) and (6.23) are exact when $g(\mathbf{X})$ is linear (Kendall and Stuart, 1969). Since third and fourth central moments for Y from the standard approach do not contain any linear correlation terms, they are same as when the primary variables are assumed to be statistically independent (Siddall, 1972). Where as, the third and fourth central moments computed from the analytical method contain the linear correlations because the variable transformation ensures that they are included in equations (6.13) and (6.14).

When primary variables are statistically independent and the number of variables is large, from the central limit theorem the derived variable should approach normality. Even with five work package costs the shape characteristics for project cost for the independent case are close to a normal distribution. When shape characteristics for project cost from the analytical method and standard approach are compared, those from the standard approach do not reflect the skewness of the work package costs, and the kurtosis is in the impossible range for a distribution. Those from the analytical method reflects skewness and kurtosis because it has included the linear correlation between the work package costs.

6.6 Summary

This chapter combined all of the developments and studies done in the previous chapters with the project economic structure to propose an analytical method for time and economic risk quantification during feasibility analysis for large engineering projects. The method computes the first four moments of derived variables at work package/revenue stream level (work package duration, cost and net revenue), project performance level (project duration, cost and revenue) and project decision level (net present value) using the moments of primary variables in their functional forms. The shape characteristics of the derived variables are used to approximate Pearson type distributions for them to quantify their uncertainty. The bounds for derived variables are obtained when transitional correlation $\rho = 1$ and $\rho = 0$.

The computed moments for derived variables at project decision and project performance level are exact. The approximations for moments are only for the derived variables at the work package/revenue stream level. The expected value, standard deviation and cumulative distribution function for project duration are obtained from modified PNET while those for project internal rate of return are obtained from a variation of Hillier's method. The concerns raised by Cooper and Chapman (1987) regarding the computational accuracy of the four moment method and treatment of interdependence between primary variables have been discussed with suggestions for further improvements.

One of the objectives of this research is to computerize the analytical method to explore its behavior, to validate it and to test its practicality in the measurement of uncertainty of performance and decision parameters. The source code for the analytical method is available in a file called TIERA (TIme and Economic Risk Analysis). It has been developed as a generalized numerical processor that has the flexibility to model general functional forms for work package durations, costs and revenue streams. See Appendix D for more details.

The developed method, while providing a consistent analytical approximation to a problem that has long relied on Monte Carlo simulations for solutions, shows that it is more appropriate for time and economic risk quantification of large engineering projects. It includes the features of a good simulation model such as: interaction of time, cost, and revenue by using a precedence network; performing sensitivity and probability analysis; treating multiple paths in network analysis; treating correlation between variables at the input level; and the quantification of risks of decision variables by developing cumulative distribution functions. In addition, it overcomes most of the constraints that exist during feasibility stage for realistic modeling of an engineering project by: requiring expert judgements as input; treating correlation between primary variables and between derived variables at all levels; obtaining intermediate milestone information necessary to set realistic targets for performance; permitting the use of unlimited number of variables to model a project; estimating bounds for decision variables; and above all having the capability to evaluate a range of alternatives economically to select the most suitable strategy to develop a project.

Chapter 7

Validations and Applications

7.1 General

The analytical method to estimate bounds on and to quantify the uncertainty in time and economic risks for large engineering projects was developed in the previous chapter. This chapter describes validation and applications of the analytical method. In most of the examples presented in this chapter it is difficult to separate the validation studies from the applications. Therefore, it will be helpful to the reader if the results from the analytical method are viewed as applications and those from Monte Carlo simulations are viewed as validations.

Monte Carlo simulations are used to validate the analytical method because at present, simulation based models are considered to be the "state-of-the-art" for quantification of time and economic risks in large engineering projects (Cain, 1980; Diekmann, 1985; Flanagan et al., 1987; Hayes et al., 1986; Jaafari, 1988a; Newendorp, 1976; Perry and Hayes, 1985b; Thompson and Wilmer, 1985). When the variables are uncorrelated, a successful validation should demonstrate that given the same problem structure, primary variables and probability distributions, the quantified uncertainty of time and economic variables from the simulation lie within the upper and lower bounds approximated from the analytical method.

Since, the analytical method treats correlations efficiently, correlations must be treated in the simulation process to permit comparisons for validations. The treatment of correlations in Monte Carlo simulations is a non-trivial task (Johnson, 1987). Even though a number of methods have been suggested for treating correlations in simulations, no method has been validated rigorously (*eg.* compared to known analytical solutions) to be considered as a bench mark for these validations. Nevertheless, a method which the author considers as the best approximation for treating correlations between variables in simulations is adopted. However, rigor in the validations similar to that of the uncorrelated situations cannot be achieved. The next section contains a brief description on Monte Carlo simulation, the theoretical basis for the method used to include correlations between primary variables, and the "acceptable" number of iterations for the simulation.

In the third section, the modified PNET algorithm is applied to the two numerical examples presented by Ang et al. (1975). The first is a road pavement project, while the second is an industrial building project. The applications show that the modified PNET algorithm which is based on the precedence network reproduces the results obtained by Ang et al. (1975) using the arrow network, thereby validating the modified algorithm. The flowchart for the modified PNET algorithm is illustrated in Appendix D.

Sections four to six describe the validation studies that were performed. In the fourth, a parallel network of identical work packages is used to validate the simulation process. This is the first of the two limiting cases that are used to validate the Monte Carlo simulation process. The fifth section uses data from an actual deterministic feasibility analysis as the first example to validate the analytical method. The first example contains the second limiting case for the Monte Carlo simulation and four simulations to validate the analytical method. The second limiting case is a single dominant path of a highly interrelated precedence network. In the first two simulations, low coefficients of variation for work package durations are used. In addition to the validation, this permits a realistic comparison with the deterministic study. The third and fourth simulations use the same numerical example with high coefficients of variations for work package durations. Since derived economic variables are dependent upon the start times, this increase permits the study of the effect of high variance on the quantification and bounding of their risks.

In the sixth section the second example that is used for the validation is presented. It is a hypothetical engineering project developed to demonstrate the full potential of the analytical method. Two complete simulation were performed. The first assumed that all the primary variables are uncorrelated, while the second assumed that the primary variables at the input level are correlated. This is the correlation treatment that can be duplicated by simulation. The example is extended to a third level where correlations at all levels of the project economic structure are treated.

In the seventh section, the different ways in which the analytical method can perform sensitivity analysis are explored. This discussion outlines how one of the sensitivity analyses can be used to distribute the contingency allocated to a derived variable at a desired probability of success, to its primary variables. The current dollar estimate for project cost is used as the derived variable.

7.2 Monte Carlo Simulation

Conceptually, performing a Monte Carlo simulation is simple. It requires a deterministic model, identification of the random variables, a probability distribution for each random variable, a random number generator, and then a sample value from each distribution for each iteration using a random number from the uniform distribution on the interval [0.1], (*i.e* $U(0,1)$), as the entry point in a cumulative distribution function of the variables (see figure 7.1). The larger the number of iterations, the more reliable are the results from the simulation (Cain, 1980; Eilon and Fowkes, 1973; Flanagan et al., 1987; Inyang, 1983; Jaafari, 1988a; Johnson, 1987; Hertz, 1964; Hull, 1977, 1980; Kalos and Whitlock, 1986; Kryzanowski et al., 1972; Newendorp, 1976; Riggs, 1989; Van Tetterode, 1971). The procedure described above however, implies that each random variable is independent of the others. In the current problem most variables are dependent (Cooper and Chapman, 1987; Inyang, 1983; Perry and Hayes, 1985b).

7.2.1 Treatment of Correlations

The importance of treating correlations between variables in Monte Carlo simulations has been long recognized (Eilon and Fowkes, 1973; Inyang, 1983; Hertz, 1964; Hull, 1977, 1980; Kryzanowski et al., 1972; Newendorp, 1976; Thompson and Wilmer, 1986; Van Tetterode, 1971). None of the suggested methods however, has been rigorously validated. After an extensive review of the available techniques, Inyang (1983) proposed the following approach to model correlations in Monte Carlo simulations for risk analysis of engineering projects.

1. Random numbers are generated for each of the variables that make up the risk analysis model. A column of random numbers is thus generated.

2. The correlation factors between variables have to be input as a matrix. The random numbers are modified depending on the correlation with each other. Any type of correlation factor (total, partial or no correlation) can be handled.

3. The value of a variable is obtained depending on the value of its modified random number as a result of the correlation between the variables.

The author agrees with Inyang (1983), that the above procedure is the most suitable approach to model correlations in simulations. However, two shortcomings have to be highlighted. First, the algorithm used for the modification of random numbers was not derived in the thesis by Inyang (1983). Second, the correlation matrices elicited for the simulation have to be positive definite (see section 4.2). The process that was used for the validation model is based on the above procedure (Inyang, 1983). However, since it is not known whether the method for treating correlations in the simulation overestimates or underestimates the effects of correlation, the simulation results provide only an approximate bench mark for the analytical treatment of correlation. The possibility thus exists that the simulation results may not be contained within the upper and lower bounds predicted by the analytical method.

The random numbers were modified by extending the algorithm developed by Van Tetterode (1971) to the multivariate situation. The random number correction is pairwise. Since the positive definite correlation matrix is used to modify the random numbers assigned to the primary variables, the multivariate situation is recognized. The random number correction is as follows (Van Tetterode, 1971).

$$RN_{ij} = RN_j + \alpha_{ij} (RN_i - RN_j) \tag{7.1}$$

where RN_i and RN_j are the i^{th} and j^{th} random numbers in the column generated from $U(0,1)$, (step 1, Inyang, 1983), RN_{ij} is the ij^{th} random number corrected for the correlation between variables i and j in the matrix of corrected random numbers, and α_{ij} is the correction factor given by,

$$\alpha_{ij} = \frac{\rho_{ij}^2 \pm \rho_{ij} \sqrt{1 - \rho_{ij}^2}}{2\,\rho_{ij}^2 - 1} \tag{7.2}$$

where ρ_{ij} is the correlation coefficient between variables i and j. The correction factor α_{ij} is lies in the interval, $0 \leq \alpha_{ij} \leq 1$ (see figure 7.2 and Appendix E for proof) for all correlation values. The modification from equation (7.1) and (7.2) ensures that the corrected random numbers are within the interval [0,1].

A small numerical example is presented to demonstrate the random number modification process. Assume a three variable model having the following correlation matrix, **R**, where

$$\mathbf{R} = \begin{bmatrix} 1.0 & -0.48 & 0.42 \\ -0.48 & 1.0 & -0.69 \\ 0.42 & -0.69 & 1.0 \end{bmatrix}$$

and the column of random numbers generated from $U(0,1)$ as $[\ 0.32\ 0.75\ 0.14]^T$. Then, the matrix of α_{ij} values from equation (7.2) and the matrix of random numbers corrected for the correlation values from equation (7.1) are given below.

$$\alpha = \begin{bmatrix} 1.0 & 0.35365 & 0.31638 \\ 0.35365 & 1.0 & 0.48804 \\ 0.31638 & 0.48804 & 1.0 \end{bmatrix} \quad ; \quad \mathbf{RN} = \begin{bmatrix} 0.32 & 0.5979 & 0.1969 \\ 0.4721 & 0.75 & 0.4377 \\ 0.2631 & 0.4523 & 0.14 \end{bmatrix}$$

For each iteration, matrix **RN** is computed from the generated column of random numbers. Then, a row selected from the matrix **RN** at random can be used as the random numbers for that iteration of the simulation.

7.2.2 The Number of Iterations

The literature is diverse on the number of iterations that should be performed for an "acceptable" simulation (Flanagan et al., 1987; Jaafari, 1988; Inyang 1983; Perry and Hayes, 1985b). The recommended numbers range from 100 to 1000 iterations. However, most of these recommendations are not supported theoretically or empirically, and may not be applicable in all situations (Inyang, 1983).

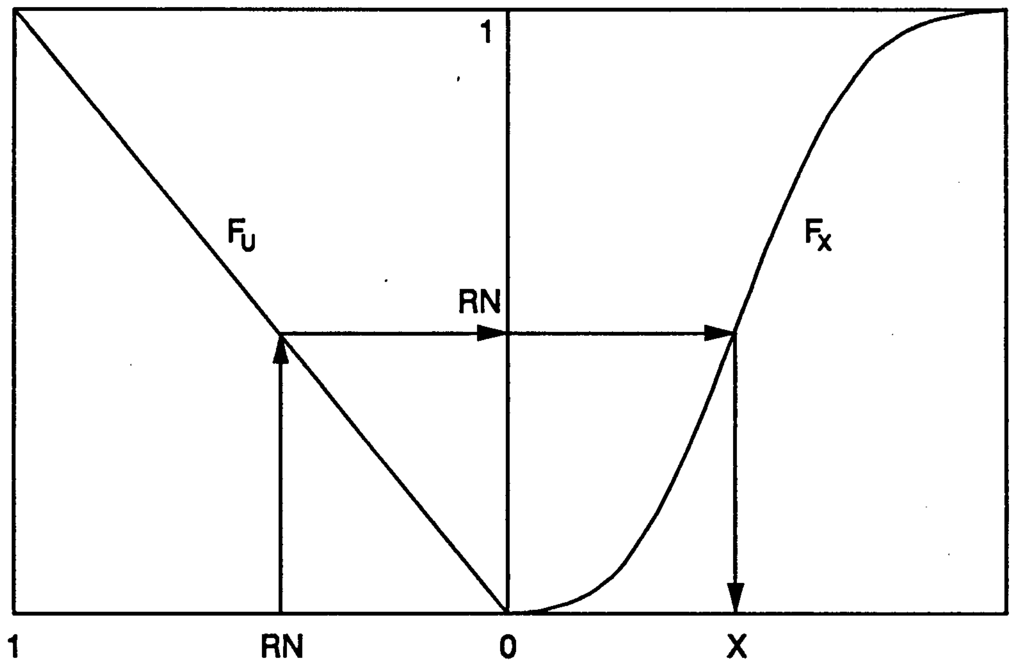

Figure 7.1: Random Variate Generation

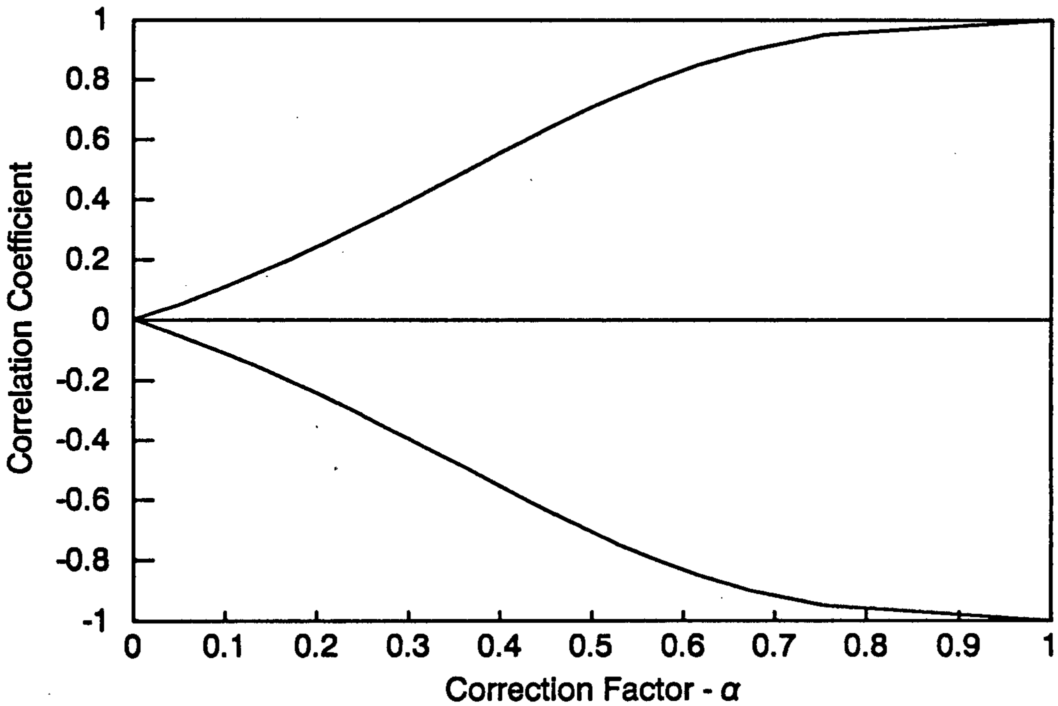

Figure 7.2: The Correction Factor α for Different Values of ρ

Bury (1975) has shown that a simulation of 1000 iterations has an error band of 4.3% at 95% confidence level. Error band is the accuracy to which the cumulative distribution function generated from the simulation approximates to the unknown cumulative distribution function of the derived variable. That is, the error band brackets the unknown cumulative distribution function in 95% (or $(1-\alpha)$ 100%) of all simulation samples. At 95% confidence level, for an error band of 2% at least 4600 iterations are required. Inyang (1983) states that at 95% probability, 1000 iterations will give a level of accuracy of 6% and 8.5% for the expected value (mean) and the standard deviation respectively.

Since simulation generates a random sample to represent the derived variable, irrespective of the number of primary variables, the larger the size of the sample the more accurate are the estimates for the expected value, standard deviation and the cumulative distribution function generated from simulation. In this thesis, when duration was the only derived variable 15,000 to 20,000 iterations were used for the simulation. For complete time and economic risk quantification 4,000 to 6,000 iterations were used. Larger simulations were used for duration because of its smaller problem structure and because of its importance as the linking variable in economic risk quantification. The comparatively large size of the simulations also permits the study of the stability of the expected value and standard deviation with increasing number of iterations.

7.3 Modified PNET Algorithm

The modified PNET algorithm is applied to the two numerical examples that were presented by Ang et al., (1975). The first example is a road pavement project, while the second is an industrial building project.

7.3.1 Road Pavement Project

This project involves the paving of 2.2 miles of roadway pavement and the construction of appurtenant drainage structures, excavation to grade, placement of macadam shoulders, erection of guardrails, and landscaping (Ang et al., 1975). The precedence network for the project used by the modified PNET, based on the logic of the arrow network given by Fig.2 of Ang et al. (1975) is shown in figure (7.3). The various activities of the project, respective mean durations and standard deviations for the activities from Table 1 of Ang et al. (1975), are given in Appendix F.

Table 2 from Ang et al., (1975), containing all nine paths of the network arranged in order of decreasing mean path durations, mean path durations (μ_T) and standard deviations (σ_T) are listed in Table 7.1. The nine paths, mean path durations and standard deviations from the modified PNET algorithm are given in Table 7.2.

Table 7.1: Ordered Paths and Duration Statistics - Table 2, Ang et al., (1975)

Path #	Activities in the Path	μ_T days	σ_T days
1	4, 7, 12, 13, 18, 20, 22, 25, 27	61	5.00
2	6, 10, 15, 19, 21, 23, 24, 26, 27	57	9.00
3	6, 11, 16, 19, 21, 23, 24, 26, 27	52	7.94
4	5, 9, 14, 19, 21, 23, 24, 26, 27	49	6.54
5	5, 8, 13, 18, 20, 22, 25, 27	42	4.00
6	3, 28, 20, 22, 25, 27	29	3.24
7	3, 1, 23, 24, 26, 27	29	5.19
8	2, 17, 28, 20, 22, 25, 27	28	3.16
9	2, 17, 1, 23, 24, 26, 27	28	5.12

The dummy activities required for the arrow network (activities 1 and 28) are not necessary for the precedence network used by the modified PNET (see paths 6, 7, 8 and 9 in Tables 7.1 and 7.2). In addition to ordering the paths in decreasing mean durations, the modified PNET orders the paths in decreasing standard deviations

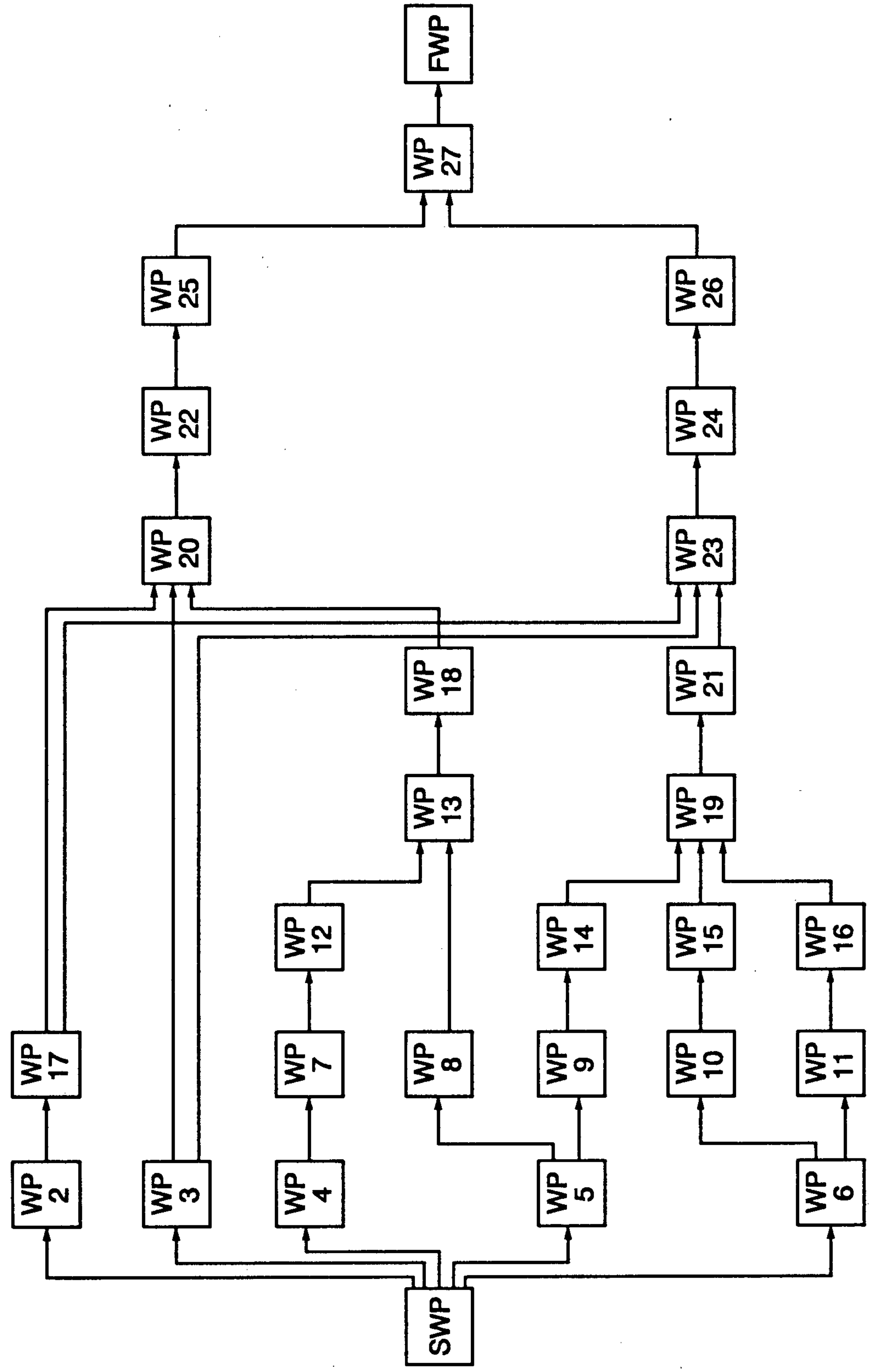

Figure 7.3: The Precedence Network for the Road Pavement Project

Table 7.2: Ordered Paths and Duration Statistics from Modified PNET

Path #	Ang et al	Activities in the Path	μ_T days	σ_T days
1	1	4, 7, 12, 13, 18, 20, 22, 25, 27	61	5.00
2	2	6, 10, 15, 19, 21, 23, 24, 26, 27	57	9.00
3	3	6, 11, 16, 19, 21, 23, 24, 26, 27	52	7.93
4	4	5, 9, 14, 19, 21, 23, 24, 26, 27	49	6.59
5	5	5, 8, 13, 18, 20, 22, 25, 27	42	4.00
6	7	3, 23, 24, 26, 27	29	5.17
7	6	3, 20, 22, 25, 27	29	3.24
8	9	2, 17, 23, 24, 26, 27	28	5.12
9	8	2, 17, 20, 22, 25, 27	28	3.16

when mean path durations are equal. This ensures the selection of the path with the highest variance as the representative path from the paths having the same mean duration (see paths 6, 7, 8 and 9 in Tables 7.1 and 7.2).

The representative paths for the transitional correlation $\rho = 0.5$ are paths 1 and 2 from PNET (Ang et al., 1975) and the modified PNET. The comparison shows that modified PNET identifies the paths correctly, evaluates the expected value (mean) and standard deviation for path durations accurately, and selects the representative paths correctly. The ordering of paths may differ because the modified PNET gives priority to the path with the higher variance when the mean durations are identical.

7.3.2 Industrial Building Project

This project involves the construction of a single-story industrial building. The building is comprised of reinforced concrete piers, frost walls, structural steel columns, and a precast roof (Ang et al., 1975). The precedence network for the project used by the modified PNET, based on the logic of the arrow network given by Fig.5 of Ang et al. (1975) is shown in figure (7.4). The various activities of the project, respective mean durations and standard deviations for the activities from Table 3 of Ang et al.

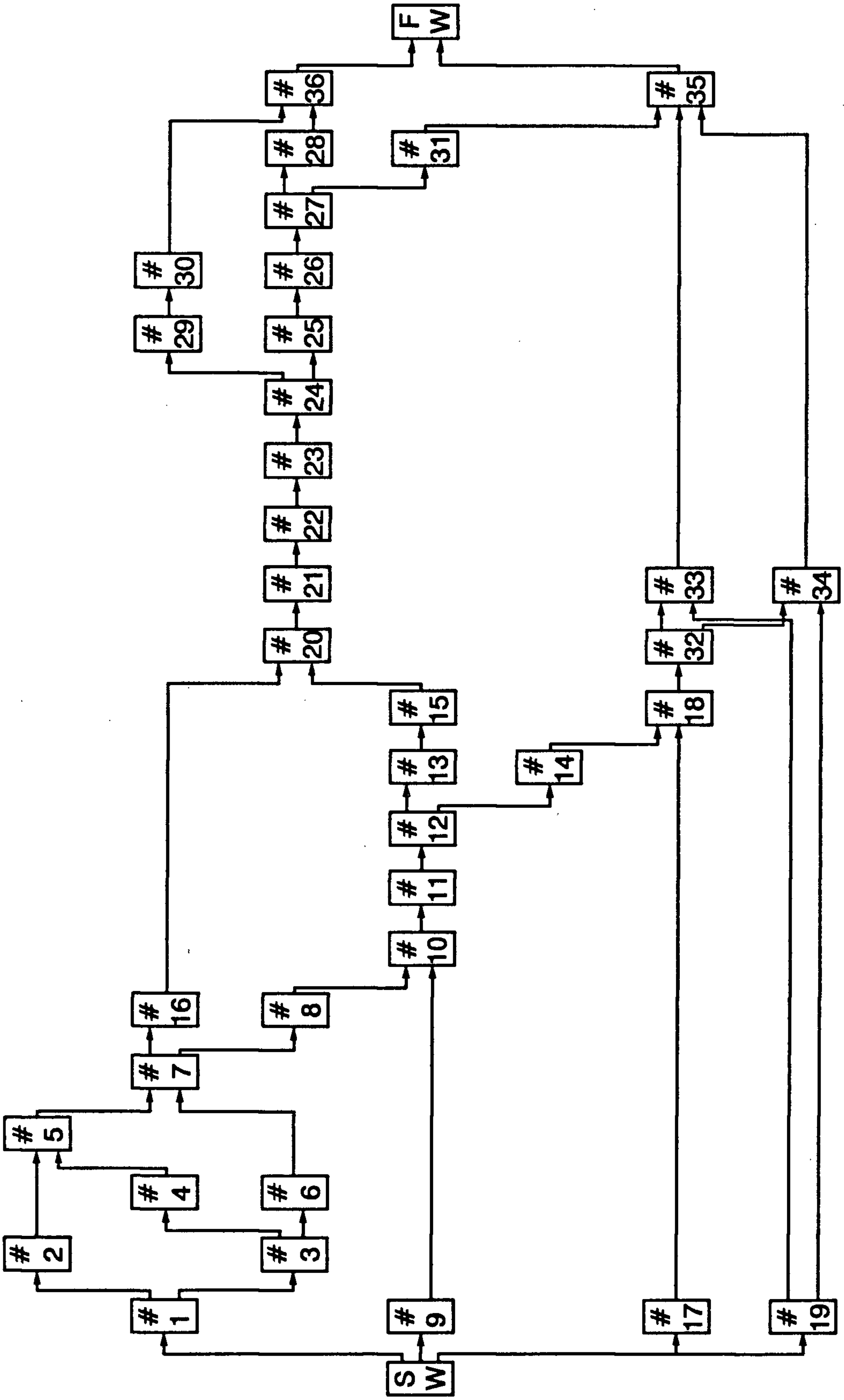

Figure 7.4: The Precedence Network for the Industrial Building Project

(1975), are given in Appendix F.

Ang et al. (1975), listed only the first ten paths arranged in decreasing mean path durations (Table 4, Ang et al., 1975). Table 7.3 lists all 33 paths in the project network as ordered by the modified PNET algorithm. The second column in Table 7.3 contains the path numbers of the ten paths listed in Table 4, Ang et al., (1975). Even though path 7 from modified PNET had the largest variance of the paths with mean duration of 66 *days*, PNET had not considered it as a major path.

The representative paths for the transitional correlation $\rho = 0.5$ are paths 1, 3 and 5 from PNET (Ang et al., 1975), while the modified PNET algorithm identifies paths 1, 3, 5, and 32. PNET considered only the first ten paths as the major paths. Even though path 32 is also a representative path by definition, it does not play a role in the completion time probability calculations because its mean path duration is insignificant when compared to the other representative paths. While PNET neglects those paths with low mean path durations, the modified PNET considers all the paths in the selection of representative paths. As shown later in the validations of the analytical method, the difference in execution time for the modified PNET routine to evaluate a single path (longest path approach) or all the paths in the project network is negligible.

The two comparisons validate the modified PNET algorithm used in the analytical method for time and economic risk quantification.

7.4 Parallel Network

A parallel network consisting of thirty five identical work packages in five parallel paths as shown in figure (7.5) is used as the first limiting case to validate the Monte Carlo simulation process. Since simulations are used to validate the analytical approach, it is essential to validate the simulation process first.

Table 7.3: Ordered Paths and Duration Statistics for the Industrial Building

Path #	Ang et al	Activities in the Path	μ_T days	σ_T days
1	1	17, 18, 32, 33, 35	78	12.20
2	2	17, 18, 32, 34, 35	76	12.20
3	3	9, ...*, 13, 15, 20, ..., 28, 36	69	12.12
4	4	9, ..., 13, 15, 20, ..., 27, 31, 35	68	12.14
5	5	1, 2, 5, 7, 8, 10, ..., 13, 15, 20, ..., 28, 36	67	3.85
6	6	1, 3, 4, 5, 7, 8, 10, ..., 13, 15, 20, ..., 28, 36	67	3.85
7	-	9, ..., 12, 14, 18, 32, 33, 35	66	12.25
8	8	9, ..., 13, 15, 20, ..., 24, 29, 30, 36	66	12.09
9	9	1, 2, 5, 7, 8, 10, ..., 13, 15, 20, ..., 27, 31, 35	66	3.87
10	10	1, 3, 4, 5, 7, 8, 10, ..., 13, 15, 20, ..., 27, 31, 35	66	3.87
11	7	1, 3, 6, 7, 8, 10, ..., 13, 15, 20, ..., 28, 36	66	3.85
12	-	1, 3, 6, 7, 8, 10, ..., 13, 15, 20, ..., 27, 31, 35	65	3.87
13	-	9, ..., 12, 14, 18, 32, 34, 35	64	12.25
14	-	1, 3, 4, 5, 7, 8, 10, ..., 12, 14, 18, 32, 33, 35	64	4.22
15	-	1, 2, 5, 7, 8, 10, ..., 12, 14, 18, 32, 33, 35	64	4.22
16	-	1, 3, 4, 5, 7, 8, 10, ..., 13, 15, 20, ..., 24, 29, 30, 36	64	3.79
17	-	1, 2, 5, 7, 8, 10, ..., 13, 15, 20, ..., 24, 29, 30, 36	64	3.79
18	-	1, 3, 6, 7, 8, 10, ..., 12, 14, 18, 32, 33, 35	63	4.22
19	-	1, 3, 6, 7, 8, 10, ..., 13, 15, 20, ..., 24, 29, 30, 36	63	3.79
20	-	1, 3, 4, 5, 7, 8, 10, ..., 12, 14, 18, 32, 34, 35	62	4.22
21	-	1, 2, 5, 7, 8, 10, ..., 12, 14, 18, 32, 34, 35	62	4.22
22	-	1, 3, 6, 7, 8, 10, ..., 12, 14, 18, 32, 34, 35	61	4.22
23	-	1, 3, 4, 5, 7, 16, 20, ..., 28, 36	59	3.73
24	-	1, 2, 5, 7, 16, 20, ..., 28, 36	59	3.73
25	-	1, 3, 4, 5, 7, 16, 20, ..., 27, 31, 35	58	3.75
26	-	1, 2, 5, 7, 16, 20, ..., 27, 31, 35	58	3.75
27	-	1, 3, 6, 7, 16, 20, ..., 28, 36	58	3.73
28	-	1, 3, 6, 7, 16, 20, ..., 27, 31, 35	57	3.75
29	-	1, 3, 4, 5, 7, 16, 20, ..., 24, 29, 30, 36	56	3.67
30	-	1, 2, 5, 7, 16, 20, ..., 24, 29, 30, 36	56	3.67
31	-	1, 3, 6, 7, 16, 20, ..., 24, 29, 30, 36	55	3.67
32	-	19, 33, 35	38	6.14
33	-	19, 34, 35	36	6.14

* Includes all intervening activities

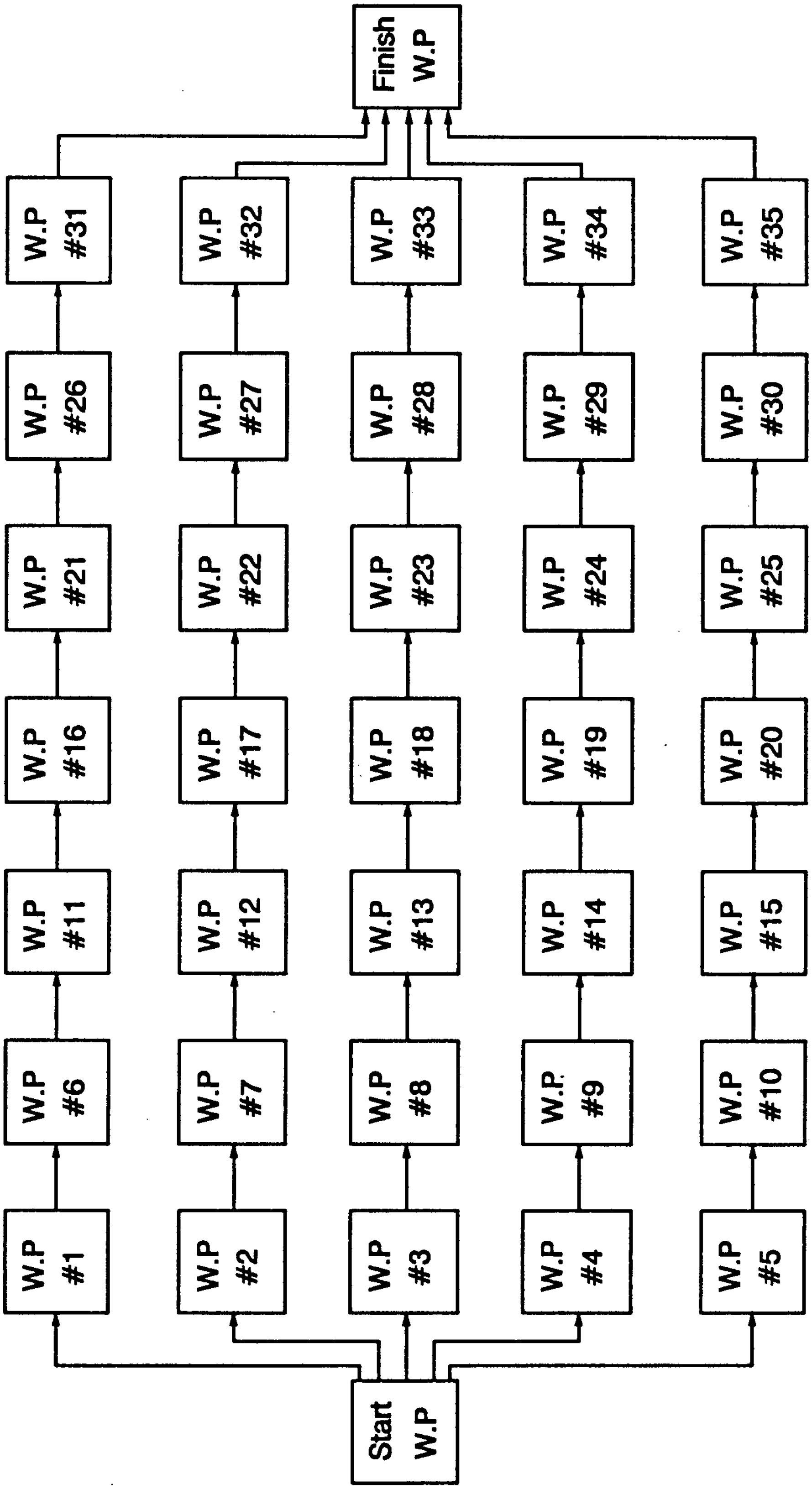

Figure 7.5: The Parallel Network

From the longest path (or PERT) approach every path in the parallel network is a critical path. Therefore, the cumulative distribution function for project duration is the cumulative distribution function from any path duration. This is the lower bound for completion time probability. From the modified PNET algorithm, completion time probability for project duration for any transitional correlation ρ (*i.e* $0 < \rho \leq 1$) is the same as the upper bound ($\rho = 1$). Therefore, the cumulative distribution function for project duration for the parallel network from a valid Monte Carlo simulation process should give the same cumulative distribution function as for the upper bound from the PNET algorithm.

The expected value, standard deviation, skewness and kurtosis for duration for all the work packages are $E[WPD] = 3.644\ months$, $\sigma_{WPD} = 0.67\ months$, $\sqrt{\beta_1} = 0.3$ and $\beta_2 = 3.5$. The statistics for project duration from the longest path (lower bound $\rho = 0$), for any transitional correlation, $0 < \rho \leq 1$, and the expected value and standard deviation from the simulation are given in Table 7.4.

Table 7.4: Statistics for Project Duration for First Limiting Case

Project Duration (months)	Expected Value	Standard Deviation	$\sqrt{\beta_1}$	β_2
Longest Path ($\rho = 0$)	25.51	1.76	0.11	3.07
When $0 < \rho \leq 1$	27.63	1.24	0.30	3.2
Monte Carlo Simulation	27.60	1.30		

The cumulative distribution functions for project duration from the longest path, when $0 < \rho \leq 1$ and a Monte Carlo simulation of 20,000 iterations is depicted in figure (7.6). This simple limiting case, while validating the Monte Carlo simulation process that is used to validate the analytical method, also confirms the theoretical postulations made by the modified PNET algorithm.

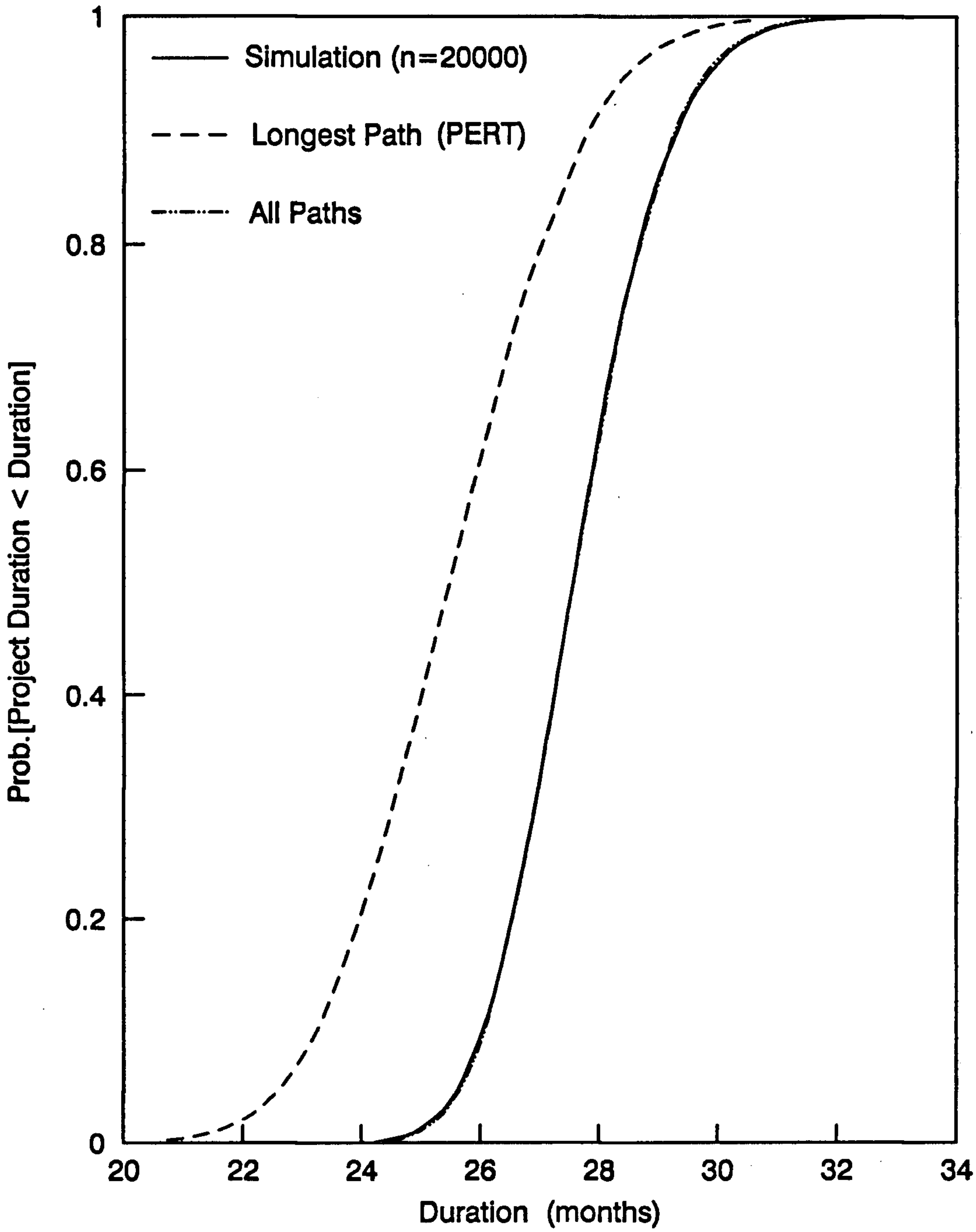

Figure 7.6: CDFs for Project Duration for the Parallel Network

7.5 First Example

This section demonstrates the second limiting case to validate the Monte Carlo simulation and the first two validations of the analytical method. The data for this example is obtained from an actual deterministic feasibility analysis conducted for a mineral project in South America. The starting point for the analysis is at the work package level. For study purposes herein, the original construction program is modified as shown in figure (7.7). The logic of the original program is maintained throughout. The work package durations are developed to correspond to the modified construction schedule. The deterministic estimates and statistics for work package durations are given in Appendix F.

7.5.1 Second Limiting Case

The precedence network depicted in figure (7.7) is highly interrelated. However, if there is one dominant path in the network then that path will dominate completion time probability of the project. Therefore, the project duration from the longest path (lower bound), all the paths (upper bound) and from the simulation should be similar. Such a path can be created by changing the statistics for duration for work package #7 to $E[WPD] = 20.01$ *months*, $\sigma_{WPD} = 1.609$ *months*, $\sqrt{\beta_1} = 0.2$ and $\beta_2 = 2.6$. Then the dominant path consists of work packages #2, #7, #20, #24, #30, #31. The expected value, standard deviation, skewness and kurtosis for project duration for the dominant path (lower bound $\rho = 0$), from all paths ($\rho = 1$) and the expected value and standard deviation from simulation are given in Table 7.5.

The cumulative distribution functions for project duration from lower and upper bounds, and a Monte Carlo simulation of 15000 iterations are depicted in figure (7.8). This limiting case also validates the Monte Carlo simulation process. In addition, it

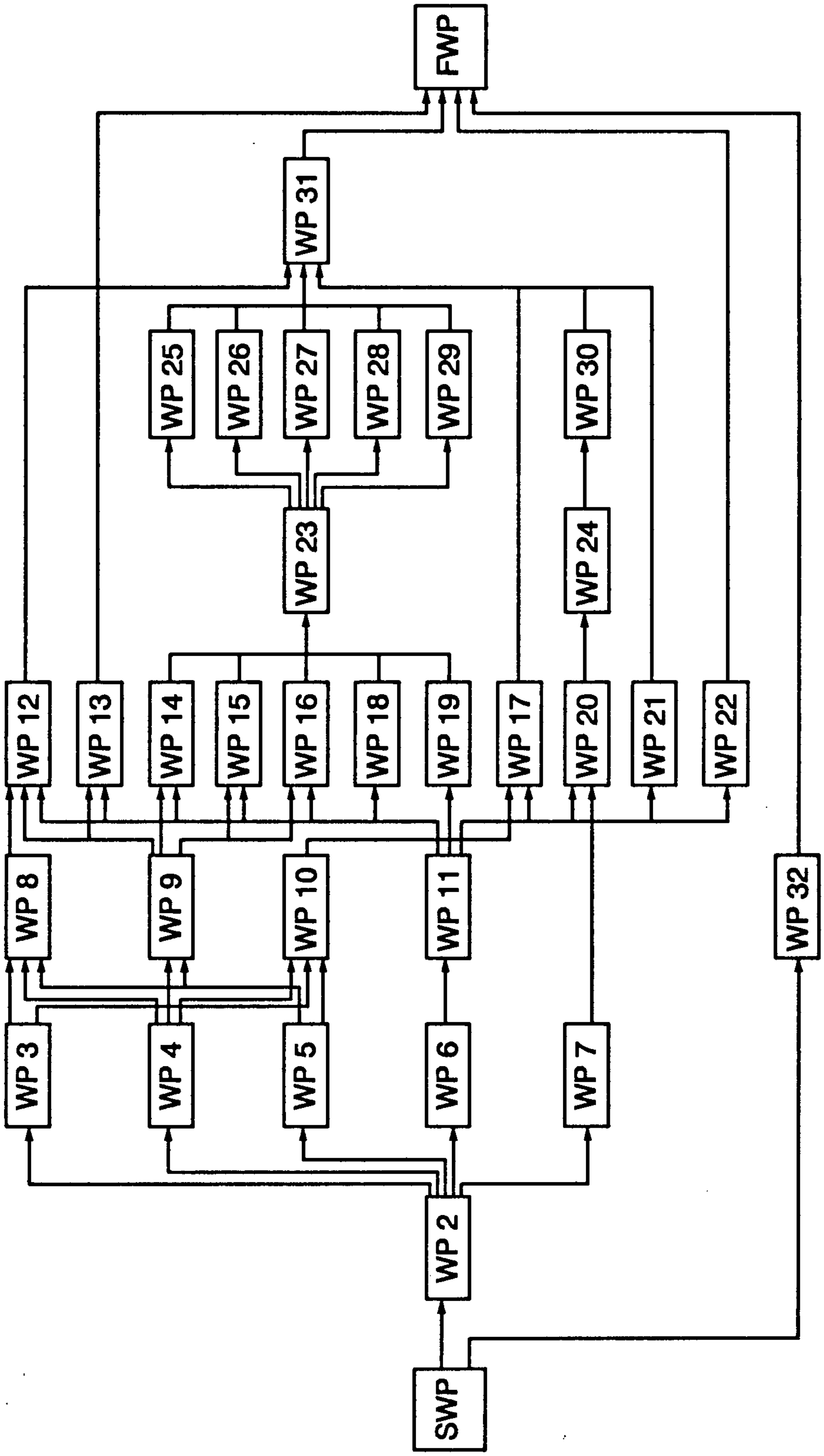

Figure 7.7: The Project Network for the First Example

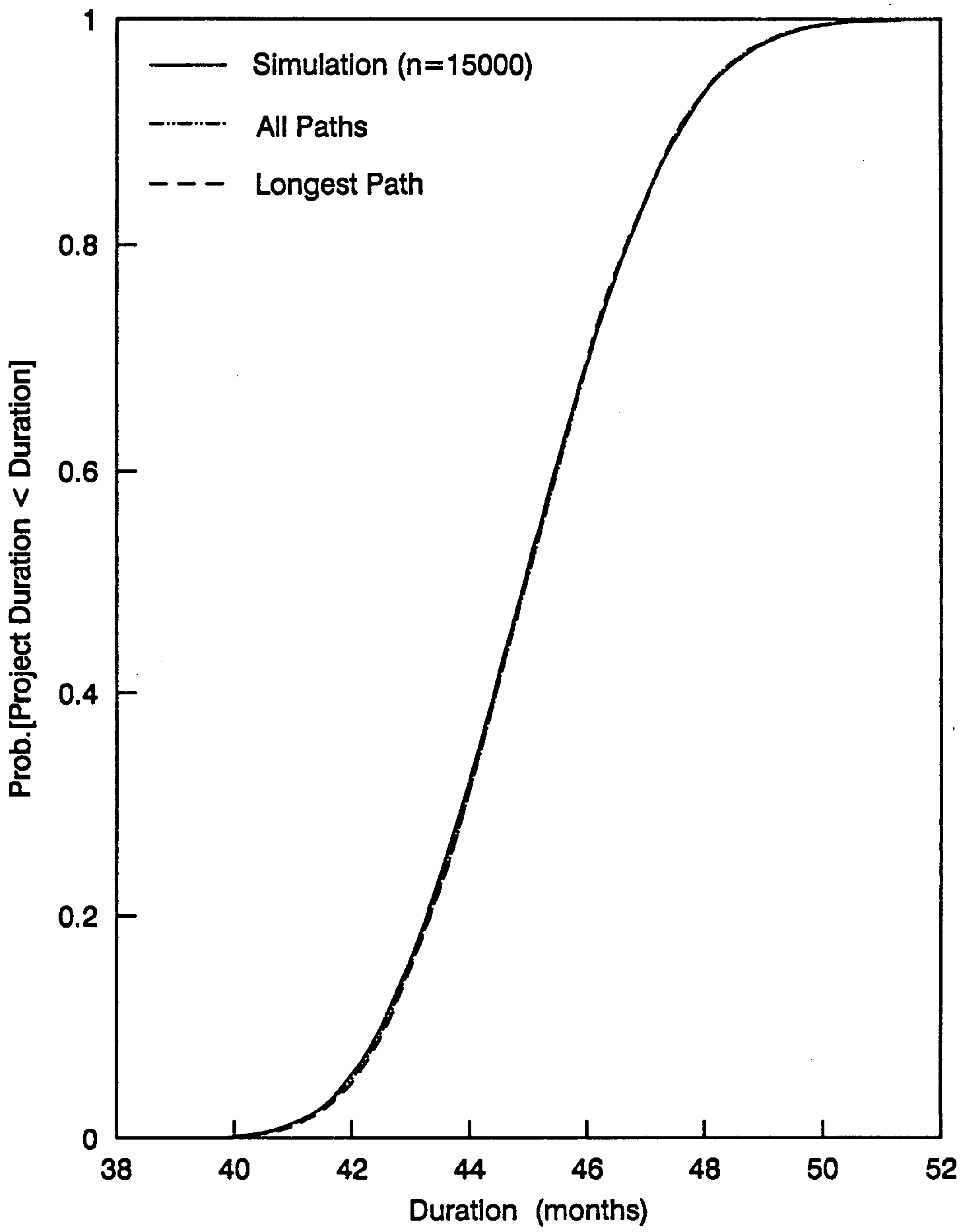

Figure 7.8: CDFs for Project Duration for the Single Dominant Path

Table 7.5: Statistics for Project Duration for Second Limiting Case

Project Duration (months)	Expected Value	Standard Deviation	$\sqrt{\beta_1}$	β_2
Longest Path ($\rho = 0$)	45.01	1.87	0.15	2.85
All the Paths ($\rho = 1$)	45.01	1.88	0.1	2.8
Monte Carlo Simulation	44.96	1.53		

confirms accuracy of the modified PNET algorithm.

7.5.2 First Validation

Two simulations were done as the first validation. The derived variable for the first simulation was only project duration. For the first validation, low coefficients of variation for work package durations are assumed. Table 7.6 contains the expected values and standard deviations for project duration from the simulation at 1000 iteration intervals, and the statistics evaluated from the analytical approach at different transitional correlations. Figure (7.9) illustrates the cumulative distribution functions for upper and lower bounds approximated from the analytical method and that generated from a simulation of 15,000 iterations. Figure (7.10) depicts, in addition to those in figure (7.9), the cumulative distribution functions for project duration at different transitional correlation (ρ) values.

The second simulation is a complete time and economic risk quantification. However, the statistics and cumulative distribution function generated for project duration are not considered because the first simulation is much larger. The work package costs of the project network depicted by figure (7.7) are estimated such that the sum of the work package costs is equivalent to the constant dollar cost estimate of the deterministic feasibility analysis. The deterministic estimates for work package costs are given in Appendix F.

Table 7.6: Statistics for Project Duration from First Validation - Ex #1

Simulation			Analytical Method				
#	$E[PD]$ mths	σ_{PD} mths	ρ	$E[PD]$ mths	σ_{PD} mths	$\sqrt{\beta_1}$	β_2
1000	37.29	1.11	0.0	36.08	1.30	0.13	2.95
2000	37.30	1.15	0.1	36.11	1.27	0.2	2.8
3000	37.28	1.40	0.2	36.11	1.27	0.2	2.8
4000	37.27	1.51	0.3	36.92	1.03	0.3	3.2
5000	37.27	1.56	0.4	36.92	1.03	0.3	3.2
6000	37.26	1.59	0.5	36.92	1.03	0.3	3.2
7000	37.24	1.61	0.6	37.31	1.02	0.3	3.3
8000	37.25	1.63	0.7	37.31	1.02	0.3	3.3
9000	37.25	1.65	0.8	37.74	0.91	0.4	3.4
10000	37.25	1.66	0.9	37.95	0.82	0.5	3.5
11000	37.25	1.66	1.0	38.34	0.68	0.5	3.4
12000	37.25	1.67					
13000	37.25	1.51					
14000	37.26	1.34					
15000	37.26	1.19					

The function for discounted work package cost (WPC_i) used for the analysis is as follows.

$$\begin{aligned} WPC_i &= f\ e^{(\theta_{C_i}-y)T_{S_{C_i}}} \int_0^{T_{Ci}} C_{0i}(\tau)\ e^{(\theta_{C_i}-y)\tau}\ d\tau \\ &\quad +(1-f)e^{(r-y)T_p} e^{\theta_{C_i} T_{S_{C_i}}} \int_0^{T_{Ci}} C_{0i}(\tau) e^{(\theta_{C_i}-r)\tau} d\tau \end{aligned} \tag{7.3}$$

where WPC_i is the discounted i^{th} work package cost, $C_{0i}(\tau)$ is the constant dollar cash flow for the i^{th} work package, $T_{S_{C_i}}$ and T_{Ci} are work package start time and duration, T_p is the time at which the repayment of interim financing is due for all work packages (assumed as the end of the construction phase), f is the equity fraction, and θ_{C_i}, r and y are inflation, interest and discount rates respectively. The time τ is measured from the start of the i^{th} work package.

The function for revenue streams (NRS_i) is as follows.

$$NRS_i = \int_{T_{S_{Ri}}}^{T_{S_{Ri}}+T_{Ri}} \left[R_{0i}(t)e^{\theta_{R_i}(t-T_{S_{Ri}})} - M_{0i}(t)e^{\theta_{M_i}t}\right] e^{-yt}dt \qquad (7.4)$$

where NRS_i is the discounted i^{th} revenue stream, $R_{0i}(t)$ and $M_{0i}(t)$ are the constant dollar cash flow for the i^{th} gross revenue and operation and maintenance cost, $T_{S_{Ri}}$ and T_{Ri} are start time and duration of the revenue stream, and $\theta_{R_i}, \theta_{M_i}$ and y are inflation and discount rates respectively.

The deterministic values for the respective primary variables (*i.e* work package durations and costs, annual revenues and operating costs, inflation and financing rates) are assumed to be the median values of their frequency distributions. The expected value, standard deviation, skewness and kurtosis for work package durations, costs, annual gross revenues and operation and maintenance costs for the revenue streams are given in Appendix F.

For illustrative purposes herein, uniform constant dollar expenditure profiles for work package costs and annual operating costs were assumed. Similarly, uniform constant dollar revenue profiles were assumed for gross annual revenue streams. A common inflation rate with the following statistics, $E[\theta_C] = 5.837\%, \sigma_{\theta_C} = 0.395\%, \sqrt{\beta_1} = 0.1$ and $\beta_2 = 2.6$ is assumed for all work package costs. A construction loan for 85% ($f = 0.15$) of the current dollar expenditure on construction is assumed. The statistics for the interest rate on the construction loan are $E[r] = 8.631\%, \sigma_r = 0.704\%, \sqrt{\beta_1} = 0.0$ and $\beta_2 = 3.6$. The minimum attractive rate of return used for the analysis and validations is 20%. All the variables in the analysis are assumed to be uncorrelated.

Tables 7.7, 7.8 and 7.9 contain results from the second simulation at 500 iteration intervals and statistics from the analytical method at different transitional correlation values for discounted project cost, discounted project revenue and project net present value.

Table 7.7: Statistics for Discounted Project Cost from First Validation - Ex #1

Simulation			Analytical Method				
#	$E[DPC]$ $	σ_{DPC} $	ρ	$E[DPC]$ $	σ_{DPC} $	$\sqrt{\beta_1}$	β_2
500	87054064	9588990	0.0	87792088	9658726	0.097	2.617
1000	86791152	9586503	0.1	87767561	9655823	0.097	2.617
1500	86764320	9739846	0.2	87766524	9655819	0.097	2.617
2000	86775360	9789910	0.3	87168757	9591506	0.097	2.617
2500	86727216	9819083	0.4	87168757	9591506	0.097	2.617
3000	86785520	9834827	0.5	87166313	9591480	0.097	2.617
3500	86804592	9757994	0.6	86901802	9562288	0.097	2.617
4000	86805648	9705724	0.7	86900640	9562167	0.097	2.617
			0.8	86598834	9529379	0.097	2.617
			0.9	86445916	9512848	0.097	2.617
			1.0	86130819	9480193	0.097	2.617

Table 7.8: Statistics for Discounted Project Revenue from First Validation-Ex #1

Simulation			Analytical Method				
#	$E[DPR]$ $	σ_{DPR} $	ρ	$E[DPR]$ $	σ_{DPR} $	$\sqrt{\beta_1}$	β_2
500	143957280	12268435	0.0	147761732	11912759	-0.051	2.744
1000	143692784	12273304	0.1	147689326	11906346	-0.051	2.744
1500	144041696	12045838	0.2	147689326	11906346	-0.051	2.744
2000	144331664	12135314	0.3	146044842	11790591	-0.051	2.744
2500	144558464	12128094	0.4	146056516	11790591	-0.051	2.744
3000	144623264	11984054	0.5	146056516	11790591	-0.051	2.744
3500	144699216	11889887	0.6	145282056	11740939	-0.051	2.744
4000	144820096	11871214	0.7	145278674	11740788	-0.051	2.744
			0.8	144429474	11682179	-0.051	2.744
			0.9	144015454	11652676	-0.051	2.744
			1.0	143232846	11598742	-0.051	2.744

Table 7.9: Statistics for Project NPV from First Validation - Ex #1

Simulation			Analytical Method				
#	$E[NPV]$	σ_{NPV}	ρ	$E[NPV]$	σ_{NPV}	$\sqrt{\beta_1}$	β_2
	$	$		$	$		
500	56903152	15845556	0.0	59969644	15336388	-0.048	2.847
1000	56909520	15262957	0.1	59921765	15329579	-0.048	2.847
1500	57275088	15202756	0.2	59922801	15329576	-0.048	2.847
2000	57546720	15573732	0.3	58879177	15199179	-0.048	2.847
2500	57817088	15527935	0.4	58887759	15199179	-0.048	2.847
3000	57820480	15474244	0.5	58890203	15199162	-0.048	2.847
3500	57875184	15318487	0.6	58380254	15142226	-0.048	2.847
4000	57993648	15290463	0.7	58378034	15142032	-0.048	2.847
			0.8	57830640	15075887	-0.048	2.847
			0.9	57569539	15042578	-0.048	2.847
			1.0	57102027	14980150	-0.048	2.847

Table 7.10 contains the expected value and standard deviation for project internal rate of return from the simulation and from the analytical method at different transitional correlation values. The analytical method develops the cumulative distribution function for internal rate of return using cumulative distribution functions for net present value at incremental discount rates. The expected value and standard deviation for internal rate of return are approximated using percentiles from that cumulative distribution function.

Figures (7.11), (7.12), (7.13) and (7.14) illustrate the cumulative distribution functions for upper and lower bounds approximated from the analytical method and those generated from a simulation of 4000 iterations for discounted project cost, discounted project revenue, net present value and internal rate of return. The cumulative distribution functions and the estimates for expected values for derived time and economic variables demonstrate that the results generated from Monte Carlo simulation are within the upper and lower bounds predicted by the analytical approximations, thereby validating the analytical method.

Table 7.10: Statistics for Project IRR from First Validation - Ex #1

Simulation			Analytical Method		
#	$E[IRR]$	σ_{IRR}	ρ	$E[IRR]$	σ_{IRR}
500	32.71	4.14	0.0	33.241	4.094
1000	32.69	3.99	0.1	33.231	4.091
1500	32.76	4.01	0.2	33.231	4.091
2000	32.81	4.11	0.3	33.019	4.034
2500	32.87	4.13	0.4	33.019	4.034
3000	32.88	4.11	0.5	33.020	4.034
3500	32.88	4.07	0.6	32.930	4.037
4000	32.89	4.03	0.7	32.930	4.037
			0.8	32.836	4.033
			0.9	32.791	4.029
			1.0	32.720	4.020

Table 7.11: Comparison of CPU times from First Validation - Ex #1

Simulation		Analytical Method	
#	CPU Sec.	ρ	CPU Sec.
500	511	0.0	34.45
1000	1021	0.1	35.03
1500	1531	0.2	35.09
2000	2042	0.3	35.18
2500	2550	0.4	35.56
3000	3059	0.5	34.89
3500	3567	0.6	35.22
4000	4079	0.7	35.75
		0.8	35.86
		0.9	36.54
		1.0	37.78

Table 7.11 contains a comparison of the execution time for simulation and the analytical method. The computational economy of the analytical method is clearly highlighted. For this example, the analytical method is about thirty times faster when compared to the generally recommended number of iterations (1000) for risk quantification using Monte Carlo simulation (Inyang, 1983; Perry and Hayes, 1985b).

Both analyses were done on an IBM 3081 mainframe computer. There are seventy three possible paths to complete the project network depicted by figure (7.7). When $\rho = 0$ only the moments on the longest path are considered to evaluate the statistics for project duration. When $\rho = 1$ the moments of all 73 paths are considered to evaluate the statistics for project duration. The comparison of the execution times however, show that the time difference to evaluate statistics and cumulative distribution functions for upper and lower bounds from the analytical method are negligible. While 73 paths is not a significant number for a large engineering project, it still demonstrates that evaluating the bounds for an alternative is not an excessive burden in terms of the computational economy when compared to simulation.

7.5.3 Second Validation

The same numerical values as for the previous case are used for the second validation. The only difference is that the coefficients of variation for work package durations are approximately 40% instead of the 3% to 13% used in the previous case. Since the derived economic variables are dependent upon time, this increase permits us to study its effect on their risk quantification. The statistics for the revised work package durations are given in Appendix F.

Two simulations were done. The derived variable for the first simulation was only project duration. Table 7.12 contains the expected values and standard deviations from the simulation and the statistics evaluated from the analytical method.

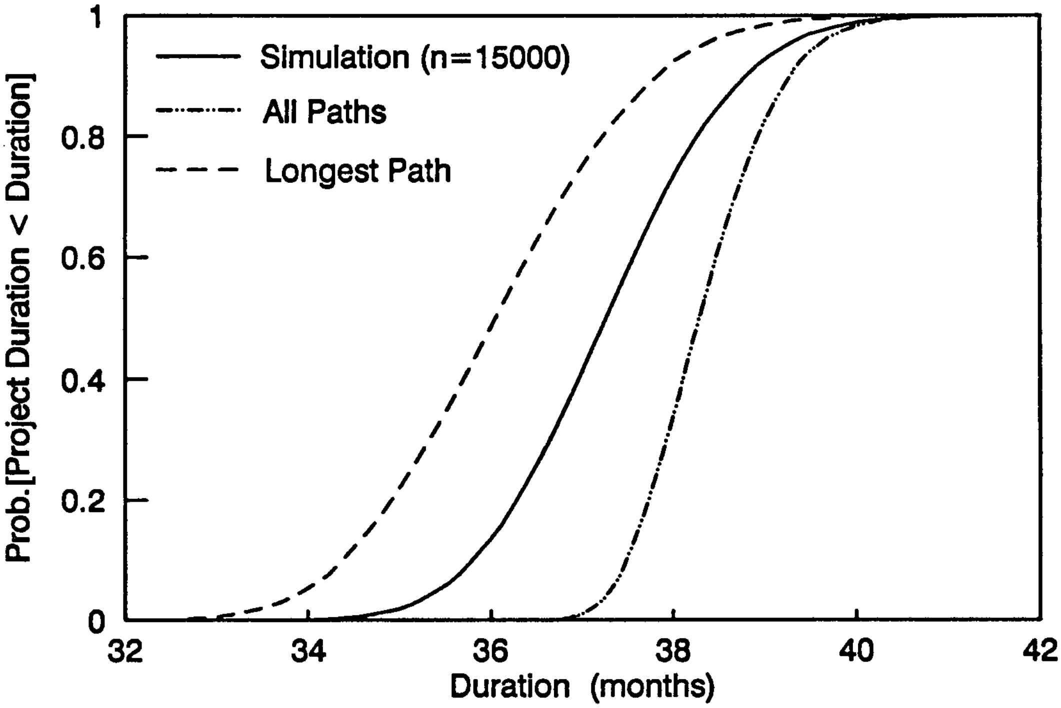

Figure 7.9: CDFs for Project Duration - First Validation - Ex #1

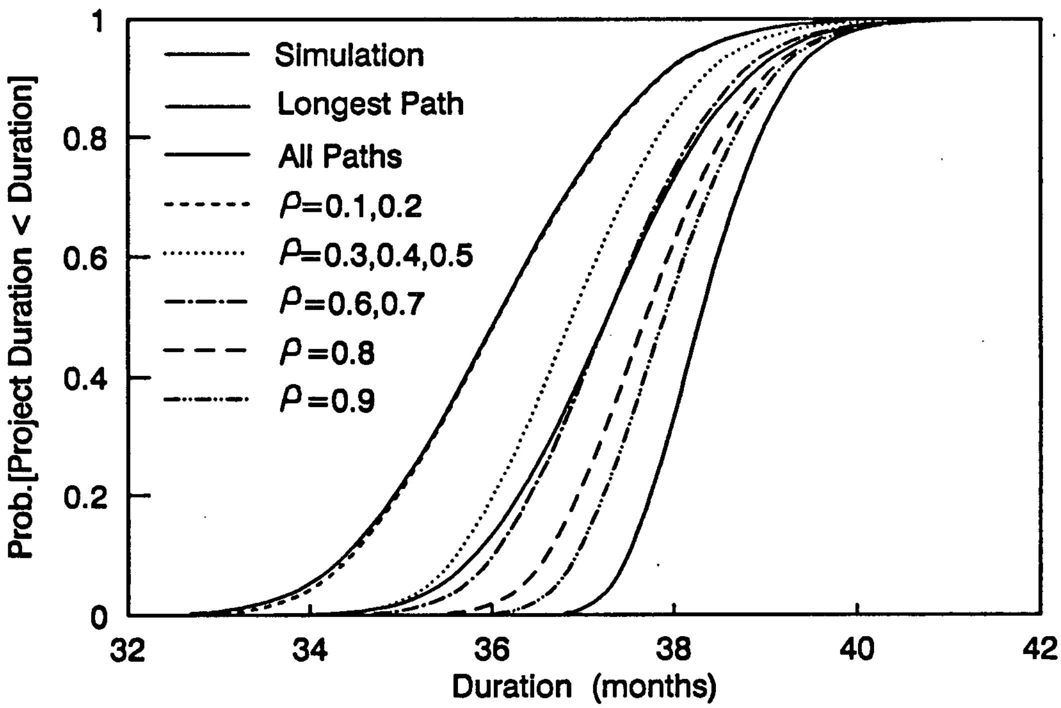

Figure 7.10: CDFs for Project Duration - First Validation - Ex #1

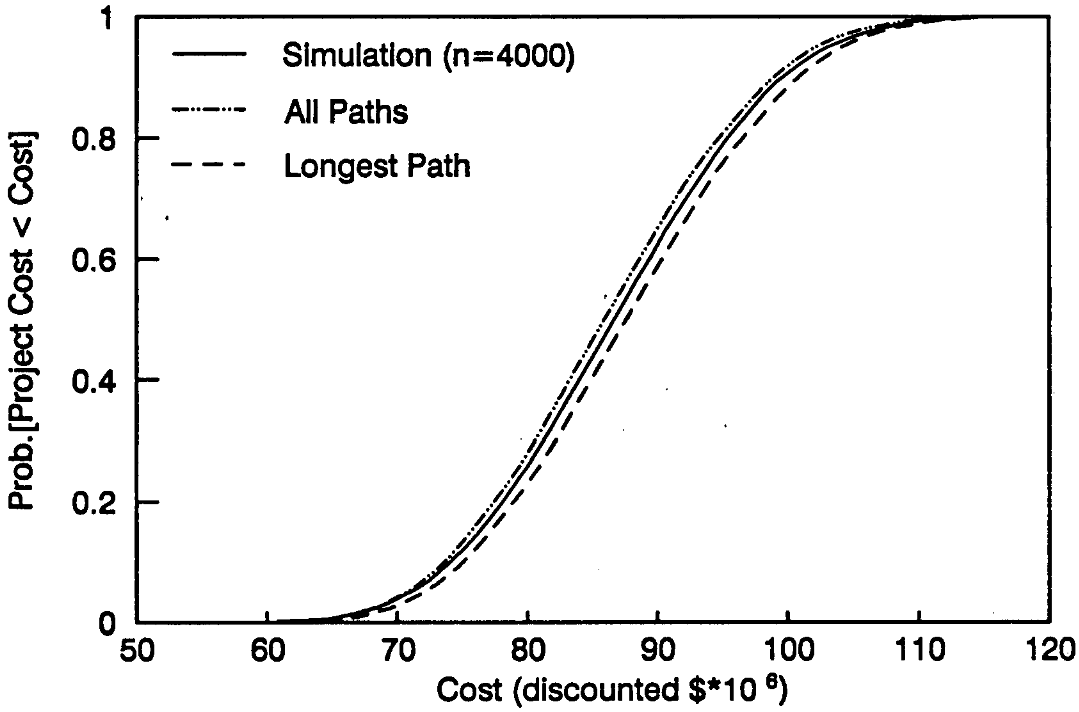

Figure 7.11: CDFs for Discounted Project Cost - First Validation - Ex #1

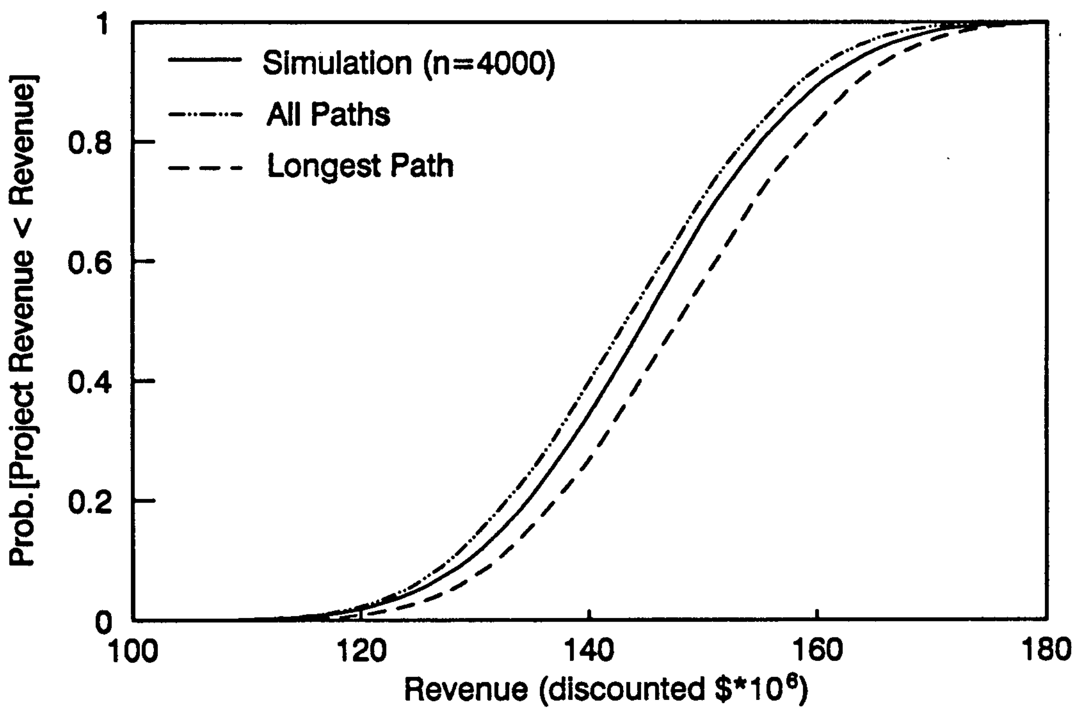

Figure 7.12: CDFs for Discounted Project Revenue - First Validation - Ex #1

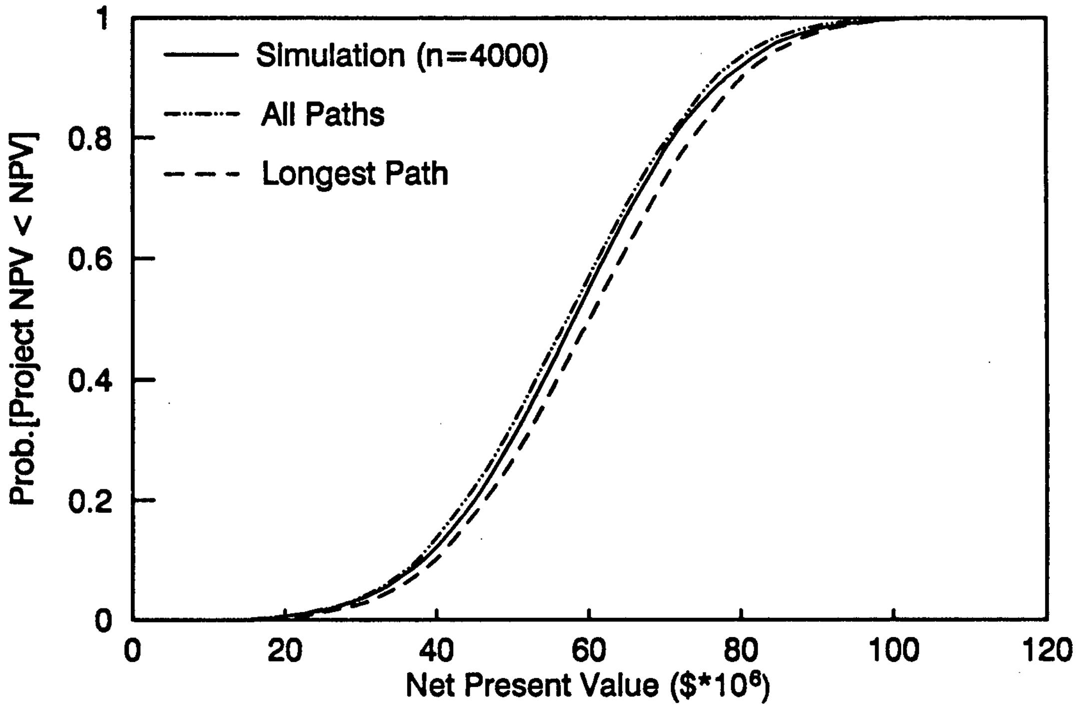

Figure 7.13: CDFs for Project Net Present Value - First Validation - Ex #1

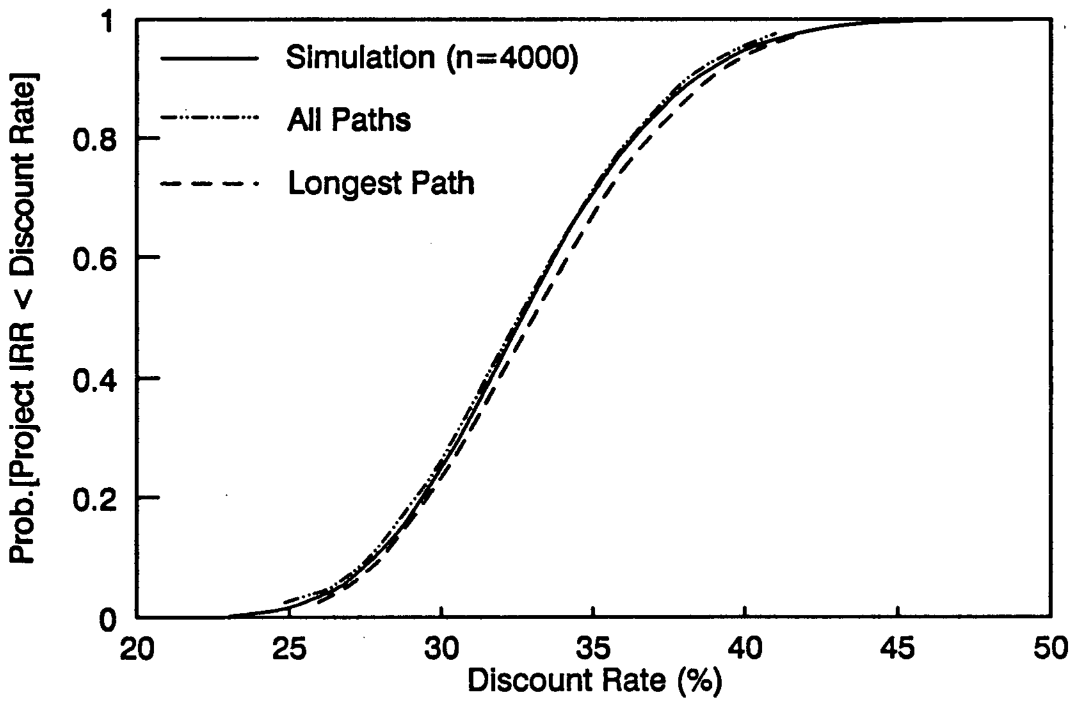

Figure 7.14: CDFs for Project Internal Rate of Return - First Validation - Ex #1

Table 7.12: Statistics for Project Duration from Second Validation - Ex #1

Simulation			Analytical Method				
#	$E[PD]$ mths	σ_{PD} mths	ρ	$E[PD]$ mths	σ_{PD} mths	$\sqrt{\beta_1}$	β_2
1000	46.01	5.19	0.0	36.31	6.56	0.12	3.09
2000	45.89	5.00	0.1	42.38	5.91	0.7	3.9
3000	45.93	5.08	0.2	42.38	5.91	0.7	3.9
4000	45.91	5.09	0.3	44.12	5.17	0.8	4.6
5000	45.88	5.10	0.4	44.45	5.02	0.8	4.5
6000	45.87	5.11	0.5	45.55	4.50	0.8	4.5
7000	45.83	5.12	0.6	46.01	4.28	0.9	5.0
8000	45.84	5.10	0.7	46.45	4.11	0.8	4.5
9000	45.83	5.02	0.8	46.84	3.91	0.8	4.4
10000	45.84	4.95	0.9	48.61	3.19	0.7	4.2
11000	45.87	4.92	1.0	49.12	2.97	0.8	5.2
12000	45.89	4.90					
13000	45.89	4.88					
14000	45.91	4.86					
15000	45.91	4.85					

Figure (7.15) illustrates the cumulative distribution functions for upper and lower bounds approximated from the analytical method and that generated from a simulation of 15,000 iterations. Figure (7.16) depicts, in addition to those in figure (7.15), the cumulative distribution functions for project duration at different transitional correlation values.

The second simulation is again a complete time and economic risk quantification. Tables 7.13, 7.14 and 7.15 contain results from the simulation at 500 iteration intervals and statistics from the analytical method at different transitional correlation values for discounted project cost, discounted project revenue and project net present value. Tables 7.16 contains the expected value and standard deviation for project internal rate of return from the simulation and the analytical method at different transitional correlation values.

Table 7.13: Statistics for Discounted Project Cost from Second Validation - Ex #1

Simulation			Analytical Method				
#	$E[DPC]$	σ_{DPC}	ρ	$E[DPC]$	σ_{DPC}	$\sqrt{\beta_1}$	β_2
	$	$		$	$		
500	81021792	9462825	0.0	87795608	9800603	0.092	2.557
1000	80835632	9537916	0.1	83698131	9322582	0.092	2.564
1500	80844544	9699357	0.2	83438139	9276978	0.094	2.574
2000	80826096	9752933	0.3	82318999	9166055	0.093	2.569
2500	80795552	9797531	0.4	82090258	9138546	0.093	2.570
3000	80809072	9791703	0.5	81377056	9049674	0.094	2.574
3500	80816496	9750850	0.6	80958126	9004202	0.094	2.574
4000	80828448	9728706	0.7	80600925	8966904	0.094	2.574
			0.8	80348004	8935649	0.094	2.575
			0.9	79175607	8800934	0.095	2.579
			1.0	78815314	8760869	0.095	2.580

Table 7.14: Statistics for Discounted Project Revenue from Second Validation-Ex #1

Simulation			Analytical Method				
#	$E[DPR]$	σ_{DPR}	ρ	$E[DPR]$	σ_{DPR}	$\sqrt{\beta_1}$	β_2
	$	$		$	$		
500	127760688	14264898	0.0	147861348	12781663	-0.038	2.720
1000	127718064	13930838	0.1	135939654	11762603	-0.046	2.724
1500	128100240	13516248	0.2	135668030	11502775	-0.045	2.728
2000	128309872	13668063	0.3	132635839	11390905	-0.046	2.727
2500	128541968	13602110	0.4	132007641	11322201	-0.046	2.727
3000	128504048	13480204	0.5	129969465	11101722	-0.046	2.728
3500	128558912	13500583	0.6	129129937	11010403	-0.046	2.729
4000	128685376	13480615	0.7	128343894	10936793	-0.046	2.729
			0.8	127624844	10861321	-0.046	2.730
			0.9	124503245	10569469	-0.045	2.731
			1.0	123611279	10487196	-0.045	2.732

Table 7.15: Statistics for Project NPV from Second Validation - Ex #1

Simulation			Analytical Method				
#	$E[NPV]$	σ_{NPV}	ρ	$E[NPV]$	σ_{NPV}	$\sqrt{\beta_1}$	β_2
	\$	\$		\$	\$		
500	46737200	15634686	0.0	60065740	16106605	-0.040	2.828
1000	46889744	14909846	0.1	52241523	15008997	-0.044	2.831
1500	47259008	14681378	0.2	52229892	14777556	-0.045	2.834
2000	47478240	15058231	0.3	50316839	14620851	-0.045	2.833
2500	47735504	15001127	0.4	49917383	14550094	-0.045	2.833
3000	47680688	14979397	0.5	48592410	14322878	-0.045	2.834
3500	47725680	14848319	0.6	48171811	14223384	-0.045	2.834
4000	47838080	14796526	0.7	47742969	14142800	-0.045	2.834
			0.8	47276840	14064641	-0.045	2.835
			0.9	45327638	13753913	-0.045	2.836
			1.0	44795965	13665069	-0.045	2.836

Table 7.16: Statistics for Project IRR from Second Validation - Ex #1

Simulation			Analytical Method		
#	$E[IRR]$	σ_{IRR}	ρ	$E[IRR]$	σ_{IRR}
500	30.83	3.94	0.0	33.457	4.756
1000	30.86	3.77	0.1	31.887	4.233
1500	30.93	3.76	0.2	31.886	4.073
2000	30.97	3.87	0.3	31.551	4.087
2500	31.03	3.89	0.4	31.472	4.056
3000	31.02	3.88	0.5	31.176	3.908
3500	31.02	3.82	0.6	31.075	3.971
4000	31.03	3.78	0.7	31.019	3.957
			0.8	30.929	3.927
			0.9	30.571	3.708
			1.0	30.477	3.702

Table 7.17: Comparison of CPU times from Second Validation - Ex #1

Simulation		Analytical Method	
#	CPU Sec.	ρ	CPU Sec.
500	551	0.0	35.04
1000	1100	0.1	35.19
1500	1642	0.2	35.25
2000	2185	0.3	35.98
2500	2728	0.4	36.20
3000	3272	0.5	36.31
3500	3816	0.6	36.27
4000	4361	0.7	36.34
		0.8	36.32
		0.9	36.89
		1.0	39.65

Figures (7.17), (7.18), (7.19) and (7.20) illustrate the cumulative distribution functions for the transitional correlation $\rho = 0.5$, upper and lower bounds approximated from the analytical method and those generated from a simulation of 4,000 iterations for discounted project cost, revenue, net present value and internal rate of return. The second validation also demonstrates that cumulative distribution functions and the estimates for time and economic variables generated from the simulation are within the upper and lower bounds predicted by the analytical method.

Table 7.17 contains a comparison of the execution time for the simulation and the analytical method. Again, the computational economy of the analytical method is highlighted. The more significant observation is the wider bounds for derived variables when compared to the previous case. The only change from the first to second validation is an increase in the coefficients for variation for work package durations. Therefore, wider bounds are a direct result of the increase in the variance for work package durations and start times. This observation highlights the significance of work package duration and start time in economic risk quantification.

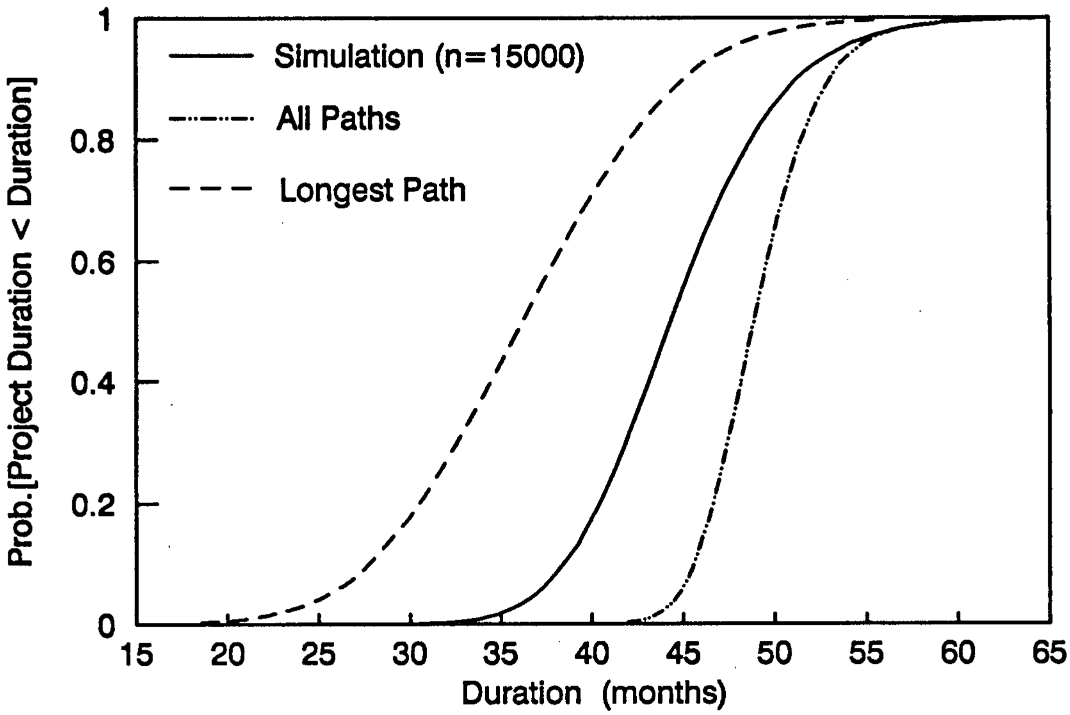

Figure 7.15: CDFs for Project Duration - Second Validation - Ex #1

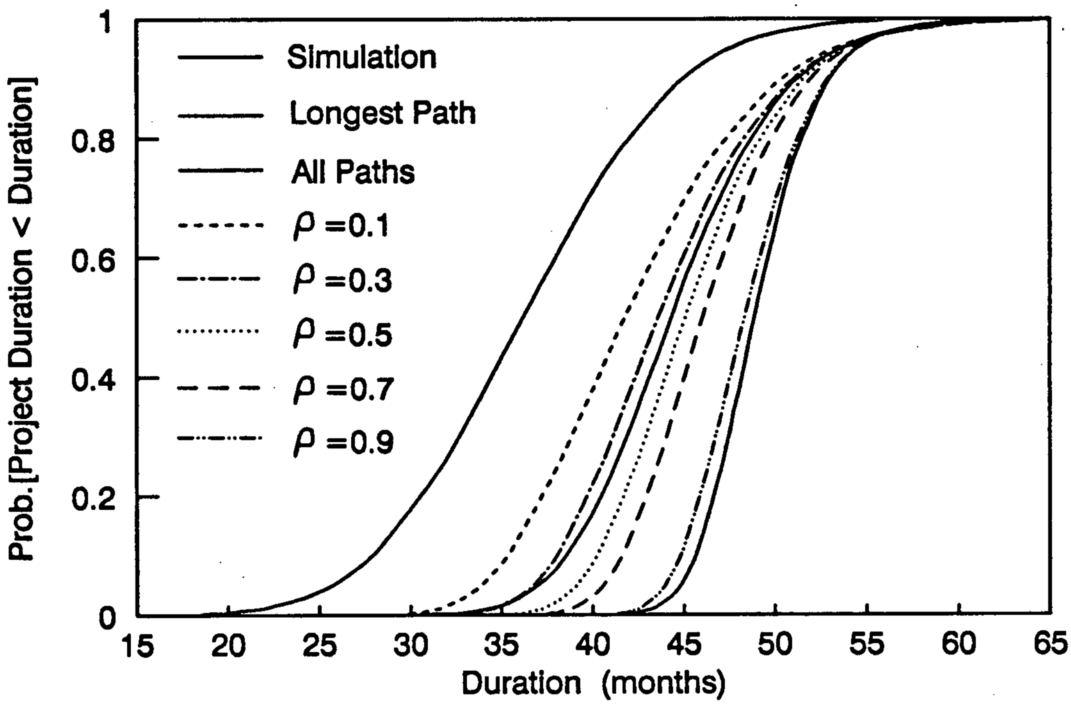

Figure 7.16: CDFs for Project Duration - Second Validation - Ex #1

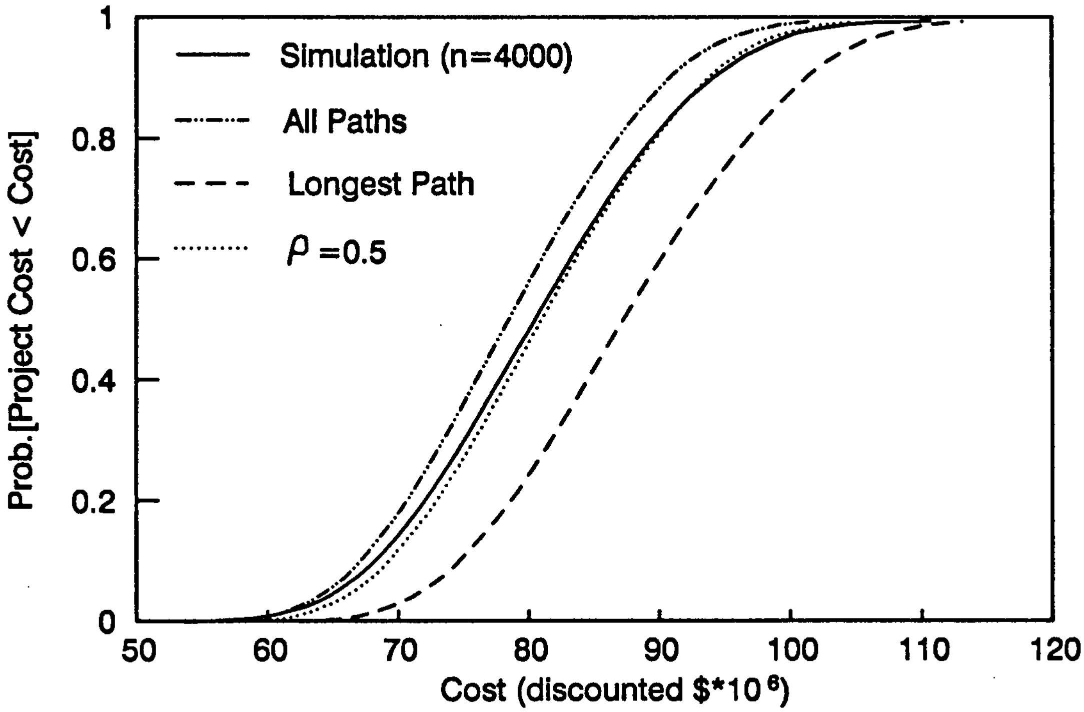

Figure 7.17: CDFs for Discounted Project Cost - Second Validation - Ex #1

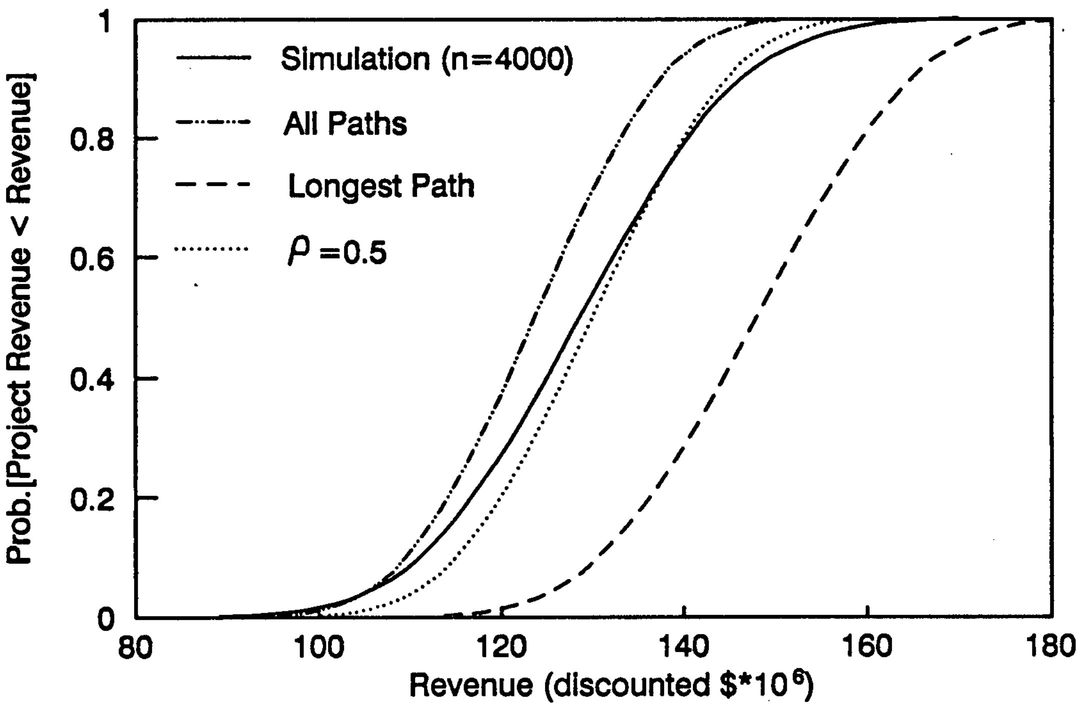

Figure 7.18: CDFs for Discounted Project Revenue - Second Validation - Ex #1

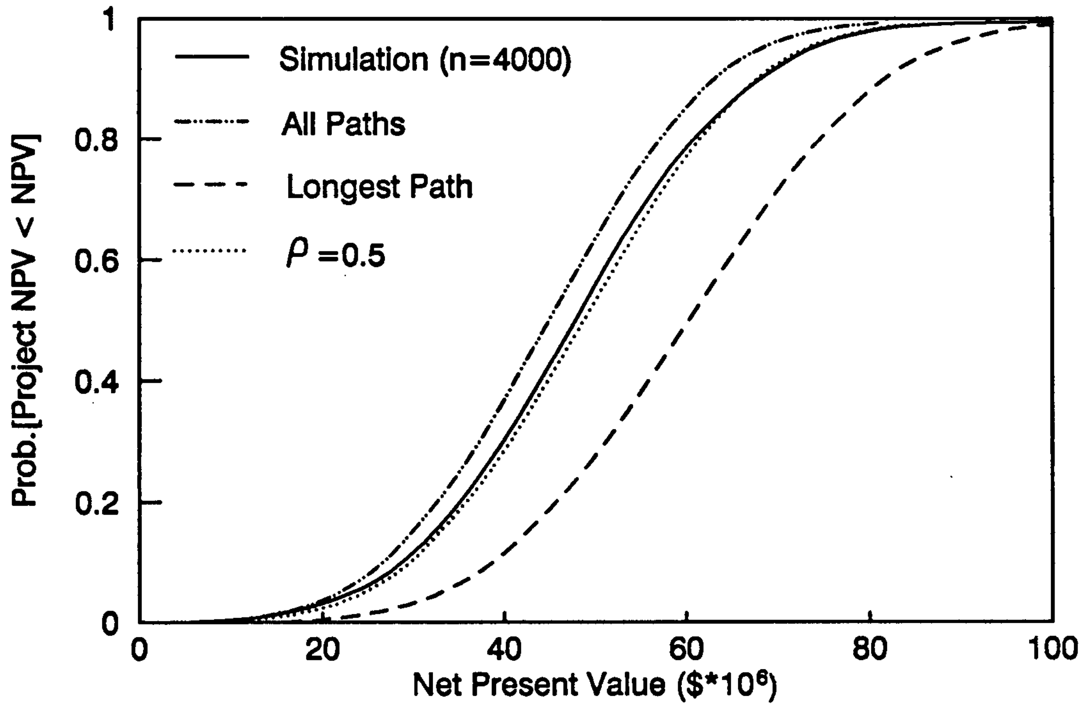

Figure 7.19: CDFs for Project Net Present Value - Second Validation - Ex #1

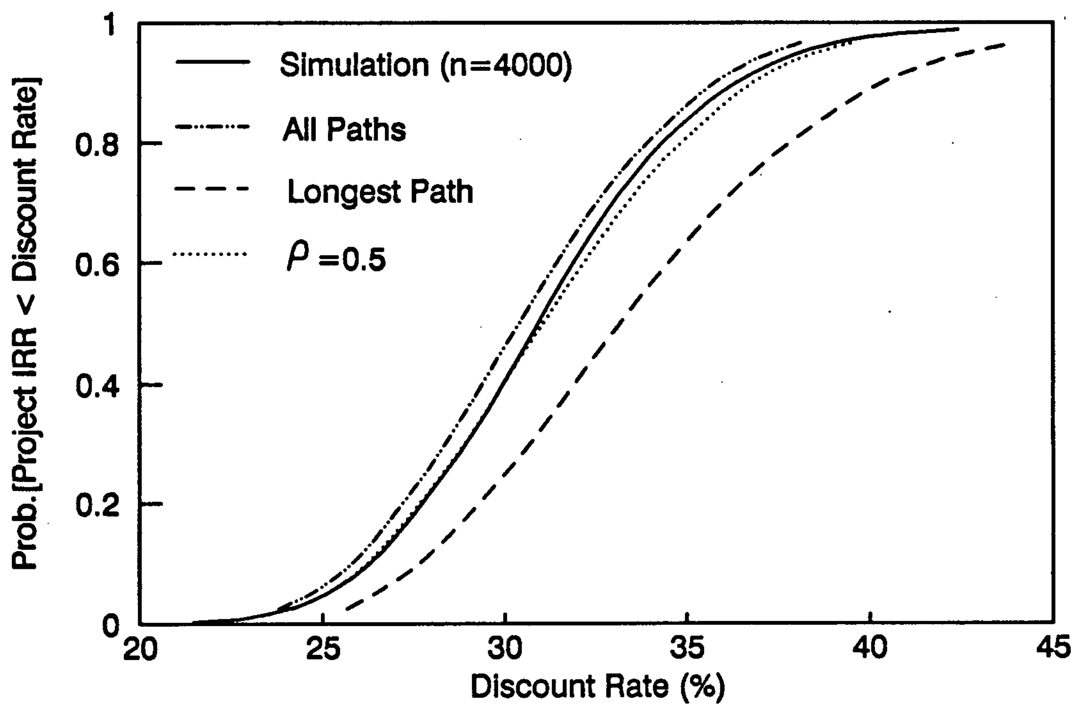

Figure 7.20: CDFs for Project Internal Rate of Return - Second Validation - Ex #1

The analytical method permits the analyst to specify a transitional correlation (ρ) for decision making. The cumulative distribution functions for time and economic variables when $\rho = 0.5$ are included to demonstrate how, in addition to providing a risk quantification at the specified ρ, the analytical method can perform the sensitivity of that quantification by approximating the bounds. It must be stressed that $\rho = 0.5$ is used only as an example, because it is not possible to recommend a single value for ρ that can be used for all risk analyses of engineering projects. The analysis however, can be conducted using the limiting values for ρ, (0,1), as well as an intermediate value (say $\rho = 0.5$). This approach provides the analyst with additional insights and as demonstrated by this example, it is still ten times faster than Monte Carlo simulation.

7.5.4 Discussion

The examples presented in this section validated the analytical method. In addition, the validations clearly demonstrated the computational economy of the analytical method when compared to Monte Carlo simulations. There were 164 random primary variables at the input level for both approaches. The two limiting cases that were used to validate the simulation process also confirmed the theoretical postulations made by the modified PNET algorithm.

The first validation demonstrated the ability of the analytical method to fit easily into the existing deterministic estimation approaches prevalent in the construction industry. This flexibility is important for a theoretical development to become a practical tool in the industry. By considering the deterministic estimates as the median values for the work packages, subjective probabilities can be elicited. This permits the analyst/experts in engineering construction to begin the risk quantification process from the familiar deterministic structure.

Table 7.18 contains a comparison of the deterministic and probabilistic estimates for constant, current and total dollar estimates for project cost. While the deterministic values and the expected values are comparable, it demonstrates that the deterministic values on which most of the decisions are based at present, only have about a 50% probability of success. The quantification of the uncertainty associated with the estimates for project cost permits the contingency to be allocated on the probability of success of the project.

Table 7.18: Deterministic and Probabilistic Analyses of Project Cost

	Deterministic	Probabilistic			
		$E[PC]$	σ_{PC}	$\sqrt{\beta_1}$	β_2
Constant Dollar Cost	124450100	126394711	14041896	0.095	2.61
Current Dollar Cost	137628834	139737616	15742136	0.093	2.59
Total Dollar Cost	151287416	153804634	17036142	0.096	2.61

In addition to validating the analytical method, the second validation demonstrated the significance of the variance of work package durations and start times to the derived economic variables. The bounds of the derived variables are wider when compared to the first validation. The cumulative distribution functions at the transitional correlation $\rho = 0.5$, illustrate the ability of the analytical method to quantify the economic variables for decision making.

Figures (7.21) and (7.22) depict the cumulative distribution functions for $\rho = 0.5$, upper and lower bounds for current and total project costs. Even though, project duration has wide bounds (see figure 7.15), the bounds for current and total dollar project costs are relatively tight. The reason for this phenomena is that, since the start time is one of the six variables in the function for work package cost its significance (sensitivity) is reduced. This is further highlighted in the next example when the start time is one of seventeen variables in the work package cost function.

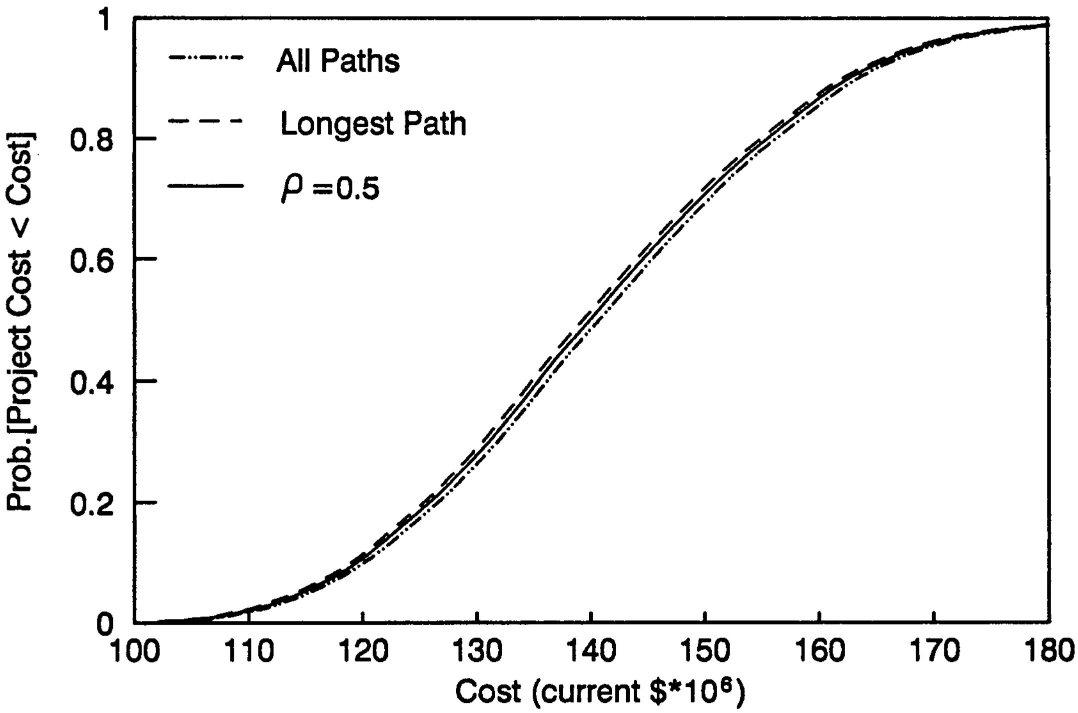

Figure 7.21: CDFs for Current Dollar Project Cost - Second Validation - Ex #1

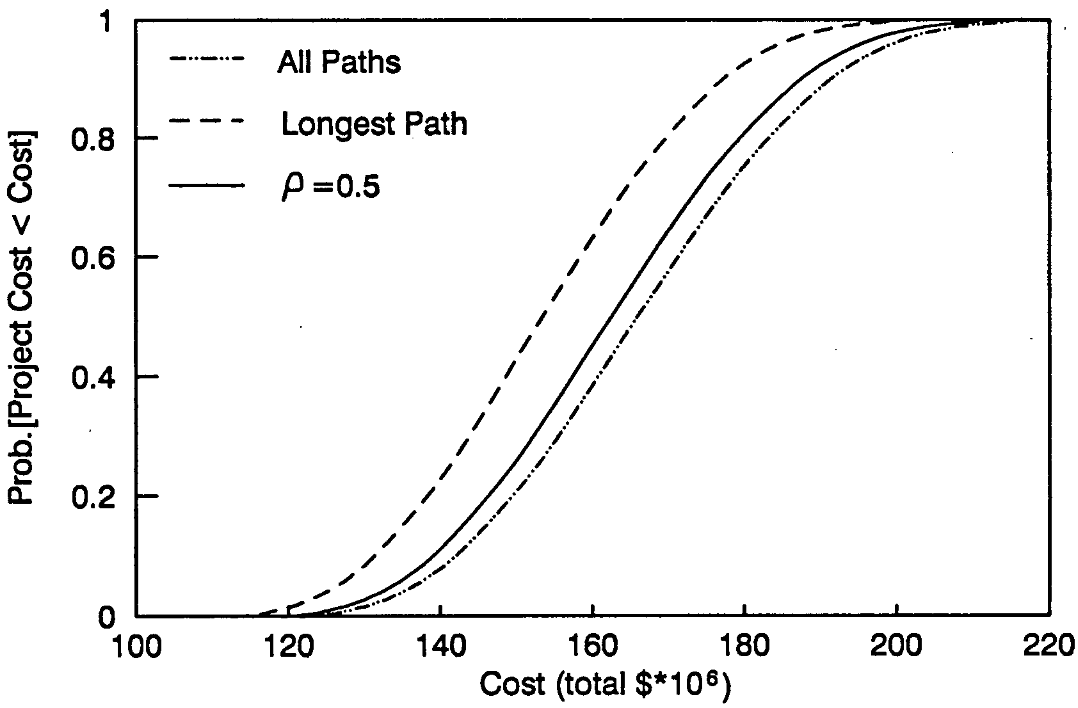

Figure 7.22: CDFs for Total Dollar Project Cost - Second Validation - Ex #1

7.6 Second Example

The second example is a hypothetical engineering project of thirteen work packages and three revenue streams. The precedence network of the work packages is shown in figure (7.23). For illustrative purposes herein the primary variables in the functions for the work package durations, costs and revenue streams are assumed to be stationary over the duration of the work package or revenue stream. In reality these primary variables (labor usage, productivity, inflation and interest rates, etc) are time dependent. The assumption allows the development of simplified but realistic models. The function for work package durations used in this example is as follows.

$$WPD_i = \frac{Q_i}{P_{L_i} \; L_i} \tag{7.5}$$

where Q_i is the quantity descriptor, P_{L_i} is the labour productivity rate and L_i is the labour usage. The function for discounted work package cost (WPC_i) is as follows.

$$\begin{aligned}
WPC_i = \; & f \Bigg[e^{(\theta_{L_i}-y)T_{S_{Ci}}} \int_0^{T_{Ci}} C_{L_i} \; L_i \; e^{(\theta_{L_i}-y)\tau} \, d\tau \\
& + \; e^{(\theta_{M_i}-y)T_{S_{Ci}}} \int_0^{T_{Ci}} C_{M_i} \; P_{L_i} \; L_i \; e^{(\theta_{M_i}-y)\tau} \, d\tau \\
& + \; e^{(\theta_{E_i}-y)T_{S_{Ci}}} \int_0^{T_{Ci}} C_{E_i} \; E_i \; e^{(\theta_{E_i}-y)\tau} \, d\tau \\
& + \; e^{(\theta_{I_i}-y)T_{S_{Ci}}} \int_0^{T_{Ci}} I_{C_i} \; e^{(\theta_{I_i}-y)\tau} \, d\tau \\
& + \; e^{(\theta_{S_i}-y)T_{S_{Ci}}} \int_0^{T_{Ci}} \frac{S_i}{T_{Ci}} \; e^{(\theta_{S_i}-y)\tau} \, d\tau \Bigg] \\
& + (1-f) e^{(r-y)T_p} \Bigg[e^{\theta_{L_i} T_{S_{Ci}}} \int_0^{T_{Ci}} C_{L_i} \; L_i \; e^{(\theta_{L_i}-r)\tau} d\tau \\
& + \; e^{\theta_{M_i} T_{S_{Ci}}} \int_0^{T_{Ci}} C_{M_i} \; P_{L_i} \; L_i \; e^{(\theta_{M_i}-r)\tau} d\tau \\
& + \; e^{\theta_{E_i} T_{S_{Ci}}} \int_0^{T_{Ci}} C_{E_i} \; E_i \; e^{(\theta_{E_i}-r)\tau} d\tau \\
& + \; e^{\theta_{I_i} T_{S_{Ci}}} \int_0^{T_{Ci}} I_{C_i} \; e^{(\theta_{I_i}-r)\tau} d\tau \\
& + \; e^{\theta_{S_i} T_{S_{Ci}}} \int_0^{T_{Ci}} \frac{S_i}{T_{Ci}} \; e^{(\theta_{S_i}-r)\tau} d\tau \Bigg]
\end{aligned} \tag{7.6}$$

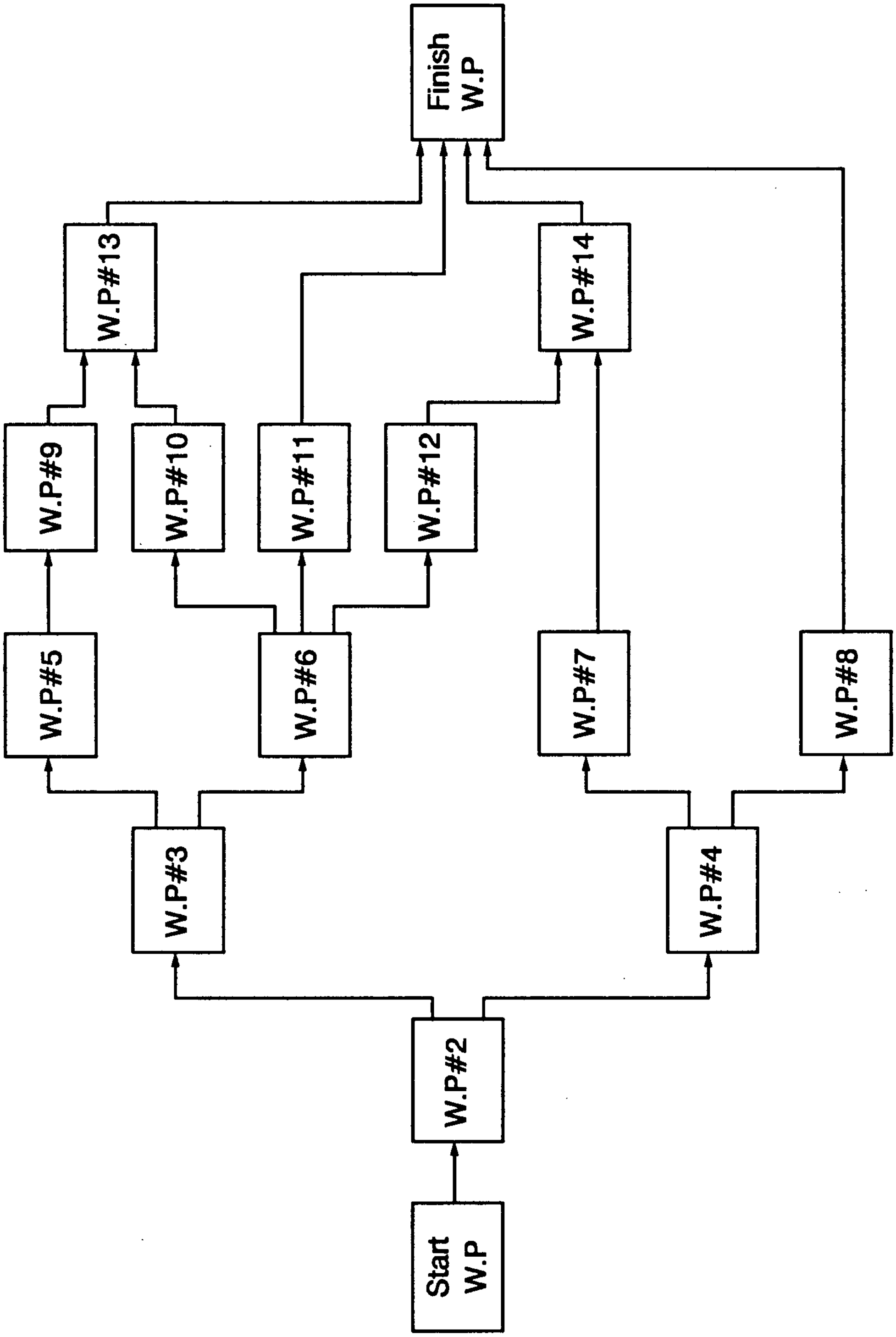

Figure 7.23: The Project Network for the Second Example

where WPC_i is the discounted i^{th} work package cost, C_{L_i}, C_{M_i}, and C_{E_i} are the unit rates for labour, materials and equipment, L_i and E_i are the labour and equipment usage profiles, P_{L_i} is the labour productivity rate, I_{C_i} and S_i are the indirect and sub contractor costs assumed as uniform constant dollar profiles, θ_{L_i}, θ_{M_i}, θ_{E_i}, θ_{I_i} and θ_{S_i} are inflation rates for labour, materials, equipment, indirect cost and sub contractor cost, and $T_{S_{Ci}}$ and T_{Ci} are work package start time and duration for the i^{th} work package respectively. T_p is the time at which the repayment of interim financing is due for all work packages (assumed as the end of the construction phase), f is the equity fraction, and r and y are interest and discount rates respectively. The time τ is measured from the start of the i^{th} work package.

The function for revenue streams (NRS_i) is as follows.

$$NRS_i = \int_{T_{S_{Ri}}}^{T_{S_{Ri}}+T_{Ri}} \left[R_{0i}e^{\theta_{R_i}(t-T_{S_{Ri}})} - M_{0i}e^{\theta_{M_i}t}\right] e^{-yt}dt \tag{7.7}$$

where NRS_i is the discounted i^{th} revenue stream, R_{0i} and M_{0i} are the constant dollar cash flow for the i^{th} gross revenue and operation and maintenance cost assumed as uniform profiles, $T_{S_{Ri}}$ and T_{Ri} are start time and duration of the revenue stream, and $\theta_{R_i}, \theta_{M_i}$ and y are inflation and discount rates respectively.

The expected value, standard deviation, skewness and kurtosis for individual primary variables in the functions for work package durations, costs, and revenue streams are given in Appendix F. A construction loan for 75% ($f = 0.25$) of the current dollar expenditure on construction is assumed. The statistics for the interest rate on the construction loan are $E[r] = 7.537\%, \sigma_r = 0.852\%, \sqrt{\beta_1} = 0.2$ and $\beta_2 = 2.5$. A minimum attractive rate of return of 9% is used for the analysis and validation.

This section demonstrates third and fourth validations of the analytical method. The third validation assumes all the primary variables to be uncorrelated. The fourth validation treats linear correlations between the primary variables. The example is extended to a third level where correlations at all levels of the project economic

Table 7.19: Statistics for Project Duration from Third Validation - Ex #2

Simulation			Analytical Method				
#	$E[PD]$ mths	σ_{PD} mths	ρ	$E[PD]$ mths	σ_{PD} mths	$\sqrt{\beta_1}$	β_2
500	30.73	4.24	0.0	29.44	4.69	0.34	2.77
1000	30.85	4.40	0.1	29.44	4.69	0.34	2.77
1500	30.89	4.37	0.2	29.44	4.69	0.34	2.77
2000	30.92	4.42	0.3	29.44	4.69	0.34	2.77
2500	30.94	4.46	0.4	29.78	4.53	0.4	2.9
3000	30.98	4.50	0.5	31.84	4.06	0.3	2.9
3500	30.98	4.55	0.6	31.84	4.06	0.3	2.9
4000	31.01	4.62	0.7	31.94	3.98	0.3	2.8
			0.8	32.04	3.90	0.4	3.1
			0.9	32.42	3.69	0.4	3.2
			1.0	32.42	3.69	0.4	3.2

structure are treated.

7.6.1 Third Validation

A simulation for complete time and economic risk quantification was done for the third validation of the analytical method. For this simulation all of the variables were assumed to be uncorrelated. Tables 7.19, 7.20, 7.21 and 7.22 contain results from the simulation at 500 iteration intervals and statistics from the analytical method at different transitional correlation values for project duration, discounted project cost, discounted project revenue and project net present value. Tables 7.23 contains the expected value and standard deviation for project internal rate of return from the simulation and the analytical method at different transitional correlation values.

Table 7.20: Statistics for Discounted Project Cost from Third Validation - Ex #2

Simulation			Analytical Method				
#	$E[DPC]$	σ_{DPC}	ρ	$E[DPC]$	σ_{DPC}	$\sqrt{\beta_1}$	β_2
	\$	\$		\$	\$		
500	47747712	7272635	0.0	47656668	6800455	0.188	2.724
1000	47549456	7255953	0.1	47656668	6800455	0.188	2.724
1500	47539104	7097718	0.2	47656668	6800455	0.188	2.724
2000	47484848	7290801	0.3	47656668	6800455	0.188	2.724
2500	47574352	7313699	0.4	47642261	6798415	0.188	2.724
3000	47586560	7287798	0.5	47548355	6784741	0.188	2.724
3500	47592672	7249368	0.6	47548355	6784741	0.188	2.724
4000	47623920	7317997	0.7	47544054	6784131	0.188	2.724
			0.8	47535878	6783307	0.188	2.724
			0.9	47519155	6780948	0.188	2.724
			1.0	47519155	6780948	0.188	2.724

Table 7.21: Statistics for Discounted Project Revenue from Third Validation - Ex #2

Simulation			Analytical Method				
#	$E[DPR]$	σ_{DPR}	ρ	$E[DPR]$	σ_{DPR}	$\sqrt{\beta_1}$	β_2
	\$	\$		\$	\$		
500	69621328	13437693	0.0	70266290	13744103	-0.431	4.718
1000	69777088	13114197	0.1	70266290	13744103	-0.431	4.718
1500	69942976	13445692	0.2	70266290	13744103	-0.431	4.718
2000	70071568	13397565	0.3	70266290	13744103	-0.431	4.718
2500	70116800	13345129	0.4	70183320	13726601	-0.431	4.717
3000	70091024	13409836	0.5	69672971	13624327	-0.429	4.706
3500	69932000	13602733	0.6	69672971	13624327	-0.429	4.706
4000	69941072	13626623	0.7	69648247	13618945	-0.429	4.706
			0.8	69624938	13613874	-0.429	4.706
			0.9	69529211	13593995	-0.428	4.705
			1.0	69529211	13593995	-0.428	4.705

Table 7.22: Statistics for Project NPV from Third Validation - Ex #2

Simulation			Analytical Method				
#	$E[NPV]$ \$	σ_{NPV} \$	ρ	$E[NPV]$ \$	σ_{NPV} \$	$\sqrt{\beta_1}$	β_2
500	21872384	15545242	0.0	22609623	15334489	-0.327	4.098
1000	22226560	15178882	0.1	22609623	15334489	-0.327	4.098
1500	22411568	15381492	0.2	22609623	15334489	-0.327	4.098
2000	22592032	15317318	0.3	22609623	15334489	-0.327	4.098
2500	22546816	15415091	0.4	22541059	15317899	-0.327	4.097
3000	22507776	15496218	0.5	22124616	15220217	-0.324	4.085
3500	22338016	15561380	0.6	22124616	15220217	-0.324	4.085
4000	22312432	15657844	0.7	22104194	15215127	-0.324	4.084
			0.8	22089060	15210220	-0.324	4.084
			0.9	22010056	15191377	-0.324	4.082
			1.0	22010056	15191377	-0.324	4.082

Table 7.23: Statistics for Project IRR from Third Validation - Ex #2

Simulation			Analytical Method		
#	$E[IRR]$	σ_{IRR}	ρ	$E[IRR]$	σ_{IRR}
500	16.13	5.07	0.0	16.305	5.061
1000	16.26	4.94	0.1	16.305	5.061
1500	16.29	5.00	0.2	16.305	5.061
2000	16.35	4.99	0.3	16.305	5.061
2500	16.32	5.02	0.4	16.268	5.044
3000	16.30	5.03	0.5	16.061	4.955
3500	16.23	5.04	0.6	16.061	4.955
4000	16.23	5.07	0.7	16.052	4.953
			0.8	16.046	4.952
			0.9	16.014	4.944
			1.0	16.014	4.944

Figure (7.24) illustrates the cumulative distribution functions for upper and lower bounds for project duration approximated from the analytical method and that generated from the simulation of 4,000 iterations. Figure (7.25) depicts, in addition to those in figure (7.24), the cumulative distribution functions for project duration at different transitional correlation values. Figures (7.26), (7.27), (7.28) and (7.29) illustrate the cumulative distribution functions for upper and lower bounds approximated from the analytical method and those generated from the simulation for discounted project cost, revenue, net present value and internal rate of return.

The third validation also demonstrates that cumulative distribution functions and the estimates for expected values for time and economic variables generated from the simulation are within the upper and lower bounds predicted by the analytical method. Thereby, validating the analytical method. The bounds for the derived economic variables are extremely tight. These bounds are the sensitivity of the derived economic variables with respect to the start times of the work packages. Table 7.24 contains the comparison of the execution time for the simulation and the analytical method. The computational economy of the analytical method is again highlighted. For this example the analytical method is about fifty times faster when compared to 1000 iterations from the Monte Carlo simulation.

7.6.2 Fourth Validation

The fourth validation of the analytical method was also a complete time and economic risk quantification. The only change from the previous section is that the primary variables in the functions for work package durations and costs are considered to be correlated. The positive definite correlation matrices for work package duration and cost functions were obtained using the process described in sections 4.2.2.

Since the function given by equation (7.5) is used to evaluate the work package

Table 7.24: Comparison of CPU times from Third Validation - Ex #2

Simulation		Analytical Method	
#	CPU Sec.	ρ	CPU Sec.
500	861	0.0	28.64
1000	1721	0.1	28.60
1500	2577	0.2	28.67
2000	3427	0.3	28.57
2500	4275	0.4	27.97
3000	5121	0.5	28.05
3500	5969	0.6	28.11
4000	6819	0.7	28.28
		0.8	28.38
		0.9	28.23
		1.0	28.10

durations, an identical positive definite correlation matrix was used for all the work package durations. This simplification is also convenient when the correlations between the derived variables are approximated using the identified common (shared) primary variables. The positive definite correlation matrix for work package durations and the positive definite correlation matrix for work package costs used for all the work packages in this application are given in Appendix F.

Even though the function for work package costs given by equation (7.6) has seventeen variables, the positive definite correlation matrix is only 14x14. The reason is because three variables - work package duration, start time, and project duration are always pre-defined in the decomposed function for work package cost. Their moments are evaluated from the modified PNET algorithm. Therefore, the correlation matrix is only of the variables that are elicited as input primary variables. The computer program 'ELICIT' ensures that there is no confusion during the elicitation process by identifying the pre-defined variables as the first three variables of the decomposed function (see equation 6.7).

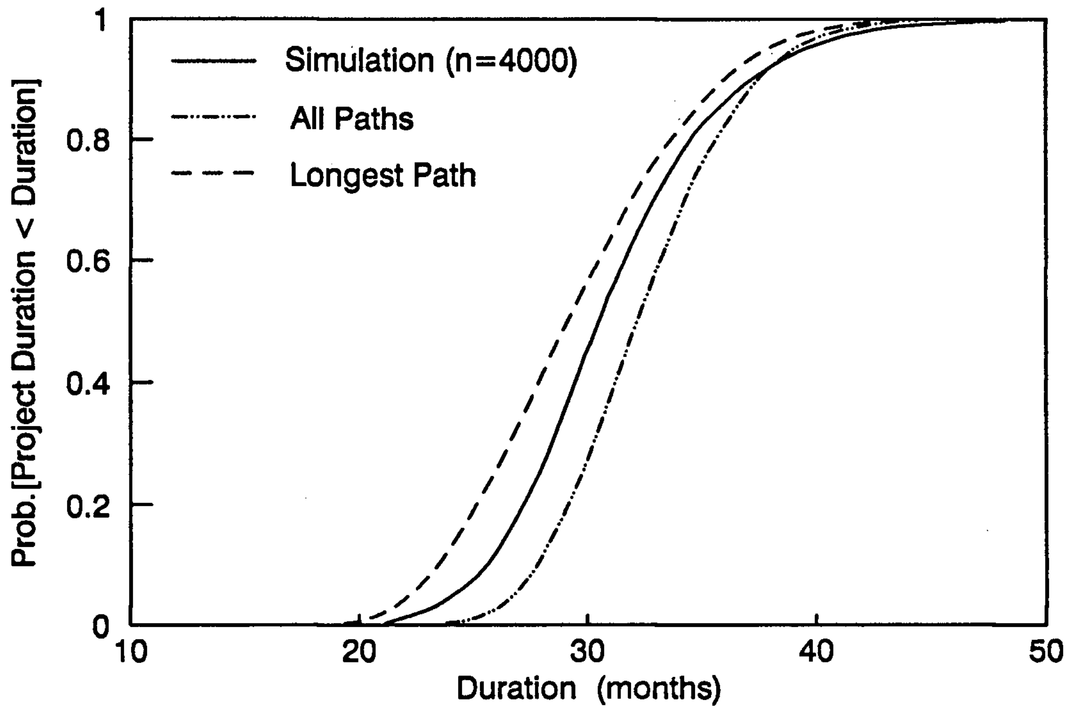

Figure 7.24: CDFs for Project Duration - Third Validation - Ex #2

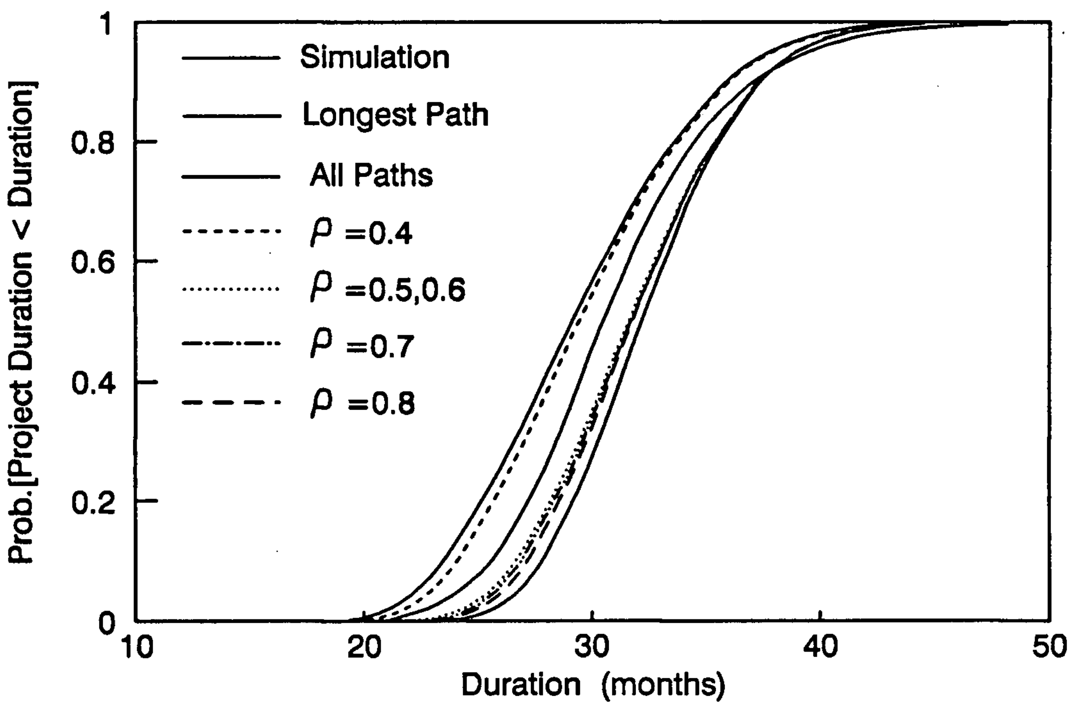

Figure 7.25: CDFs for Project Duration - Third Validation - Ex #2

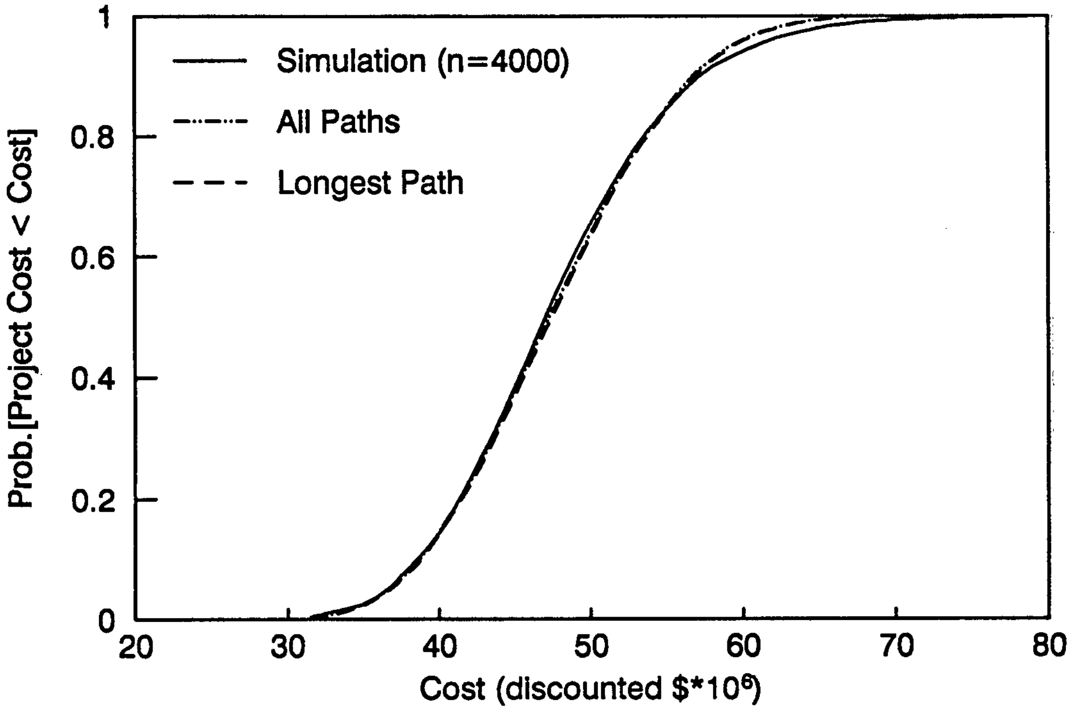

Figure 7.26: CDFs for Discounted Project Cost - Third Validation - Ex #2

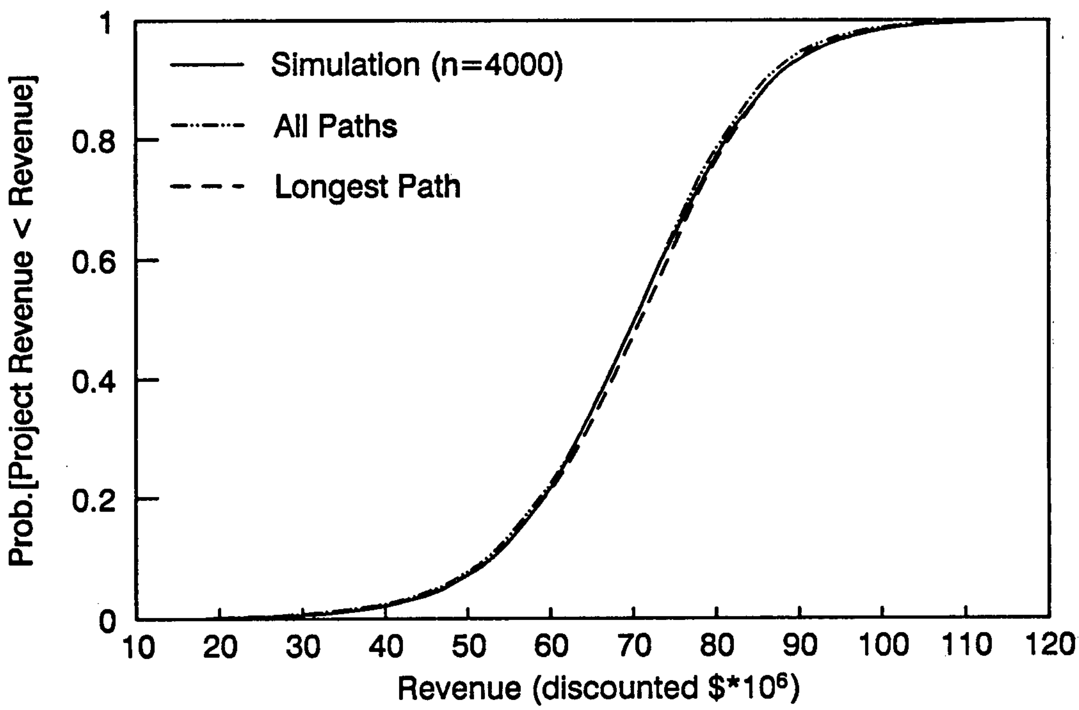

Figure 7.27: CDFs for Discounted Project Revenue - Third Validation - Ex #2

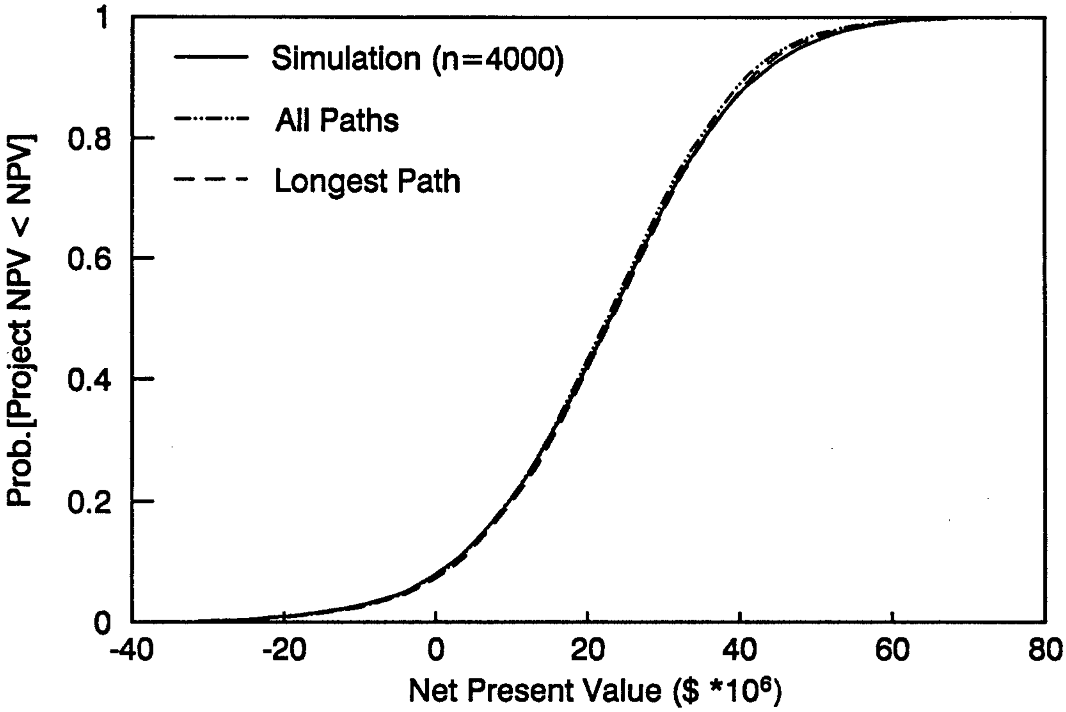

Figure 7.28: CDFs for Project Net Present Value - Third Validation - Ex #2

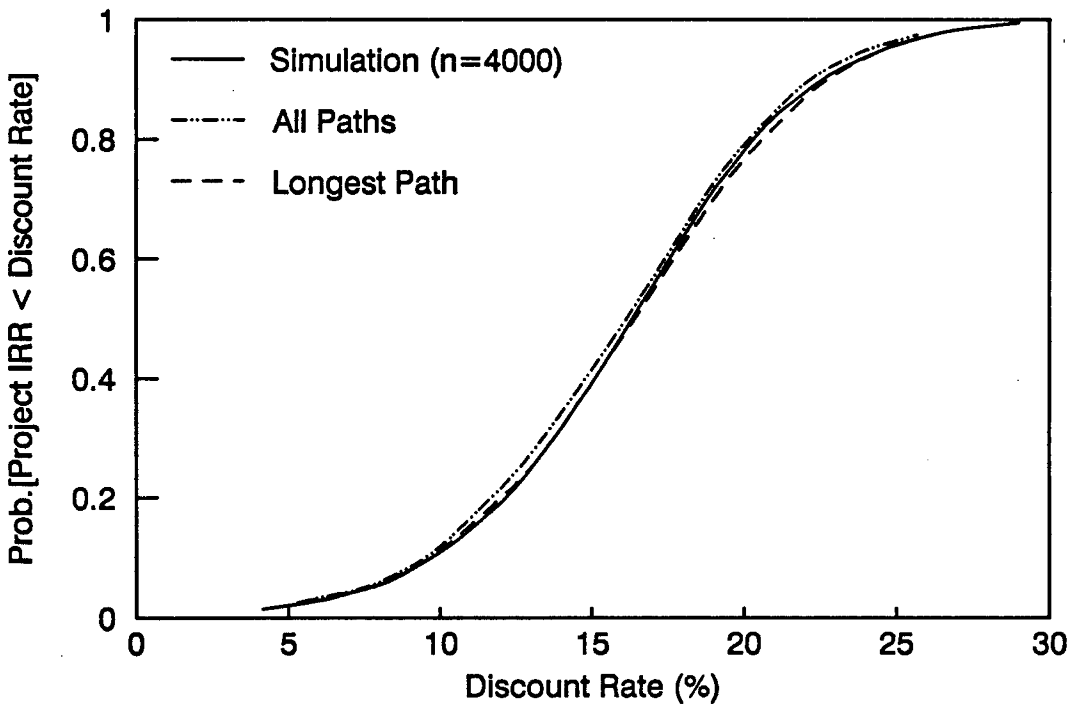

Figure 7.29: CDFs for Project Internal Rate of Return - Third Validation - Ex #2

Table 7.25: Statistics for Project Duration from Fourth Validation - Ex #2

Simulation			Analytical Method				
#	$E[PD]$ *mths*	σ_{PD} *mths*	ρ	$E[PD]$ *mths*	σ_{PD} *mths*	$\sqrt{\beta_1}$	β_2
1000	29.34	3.05	0.0	29.31	4.59	0.37	2.81
2000	29.41	3.00	0.1	29.31	4.59	0.37	2.81
3000	29.44	3.04	0.2	29.31	4.59	0.37	2.81
4000	29.48	3.17	0.3	29.31	4.59	0.37	2.81
5000	29.51	3.22	0.4	29.57	4.46	0.4	2.9
6000	29.47	3.20	0.5	31.65	4.03	0.3	2.9
			0.6	31.65	4.03	0.3	2.9
			0.7	31.77	3.94	0.4	3.1
			0.8	31.83	3.88	0.4	3.0
			0.9	32.21	3.68	0.4	3.1
			1.0	32.21	3.68	0.4	3.1

Tables 7.25 contains results from the simulation at 1000 iteration intervals and statistics from the analytical method at different transitional correlation values for project duration. Figure (7.30) illustrates the cumulative distribution functions for upper and lower bounds for project duration and that generated from the simulation of 6,000 iterations. Figure (7.31) depicts the cumulative distribution functions for project duration at different transitional correlation values.

The cumulative distribution function from the simulation is comparatively tight with the upper part outside of the bounds predicted from the analytical method. When the standard deviations from Table 7.19 and 7.25 are compared, the analytical method shows a small reduction while the simulation shows a significant dampening which can be attributed to the approach used to treat correlations. The approach used does not distinguish between positive and negative correlations. This dampening caused the distribution to be outside the bounds. However, a study of the individual work package durations show that there should not be a significant reduction in the variance for project duration.

Table 7.26 contains the expected values, standard deviations and the differences when the primary variables are uncorrelated and correlated for individual work packages. The difference of all the expected values are negligible. Except for work packages # 6, 7, 8 and 10 the difference in the standard deviations are small. When these work packages are studied in the context of the paths in the project network (see next section) and their contributions to path variances, except for the third longest path which has work packages 6 and 10, none of the others have more than one of the above four. In addition, their contributions to path variances are small. Hence, none of the paths can have a significant reduction in variance from the uncorrelated case to the correlated case.

Table 7.26: Statistics for Project Variables

	Expected Value (months)			Standard Deviation (months)		
WP#	Uncor	Corr	Differ	Uncor	Corr	Differ
02	7.715	7.681	-0.44%	2.717	2.625	- 3.38%
03	4.971	4.952	-0.38%	1.311	1.267	- 3.35%
04	6.949	6.918	-0.45%	2.451	2.368	- 3.38%
05	3.363	3.356	-0.21%	1.046	1.047	0.1 %
06	3.44	3.363	-0.22%	1.118	0.90	-19.38%
07	1.732	1.687	-0.26%	0.667	0.56	-16.04%
08	6.634	6.442	-0.29%	2.251	1.702	-24.39%
09	5.812	5.795	-0.29%	2.039	1.999	- 1.96%
10	2.748	2.686	-0.23%	0.899	0.725	-19.35%
11	4.571	4.55	-0.46%	1.185	1.114	- 6.0 %
12	6.979	6.959	-0.29%	2.453	2.405	- 1.95%
13	6.797	6.794	-0.05%	2.399	2.442	1.79%
14	6.337	6.349	0.18%	2.372	2.442	2.95%

Tables 7.27, 7.28 and 7.29 contain results from the simulation at 1000 iteration intervals and statistics from the analytical method at different transitional correlation values for discounted project cost, discounted project revenue and project net present value. Tables 7.30 contains the expected value and standard deviation for project

Table 7.27: Statistics for Discounted Project Cost from Fourth Validation-Ex #2

Simulation			Analytical Method				
#	$E[DPC]$ \$	σ_{DPC} \$	ρ	$E[DPC]$ \$	σ_{DPC} \$	$\sqrt{\beta_1}$	β_2
1000	46114784	5724537	0.0	46808547	6422711	0.214	2.742
2000	46077568	5668257	0.1	46808547	6422711	0.214	2.742
3000	46144304	5705700	0.2	46808547	6422711	0.214	2.742
4000	46200736	5750056	0.3	46808547	6422711	0.214	2.742
5000	46197600	5783185	0.4	46797312	6421195	0.214	2.742
6000	46228352	5786760	0.5	46705300	6408483	0.214	2.742
			0.6	46705300	6408483	0.214	2.742
			0.7	46700307	6407809	0.214	2.742
			0.8	46693962	6407225	0.214	2.742
			0.9	46678295	6405120	0.214	2.742
			1.0	46678295	6405120	0.214	2.742

internal rate of return from the simulation and the analytical method at different transitional correlation values. Figures (7.32), (7.33), (7.34) and (7.35) illustrate the cumulative distribution functions for upper and lower bounds approximated from the analytical method and those generated from the simulation for discounted project cost, revenue, net present value and internal rate of return.

The estimates and cumulative distribution functions for time and economic variables are reasonably close to the predicted envelope of bounds. It must be noted that the bounds are again extremely tight. In addition to start time being one of the seventeen variables in the work package cost functions, the project network given by figure (7.23) is also small, with few interrelationships between work packages. The combination of these two factors increase the tightness of the bounds.

Table 7.31 contains the comparison of the execution time for the simulation and the analytical method. The computational economy of the analytical method is again highlighted. The execution times for Monte Carlo simulation from third and fourth validations are similar. The reason for the similarity is because the same computer

Table 7.28: Statistics for Discounted Project Revenue from Fourth Validation-Ex #2

Simulation			Analytical Method				
#	$E[DPR]$	σ_{DPR}	ρ	$E[DPR]$	σ_{DPR}	$\sqrt{\beta_1}$	β_2
	\$	\$		\$	\$		
1000	70145200	13176377	0.0	70299596	13749690	-0.431	4.720
2000	70439824	13459856	0.1	70299596	13749690	-0.431	4.720
3000	70465632	13466430	0.2	70299596	13749690	-0.431	4.720
4000	70314480	13685384	0.3	70299596	13749690	-0.431	4.720
5000	70220368	13688517	0.4	70233575	13735779	-0.431	4.719
6000	70297504	13691024	0.5	69720113	13633218	-0.429	4.708
			0.6	69720113	13633218	-0.429	4.708
			0.7	69690857	13626900	-0.429	4.708
			0.8	69674338	13623299	-0.429	4.708
			0.9	69582818	13604284	-0.428	4.707
			1.0	69582818	13604284	-0.428	4.707

Table 7.29: Statistics for Project NPV from Fourth Validation - Ex #2

Simulation			Analytical Method				
#	$E[NPV]$	σ_{NPV}	ρ	$E[NPV]$	σ_{NPV}	$\sqrt{\beta_1}$	β_2
	\$	\$		\$	\$		
1000	24029376	14524955	0.0	23419049	15175809	-0.337	4.151
2000	24368384	14635079	0.1	23419049	15175809	-0.337	4.151
3000	24323088	14750314	0.2	23419049	15175809	-0.337	4.151
4000	24106912	14934437	0.3	23419049	15175809	-0.337	4.151
5000	24010976	14976197	0.4	23436263	15162564	-0.337	4.150
6000	24053856	14972542	0.5	23014812	15064305	-0.334	4.137
			0.6	23014812	15064305	-0.334	4.137
			0.7	22990550	15058301	-0.334	4.137
			0.8	22980376	15054793	-0.334	4.137
			0.9	22904523	15036692	-0.334	4.135
			1.0	22904523	15036692	-0.334	4.135

Table 7.30: Statistics for Project IRR from Fourth Validation - Ex #2

Simulation			Analytical Method		
#	$E[IRR]$	σ_{IRR}	ρ	$E[IRR]$	σ_{IRR}
1000	16.90	4.69	0.0	16.654	5.026
2000	16.97	4.73	0.1	16.654	5.026
3000	16.94	4.75	0.2	16.654	5.026
4000	16.86	4.80	0.3	16.654	5.026
5000	16.83	4.82	0.4	16.628	5.014
6000	16.83	4.81	0.5	16.427	4.912
			0.6	16.427	4.912
			0.7	16.413	4.902
			0.8	16.408	4.898
			0.9	16.366	4.869
			1.0	16.366	4.869

Table 7.31: Comparison of CPU times from Fourth Validation - Ex #2

Simulation		Analytical Method	
#	CPU Sec.	ρ	CPU Sec.
1000	1702	0.0	29.00
2000	3395	0.1	29.03
3000	5085	0.2	29.12
4000	6775	0.3	29.08
5000	8465	0.4	29.11
6000	10168	0.5	28.41
		0.6	28.40
		0.7	28.61
		0.8	28.72
		0.9	28.71
		1.0	28.83

program was used for both simulations. The only difference in the input was the use of identity matrices as the correlation matrices for the third validation and positive definite correlation matrices for the fourth. One could argue that if the random number modification algorithm was not included in the computer program for the third validation, the simulation could have been more efficient.

7.6.3 Correlations at All Levels of the Project

The analytical method can approximate the linear correlations between derived variables using the linear correlations between primary variables when the common (shared) variables are identified (see section 4.3). The common variables are defined as those variables of the same type having the same first four moments in the functional forms for two or more derived variables. Hence, correlation between work package durations, costs and revenue streams can be treated in the evaluation of the first four moments of project duration, cost and revenue. As demonstrated in this section, the contribution to the moments of the derived variables from these correlations can be significant.

It is not possible to duplicate this treatment for the Monte Carlo simulation from the available information. The simulation assumes that only the linear correlations between the primary variables in the functions are available (see figure 7.36). However, if it is possible to obtain the correlation matrix for the complete system, then simulation can treat all the correlations. For project cost in this example, one would have to elicit a $(182 - l)\text{x}(182 - l)$ positive definite correlation matrix, where l is the total number of common variables in the functions for derived variables minus the number of sets of common variables.

Even though linear correlation coefficients between all the work package durations are approximated, only the positive definite correlation matrices for the individual

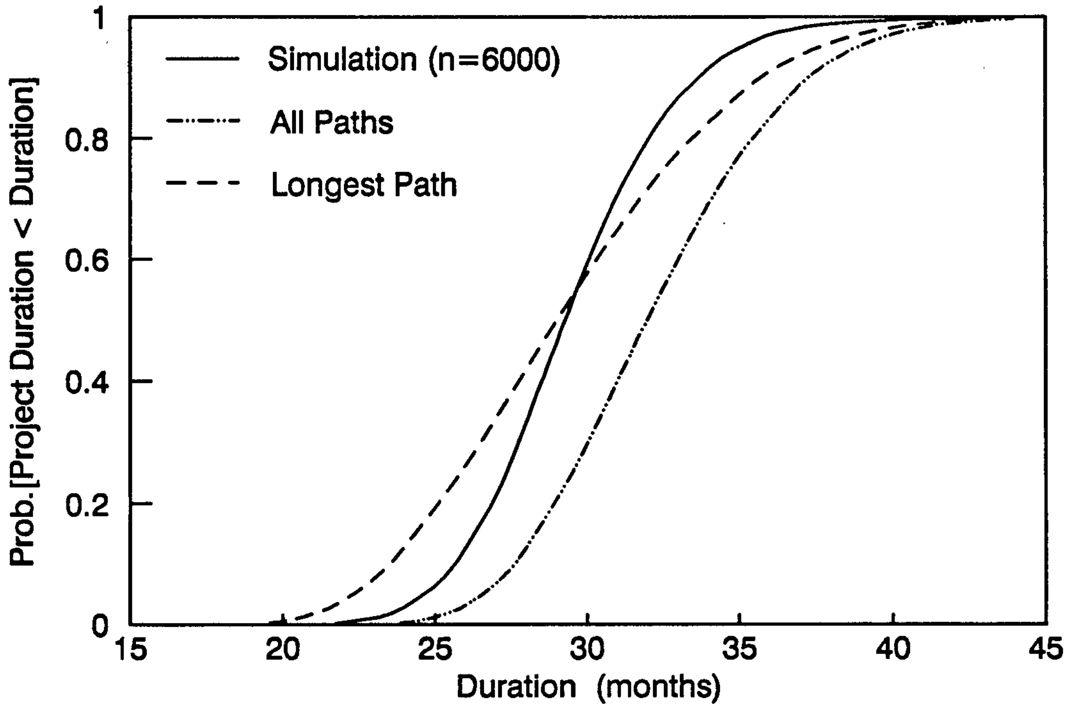

Figure 7.30: CDFs for Project Duration - Fourth Validation - Ex #2

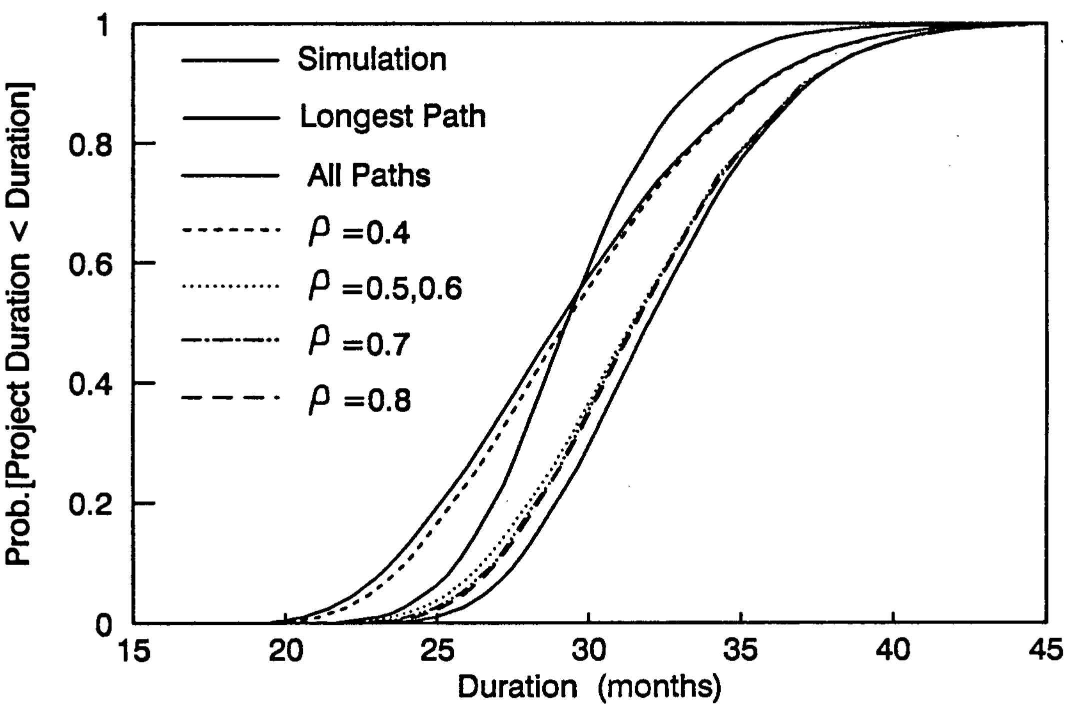

Figure 7.31: CDFs for Project Duration - Fourth Validation - Ex #2

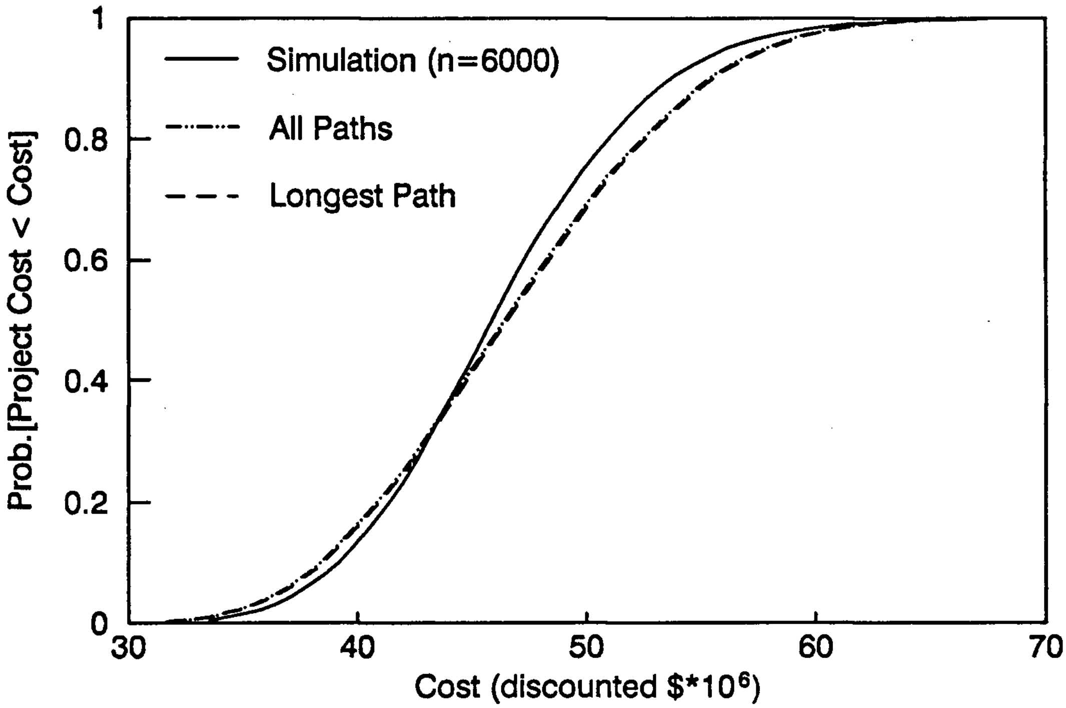

Figure 7.32: CDFs for Discounted Project Cost - Fourth Validation - Ex #2

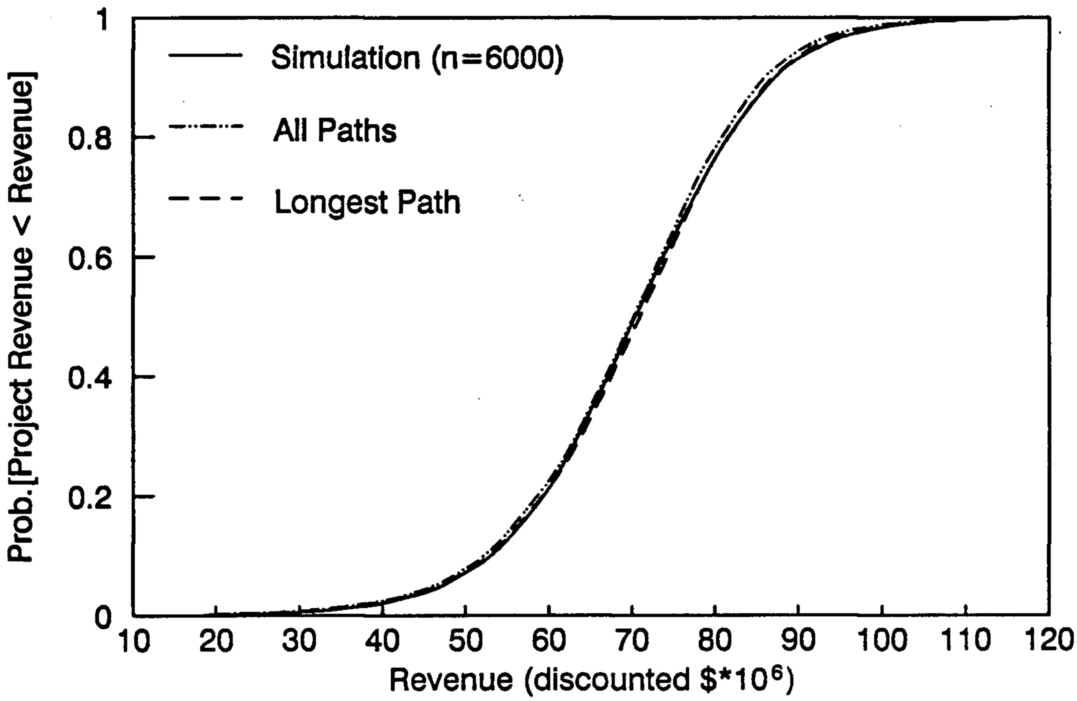

Figure 7.33: CDFs for Discounted Project Revenue - Fourth Validation - Ex #2

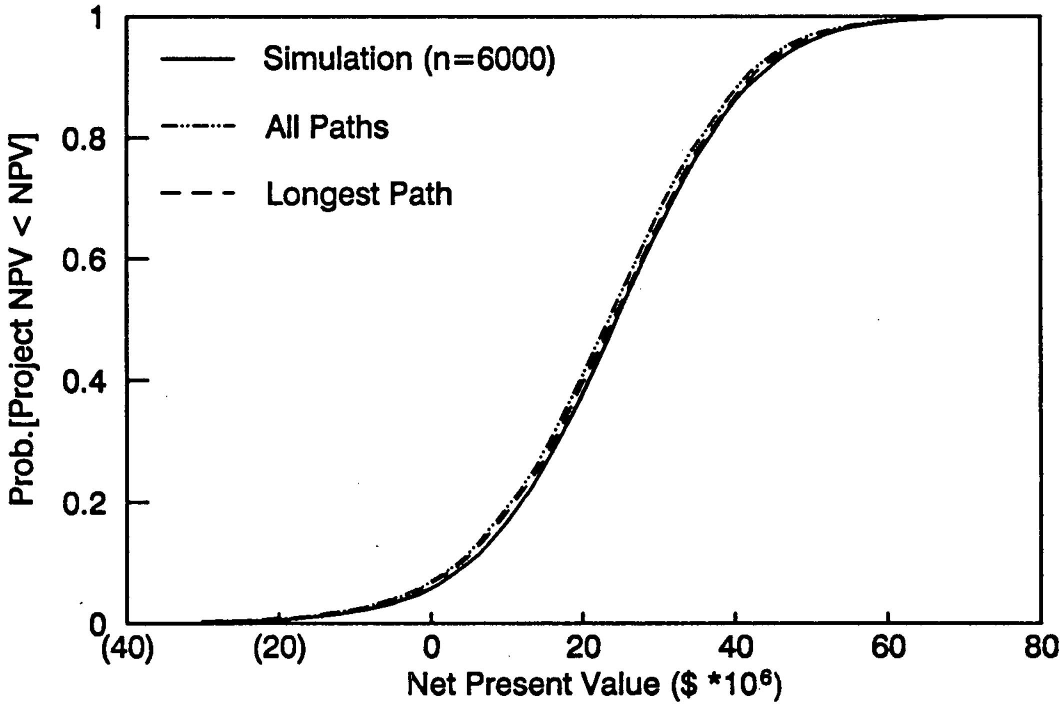

Figure 7.34: CDFs for Project Net Present Value - Fourth Validation - Ex #2

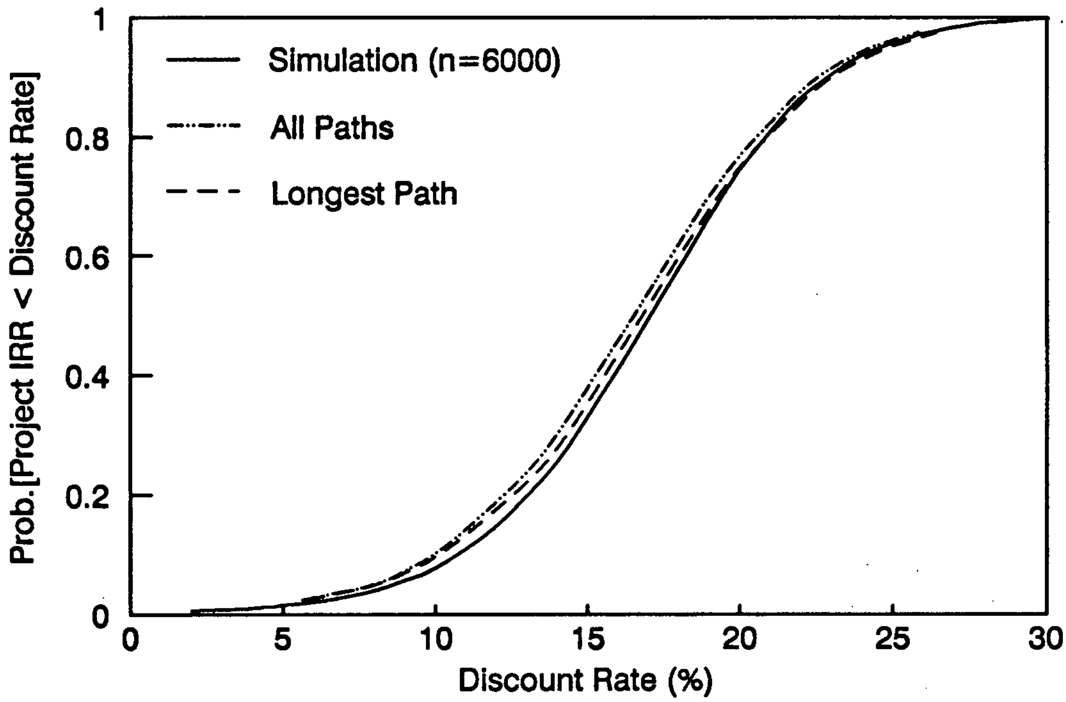

Figure 7.35: CDFs for Project Internal Rate of Return - Fourth Validation - Ex #2

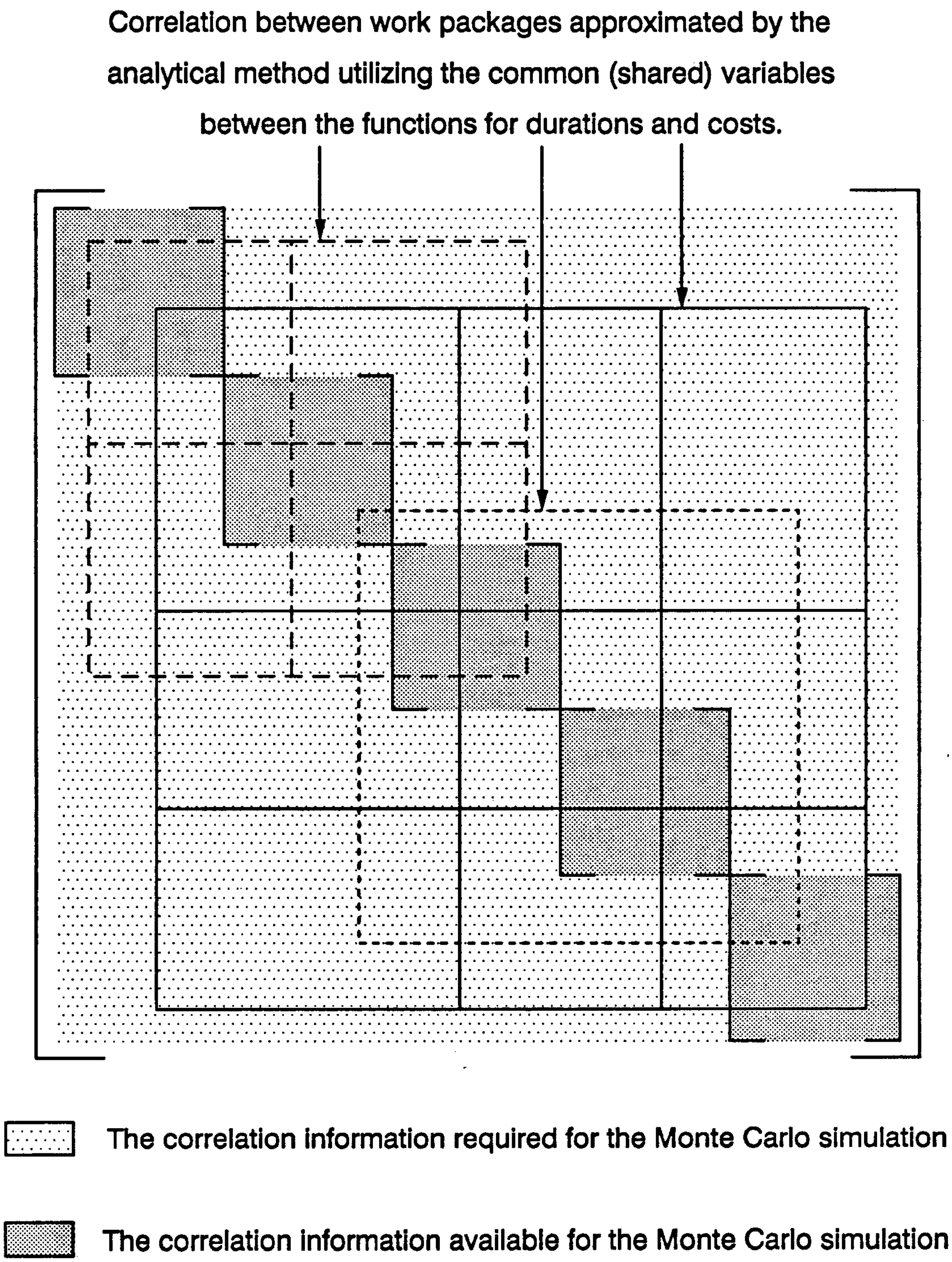

Figure 7.36: Correlation Matrix for the Complete System

paths to a work package are necessary to evaluate the first four moments of its start time. For example, there are six paths to complete the project illustrated by figure (7.23). The positive definite correlation matrices for the six paths approximated by the analytical method are as follows. (The paths are ordered in decreasing mean path durations).

Path #1 - Work Packages # 2, # 3, # 6, # 12, # 14.

$$\mathbf{R}_{\mathbf{Path\#1}} = \begin{bmatrix} 1.0 & 0.0 & 0.0 & 0.3 & 0.0 \\ 0.0 & 1.0 & 0.53 & 0.0 & 0.59 \\ 0.0 & 0.53 & 1.0 & 0.0 & 0.45 \\ 0.3 & 0.0 & 0.0 & 1.0 & -0.19 \\ 0.0 & 0.59 & 0.45 & -0.19 & 1.0 \end{bmatrix}$$

Path #2 - Work Packages # 2, # 4, # 7, # 14.

$$\mathbf{R}_{\mathbf{Path\#2}} = \begin{bmatrix} 1.0 & 0.3 & 0.0 & 0.0 \\ 0.3 & 1.0 & 0.0 & 0.79 \\ 0.0 & 0.0 & 1.0 & 0.42 \\ 0.0 & 0.79 & 0.42 & 1.0 \end{bmatrix}$$

Path #3 - Work Packages # 2, # 3, # 6, # 10, # 13.

$$\mathbf{R}_{\mathbf{Path\#3}} = \begin{bmatrix} 1.0 & 0.0 & 0.0 & 0.0 & 0.0 \\ 0.0 & 1.0 & 0.53 & 0.53 & 0.0 \\ 0.0 & 0.53 & 1.0 & 0.36 & 0.0 \\ 0.0 & 0.53 & 0.36 & 1.0 & 0.0 \\ 0.0 & 0.0 & 0.0 & 0.0 & 1.0 \end{bmatrix}$$

Path #4 - Work Packages # 2, # 3, # 5, # 9, # 13.

$$\mathbf{R}_{\mathbf{Path\#4}} = \begin{bmatrix} 1.0 & 0.0 & 0.0 & 0.92 & 0.0 \\ 0.0 & 1.0 & 0.0 & 0.0 & 0.0 \\ 0.0 & 0.0 & 1.0 & 0.0 & 0.0 \\ 0.92 & 0.0 & 0.0 & 1.0 & 0.0 \\ 0.0 & 0.0 & 0.0 & 0.0 & 1.0 \end{bmatrix}$$

Path #5 - Work Packages # 2, # 3, # 6, # 11.

$$\mathbf{R}_{\mathbf{Path\#5}} = \begin{bmatrix} 1.0 & 0.0 & 0.0 & 0.0 \\ 0.0 & 1.0 & 0.53 & 0.0 \\ 0.0 & 0.53 & 1.0 & 0.0 \\ 0.0 & 0.0 & 0.0 & 1.0 \end{bmatrix}$$

Path #6 - Work Packages # 2, # 4, # 8.

$$\mathbf{R}_{\mathbf{Path\#6}} = \begin{bmatrix} 1.0 & 0.3 & 0.0 \\ 0.3 & 1.0 & 0.0 \\ 0.0 & 0.0 & 1.0 \end{bmatrix}$$

The correlation matrix for the work package costs is however a 13x13 matrix. This is because all the work package costs are summed to evaluate the project cost. The positive definite correlation matrix for work package costs is as follows.

$$\mathbf{R} = \begin{bmatrix} 1.0 & .29 & .24 & .26 & .25 & .23 & .23 & .25 & .25 & .29 & .24 & .23 & .23 \\ .29 & 1.0 & .25 & .30 & .28 & .26 & .27 & .29 & .29 & .35 & .29 & .27 & .35 \\ .24 & .25 & 1.0 & .26 & .25 & .23 & .23 & .25 & .25 & .29 & .25 & .23 & .23 \\ .26 & .30 & .26 & 1.0 & .27 & .24 & .24 & .26 & .27 & .47 & .26 & .24 & .24 \\ .25 & .28 & .25 & .27 & 1.0 & .15 & .24 & .25 & .16 & .30 & .25 & .24 & .23 \\ .23 & .26 & .23 & .24 & .15 & 1.0 & .21 & .23 & .14 & .28 & .23 & .22 & .21 \\ .23 & .27 & .23 & .24 & .24 & .21 & 1.0 & .23 & .24 & .28 & .22 & .22 & .21 \\ .25 & .29 & .25 & .26 & .25 & .23 & .23 & 1.0 & .25 & .30 & .25 & .23 & .23 \\ .25 & .29 & .25 & .27 & .16 & .14 & .24 & .25 & 1.0 & .31 & .25 & .24 & .23 \\ .29 & .35 & .29 & .47 & .30 & .28 & .28 & .30 & .31 & 1.0 & .29 & .28 & .28 \\ .24 & .29 & .25 & .26 & .25 & .23 & .22 & .25 & .25 & .29 & 1.0 & .23 & .20 \\ .23 & .27 & .23 & .24 & .24 & .22 & .22 & .23 & .24 & .28 & .23 & 1.0 & .22 \\ .23 & .35 & .23 & .24 & .23 & .21 & .21 & .23 & .23 & .28 & .20 & .22 & 1.0 \end{bmatrix}$$

Tables 7.32 contains the expected values and standard deviations approximated by the analytical method for project duration, total dollar project cost, project net present value and internal rate of return for the transitional correlations $\rho = 0$, $\rho = 0.5$ and $\rho = 1.0$. The revenue streams were assumed to be uncorrelated because of the difficulty in identifying common variables. The correlation matrices given above were used in the evaluation of moments for project duration and costs.

Table 7.32: Statistics for Project Variables

Project Variable	$\rho = 0$		$\rho = 0.5$		$\rho = 1.0$	
	$E[PV]$	σ_{PV}	$E[PV]$	σ_{PV}	$E[PV]$	σ_{PV}
Duration	29.31	5.43	32.57	4.75	33.06	4.46
Cost (Tot$)	56955300	13650279	57850640	13869375	57984776	13902068
NPV	23494481	17758202	22838442	17612327	22738804	17589760
IRR	17.251	6.743	16.923	6.616	16.860	6.573

Table 7.33 and 7.34 contain comparisons of the expected values and standard deviations approximated by the analytical method at the transitional correlations $\rho = 0$, $\rho = 0.4$ and $\rho = 1.0$ for project duration and current dollar project costs when all the variables are uncorrelated (third validation case), primary variables in the functions for work package durations and costs are correlated (fourth validation case) and when the primary variables, work package durations and costs are correlated. The $\rho = 0.4$ is used because the current dollar project cost at $\rho \geq 0.5$ is same as the upper bound ($\rho = 1$) estimate.

There is only a marginal difference in the statistics for the project duration and the project cost from the first and second cases. Even though the expected value for project duration is slightly larger when $\rho = 0.4$ and $\rho = 1$ for the third case, the expected values for project cost are similar for all three situations. Since project cost is the linear addition of work package costs, there is no effect from the correlation between work package costs. Hence, the identical expected values for project cost

Table 7.33: Comparison of the Statistics for Project Duration

Type of the Correlations	$\rho = 0$		$\rho = 0.4$		$\rho = 1.0$	
	$E[PD]$	σ_{PD}	$E[PD]$	σ_{PD}	$E[PD]$	σ_{PD}
Uncorrelated	29.44	4.69	29.78	4.53	32.42	3.69
Primary only	29.31	4.59	29.57	4.46	32.21	3.68
All Variables	29.31	5.43	32.22	4.82	33.06	4.46

Table 7.34: Comparison of the Statistics for Current Dollar Project Cost

Type of the Correlations	$\rho = 0$		$\rho = 0.4$		$0.5 \leq \rho \leq 1.0$	
	$E[PC]$	σ_{PC}	$E[PC]$	σ_{PC}	$E[PC]$	σ_{PC}
Uncorrelated	54129602	7584872	54145288	7588666	54158049	7589427
Primary only	53485125	7285335	53500807	7289218	53513534	7289989
All Variables	53485125	12776812	53500807	12781905	53513534	12784827

from the second and third cases. However, there is nearly 18% and 75% increases in the standard deviations for project duration and cost due to the correlations between work package durations and costs. The correlations between work package costs are relatively small. This clearly illustrates that these correlations can be significant.

Figure (7.37) depicts the cumulative distribution functions for project duration at different transitional correlation values approximated from the analytical method. Figure (7.38), (7.39) and (7.40) depict, the cumulative distribution functions for upper and lower bounds approximated from the analytical method for total dollar project cost, project net present value and internal rate of return. The linear correlations between the primary variables in the functions for work package durations and costs, and the linear correlations between work package durations and work package costs respectively are treated.

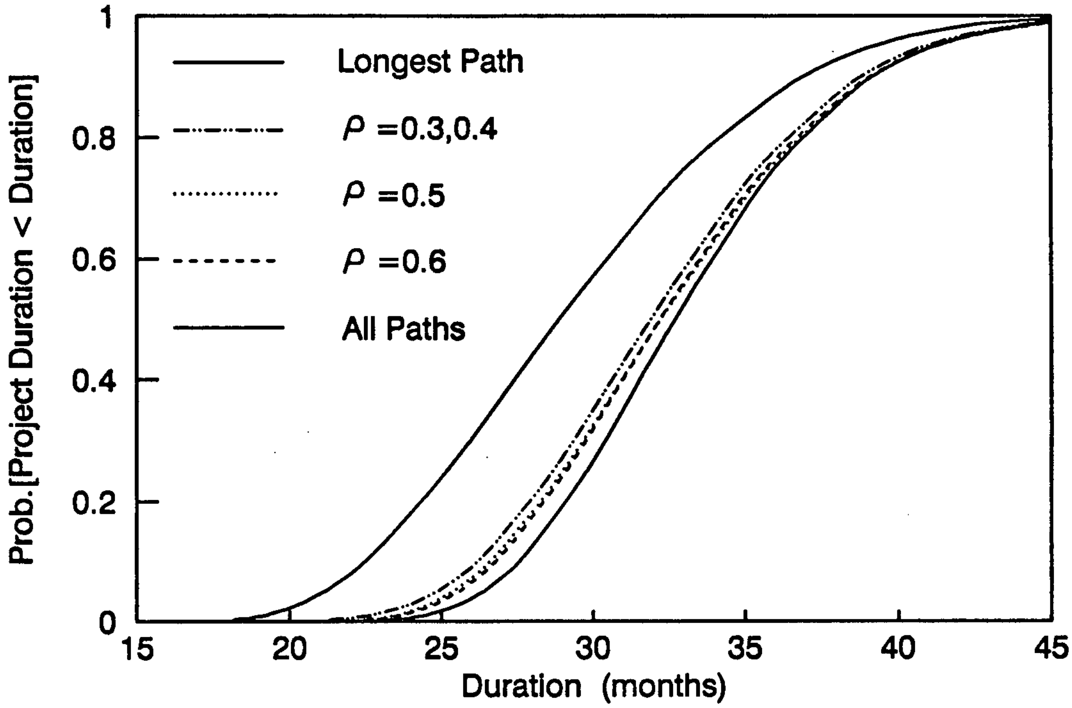

Figure 7.37: CDFs for Project Duration

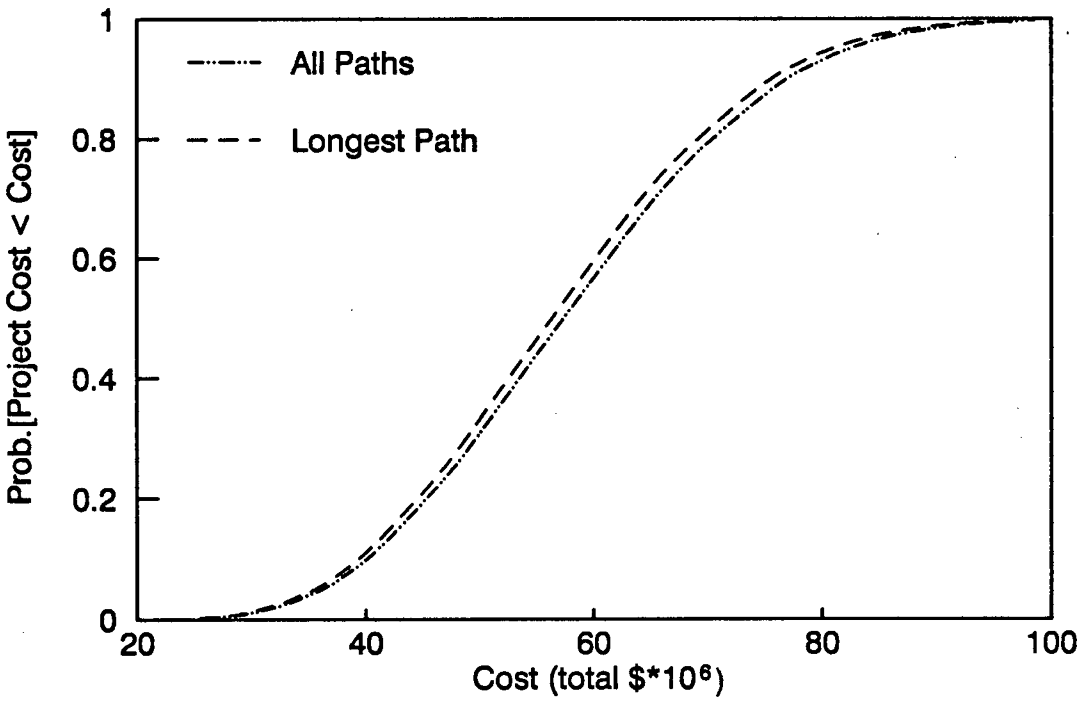

Figure 7.38: CDFs for Total Dollar Project Cost

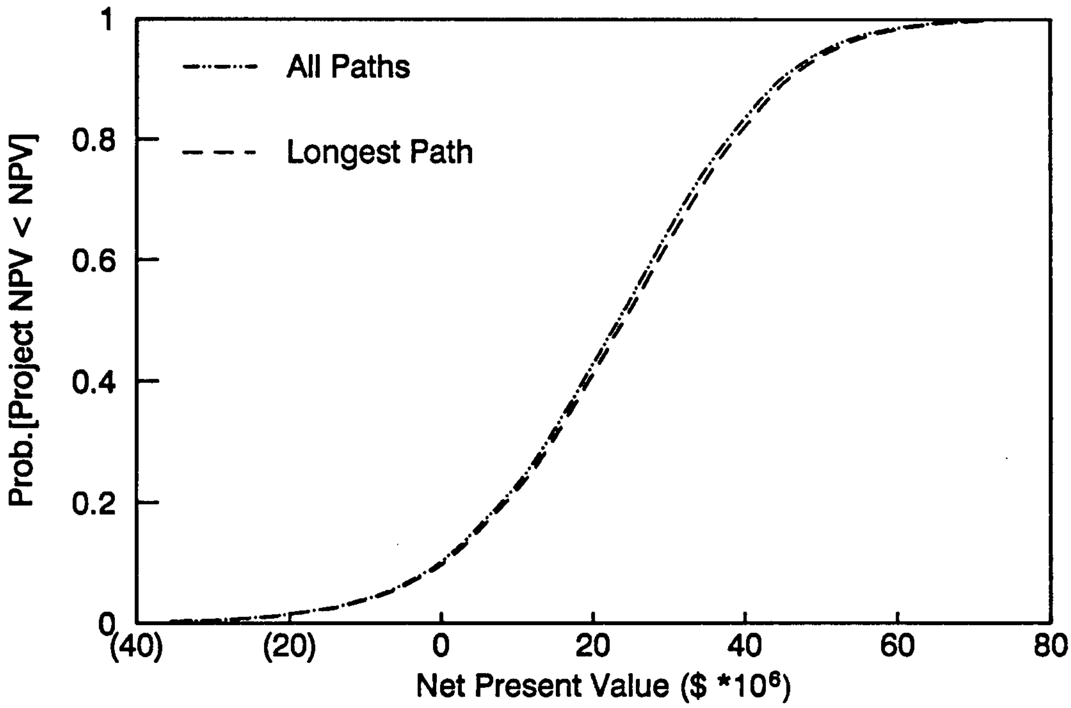

Figure 7.39: CDFs for Project Net Present Value

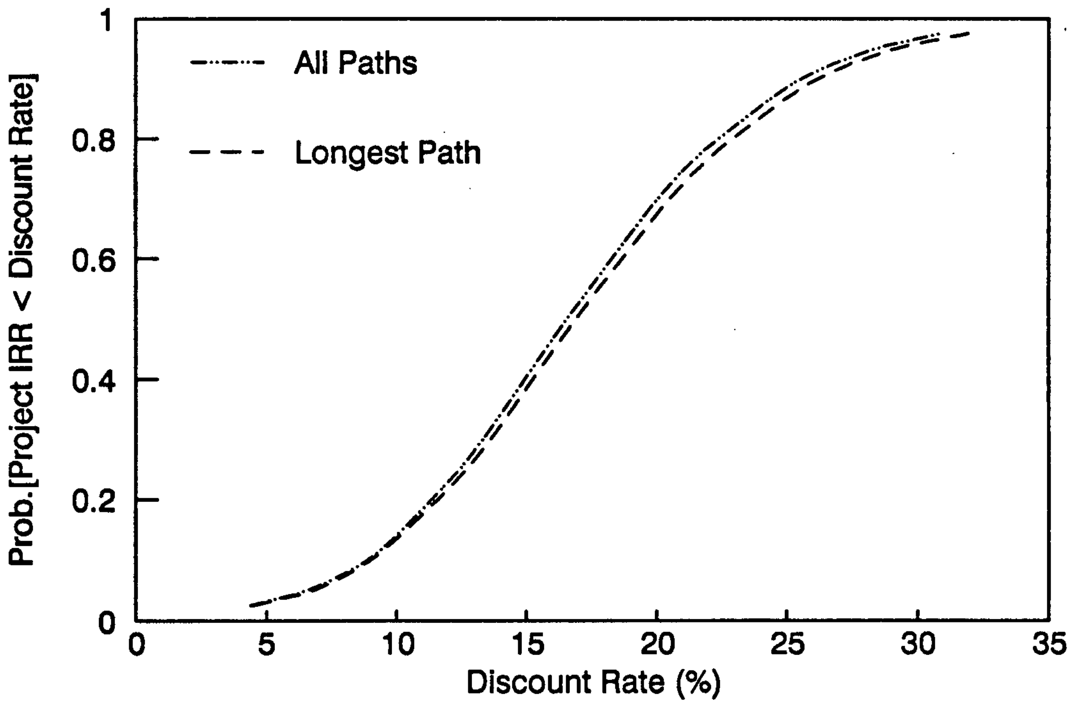

Figure 7.40: CDFs for Project Internal Rate of Return

7.6.4 Discussion

The precedence network for the engineering project used as the example in this section was small (see figure 7.23), with few interrelationships between work packages. This feature had both advantages and disadvantages. The advantages were, because of its size it was possible to elaborate the work package durations and costs to detailed functions, thereby demonstrating the full potential of the analytical method. Also, it was possible to illustrate the treatment of correlations at all levels of the project economic structure. For example, if the precedence network was highly interrelated with a large number of paths to complete the project as in the first example, it would not have been feasible to illustrate the positive definite correlation matrices for work package durations on individual paths and for work package costs. The disadvantage is that the elaboration of work package cost functions and the few interrelationships between work packages, combined to approximate extremely tight bounds for economic variables. Thereby, hampering the validation process.

There were 210 random primary variables at the input level. The comparisons of execution times for the simulation and the analytical method highlighted the computational economy of the analytical method. The treatment of correlations between work package durations and between work package costs clearly demonstrated their significance.

7.7 Sensitivity Analysis and Contingency

This section will briefly discuss the different ways in which the analytical method can perform sensitivity analysis and use one of them to outline a method to distribute the contingency allocated to a derived variable to its primary variables. Current dollar project cost is used as the example for the derived variable.

7.7.1 Sensitivity Analysis

The concept of sensitivity analysis is simple. If a change in a primary variable has little effect on the derived variable, then the estimate for the derived variable is not likely to depend to any great extent on the accuracy of the estimate for that primary variable. On the other hand, if a change in a primary variable produces a large change in the estimate for the derived variable, then the uncertainty surrounding that primary variable may well be a significant consideration when evaluating the derived variable. The sensitivity of a primary variable is measured by the total sensitivity coefficient for that variable.

For a functional relationship given by, $Y = g(\mathbf{X})$, the sensitivity of the derived variable with respect to the primary variables is given by (Russell, 1985),

$$\frac{\Delta Y}{Y} \approx \sum_i S_i \frac{\Delta X_i}{X_i} \tag{7.8}$$

where $\frac{\Delta Y}{Y}$ and $\frac{\Delta X_i}{X_i}$ are the percent changes in Y and X_i respectively, and S_i is the total sensitivity coefficient of X_i. For the sensitivity plot, S_i is the gradient of the sensitivity line relating percent change of X_i to percent change in Y. The total sensitivity coefficient S_i is defined as (Russell, 1985),

$$S_i = \frac{\partial Y}{\partial X_i} \frac{X_i}{Y} \tag{7.9}$$

where $\frac{\partial Y}{\partial X_i}$ is the sensitivity coefficient of Y with respect to X_i.

Since moment analysis is based on the truncated Taylor series expansion of $g(\mathbf{X})$, the partial derivatives with respect to primary variables should be evaluated. However, the analytical method transforms the primary variables $\mathbf{X}$ to $\mathbf{Z}$ and $g(\mathbf{X})$ to $G(\mathbf{Z})$ prior to using the Taylor series expansion. Even though the sensitivity coefficients

$\frac{\partial g}{\partial X_i}$ are not evaluated by the analytical method it still evaluates $\frac{\partial G}{\partial Z_i}$, sensitivity coefficients with respect to the transformed variables. Hence, the analytical method has an in-built sensitivity analysis process, whereby the sensitivity coefficients either increase or decrease the contribution of each term, depending on the importance of each transformed variable to the derived variable.

Nevertheless, the sensitivity plot of $\frac{\Delta Y}{Y}$ versus $\frac{\Delta X_i}{X_i}$ can be developed by obtaining a range of outputs at different percent changes of X_i. This can be a rather long process. However, since the analytical method is efficient and computationally economical, if desired it can be developed. Similar sensitivity analysis can be performed on the subjective estimates for primary variables. If the analyst requires a sensitivity analysis on the subjective estimates, again a sensitivity plot can be developed from a range of outputs at percent changes of subjective estimates. Since the objectives of this thesis do not require the validation of the input primary variables, such a study is not presented.

The third sensitivity analysis performed by the analytical method is on the transitional correlation (ρ) specified by the analyst. The bounds for time and economic variables recognize the high degree of uncertainty associated with the decisions that have to be made during the feasibility analysis. By the definition of risk analysis, the quantification of risk for a specified transitional correlation should encompass the uncertainty of the assumed scenarios. However, the bounds add further reliability to the quantification because they are the true analytical bounds for those assumed scenarios.

The fourth way in which the analytical method can perform sensitivity analysis is the basis for the method to distribute the contingency allocated to a derived variable to its primary variables. The derived variables at the project performance level are all linear additive and the sensitivity coefficients with respect to primary variables are equal to one. When the uncertainty in the derived variable is due to the uncertainty of the primary variables alone (no effects due to correlation), the variance of Y is the summation of the variances of the primary variables. The allocated contingency can then be distributed to the primary variables on the basis of the individual percentage contributions to the total uncertainty of the derived variable.

7.7.2 Distribution of Contingency

The contingency is generally defined as the amount included in an estimate to cover the overruns due to unforeseen items and events in the defined project scope. Since, this allocation is done for derived variables such as project duration or cost, its management is important. The ability to distribute the contingency to work packages provide a logical basis to manage it. The objective of this section is to demonstrate an analytical method to distribute the contingency.

Inyang (1983) derived the contingency (C) as,

$$C = X_C - E_B \tag{7.10}$$

where X_C is the target cost and E_B is the base estimate cost. He preferred the base estimate cost to the expected value used by Yeo (1982), because it was necessary to assume that the project cost was normally distributed to derive the contingency and because the base estimate cost is always smaller than the expected value. However, the target cost X_C was not related to any probability of success (or failure). As highlighted in the introduction, institutions such as the World Bank now recommend the use of probabilities of success (or failure) for performance variables.

This thesis derives the contingency (C) as,

$$C = X_P - E[PC] \tag{7.11}$$

where X_P is the cost estimate to achieve a desired probability of success and $E[PC]$ is the expected value of project cost.

Consider the current dollar project cost at the transitional correlation of $\rho = 0.5$ for the second example. Table 7.35 contains the values for X_P, C and the percentage of the contingency to the expected value ($\frac{C}{E[PC]} \times 100$) for $Pr.[PC] \leq 0.75$, $Pr.[PC] \leq 0.9$ and $Pr.[PC] \leq 0.95$ for two cases. The first case has considered only the correlation between primary variables (the correlation treatment that the Monte Carlo simulation can duplicate) and the second has treated the correlation between primary variables, work package durations and work package costs.

The expected values for project cost from both cases are identical (see Table 7.34). However, the variance for the second case has increased by about 200%. This is reflected in the values for X_P and C, where to achieve the same probability of success the contingencies have to be increased by about 80%. For example, if the contingency was set at a 90% probability of success using the results from the first case, in reality the project cost has only about a 75% probability of success. This example, again highlights the significance of the correlation between work package costs. When the contingencies are compared as percentages of the expected value, the insufficiency of the traditional allocations of 10% to 15% is clearly demonstrated.

Table 7.35: X_P, C and $\frac{C}{E[PC]}$ for Different Probabilities of Success

Correlations	Primary Variables Only			Primary & Derived Variables		
Scenario	X_P	C	%	X_P	C	%
$Pr.[PC] \leq 0.75$	58470505	4956971	9.26	62088867	8575333	16.0
$Pr.[PC] \leq 0.9$	63243785	9730251	18.2	70581087	17067553	31.9
$Pr.[PC] \leq 0.95$	66037795	12524261	23.4	75632830	22119296	41.3

The main advantage of this definition is that the contingency distributed to individual work packages can be used to predict their probabilities of success (or failure).

This provides the bench mark to manage the contingency. Since the analytical method evaluates the expected value of project cost as the summation of all the expected values of work package costs, equation (7.11) can be re-written as,

$$X_p = \sum_{i=1}^{n}(E[WPC_{CDi}] + CON_i) \tag{7.12}$$

where $E[WPC_{CDi}]$ is the expected value of the i^{th} work package cost and CON_i is the contingency distributed to the i^{th} work package on the basis of its percentage contribution to the variance of project cost. Then, $E[WPC_{CDi}] + CON_i$, is the amount available for the cost of defined scope and unforeseen items and events of the i^{th} work package. The probability of success of this amount can be measured from cumulative distribution function for the i^{th} work package cost. Not only does it provide a bench mark to manage the contingency but also allows the project manager to transfer contingency between work packages on a logical basis.

Consider the second case in Table 7.35. The derived variable is the current dollar project cost while the primary variables are work package costs. The expected values, standard deviations, coefficients of variations, skewness, kurtosis and the percentage contributions to the variance of the project cost from individual work package costs are given in Table 7.36. The percentage contributions to variance of project cost is evaluated from the following function.

$$\%\ Contribution\ from\ WPC_{CDi} = \frac{\mu_2(WPC_{CDi})}{\sum_{i=1}^{n} \mu_2(WPC_{CDi})} \tag{7.13}$$

where $\mu_2(WPC_{CDi})$ is the variance of the i^{th} work package cost.

Figure (7.41) depicts the cumulative distribution function for the current dollar project cost at $\rho = 0.5$, and the X_P values given in Table 7.35. The values for the

Table 7.36: Statistics for Current Dollar Work Package Cost - Ex #2

WP#	$E[WPC]$	σ_{WPC}	C.O.V %	$\sqrt{\beta_1}$	β_2	% Cont.
02	3894677	1575760	40.46	0.381	2.174	4.67
03	6137617	2237672	36.45	0.818	2.802	9.42
04	7877801	3215341	40.81	0.387	2.179	19.45
05	1723771	701822	40.71	0.503	2.304	0.93
06	917295	353412	38.52	1.317	4.082	0.25
07	2146194	959195	44.69	1.001	3.204	1.73
08	3667912	1486228	40.52	1.572	4.967	4.15
09	4080786	1677239	41.10	0.406	2.198	5.29
10	1728653	697393	40.34	1.256	3.892	0.92
11	2370725	852850	35.97	0.344	2.142	1.37
12	8220980	3370799	41.00	0.404	2.196	21.38
13	2397196	1037134	43.26	0.859	2.885	2.02
14	8349926	3886408	46.54	0.578	2.401	28.42

contingencies distributed on the basis of percentage contributions, the total amount available for the defined scope and unforeseen items, and the probabilities of success for individual work packages based on those total amounts for the three scenarios of $Pr.[PC] \leq 0.75$, $Pr.[PC] \leq 0.9$ and $Pr.[PC] \leq 0.95$ are given in Table 7.37. Figure (7.42) depicts the values for work package # 4 from Table 7.37. This work package was selected because it is an early work package in the network that has a high contribution to variance of project cost. A typical example of where things could go wrong.

The values in Table 7.37 show that allocating contingency on the probability of success (or failure) of a global criterion such as project cost may not necessarily reflect the true situation because none of the work packages costs achieved the probability of success desired for the project cost. Analytically this can be reasoned that risks decrease when they are aggregated. In terms of practical situations, the importance of distributing the contingency becomes apparent.

The contingency was distributed on the assumption that those work package cost

Table 7.37: Distributed Contingency and Probability of Success

#	$Pr.[PC] \leq 0.75$			$Pr.[PC] \leq 0.9$			$Pr.[PC] \leq 0.95$		
	Contn.	Total	$\leq$	Contn.	Total	$\leq$	Contn.	Total	$\leq$
02	400468	4295145	.61	797055	4691732	.68	1032971	4927648	.72
03	807796	6945413	.67	1607763	7745380	.76	2083637	8221254	.80
04	1667902	9545703	.68	3319639	11197440	.81	4302203	12180004	.87
05	79752	1803523	.58	158728	1882499	.61	205710	1929481	.63
06	21440	938735	.62	42672	959967	.64	55300	972595	.65
07	148353	2294547	.63	295269	2441463	.67	382664	2528858	.69
08	355876	4023788	.68	708303	4376215	.75	17951	4585863	.77
09	453635	4534421	.61	902873	4983659	.69	1170110	5250896	.74
10	78893	1807546	.63	157021	1885674	.66	203497	1932150	.68
11	117482	2488207	.57	233825	2604550	.61	303034	2673759	.63
12	1833406	10054386	.69	3649043	11870023	.82	4729105	12950085	.89
13	173221	2570417	.62	344764	2741960	.66	446810	2844006	.69
14	2437109	10787035	.73	4850598	13200524	.85	6286304	14636230	.91

variances which contribute most to the variance of project cost cause most of the uncertainty in the project cost. Therefore, the distribution ensured that the work packages with higher contributions had the greater probability of success and vice versa. Having got the initial bench marks to reflect the reasoning for the distribution, it is now possible to transfer some of the contingency from work package costs that have a greater probability of success to those which have a greater probability of failure.

Unlike project cost, the distribution of contingency for project duration is not straightforward because the project duration is not a summation of all the work package durations. However, the modified PNET algorithm does permit a basis for an approach. Since the variances of all the paths to complete the project are evaluated when the transitional correlation $\rho = 1$, a sensitivity analysis similar to that adopted for the project cost can be utilized.

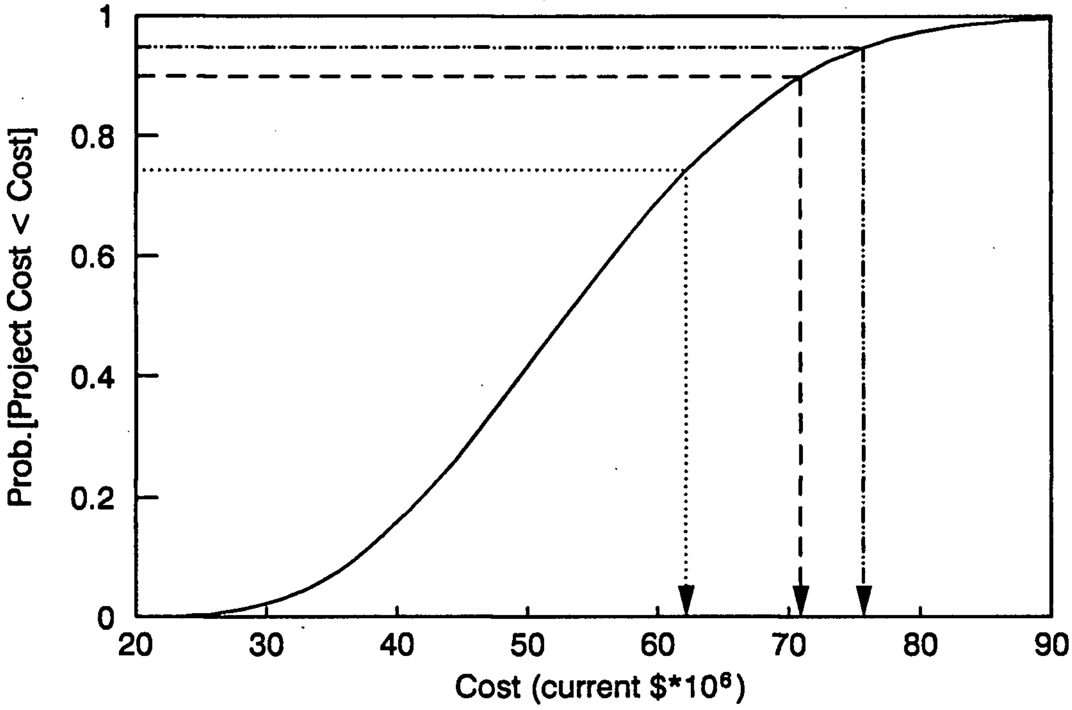

Figure 7.41: CDF for Current Dollar Project Cost

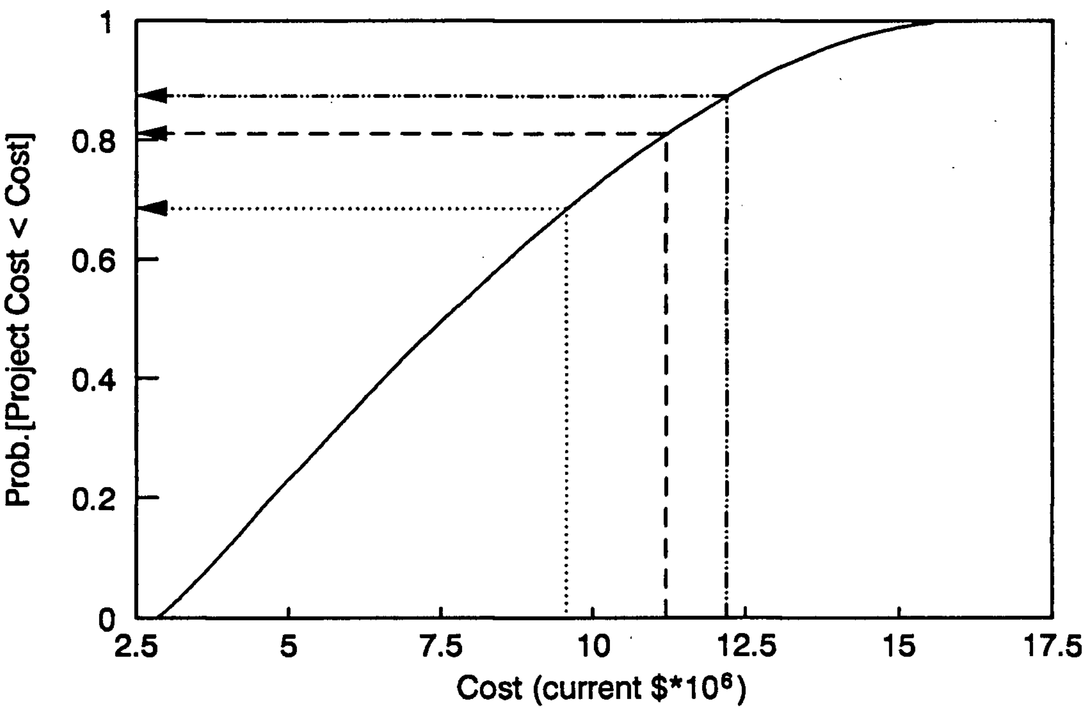

Figure 7.42: CDF for Current Dollar Cost for Work Package #4

Consider the j^{th} path of the precedence network. The variance for the j^{th} path when the correlations between work package durations are not considered is given by,

$$\mu_2(T_j) = \sum_{i=1}^{n} \mu_2(WPD_{ij}) \tag{7.14}$$

where $\mu_2(WPD_{ij})$ is the variance of the i^{th} work package duration on the j^{th} path. The contingency allocated for project duration can then be distributed to the work package durations on the j^{th} path based on the individual percent contributions to the variance of j^{th} path duration. Then similar to work package costs it is possible to measure the probability of success of individual work package durations for defined scope and unforeseen items and events.

However, since a work package can be on more than one path it can have a number of distributed contingency durations. In such situations the lowest distributed contingency should be assumed as the duration for unforeseen items and events. The measured probabilities of success will then be the lowest for every work package duration in the network. Again, providing a bench mark to manage the contingency allocated for project duration.

7.8 Summary

This chapter described the validations and the applications of the analytical method developed in the previous chapter. The validations of the analytical method was performed by using Monte Carlo simulations. The simulations were used because at present, simulation based models are considered to be the "state-of-the-art" for quantification of time and economic risks in large engineering projects.

The second section of this chapter derived the random number modification process used to treat correlations between variables in Monte Carlo simulation. The number of iterations for an "acceptable" simulation should be based on the standard error for the expected value and standard deviation, and the error band for the cumulative distribution function generated from the simulation.

The modified PNET algorithm was validated by solving the two examples presented by Ang et al., (1975). The comparison of the results showed that modified PNET identifies the individual paths correctly, evaluates the expected value (mean) and standard deviation for path durations accurately, and selects the representative paths correctly. The ordering of paths may differ because in addition to ordering the paths in decreasing mean durations, the modified PNET orders the paths in decreasing standard deviations when mean path durations are equal. This ensures the selection of the path with the highest variance as the representative path from the paths having the same mean duration.

The project duration of a parallel network was used as the first limiting case to validate the Monte Carlo simulation process. When simulations are used to validate an analytical approach it is essential that the simulation process is first validated. A single dominant path of a highly interrelated network was used as the second limiting case. The simulations for both cases behaved as predicted thereby validating the Monte Carlo simulation process.

The first example for validating the analytical method was an actual feasibility study of a mineral project. This example had 164 random primary variables at the input level and four simulations were performed. Two were for the case when the coefficients of variations for work package durations were low (first validation), while the others were for the case when the coefficients of variations for work package durations were approximately 40% (second validation).

In both instances, the cumulative distribution functions and the estimates for expected values for derived time and economic variables generated from simulations were within the upper and lower bounds predicted by the analytical method. Thereby, validating the analytical method. The computational economy of the analytical method was highlighted from the comparisons of execution times. A brief comparison with deterministic results and when $\rho = 0.5$ were done for the first and second validations respectively.

The second example was an hypothetical engineering project developed to demonstrate the full potential of the analytical method. The project network was small (thirteen work packages) with few interrelationships between work packages. Elaborate functions were used for work package durations, costs and revenue streams. There were 210 random primary variables at the input level. Two complete simulations for time and economic risk quantifications were done. The first assumed that all the primary variables were uncorrelated (third validation) while the second assumed that the primary variables in the functions for work package durations and costs were correlated (fourth validation).

The bounds for derived economic variables predicted by the analytical method were extremely tight, because the work package start time was one of the seventeen variables in the function for work package cost and because of the few interrelationships between work packages. The simulations demonstrated the validity of the analytical method by generating estimates and cumulative distribution functions similar to those from the analytical method.

The third section demonstrated the treatment of linear correlations between work package durations and between work package costs when evaluating the moments for project duration and costs. The analytical method approximates these correlations using the linear correlations between the primary variables when the common (shared) variables are identified. The positive definite correlation matrices developed by the analytical method were given. The Monte Carlo simulation does not have the capability to duplicate this treatment. A comparison of the statistics and the contingencies that should be allocated to achieve a desired probability of success highlighted the significance of the effect of these correlations. The standard deviations for project duration and current dollar cost increased nearly 18% and 75% respectively.

The final section described the four ways in which the analytical method can and/or do perform sensitivity analyses, and used the fourth approach to develop an analytical basis to distribute the contingency allocated at a desired probability of success. The distribution of contingencies allocated for $Pr.[PC] \leq 0.75$, $Pr.[PC] \leq 0.9$ and $Pr.[PC] \leq 0.95$ of current dollar project cost showed that none of the work packages achieved the same probability of success. This distribution is biased towards the work packages costs with variances that contribute most to project variance, giving them the greatest probability of success.

Overall, this chapter demonstrated the validity and the computational economy of the analytical method in the quantification of time and economic variables in large engineering projects.

Chapter 8

Conclusions and Recommendations

8.1 Conclusions

The primary objectives of this thesis were to develop an analytical method for economic risk quantification during feasibility analysis for large engineering projects and to computerize the method to explore its behavior, to validate it and to test its practicality in the measurement of uncertainty of decision variables. The secondary objective was to lay the foundation for obtaining the input data necessary to make the analytical method a practical tool for the construction industry. The main conclusions from the developments of this thesis are as follows.

1. The analytical method is a comprehensive alternative to Monte Carlo simulation for the quantification of time and economic risks in large engineering projects.

2. The start times of work packages and revenue streams evaluated from the analysis of the precedence network provided the link to model the interaction of time, cost and revenue throughout the life cycle of a project.

3. The definition of the project economic structure and the freedom to use any type of functional form for work package durations, costs and revenue streams provided the freedom to model a project realistically to any level of detail using any number of variables.

4. The risk measurement framework is suitable for systems where predetermined functional forms are available, data limitations exist and the decisions are not based on extreme probabilities.

5. The reliance on subjective probabilities to obtain data for the primary variables at the input level recognized the data limitations that exist during the feasibility analysis. The elicitation of accurate, calibrated and coherent subjective probabilities as the measurement of expert belief incorporated the theoretical requirements into the practical process.

6. It was concluded that when eliciting subjective estimates for duration, neither the holistic nor the decomposed estimation was the "better" approach.

7. The consideration of multiple paths of the project network provided a more realistic evaluation of the statistics for work package start time. In addition, the modified PNET algorithm provided the basis to evaluate the true analytical bounds for derived time and economic variables.

8. The uncertainty of the project performance and decision variables were quantified by consistently utilizing the moment analysis approach with the Pearson family of distributions.

9. The correlations between variables was identified as an important feature of this problem. The variable transformation approach developed to treat the correlations between primary variables and between derived variables was found to be accurate and robust.

10. The elicitation of positive definite correlation matrices for primary variables in the functional forms at the input level incorporated an important theoretical requirement into the practical application.

11. The approximation and the treatment of correlations between work package durations, between work package costs and between revenue streams demonstrated the ability of the analytical method to go beyond the capabilities of the Monte Carlo simulation process.

12. It was concluded that during the feasibility analysis work package concept can be utilized as the approach to obtain intermediate milestone information to set realistic targets for performance.

13. The quantification of uncertainty of project time and economic variables provided the basis to answer such strategic questions as setting up of the contingency for a probability of success (or failure) and the reliability of the "go - no go" decision. The individual contributions to the overall uncertainty was used to distribute the contingency for project variables to work packages. It was found that the probability of success predicted at the project level was not achieved at the work package level.

14. When the starting points were identical, the results from the Monte Carlo simulations were within the upper and lower bounds predicted by the analytical method. From the validations it was concluded that the analytical method had the flexibility to model and evaluate the derived time and economic variables of a project accurately and economically.

The analytical method and the computer programs developed from this research achieved the objectives of this thesis. However, in terms of the total process of risk management for large engineering projects, these developments are only the beginning. Unless there is an efficient approach to quantify time and economic risks it is impossible to respond to the identified risks. On other hand, until the area of risk response is developed the quantifications are of little use. The recommendations for future research briefly highlight the scope of the work that is necessary to make this development a practical tool for engineering construction.

8.2 Recommendations for Future Work

Recommendations for future work are identified under three sections, namely, the analytical method, computer programs and the risk management process.

8.2.1 Analytical Method

The analytical method developed in this thesis consisted of a number of major building blocks - project economic structure; risk measurement framework; elicitation of subjective probabilities and positive definite correlation matrices; treatment of correlations between variables; and the modified PNET algorithm.

1. **Project Economic Structure** : It is recommended that a suite of time, cost and revenue estimating relationships at the work package/revenue stream level be developed. This would significantly increase the ability of the analytical method to model large engineering projects realistically. The publications by Tanchoco et al., (1981) and Buck, (1989) are useful starting points.

2. **Risk Measurement Framework** : At present the approximations for the first four moments of the derived variable consider terms only up to the fourth order. However, since all the primary variables are approximated to Pearson type distributions, it is possible to obtain the higher order moments from the recurrence property of the Pearson family. It is recommended that with practical experience in the elicitation of subjective probabilities terms up to the eighth order be included. This would

ensure more accurate approximations for the first four moments of the derived variables at the work package/revenue stream level because all of the necessary terms are included.

3. **Elicitation of Subjective Probabilities** : The development in this thesis was the foundation for obtaining input data. The next stage should concentrate on building up experience from field applications, refining and validating the elicitation approach. To this end, a complete automation and pre-testing of the process is recommended. With experience from field applications, the process can be refined based on the performance of analysts and experts. Also, the development of calibration curves for validation is recommended. The publications by Budescu and Wallsten, (1987), Phillips, (1987), Wallsten and Budescu, (1983), Murphy and Winkler, (1984), and Wright and Ayton, (1987) are useful starting points.

4. **Elicitation of Correlation Matrices** : A more consistent approach to elicit the correlation coefficients between variables is necessary. It is recommended that effort should be devoted towards developing questions that would better capture the expert's knowledge about correlated variables. The publications by Inyang (1983), Hull (1977), Kadane et al., (1980) and Keefer and Bodily, (1983) are good starting points. The routine in the interactive computer program "ELICIT" that ensures the positive definiteness of the correlation matrix should be further refined with a better user interface.

5. **Correlations between Variables** : The variable transformation approach described in this thesis was found to be both accurate and robust in the treatment of correlations between variables. More studies to test

and to further understand the transformation is recommended. The approximation and the treatment of correlations between derived variables require further studies to understand the benefits from the transformation.

6. **Modified PNET Algorithm** : At present the modified algorithm treats only *finish to start* = 0 relationships. The extension of the modified PNET algorithm to treat overlapping relationships will increase the versatility of the modelling capability. This extension however, requires the development of an algorithm to treat the correlations between work package durations in overlapping relationships. It is strongly recommended that future efforts be devoted to developing such an algorithm.

   Ideally, a single value for the transitional correlation ρ that can be used for all economic risk analyses of engineering projects should be recommended. However, at present it is not possible to make this recommendation. Efforts should be devoted to deriving such a value. The study of the behavior of the modified PNET algorithm for highly interrelated networks versus those with few relationships (linear networks in pipeline or highway projects) is the logical starting point.

8.2.2 Computer Programs

One of the primary objectives of this thesis was to computerize the analytical method to explore its behavior, to validate it and to test its practicality in the measurement of uncertainty. The two computer programs developed by this research facilitated the achievement of this objective. While the computer programs were not meant to be software development, they are a useful starting point for a software package.

At present, both programs, "ELICIT" and "TIERA", lack sophistication especially in the area of user friendliness. It is recommended that efforts be devoted to achieving a higher degree sophistication, for this development to become a practical tool for decision makers in engineering construction.

8.2.3 Risk Management Process

In the introduction, the process of risk identification, risk quantification and risk response was identified as the most suitable approach for risk management in engineering construction. This thesis presented a computationally economical approach that can be used develop the basis for decision makers to respond to the identified risks. Until the area of risk response is developed the quantifications of derived time and economic risks are of little use.

The next stage of this research should concentrate on developing strategies to respond to the quantified risks. It is strongly recommended that efforts be devoted towards this end. The extensive research done by the Project Management Group at the University of Manchester Institute of Science and Technology (UMIST) can be used as the starting point. The publications by Perry and Hayes, (1985a), (1985b), Hayes et al, (1986), and Howard (1988) should help this process. It is strongly recommended that efforts should be devoted towards obtaining industry collaboration for applications of this development.

Bibliography

[1] **Ahuja, H.N., & Arunachalam, V. (1984).**, "Risk Evaluation in Resource Allocation.", *Journal of Construction Engineering & Management*, Vol.110, No.3, pp.324-336.

[2] **Ahuja, H.N., & Nandakumar, V. (1985).**, "Simulation Model to Forecast Project Completion Time " , *Journal of Construction Engineering & Management*, Vol.111, No.4, pp.325-342.

[3] **Amos, D.E., & Daniel, S.L. (1971).**, "Tables of Percentage Points of Standardized Pearson Distributions", *Research Report # SC-RR - 710348*, NTIS, Virginia, U.S.A.

[4] **Ang, A.H-S., Abdelnour, J., & Chaker, A.A. (1975).**, "Analysis of Activity Networks under Uncertainty.", *Journal of Engineering Mechanics Division, ASCE* , Vol.101 No.EM4, pp.373-387.

[5] **Ang, A.H-S., & Tang, W.H. (1975).**, *Probability Concepts in Engineering Planning and Design*, John Wiley & Sons, New York.

[6] **Ashley, D.B. (1980a).**, "Construction Joint Ventures", *Journal of the Construction Division, ASCE*, Vol.106, No.CO3, pp.267-280.

[7] **Ashley, D.B. (1980b).**, "Coordinated Insurance for Major Construction Project", *Journal of the Construction Division, ASCE*, Vol.106, No.CO3, pp.307-313.

[8] **Ashley, D.B., & Bonner, J.J. (1987).**, "Political Risks in International Construction" , *Journal of Construction Engineering & Management*, Vol.113, No.3, pp.447-467.

[9] **Ashton, A.H., & Ashton, R.H. (1985).**, "Aggregating Subjective Forecasts : Some Empirical Results", *Management Science*, Vol.31, No.12, pp.1499-1508.

[10] **Ashton R.H., (1986).**, "Combining Judgements of Experts: How Many and Which Ones ?", *Organisational Behavior & Human Decision Processes*, Vol.38, pp.405-414.

[11] **Au, T. (1988).**, "Profit Measures and Methods of Economic Analysis for Capital Project Selection", *Journal of Management in Engineering*, Vol.4, No.3, pp.217-228.

[12] **Baker, R.W. (1986).**, "Handling Uncertainty", *International Journal of Project Management*, Vol.4, No.4, pp.205-210.

[13] **Bacharach, M. (1975).**, "Group Decisions in the Face of Differences of Opinion", *Management Science*, Vol.22, No.2, pp.182-191.

[14] **Beach, B.H. (1975).**, "Expert Judgements About Uncertainty: Bayesian Decision Making in Realistic Settings", *Organisational Behavior & Human Performance*, Vol.14, pp.10-59.

[15] **Beach, L.R., Christensen-Szalanski, J.J.J., & Barnes, V. (1987).**, "Assessing Human Judgement: Has it Been Done, Can it Be Done, Should it Be Done?", *Judgmental Forecasting*, Edited by G. Wright and P. Ayton, John Wiley & Sons, N.Y.. pp.49-62.

[16] **Bellman, R. (1970).**, *Introduction to Matrix Analysis* (2nd Edition), McGraw-Hill, New York.

[17] **Benjamin, J.R., & Cornell, C.A. (1970).**, *Probability, Statistics and Decision for Civil Engineers*, McGraw-Hill Book Co., New York.

[18] **Bjornsson, H.C. (1977).**, "Risk Analysis of Construction Cost Estimates", *Transactions, American Association of Cost Engineers*, pp.182-189.

[19] **Bonini, C.P. (1975).**, "Risk Evaluation of Investment Projects", *OMEGA, International Journal of Management Science*, Vol.3, No.6 , pp.735-750.

[20] **Borcherding, J.D. (1977).**, "Cost Control Simulation and Decision Making", *Journal of the Construction Division, ASCE*, Vol.103, No.CO4, pp.577-591.

[21] **Bordley, R.F. (1982).**, "A Multiplicative Formula for Aggregating Probability Assessments", *Management Science*, Vol.28, No.10, pp.1137-1148.

[22] **Bordley, R.F., & Wolff, R.W. (1981).**, "On the Aggregation of Individual Probability Estimates", *Management Science*, Vol.27, No.8, pp.959-964.

[23] **Britney, R.B. (1976).**, "Bayesian Point Estimation and the PERT Scheduling of Stochastic Activities", *Management Science*, Vol.22, No.9, pp.938-948.

[24] **Buck, J.R. (1989).**, *Economic Risk Decisions in Engineering and Management*, Iowa State University Press, Ames.

[25] **Budescu, L.R., & Wallsten, T.S. (1987).**, "Subjective Estimation of Precise and Vague Uncertainties", *Judgmental Forecasting*, Edited by G. Wright and P. Ayton, John Wiley & Sons, N.Y., pp.63-82.

[26] **Bunn, D.W. (1975).**, "Anchoring Bias in the Assessment of Subjective Probability", *Operation Research Quarterly*, Vol.26, No.2, ii, pp.449-454.

[27] **Bunn, D.W. (1979a).**, "A Perspective on Subjective Probability for Prediction and Decision", *Technical Forecasting & Social Change*, Vol.14, pp.39-45.

[28] **Bunn, D.W. (1979b).**, "Estimation of Subjective Probability Distributions in Forecasting and Decision Making", *Technical Forecasting & Social Change*, Vol.14, pp.205-216.

[29] **Bury, K.V. (1975).**, *Statistical Models in Applied Science*, John Wiley & Sons, New York.

[30] **Cain, W.O. Jr (1980).**, "Risk Analysis Enhances Exploration", *Oil & Gas Journal*, Nov 17, 1980, pp.119-132.

[31] **Carr, R.I., Brightman, T.O., & Johnson, F.B. (1974).** "Progress Model for Construction Activity", *Journal of Construction Division, ASCE*, Vol. 100, No. CO1, pp.59-64.

[32] **Chapman, C.B. (1979).**, "Large Engineering Project Risk Analysis", *IEEE Transactions on Engineering Management*, Vol.EM 26, No.3, pp.78-86.

[33] **Chapman, C.B. & Cooper, D.F. (1983).**, "Risk Engineering: Basic Controlled Interval and Memory Models", *Journal of the Operational Research Society*, Vol.34, No.1, pp.51-60.

[34] **Chapman, C.B., Phillips, E.D., Cooper, D.F. & Lightfoot, L. (1985).**, "Selecting an Approach to Project Time and Cost Planning", *International Journal of Project Management*, Vol.3, No.1, pp.19-26.

[35] **Chesley, G.R. (1975).**, "Elicitation of Subjective Probabilities: A Review", *The Accounting Review*, Vol.50, pp.325-337.

[36] **Chicken, J.C., & Hayns, M.R. (1989).**, *The Risk Ranking Techniques in Decision Making*, Pergamon Press, U.K.

[37] **Christensen-Szalanski, J.J.J. & Beach, L.R., (1984).**, "The Citation Bias: Fad and Fashion in the Judgement and Decision Literature", *American Psychologist*, Vol.39, pp.75-78.

[38] **Clemen, R.T. (1986).**, "Calibration and the Aggregation of Probabilities", *Management Science*, Vol.32, No.3, pp.312-314.

[39] **Cooper, D.F., MacDonald, D.H., & Chapman, C.B. (1985).**, "Risk Analysis of a Construction Cost Estimate", *International Journal of Project Management*, Vol.3, No.1, pp.19-26.

[40] **Cooper, D.F., & Chapman, C.B. (1987).**, *Risk Analysis for Large Projects, Models, Methods and Cases*, John Wiley & Sons, N.Y.

[41] **Crandall, K.C. (1976).**, "Probabilistic Time Scheduling", *Journal of the Construction Division, ASCE*, Vol.102, No.CO3, pp.415-423.

[42] **Crandall, K.C. (1977).**, "Analysis of Schedule Simulations", *Journal of the Construction Division, ASCE*, Vol.103, No.CO3, pp.387-394.

[43] **Crandall, K.C., & Woolery, J.C. (1982).**, "Schedule Development under Stochastic Scheduling", *Journal of Construction Division, ASCE*, Vol.108, No.CO2, pp.321-329.

[44] **Davidson, L.B., & Cooper, D.O. (1976).**, "A Simple Way of Developing a Probability Distribution of Present Value", *Journal of Petroleum Technology*, pp.1069-1078.

[45] **DeCoster, D.T. (1970).**, "PERT/COST : The Challenge", *Information for Decision Making : Quantitative and Behavioral Dimensions*, A. Rappaport, Prentice-Hall, N.J., pp.249-257.

[46] **DeFinetti, B. (1970).**, *Theory of Probability*, John Wiley & Sons, New York.

[47] **DeGroot, M.H. (1970).**, *Optimal Statistical Decision*, McGraw-Hill Book Co., New York.

[48] **DeGroot, M.H. (1975).**, *Probability and Statistics*, Addison-Wesley, Reading, Mass.

[49] **DeGroot, M.H. (1979).**, "Improving Predictive Distributions ", *Bayesian Statistics* , J. M. Bernardo et al. (Eds.), University Press, Valencia, Spain., Vol.1, pp.385-395.

[50] **Delbecq, A.L., Van de Ven, A.H., & Gustafson, D.H. (1975).**, *Group Techniques for Program Planning*, Scott Foresman, Glenview, Illinois.

[51] **Der Kiureghian, A., & Liu, P-L. (1986).**, "Structural Reliability under Incomplete Probability Information", *Journal of Engineering Mechanics, ASCE*, Vol.112, No.1, pp.85-104.

[52] **Deshmukh, S.S. (1976).**, "Risk Analysis", *Transactions American Association of Cost Engineers*, pp.118-121.

[53] **Devore, J.L. (1982).**, *Probability & Statistics for Engineering and the Sciences* , Brooks / Cole Publishing Company, Monterey, California.

[54] **Diaconis, P., & Ylvisaker, D. (1985).**, "Quantifying Prior Opinion", *Bayesian Statistics 2*, J.M. Bernardo et al. (Eds.) Elsevier Science Publishing Co., N.Y., Vol.2, pp.133-156. 38.

[55] **Dickey, J.M. (1979).**, "Beliefs about Beliefs, A Theory for Stochastic Assessments of Subjective Probabilities", *Bayesian Statistics*, J.M. Bernardo et al. (Eds.), University Press, Valencia, Spain., Vol.1, pp.471-488.

[56] **Dickey, J.M., & Chen, C-H. (1985).**, "Direct Subjective Probability Modelling Using Ellipsoidal Distributions", *Bayesian Statistics 2*, J.M. Bernardo et al. (Eds.) Elsevier Science Publishing Co., N.Y., Vol.2, pp.157-182.

[57] **Dickey, J.M., Dawid, A.P., & Kadane, J.B., (1986).** "Subjective Probability Assessment Methods for Multivariate-t and Matrix-t Models", Chap 12, *Bayesian Inference and Decision Techniques: Studies in Bayesian Econometrics and Statistics*, Elsevier Science Publishing Co., New York, Vol.6, pp.177-195.

[58] **Diekmann, J.E. (1983).**, "Probabilistic Estimating : Mathematics and Applications " , *Journal of Construction Engineering & Management*, Vol.109, No.3, pp.297-307.

[59] **Eilon, S. & Fowkes, T.R. (1973).**, "Sampling Procedures for Risk Simulation", *Operation Research Quarterly*, Vol.24, No.2, pp.241-252.

[60] **Einhorn, H.J. (1972).**, "Expert Measurement and Mechanical Combination", *Organisational Behavior & Human Performance*, Vol.7, pp.86-106.

[61] **Elmaghraby, S.E. (1977).**, "Probabilistic Activity Networks (PANs): A Critical Evaluation and Extension of the PERT Model-Chap 4", *Activity Networks: Project Planning and Control by Network Models*, John Wiley & Sons., pp.228-320.

[62] **Farid, F., Boyer, L.T., & Kangari, R., (1989).** "Required Return on Investments in Construction" , *Journal of Construction Engineering & Management*, Vol.115, No.1, pp.109-125.

[63] **Flanagan, R. & Norman, G. (1980).**, "Risk Analysis - an Extension of Price Prediction Techniques for Building Work", *Construction Papers*, Vol.1, No.3 pp.27-34.

[64] **Flanagan, R., Kendall, A., Norman, G., & Robinson, G.D. (1987).**, "Life Cycle Costing and Risk Management", *Journal of Construction Management and Economics*, Vol.5, pp.S53-S71.

[65] **French, S. (1980).**, "Updating of Belief in the Light of Someone Else's Opinion", *Journal of Royal Statistical Society A*, Vol.143 Pt.1, pp.43-48.

[66] **French, S. (1982).**, "On the Axiomatisation of Subjective Probabilities", *Theory & Decision*, Vol.14, pp.19-33.

[67] **French, S. (1985).**, " Group Consensus Probability Distributions : A Critical Survey " , *Bayesian Statistics 2*, J.M. Bernardo et al. (Eds.), Elsevier Science Publishing Co., N.Y., Vol.2, pp.183-202.

[68] **French, S. (1986).**, "Calibration and the Expert Problem", *Management Science*, Vol. 32, No.3, pp.315-321.

[69] **Gates, M. (1971).**, "Bidding Contingencies and Probabilities", *Journal of the Construction Division, ASCE*, Vol.97, No.CO2, pp.277-303.

[70] **Graybill, F.A. (1983).**, *Matrices with Applications in Statistics*, Wadsworth International Group, CA.

[71] **Green, P.E. (1967).**, "Critique of: Ranking Procedures and Subjective Probability Distributions", *Management Science*, Vol.14, No.4, pp.B250-B251.

[72] **Gustafson, D.H., Shukla, R.K., Delbecq, A., & Walster, G.W. (1973).**, "A Comparative Study of Differences in Subjective Likelihood Estimates Made by Individuals, Interacting Groups, Delphi Groups, and Nominal Groups", *Organisational Behavior & Human Performance*, Vol.9, pp.281-291.

[73] **Hall, J.N. (1986).**, "Use of Risk Analysis in North Sea Projects", *International Journal of Project Management*, Vol.4, No.4, pp.217-222.

[74] **Hampton, J.M., Moore, P.G., & Thomas, H. (1973).** "Subjective Probability and its Measurement", *Journal of Royal Statistical Society, A*, Vol.136, Pt.1, pp.21-42.

[75] **Harr, M.E. (1987).**, *Reliability Based Design in Civil Engineering*, McGraw-Hill Book Co., New York.

[76] **Harris, R.B. (1978).**, *Precedence and Arrow Networking Techniques for Construction*, John Wiley & Sons, New York.

[77] **Hayes, R.W., Perry, J.G., Thompson, P.A., & Wilmer, G. (1986).**, "Risk Management in Engineering Construction, Implications for Project Managers", *Project Management Group, UMIST*, Thomas Telford Ltd.,London., pp.5-36.

[78] **Hemphill, R.B. (1968).**, "A Method for Predicting the Accuracy of a Construction Cost Estimate " , *Transactions of American Association of Cost Engineers* , pp.20-1-20-18.

[79] **Hendrickson, C., Martinelli, D., & Rehak, D. (1987).** "Hierarchical Rule-Based Activity Duration Estimation", *Journal of Construction Engineering & Management*, Vol.113, No.2, pp.288-301.

[80] **Hertz, D.B. (1964).**, "Risk Analysis in Capital Investment", *Harvard Business Review*, Vol.42, No.1, pp.95-106.

[81] **Hilliard, J.E., & Leitch, R.A. (1975).**, "Cost Volume Profit Analysis under Uncertainty", *Accounting Review*, Vol.50, No.1, pp.69-80.

[82] **Hillier, F.S. (1963).**, "The Derivation of Probabilistic Information for the Evaluation of Risky Investments", *Management Science*, Vol.9 , pp.443-457.

[83] **Hillier, F.S. (1969).**, *The Evaluation of Risky Interrelated Investments*, North Holland, Amsterdam.

[84] **Hogarth, R.M. (1975).**, "Cognitive Processes and the Assessment of Subjective Probability Distributions", *Journal of the American Statistical Association, Applied Section*, Vol.70, No.350 , pp.271-294.

[85] **Holt, C.A., Jr. (1986).**, "Scoring-Rule Procedures for Eliciting Subjective Probability and Utility Functions", Chap. 18, *Bayesian Inference and Decision Techniques: Studies in Bayesian Econometrics and Statistics*, Elsevier Science Publishing Co., New York, Vol.6, pp.279-290.

[86] **Howard, R.A. (1971).**, "Proximal Decision Analysis", *Management Science*, Vol.17, No.9, pp.507-541.

[87] **Howard, R.A. (1988).**, "Decision Analysis: Practice and Promise", *Management Science*, Vol.34, No.6, pp.679-695.

[88] **Huber, G.P. (1974).**, " Methods for Quantifying Subjective Probabilities and Multi - Attribute Utilities " , *Decision Science*, Vol.5, No.3, pp.430-458.

[89] **Hull, J.C. (1977).**, "Dealing with Dependence in Risk Simulation", *Operation Research Quarterly*, Vol.28, No.1, ii, pp.201-213.

[90] **Hull, J.C. (1978).**, "The Accuracy of the Means and Standard Deviations of Subjective Probability Distributions", *Journal of Royal Statistical Society, A*, Vol.141, Pt.1, pp.79-85.

[91] **Hull, J.C. (1980).**, *The Evaluation of Risk in Business Investment*, Pergamon Press.

[92] **Inyang, E.D. (1983).**, *Some Aspects of Risk Analysis for Decision Making in Engineering Project Management*, Ph.D. Thesis, University of Manchester, UMIST, U.K.

[93] **Jaafari, A. (1984).**, "Criticism of CPM for Project Planning Analysis", *Journal of Construction Engineering & Management*, Vol.110, No.2, Jun 1984, pp.222-233.

[94] **Jaafari, A. (1986).**, "Strategic Issues in Formulation and Management of Megaprojects in Australia", *International Journal of Project Management*, Vol.4, No.2, pp.198-206.

[95] **Jaafari, A. (1987).**, "Genesis of Management Confidence Technique", *Journal of Management in Engineering*, Vol.3, No.1, pp.60-80.

[96] **Jaafari, A. (1988a).**, "Project Viability and Economic Risk Analysis", *Journal of Management in Engineering*, Vol.4, No.1, pp.29-45.

[97] **Jaafari, A. (1988b).**, "New Vistas in Strategic Assessment of Projects", *Journal of Management in Engineering*, Vol.4, No.2, pp.122-138.

[98] **Jaafari, A. (1988c).**, "Probabilistic Unit Cost Estimation for Project Configuration Optimization", *International Journal of Project Management*, Vol.6, No.4, pp.226-234.

[99] **Jackson, P.S. (1982).**, "A Second - Order Moments Method for Uncertainty Analysis", *IEEE Transactions on Reliability*, Vol. R-31, No.4, pp.382-384.

[100] **Jaedicke, R.K., & Robichek, A.A. (1964).**, "Cost Volume Profit Analysis under Uncertainty", *Accounting Review* Vol.39, No.4, pp.917-926.

[101] **Johnson, M.E. (1987).**, *Multivariate Statistical Simulation*, John Wiley & Sons, New York.

[102] **Johnson, N.L., Nixon, E., & Amos, D.E. (1963).**, " Table of Percentage Points of Pearson Curves , for Given $\sqrt{\beta_1}$ and β_2 , Expressed in Standard Measure " , *Biometrika*, Vol.50, 3&4, pp.459-498.

[103] **Kadane, J.B., Dickey, J.M., Winkler, R.L., Smith, W.S., & Peters, S.C. (1980).**, "Interactive Elicitation of Opinion for a Normal Linear Model", *Journal of the American Statistical Association, T&M Section*, Vol.75, No.372 , pp.845-854.

[104] **Kalos, M.H. & Whitlock, P.A., (1986).**, *Monte Carlo Methods*, John Wiley & Sons, New York.

[105] **Keefer, D.L., & Bodily, S.E. (1983).**, "Three-Point Approximations for Continuous Random Variables", *Management Science*, Vol.29, No.5, pp.595-609.

[106] **Kendall, M.G., and Stuart, A. (1969).**, *The Advanced Theory of Statistics*, Vol.1, Third Edition, Hafner Publishing Co., New York.

[107] **Kennedy, K.W., & Thrall, R.M. (1976).**, " PLANET : A Simulation Approach to PERT " , *Computer & Operation Research*, Vol.3, pp.313-325.

[108] **King, W.R., & Wilson, T.A. (1967).**, "Subjective Time Estimates in Critical Path Planning - A Preliminary Analysis", *Management Science*, Vol.13, No.5, pp.307-320.

[109] **King, W.R., Wittevrongel, D.M., & Hezel, K.D. (1967).** "On the Analysis of Critical Path Time Estimating Behavior", *Management Science*, Vol.14, No.1, pp.79-84.

[110] **King, W.R., & Lukas, P.A. (1973).**, "An Experimental Analysis of Network Planning", *Management Science*, Vol.19, No.12, pp.1423-1432.

[111] **Kottas, J.F., & Lau, H-S. (1978).**, "Stochastic Breakeven Analysis", *Journal of the Operational Research Society*, Vol.29, No.3, pp.251-257.

[112] **Kottas, J.F., & Lau, H-S. (1980).**, "A Review of the Statistical Foundation of a Class of Probabilistic Planning Models", *Computer & Operation Research*, Vol.7, pp.277-284.

[113] **Kottas, J.F., & Lau, H-S. (1982).**, "A Four-Moments Alternative to Simulation for a Class of Stochastic Management Models", *Management Science*, Vol.28, No.7, pp.749-758.

[114] **Kryzanowski, L., Lusztig, P., & Schwab, B. (1972).** "Monte Carlo Simulation and Capital Expenditure Decisions - A Case Study", *The Engineering Economist* , Vol.18, No.1, pp.31-48.

[115] **Lindley, D.V. (1982).**, "The Improvement of Probability Judgements", *Journal of Royal Statistical Society A*, Vol.145 Pt.1, pp.117-126.

[116] **Lindley, D.V. (1986).**, "Another Look at an Axiomatic Approach to Expert Resolution", *Management Science*, Vol.32, No.3, pp.303-306.

[117] **Lindley, D.V., Tversky, A., & Brown, R.V. (1979).**, "On the Reconciliation of Probability Assessments", *Journal of Royal Statistical Society A*, Vol.142 Pt.2, pp.146-180.

[118] **Lock, A. (1987).**, "Integrating Group Judgements in Subjective Forecasts", *Judgmental Forecasting*, Edited by G. Wright and P. Ayton, John Wiley & Sons, N.Y., pp.109-127.

[119] **Ludke, R.L., Stauss, F.F., & Gustafson, D.H. (1977).** "Comparison of Five Methods for Estimating Subjective Probability Distributions", *Organisational Behavior & Human Performance*, Vol.19, pp.162-179.

[120] **Malcolm, D.G., Roseboom, J.H., Clark, C.E., & Fazar, W., (1959).**, "Applications of a Technique for Research and Development Program Evaluation (PERT) " , *Operations Research*, Vol.7, No.5, pp.646-669.

[121] **Makridakis, S., & Winkler, R.L. (1983).**, "Averages of Forecasts: Some Empirical Results", *Management Science*, Vol.29, No.9, pp.987-996.

[122] **McGough, E.H. (1982).**, "Scheduling: Effective Methods and Techniques", *Journal of the Construction Division, ASCE*, Vol.108, No.CO1, pp.75-84.

[123] **Milkovich, G.T., Annoni, A.J., & Mahoney, T.A. (1972).** "The Use of the Delphi Procedures in Manpower Forecasting", *Management Science*, Vol.19, No.4, Pt.1, pp.381-388.

[124] **Mirchandani, P.B. (1976).**, "Shortest Distance and Reliability of Probabilistic Networks", *Computer & Operation Research*, Vol.3, pp.347-355.

[125] **Moder, J.J. & Rodgers, E.G. (1968).**, "Judgement Estimates of the Moments of PERT Type Distributions", *Management Science*, Vol.15, No.2, pp.B77-B83.

[126] **Moeller, P.E. (1972).**, "VERT - A Tool to Assess Risk", *AIIE, Technical Papers*, pp.211-222.

[127] **Moore, P.G. (1977).**, "The Manager's Struggles with Uncertainty", *Journal of Royal Statistical Society A*, Vol.140 Pt.2, pp.129-165.

[128] **Morris, P.A. (1974).**, "Decision Analysis Expert Use", *Management Science*, Vol.20, No.9, pp.1233-1241.

[129] **Morris, P.A. (1977).**, "Combining Expert Judgements: A Bayesian Approach", *Management Science*, Vol.23, No.7, pp.679-693.

[130] **Morris, P.A. (1983).**, "An Axiomatic Approach to Expert Resolution", *Management Science*, Vol.29, No.1, pp.24-32.

[131] **Morris, P.A. (1986).**, "Observations on Expert Aggregation" , *Management Science*, Vol.32, No.3, pp.321-328.

[132] **Morrison, D.G. (1967).**, "Critique of: Ranking Procedures and Subjective Probability Distributions", *Management Science*, Vol.14, No.4, pp.B253-B254.

[133] **Murphy, A.H., & Winkler, R.L. (1971a).**, "Forecasters and Probability Forecasts: The Responses to a Questionnaire", *Bulletin of the American Meteorological Society*, Vol.52, No.3 , pp.158-165.

[134] **Murphy, A.H., & Winkler, R.L. (1971b).**, "Forecasters and Probability Forecasts: Some Current Problems", *Bulletin of the American Meteorological Society*, Vol.52, No.4, pp.239-247.

[135] **Murphy, A.H., & Winkler, R.L. (1975).**, "Subjective Probability Forecasting: Some Real World Experiments", *Utility, Probability & Human Decision*, D. Wendt & C.A.J. Vlek (Editors), Reidel, Dordrecht, pp.177-198.

[136] **Murphy, A.H., & Daan, H. (1984).**, "Impacts of Feedback and Experience on the Quality of Subjective Probability Forecasts : Comparison of Results from First and Second Years of the Zierikzee Experiment", *Monthly Weather Review*, Vol.112, No.3, pp.413-423.

[137] **Murphy, A.H., & Winkler, R.L. (1984).**, "Probability Forecasting in Meteorology", *Journal of the American Statistical Association, T&M Sect.*, Vol.79, No.387 , pp.489-500.

[138] **Murphy, A.H., Hsu, W-R., Winkler, R.L., & Wilks, D.S. (1985).**, "Use of Probabilities in Subjective Quantitative Precipitation Forecasts : Some Experimental Results", *Monthly Weather Review*, Vol.113, No.12, pp.2075-2089.

[139] **Myers, R.H. (1986).**, *Classical and Modern Regression with Applications*, Duxbury Press, Boston, Mass.

[140] **Newendorp, P.D. (1976).**, "A Method for Treating Dependencies Between Variables in Simulation Risk-Analysis Models", *Journal of Petroleum Technology*, Oct 1976, pp.1145-1150.

[141] **Ord, J.K. (1972).**, *Families of Frequency Distributions*, Charles Griffin & Co., London.

[142] **Ord, J.K. (1985).**, "Pearson System of Distributions", *Encyclopedia of Statistical Sciences*, Vol.6, pp.655-659.

[143] **Pearson, E.S. (1963).**, "Some Problems Arising in Approximating to Probability Distribution, Using Moments", *Biometrika*, Vol.50, 1&2, pp.95-112.

[144] **Pearson, E.S., & Tukey, J.W. (1965).**, "Approximate Means and Standard Deviations Based on Distances Between Percentage Points of Frequency Curves", *Biometrika*, Vol.52, 3&4, pp.533-546.

[145] **Perry, C., & Greig, I.D. (1975).**, "Estimating the Mean and Variance of Subjective Distributions in PERT and Decision Analysis", *Management Science*, Vol.21, No.12, pp.1477-1480.

[146] **Perry, J.G. (1986).**, "Risk Management - an Approach for Project Managers", *International Journal of Project Management*, Vol.4, No.4, pp.211-216.

[147] **Perry, J.G., et al. (1983).**, "Mersey Barrage Pre-Feasibility Study", *Marintech North West*, University of Manchester, Vol 1, 2, & 3.

[148] **Perry, J.G., & Hayes, R.W. (1985a).**, "Construction Projects - Knows the Risk", *Charted Mechanical Engineer*, February, pp.42-45.

[149] **Perry, J.G., & Hayes, R.W. (1985b).**, "Risk and its Management in Construction Projects", *Proceedings, Institution of Civil Engineers, U.K*, Part 1, Vol.78, pp.499-521.

[150] **Phillips, L.D. (1987).**, "On the Adequacy of Judgmental Forecasts", *Judgmental Forecasting*, Edited by G. Wright and P. Ayton, John Wiley & Sons, N.Y., pp.11-30.

[151] **Pouliquen, L.Y., (1970).**, *Risk Analysis in Project Appraisal*, The Johns Hopkins Press, Baltimore.

[152] **Pratt, J.W., Raiffa, H., & Schlaifer, R. (1964).** "The Foundations of Decision Under Uncertainty : An Elementary Exposition", *Journal of the American Statistical Association, T&M Sect.*, Vol.59, pp.353-375.

[153] **Pratt, J.W., & Schlaifer, R. (1985).**, "Repetitive Assessment of Judgmental Probability Distributions: A Case Study", *Bayesian Statistics 2*, J.M. Bernardo et al. (Eds.), Elsevier Science Publishing Co., N.Y., Vol.2, pp.393-424.

[154] **Press, S.J. (1979).**, "Bayesian Inference in Group Judgement Formulation and Decision Making using Qualitative Controlled Feedback " , *Bayesian Statistics*, J.M. Bernardo et al. (Eds.), University Press, Valencia, Spain, Vol.1, pp.397-415.

[155] **Press, S.J. (1985).**, "Multivariate Group Assessment of Probabilities of Nuclear War " , *Bayesian Statistics 2*, J.M. Bernardo et al. (Eds.), Elsevier Science Publishing Co., N.Y., Vol.2, pp.425-462.

[156] **Pritsker, A.A.B., & Happ, W.W. (1966).**, "GERT: Graphical Evaluation and Review Technique, Part I. Fundamentals", *The Journal of Industrial Engineering*, Vol.17, No.5,pp.267-274.

[157] **Pritsker, A.A.B., & Whitehouse, G.E. (1966).**, "GERT: Graphical Evaluation and Review Technique, Part II. Probabilistic and Industrial Engineering Applications", *The Journal of Industrial Engineering*, Vol.17, No.6, pp.293-301.

[158] **Project Performance Results for 1986 (1988)**, *Operations Evaluation Department*, The World Bank, Washington, D.C.

[159] **Ranasinghe, K.A.M.K, (1987)**, *Quantification of Risks During Feasibility Analysis for Capital Projects*, M.A.Sc. Thesis, University of British Columbia, Vancouver, Canada.

[160] **Ravinder, H.V., Kleinmuntz, D.W., & Dyer, J.S. (1988).** "The Reliability of Subjective Probabilities Obtained Through Decomposition", *Management Science*, Vol.34, No.2,pp.306-312.

[161] **Reinschmidt, K.F., & Frank, W.E. (1976).**, "Construction Cost Flow Management System", *Journal of the Construction Division, ASCE*, Vol.102, No.CO4, pp.615-627.

[162] **Reutlinger, S., (1970).**, *Techniques for Project Appraisal Under Uncertainty*, The Johns Hopkins Press, Baltimore.

[163] **Riggs, L.S. (1989).**, "Numerical Approach for Generating Beta Random Variates", *Journal of Computing in Civil Engineering*, Vol.3, No.2, pp.183-191.

[164] **Russell, A.D. (1985)**, *CIVL 520 - Course Notes*, Department of Civil Engineering, University of British Columbia, Vancouver.

[165] **Savage, L.J. (1954).**, *The Foundations of Statistics*, John Wiley & Sons, New York.

[166] **Savage, L.J. (1971).**, "Elicitation of Personal Probabilities and Expectations", *Journal of the American Statistical Association, T&M Section*, Vol.66, No.336 , pp.783-801.

[167] **Schervish, M.J. (1986).**, "Comments on Some Axioms for Combining Expert Judgements", *Management Science*, Vol.32, No.3, pp.306-312.

[168] **Seaver, D.A. (1977).**, "How Groups Can Assess Uncertainty: Human Interaction Versus Mathematical Models", *IEEE, International Conference on Cybernetics*, pp.185-190.

[169] **Seaver, D.A., Von Winterfeldt, D., & Edwards, W. (1978).** "Eliciting Subjective Probability Distributions on Continuous Variables", *Organisational Behavior & Human Performance*, Vol.21, pp.379-391.

[170] **Selvidge, J. (1975).**, "A Three Step Procedure for Assigning Probabilities to Rare Events", *Utility, Probability & Human Decision*, D. Wendt & C.A.J. Vlek (Editors), Reidel, Dordrecht, pp.199-216.

[171] **Siddall, J.N. (1972).**, *Analytical Decision-Making in Engineering Design*, Prentice-Hall, Englewood Cliffs, N.J.

[172] **Shafer, S.L. (1974)** "Risk Analysis for Capital Projects Using Risk Elements", *Transactions American Association of Cost Engineers*, pp.218-223.

[173] **Smith, D.E. (1971).**, " A Taylor's Theorem - Central Limit Theorem Approximation: It's use in Obtaining the Probability Distribution of Long-Range Profit", *Management Science*, Vol.18, No.4, Part I, pp.B214-B219.

[174] **Smith, K.A., & Thoem, R.L. (1976).**, "Project Cost Evaluation Using Probability Concepts", *Transactions American Association of Cost Engineers*, pp.275-279.

[175] **Smith, L.H. (1967).**, " Ranking Procedures and Subjective Probability Distributions " , *Management Science*, Vol.14, No.4, pp.B236-B249.

[176] **Spetzler, C.S & Stael von Holstein, C-A. (1975).** "Probability Encoding in Decision Analysis", *Management Science*, Vol.22, No.3, pp.340-358.

[177] **Spooner, J.E. (1974).**, " Probabilistic Estimating " , *Journal of the Construction Division , ASCE* , Vol.100, No. CO1, pp.65-77.

[178] **Stael von Holstein, C-A.S. (1971).**, "An Experiment in Probabilistic Weather Forecasting", *Journal of Applied Meteorology*, Vol.10, pp.635-645.

[179] **Stael von Holstein, C-A.S. (1972).**, "Probabilistic Forecasting: An Experiment Related to the Stock Market", *Organisational Behavior & Human Performance*, Vol.8, pp.139-158.

[180] **Starr, M.K., & Tapiero, C.S. (1975).**, "Linear Breakeven Analysis Under Risk", *Journal of the Operational Research Society*, Vol.26, No.4, pp.847-856.

[181] **Suppes, P. (1975).**, "The Measurement of Belief", *Journal of Royal Statistical Society B*, Vol.36, pp.160-191.

[182] **Tanchoco, J.M.A., Buck, J.R., & Leung, L.C. (1981).** "Modeling and Discounting of Continuous Cash Flow Under Risk", *Engineering Costs and Production Economics*, Vol.5, pp.205-216.

[183] **Taylor, P. (1988).**, "Superiority of NPV over NPV/K, and Other Criteria in the Financial Appraisal of Projects: The Case of Energy Conservation", *International Journal of Project Management*, Vol.6, No.4, pp.223-225.

[184] **The Twelfth Annual Review of Project Performance Results (1987)**, *Operations Evaluation Department*, The World Bank, Washington, D.C.

[185] **Thompson, P.A. (1981).**, *Organisation and Economics of Construction*, McGraw-Hill Book Company, U.K.

[186] **Thompson, P.A. & Whitman, J.D. (1973).**, "Project Appraisal by Cost Model", *Management Decision*, Vol.11, pp.301-308.

[187] **Thompson, P.A. & Whitman, J.D. (1974).**, "A Network Based Financial Modelling System Linking Project Appraisal, Construction and Operation", *4th International INTERNET Congress*, Paris, pp.329-336.

[188] **Thompson, P.A. et al. (1980).**, *Severn Tidal Power, Sensitivity and Risk Analysis.*, Project Management Group, University of Manchester, UMIST.

[189] **Thompson, P.A. & Wilmer, G. (1985).**, "CASPAR - A Program for Engineering Project Appraisal and Management", *2nd International Conference on Civil & Structural Engineering Computing*, Vol.1, pp.75-81.

[190] **Tukey, J.W. (1954).**, "The Propagation of Errors, Fluctuations and Tolerances, Basic Generalized Formulas", *Technical Report No. 10*, Statistical Techniques Research Group, Princeton University, Princeton, N.J.

[191] **Tversky, A. & Kahneman, D. (1984).**, *Judgement under Uncertainty: Heuristics and Biases*, Cambridge University Press.

[192] **Van Tetterode, L.M. (1971).**, "Risk Analysis ...or Russian Roulette ?", *Transactions American Association of Cost Engineers*, pp.185-191.

[193] **Vergara, A.J. & Boyer, L.T. (1974).**, "Probabilistic Approach to Estimating and Cost Control", *Journal of the Construction Division, ASCE*, Vol.100, No.CO4, pp.543-552.

[194] **Wagle, B. (1967).**, "A Statistical Analysis of Risk in Capital Investment Projects", *Operation Research Quarterly*, Vol.18, No.1, pp.13-33.

[195] **Wallace, D.M. (1977).**, "Statistical Analysis of Cost Data", *Transactions American Association of Cost Engineers*, pp.286-296

[196] **Wallsten, T.S., & Budescu, D.V. (1983).**, "Encoding Subjective Probabilities: A Psychological and Psychometric Review", *Management Science*, Vol.29, No.2, pp.151-173.

[197] **Wilkinson, J.H. (1965).**, *The Algebraic Eigenvalue Problem*, Clarendon Press, Oxford, U.K.

[198] **Winkler, R.L. (1967a).**, "The Assessment of Prior Distributions in Bayesian Analysis", *Journal of the American Statistical Association, T&M Section*, Vol.62, pp.776-800.

[199] **Winkler, R.L. (1967b).**, "The Quantification of Judgement : Some Methodological Suggestions", *Journal of the American Statistical Association, T&M Sect.*, Vol.62, pp.1105-1120.

[200] **Winkler, R.L. (1968).**, "The Consensus of Subjective Probability Distributions", *Management Science*, Vol.15, No.2, pp.B61-B75.

[201] **Winkler, R.L. (1971).**, "Probabilistic Prediction: Some Experimental Results", *Journal of the American Statistical Association, T&M Section*, Vol.66, No.336, pp.675-685.

[202] **Winkler, R.L. (1981).**, "Combining Probability Distributions from Dependent Information Sources", *Management Science*, Vol.27, No.4, pp.479-488.

[203] **Winkler, R.L. (1983).**, "Judgement Under Uncertainty", *Encyclopedia of Statistical Sciences*, Vol.4, pp.332-336.

[204] **Winkler, R.L. (1986a).**, "Expert Resolution", *Management Science*, Vol.32, No.3, pp.298-303.

[205] **Winkler, R.L. (1986b).**, "On 'Good Probability Appraisers", Chap 17, *Bayesian Inference and Decision Techniques: Studies in Bayesian Econometrics and Statistics*, Elsevier Science Publishing Co., New York, Vol.6, pp.265-278.

[206] **Woolery, J.C., & Crandall, K.C. (1983).**, "Stochastic Network Model for Planning Scheduling", *Journal of Construction Engineering & Management*, Vol.109, No.3, pp.342-354.

[207] **Wright, G. & Ayton, P. (1987).**, "The Psychology of Forecasting", *Judgmental Forecasting*, Edited by G. Wright and P. Ayton, John Wiley & Sons, N.Y., pp.83-105.

[208] **Yeo, K.T. (1982).**, *A Systems Approach to Cost Management of Large Scale Offshore Oil Projects*, Ph.D. Thesis, University of Manchester, UMIST, U.K.

[209] **Youker, R. (1989).**, "Managing the Project Cycle for Time, Cost and Quality: Lessons from World Bank Experience", *International Journal of Project Management*, Vol.7, No.1, pp.43-48.

[210] **Zinn, C.D., Lesso, W.G., & Motazed, B. (1977).** "A Probabilistic Approach to Risk Analysis in Capital Investment Projects", *The Engineering Economist*, Vol.22, No.4, pp.239-260.

Appendix A

The First Four Moments

A.1 General

This appendix derives the functions for first four moments of the correlated primary variables and the transformed uncorrelated variables from their definitions.

Let a derived variable be described by the function $Y = g(\mathbf{X})$, where $\mathbf{X}$ is the vector of its correlated primary variables. The truncated second order Taylor series expansion of $g(\mathbf{X})$ about the mean values $\bar{\mathbf{X}}$ is given by (equation 2.19),

$$\begin{aligned} g(\mathbf{X}) \approx \quad & g(\bar{\mathbf{X}}) + \sum_{i=1}^{n} \frac{\partial g}{\partial X_i} (X_i - \bar{X}_i) \\ & + \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \frac{\partial^2 g}{\partial X_i \partial X_j} (X_i - \bar{X}_i)(X_j - \bar{X}_j) \end{aligned} \tag{A.1}$$

The function is transformed to the uncorrelated space by equation (2.35). Then the function becomes $Y = G(\mathbf{Z})$ where $\mathbf{Z}$ is the vector of transformed uncorrelated variables.

A.2 Expected Value

From equation (2.20) the expected value of Y is,

$$E[Y] \approx g(\bar{\mathbf{X}}) + \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \frac{\partial^2 g}{\partial X_i \partial X_j} cov(X_i, X_j) \tag{A.2}$$

In the transformed system the variables are uncorrelated. When using the transformed system function $G(\mathbf{Z})$ the expected value of Y is (equation 2.36),

$$E[Y] \approx G(\bar{\mathbf{Z}}) + \frac{1}{2}\sum_{i=1}^{n}\frac{\partial^2 G}{\partial Z_i^2}\,\mu_2(Z_i) \tag{A.3}$$

A.3 Second Central Moment

From equation (2.21) the second central moment of Y is,

$$\begin{aligned}\mu_2(Y) = E\Bigg[\Bigg(&\sum_{i=1}^{n}\frac{\partial g}{\partial X_i}(X_i - \bar{X}_i)\\ &+\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}\frac{\partial^2 g}{\partial X_i \partial X_j}(X_i - \bar{X}_i)(X_j - \bar{X}_j)\\ &-\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}\frac{\partial^2 g}{\partial X_i \partial X_j}\,cov(X_i, X_j)\Bigg)^2\Bigg]\end{aligned} \tag{A.4}$$

$$\begin{aligned}\mu_2(Y) = E\Bigg[&\sum_{i=1}^{n}\sum_{j=1}^{n}\frac{\partial g}{\partial X_i}\frac{\partial g}{\partial X_j}(X_i - \bar{X}_i)(X_j - \bar{X}_j)\\ &+\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n}\frac{\partial g}{\partial X_i}\frac{\partial^2 g}{\partial X_j \partial X_k}(X_i - \bar{X}_i)(X_j - \bar{X}_j)(X_k - \bar{X}_k)\\ &-\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n}\frac{\partial g}{\partial X_i}\frac{\partial^2 g}{\partial X_j \partial X_k}(X_i - \bar{X}_i)\,cov(X_j, X_k)\\ &+\frac{1}{4}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n}\sum_{l=1}^{n}\frac{\partial^2 g}{\partial X_i \partial X_j}\frac{\partial^2 g}{\partial X_k \partial X_l}(X_i - \bar{X}_i)(X_j - \bar{X}_j)\\ &\qquad\times(X_k - \bar{X}_k)(X_l - \bar{X}_l)\\ &-\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n}\sum_{l=1}^{n}\frac{\partial^2 g}{\partial X_i \partial X_j}\frac{\partial^2 g}{\partial X_k \partial X_l}(X_i - \bar{X}_i)(X_j - \bar{X}_j)\\ &\qquad\times cov(X_k, X_l)\\ &+\frac{1}{4}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n}\sum_{l=1}^{n}\frac{\partial^2 g}{\partial X_i \partial X_j}\frac{\partial^2 g}{\partial X_k \partial X_l}\,cov(X_i, X_j)\,cov(X_k, X_l)\Bigg]\end{aligned} \tag{A.5}$$

Neglecting the cross moment terms in equation (A.5) that cannot be defined due to the lack of moment information,

$$\begin{aligned}
\mu_2(Y) \approx\ & \sum_{i=1}^{n} \left[\frac{\partial g}{\partial X_i}\right]^2 \mu_2(X_i) \\
& + 2 \sum_{i=1}^{n} \sum_{j=i+1}^{n} \frac{\partial g}{\partial X_i} \frac{\partial g}{\partial X_j} cov(X_i, X_j) \\
& + \sum_{i=1}^{n} \frac{\partial g}{\partial X_i} \frac{\partial^2 g}{\partial X_i^2} \mu_3(X_i) \\
& + \frac{1}{4} \sum_{i=1}^{n} \left[\frac{\partial^2 g}{\partial X_i^2}\right]^2 \mu_4(X_i) \\
& - \frac{1}{2} \sum_{i=1}^{n} \left[\frac{\partial^2 g}{\partial X_i^2}\right]^2 [\mu_2(X_i)]^2 \\
& - 2 \sum_{i=1}^{n} \sum_{j=i+1}^{n} \left[\frac{\partial^2 g}{\partial X_i \partial X_j}\right]^2 [cov(X_i, X_j)]^2 \\
& + \frac{1}{4} \sum_{i=1}^{n} \left[\frac{\partial^2 g}{\partial X_i^2}\right]^2 [\mu_2(X_i)]^2 \\
& + \sum_{i=1}^{n} \sum_{j=i+1}^{n} \left[\frac{\partial^2 g}{\partial X_i \partial X_j}\right]^2 [cov(X_i, X_j)]^2
\end{aligned} \tag{A.6}$$

$$\begin{aligned}
\mu_2(Y) \approx\ & \sum_{i=1}^{n} \left[\frac{\partial g}{\partial X_i}\right]^2 \mu_2(X_i) \\
& + 2 \sum_{i=1}^{n} \sum_{j=i+1}^{n} \frac{\partial g}{\partial X_i} \frac{\partial g}{\partial X_j} cov(X_i, X_j) \\
& + \sum_{i=1}^{n} \frac{\partial g}{\partial X_i} \frac{\partial^2 g}{\partial X_i^2} \mu_3(X_i) \\
& + \frac{1}{4} \sum_{i=1}^{n} \left[\frac{\partial^2 g}{\partial X_i^2}\right]^2 \left[\mu_4(X_i) - [\mu_2(X_i)]^2\right] \\
& - \sum_{i=1}^{n} \sum_{j=i+1}^{n} \left[\frac{\partial^2 g}{\partial X_i \partial X_j}\right]^2 [cov(X_i, X_j)]^2
\end{aligned} \tag{A.7}$$

In the transformed system the variables are uncorrelated. When using the transformed system function $G(\mathbf{Z})$ the second central moment of Y is (equation 2.37),

$$
\begin{aligned}
\mu_2(Y) \approx\; & \sum_{i=1}^{n} \left[\frac{\partial G}{\partial Z_i}\right]^2 \mu_2(Z_i) \\
& + \sum_{i=1}^{n} \frac{\partial G}{\partial Z_i} \frac{\partial^2 G}{\partial Z_i^2} \mu_3(Z_i) \\
& + \frac{1}{4} \sum_{i=1}^{n} \left[\frac{\partial^2 G}{\partial Z_i^2}\right]^2 \left[\mu_4(Z_i) - [\mu_2(Z_i)]^2\right]
\end{aligned}
\tag{A.8}
$$

If all the correlated primary variables are normally distributed or if it is assumed that there are no non-linear correlations between the transformed variables,

$$
\begin{aligned}
\mu_2(Y) \approx\; & \sum_{i=1}^{n} \left[\frac{\partial G}{\partial Z_i}\right]^2 \mu_2(Z_i) \\
& + \sum_{i=1}^{n} \frac{\partial G}{\partial Z_i} \frac{\partial^2 G}{\partial Z_i^2} \mu_3(Z_i) \\
& + \frac{1}{4} \sum_{i=1}^{n} \left[\frac{\partial^2 G}{\partial Z_i^2}\right]^2 \left[\mu_4(Z_i) - [\mu_2(Z_i)]^2\right] \\
& + \sum_{i=1}^{n} \sum_{j=i+1}^{n} \left[\frac{\partial^2 G}{\partial Z_i \partial Z_j}\right]^2 \mu_2(Z_i)\, \mu_2(Z_j)
\end{aligned}
\tag{A.9}
$$

A.4 Third Central Moment

From equation (2.21) the third central moment of Y is,

$$
\begin{aligned}
\mu_3(Y) = E\Bigg[\Bigg(& \sum_{i=1}^{n} \frac{\partial g}{\partial X_i} (X_i - \bar{X}_i) \\
& + \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \frac{\partial^2 g}{\partial X_i \partial X_j} (X_i - \bar{X}_i)(X_j - \bar{X}_j) \\
& - \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \frac{\partial^2 g}{\partial X_i \partial X_j} cov(X_i, X_j)\Bigg)^3\Bigg]
\end{aligned}
\tag{A.10}
$$

Neglecting fifth and higher order terms in equation (A.10),

$$\begin{aligned}
\mu_3(Y) \;\approx\; & E\left[\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n}\frac{\partial g}{\partial X_i}\frac{\partial g}{\partial X_j}\frac{\partial g}{\partial X_k}(X_i - \bar{X}_i)(X_j - \bar{X}_j)(X_k - \bar{X}_k)\right. \\
& + \frac{3}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n}\sum_{l=1}^{n}\frac{\partial g}{\partial X_i}\frac{\partial g}{\partial X_j}\frac{\partial^2 g}{\partial X_k \partial X_l}(X_i - \bar{X}_i)(X_j - \bar{X}_j) \\
& \qquad \times (X_k - \bar{X}_k)(X_l - \bar{X}_l) \\
& - \frac{3}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n}\sum_{l=1}^{n}\frac{\partial g}{\partial X_i}\frac{\partial g}{\partial X_j}\frac{\partial^2 g}{\partial X_k \partial X_l}(X_i - \bar{X}_i)(X_j - \bar{X}_j) \\
& \qquad \left. \times\, cov(X_k, X_l)\right]
\end{aligned} \tag{A.11}$$

Neglecting the cross moment terms in equation (A.11) that cannot be defined due to the lack of moment information,

$$\begin{aligned}
\mu_3(Y) \;\approx\; & \sum_{i=1}^{n}\left[\frac{\partial g}{\partial X_i}\right]^3 \mu_3(X_i) \\
& + \frac{3}{2}\sum_{i=1}^{n}\left[\frac{\partial g}{\partial X_i}\right]^2 \frac{\partial^2 g}{\partial X_i^2}\,\mu_4(Z_i) \\
& - \frac{3}{2}\sum_{i=1}^{n}\left[\frac{\partial g}{\partial X_i}\right]^2 \frac{\partial^2 g}{\partial X_i^2}\,[\mu_2(Z_i)]^2 \\
& - 6\sum_{i=1}^{n}\sum_{j=i+1}^{n}\frac{\partial g}{\partial X_i}\frac{\partial g}{\partial X_j}\frac{\partial^2 g}{\partial X_i \partial X_j}\,[cov(X_i, X_j)]^2
\end{aligned} \tag{A.12}$$

$$\begin{aligned}
\mu_3(Y) \;\approx\; & \sum_{i=1}^{n}\left[\frac{\partial g}{\partial X_i}\right]^3 \mu_3(X_i) \\
& + \frac{3}{2}\sum_{i=1}^{n}\left[\frac{\partial g}{\partial X_i}\right]^2 \frac{\partial^2 g}{\partial X_i^2}\left[\mu_4(Z_i) - [\mu_2(Z_i)]^2\right] \\
& - 6\sum_{i=1}^{n}\sum_{j=i+1}^{n}\frac{\partial g}{\partial X_i}\frac{\partial g}{\partial X_j}\frac{\partial^2 g}{\partial X_i \partial X_j}\,[cov(X_i, X_j)]^2
\end{aligned} \tag{A.13}$$

In the transformed system the variables are uncorrelated. When using the transformed system function $G(\mathbf{Z})$ the third central moment of Y is (equation 2.38),

$$\begin{aligned}\mu_3(Y) \approx & \sum_{i=1}^{n} \left[\frac{\partial G}{\partial Z_i}\right]^3 \mu_3(Z_i) \\ & + \frac{3}{2}\sum_{i=1}^{n} \left[\frac{\partial G}{\partial Z_i}\right]^2 \frac{\partial^2 G}{\partial Z_i^2} \left[\mu_4(Z_i) - [\mu_2(Z_i)]^2\right]\end{aligned} \tag{A.14}$$

If all the correlated primary variables are normally distributed or if it is assumed that there are no non-linear correlations between the transformed variables,

$$\begin{aligned}\mu_3(Y) \approx & \sum_{i=1}^{n} \left[\frac{\partial G}{\partial Z_i}\right]^3 \mu_3(Z_i) \\ & + \frac{3}{2}\sum_{i=1}^{n} \left[\frac{\partial G}{\partial Z_i}\right]^2 \frac{\partial^2 G}{\partial Z_i^2} \left[\mu_4(Z_i) - [\mu_2(Z_i)]^2\right] \\ & + 6 \sum_{i=1}^{n} \sum_{j=i+1}^{n} \frac{\partial G}{\partial Z_i} \frac{\partial G}{\partial Z_j} \frac{\partial^2 G}{\partial Z_i \partial Z_j} \mu_2(Z_i)\, \mu_2(Z_j)\end{aligned} \tag{A.15}$$

A.5 Fourth Central Moment

From equation (2.23) the fourth central moment of Y is,

$$\begin{aligned}\mu_4(Y) = E\Bigg[\Bigg(& \sum_{i=1}^{n} \frac{\partial g}{\partial X_i} (X_i - \bar{X}_i) \\ & + \frac{1}{2}\sum_{i=1}^{n} \sum_{j=1}^{n} \frac{\partial^2 g}{\partial X_i \partial X_j} (X_i - \bar{X}_i)(X_j - \bar{X}_j) \\ & - \frac{1}{2}\sum_{i=1}^{n} \sum_{j=1}^{n} \frac{\partial^2 g}{\partial X_i \partial X_j} cov(X_i, X_j)\Bigg)^4\Bigg]\end{aligned} \tag{A.16}$$

Neglecting fifth and higher order terms in equation (A.16),

$$\mu_4(Y) \approx E\left[\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n}\sum_{l=1}^{n} \frac{\partial g}{\partial X_i}\frac{\partial g}{\partial X_j}\frac{\partial g}{\partial X_k}\frac{\partial g}{\partial X_l}(X_i - \bar{X}_i)(X_j - \bar{X}_j) \times (X_k - \bar{X}_k)(X_l - \bar{X}_l)\right] \tag{A.17}$$

Neglecting the cross moment terms in equation (A.17) that cannot be defined,

$$\mu_4(Y) \approx \sum_{i=1}^{n}\left[\frac{\partial g}{\partial X_i}\right]^4 \mu_4(X_i) \tag{A.18}$$

In the transformed system the variables are uncorrelated. When using the transformed system function $G(\mathbf{Z})$ the fourth central moment of Y is (equation 2.39),

$$\mu_4(Y) \approx \sum_{i=1}^{n}\left[\frac{\partial G}{\partial Z_i}\right]^4 \mu_4(Z_i) \tag{A.19}$$

If all the correlated primary variables are normally distributed or if it is assumed that there are no non-linear correlations between the transformed variables,

$$\mu_4(Y) \approx \sum_{i=1}^{n}\left[\frac{\partial G}{\partial Z_i}\right]^4 \mu_4(Z_i) \tag{A.20}$$

$$+ 6\sum_{i=1}^{n}\sum_{j=i+1}^{n}\left[\frac{\partial G}{\partial Z_i}\right]^2\left[\frac{\partial G}{\partial Z_j}\right]^2 \mu_2(Z_i)\,\mu_2(Z_j)$$

A.6 Note : Higher Order Moments

The fifth and higher order terms are neglected in the approximations for the third and fourth central moments of a derived variable because moment information of the primary variables are not available beyond the fourth order (section 2.3.2). However, since all of the primary variables are approximated by Pearson type distributions, if required, it is possible to generate higher order central moments for the primary variables from the recurrence property of the Pearson family (Kendall and Stuart, 1969). The recurrence relationship for fifth and higher order central moments for primary variables approximated by Pearson type distributions is (Pearson, 1963; Kendall and Stuart, 1969),

$$\mu_{n+1} = \frac{[a - (n+1)\,b_1]\,\mu_n - n\,b_0\,\mu_{n-1}}{(n+2)\,b_2 + 1} \tag{A.21}$$

where

$$a = -\frac{\mu_3\,(\mu_4 + 3\mu_2^2)}{10\mu_4\mu_2 - 18\mu_2^3 - 12\mu_3^2} \tag{A.22}$$

$$b_0 = -\frac{\mu_2\,(4\mu_2\mu_4 - 3\mu_3^2)}{10\mu_4\mu_2 - 18\mu_2^3 - 12\mu_3^2} \tag{A.23}$$

$$b_1 = -\frac{\mu_3\,(\mu_4 + 3\mu_2^2)}{10\mu_4\mu_2 - 18\mu_2^3 - 12\mu_3^2} \tag{A.24}$$

$$b_2 = -\frac{(2\mu_2\mu_4 - 3\mu_3^2 - 6\mu_2^3)}{10\mu_4\mu_2 - 18\mu_2^3 - 12\mu_3^2} \tag{A.25}$$

Appendix B

Investigation of $\mathbf{R_0}$

This appendix investigates the positive semi-definite correlation matrix $\mathbf{R_0}$ discussed in section (4.2.1). The correlation matrix is as follows.

$$\mathbf{R_0} = \begin{bmatrix} 1.0 & 0.5 & 0.5 \\ 0.5 & 1.0 & -0.5 \\ 0.5 & -0.5 & 1.0 \end{bmatrix} \tag{B.1}$$

Since the correlation matrix is in the standardized space, $E\left[X_i\right] = 0$ and $\mu_2(X_i) = 1$. The variances of the linear combinations of variables are,

$$\mu_2(X_1 + X_2) = \mu_2(X_1) + \mu_2(X_2) + 2\,Cov(X_1, X_2) \neq 1 \tag{B.2}$$

$$\mu_2(X_1 + X_3) = \mu_2(X_1) + \mu_2(X_3) + 2\,Cov(X_1, X_3) \neq 1 \tag{B.3}$$

$$\mu_2(X_2 + X_3) = \mu_2(X_2) + \mu_2(X_3) + 2\,Cov(X_2, X_3) = 1 \tag{B.4}$$

Therefore, only the linear combination of variables 2 and 3 is valid. Hence,

$$\rho_{1,2+3} = \frac{Cov(X_1, X_2 + X_3)}{\sqrt{\mu_2(X_1)\,\mu_2(X_2 + X_3)}} \tag{B.5}$$

Since $\mu_2(X_1) = 1$ and $\mu_2(X_2 + X_3) = 1$

$$\rho_{1,2+3} = Cov(X_1, X_2 + X_3) \tag{B.6}$$

From definition,

$$Cov(X_1, X_2 + X_3) = E[X_1(X_2 + X_3)] - E[X_1]\, E[X_2 + X_3] \tag{B.7}$$

$$= E[X_1\, X_2] + E[X_1\, X_3]$$

$$= E[X_1]E[X_2] + Cov(X_1, X_2) + E[X_1]E[X_3] + Cov(X_1, X_3)$$

Therefore,

$$Cov(X_1, X_2 + X_3) = Cov(X_1, X_2) + Cov(X_1, X_3) \tag{B.8}$$

From definition,

$$Cov(X_1, X_2) = \rho_{1,2} \sqrt{\mu_2(X_1)\, \mu_2(X_2)} = 0.5 \tag{B.9}$$

$$Cov(X_1, X_3) = \rho_{1,3} \sqrt{\mu_2(X_1)\, \mu_2(X_3)} = 0.5 \tag{B.10}$$

From equations (B.6), (B.8), (B.9) and (B.10),

$$\rho_{1,2+3} = Cov(X_1, X_2) + Cov(X_1, X_3) = 1 \tag{B.11}$$

Variable 1 is perfectly correlated with the linear combination of Variables 2 and 3.

Appendix C

Bounds for a Correlation Coefficient

This appendix derives the bounds for a correlation coefficient to exist in a positive definite correlation matrix. The bounds are derived from the necessary condition on vector $\mathbf{b}$ (see equation C.17) given by Kadane et al., (1980), for a nxn correlation matrix $\mathbf{R_n}$ to be positive definite when $\mathbf{R_{n-1}}$ is positive definite. The proof for that condition was derived by Dr. Ricardo O. Foschi during the review period of this thesis. His contribution which is described in the first section is gratefully acknowledged.

C.1 The Proof

Let $\mathbf{R_n}$ be a nxn correlation matrix partitioned as,

$$\mathbf{R_n} = \begin{bmatrix} \mathbf{R_{n-1}} & \mathbf{b} \\ \mathbf{b^T} & 1 \end{bmatrix} \tag{C.1}$$

where $\mathbf{R_{n-1}}$ is $(n-1)$x$(n-1)$ correlation matrix for $n = 2, 3, ...$ and $\mathbf{b^T} = [\rho_{1n} \ \ \rho_{2n} \ .... \ \rho_{n-1n}]$.

Let $\mathbf{R_{n-1}}$ be positive definite. For any vector of n scalars $\mathbf{x}$

$$[\mathbf{x_{n-1}^T} \quad x_n] \begin{bmatrix} \mathbf{R_{n-1}} & \mathbf{b} \\ \mathbf{b^T} & 1 \end{bmatrix} \begin{bmatrix} \mathbf{x_{n-1}} \\ x_n \end{bmatrix} > 0 \tag{C.2}$$

where $\mathbf{x_{n-1}^T} = [x_1 \ x_2 \ .... \ x_{n-1}]$.

Expanding equation (C.2),

$$\mathbf{x_{n-1}^T} \ \mathbf{R_{n-1}} \ \mathbf{x_{n-1}} \ + \ 2 \ x_n \ \mathbf{x_{n-1}^T} \ \mathbf{b} \ + \ x_n^2 \ > \ 0 \tag{C.3}$$

The roots for x_n from equation (C.3) are,

$$x_n = -\mathbf{x}_{n-1}^T \mathbf{b} \pm \sqrt{(\mathbf{x}_{n-1}^T \mathbf{b})^2 - \mathbf{x}_{n-1}^T \mathbf{R}_{n-1} \mathbf{x}_{n-1}} \tag{C.4}$$

For imaginary roots

$$\mathbf{x}_{n-1}^T \mathbf{b} \mathbf{b}^T \mathbf{x}_{n-1} - \mathbf{x}_{n-1}^T \mathbf{R}_{n-1} \mathbf{x}_{n-1} < 0 \tag{C.5}$$

Rewriting equation (C.5),

$$\mathbf{x}_{n-1}^T \left[\mathbf{R}_{n-1} - \mathbf{b} \mathbf{b}^T\right] \mathbf{x}_{n-1} > 0 \tag{C.6}$$

Therefore, for $\mathbf{R_n}$ to be positive definite $(\mathbf{R}_{n-1} - \mathbf{b}\mathbf{b}^T)$ must be positive definite. Then, for any vector of $n-1$ scalars $\mathbf{y}$,

$$\mathbf{y}^T \left[\mathbf{R}_{n-1} - \mathbf{b} \mathbf{b}^T\right] \mathbf{y} > 0 \tag{C.7}$$

Since $\mathbf{R}_{n-1}$ is symmetric $\mathbf{R}_{n-1}^{-1}$ is also symmetric. Choosing $\mathbf{y} = \mathbf{R}_{n-1}^{-1} \mathbf{b}$

$$\mathbf{y}^T = \mathbf{b}^T (\mathbf{R}_{n-1}^{-1})^T = \mathbf{b}^T \mathbf{R}_{n-1}^{-1} \tag{C.8}$$

Substituting in equation (C.7),

$$\mathbf{b}^T \mathbf{R}_{n-1}^{-1} \left[\mathbf{R}_{n-1} - \mathbf{b} \mathbf{b}^T\right] \mathbf{R}_{n-1}^{-1} \mathbf{b} > 0 \tag{C.9}$$

Expanding equation (C.9),

$$\left[1 - \mathbf{b}^T \mathbf{R}_{n-1}^{-1} \mathbf{b}\right] \mathbf{b}^T \mathbf{R}_{n-1}^{-1} \mathbf{b} > 0 \tag{C.10}$$

Since $\mathbf{R}_{n-1}$ is positive definite, from Cholesky decomposition,

$$\mathbf{R}_{n-1} = \mathbf{L}\mathbf{L}^T \tag{C.11}$$

Hence,

$$\mathbf{R}_{n-1}^{-1} = \left(\mathbf{L}^{-1}\right)^T \mathbf{L}^{-1} \tag{C.12}$$

Therefore,

$$\mathbf{b^T\,R_{n-1}^{-1}\,b} \;=\; \mathbf{b^T}\left(\mathbf{L}^{-1}\right)^{\mathbf{T}}\mathbf{L}^{-1}\,\mathbf{b} \tag{C.13}$$

Substituting $\mathbf{z} = \mathbf{L}^{-1}\,\mathbf{b}$ and $\mathbf{z^T} = \mathbf{b^T}\left(\mathbf{L}^{-1}\right)^{\mathbf{T}}$ in equation (C.13)

$$\mathbf{b^T\,R_{n-1}^{-1}\,b} \;=\; \mathbf{z^T\,z} > 0 \tag{C.14}$$

Hence, $\mathbf{R_{n-1}^{-1}}$ is also positive definite. From equation (C.10)

$$1 - \mathbf{b^T\,R_{n-1}^{-1}\,b} > 0 \tag{C.15}$$

Therefore, the necessary condition on $\mathbf{b}$ when $\mathbf{R_{n-1}}$ is positive definite is

$$\mathbf{b^T\,R_{n-1}^{-1}\,b} < 1. \tag{C.16}$$

C.2 The Bounds

$\mathbf{R_n}$ is positive definite if $\mathbf{R_{n-1}}$ is positive definite and,

$$\mathbf{b^T\;R_{n-1}^{-1}\;b} < 1 \tag{C.17}$$

Partitioning $\mathbf{R_{n-1}^{-1}}$ and $\mathbf{b}$ matrices in equation (C.17),

$$\begin{bmatrix} \mathbf{B_1^T} & \vdots & \Gamma & \vdots & \mathbf{B_2^T} \end{bmatrix} \begin{bmatrix} \mathbf{S_1} \\ \cdots \\ \mathbf{S_j} \\ \cdots \\ \mathbf{S_2} \end{bmatrix} \begin{bmatrix} \mathbf{B_1} \\ \cdots \\ \Gamma \\ \cdots \\ \mathbf{B_2} \end{bmatrix} < 1 \tag{C.18}$$

where Γ is the correlation coefficient (ρ_{jn}) for which bounds are required,

$\mathbf{S_1}$ is a $(j-1)$x$(n-1)$ matrix and $\mathbf{S_2}$ is a $(n-1-j)$x$(n-1)$ matrix,

$\mathbf{B_1^T}$ and $\mathbf{B_2^T}$ are 1x$(j-1)$ and 1x$(n-1-j)$ row matrices, and

$\mathbf{S_j}$ is a 1x$(n-1)$ row matrix.

Multiplying $\mathbf{b^T}$ by $\mathbf{R_{n-1}^{-1}}$ in equation (C.18),

$$\begin{bmatrix} \mathbf{C_1} & \vdots & \Gamma\, \mathbf{S_j} & \vdots & \mathbf{C_2} \end{bmatrix} \begin{bmatrix} \mathbf{B_1} \\ \cdots \\ \Gamma \\ \cdots \\ \mathbf{B_2} \end{bmatrix} < 1 \tag{C.19}$$

where $\mathbf{C_1} = \mathbf{B_1^T}\, \mathbf{S_1}$ and $\mathbf{C_2} = \mathbf{B_2^T}\, \mathbf{S_2}$.

Since, $\mathbf{S_j}$, $\mathbf{C_1}$ and $\mathbf{C_2}$ are $1x(n-1)$ row matrices the quadratic equation (4.13) for real bounds from equation (C.3) is,

$$S_{jj}\,\Gamma^2 + [C_{1j} + C_{2j} + \sum_{i=1}^{j-1} S_{ji}\,B_{1i} + \sum_{i=j+1}^{n-1} S_{ji}\,B_{2i}]\,\Gamma$$

$$+ \sum_{i=1}^{j-1} (C_{1i} + C_{2i})\,B_{1i} + \sum_{i=j+1}^{n-1} (C_{1i} + C_{2i})\,B_{2i} - 1 < 0$$

Appendix D

The Computer Programs

D.1 General

One of the primary objectives of this research was to computerize the analytical method for economic risk analysis. The computer programs could then be used to explore its behavior, to verify it and to test its practicality in the measurement of uncertainty of performance and decision variables. This appendix describes the two computer programs, "ELICIT" and "TIERA", developed to achieve this objective. The two programs written in FORTRAN 77 can be executed together or separately.

D.2 ELICIT - Program to Obtain Input Data

ELICIT is an interactive program, to ensure that the subjective probabilities elicited for primary variables at the input level are coherent (section 3.7), their correlation matrices are positive definite (section 4.2.2), and to elicit the common (shared) variables in the functional forms. The flowchart for ELICIT is depicted in Figure (D.1).

The objective of ELICIT is to obtain interactively all of the information necessary to set up the input files required to execute TIERA. ELICIT is developed in three sections - work package durations, work package costs and revenue streams. The program begins with work package durations and proceeds to the next module only if it is asked to. The output from the three sections are written to data files in Units

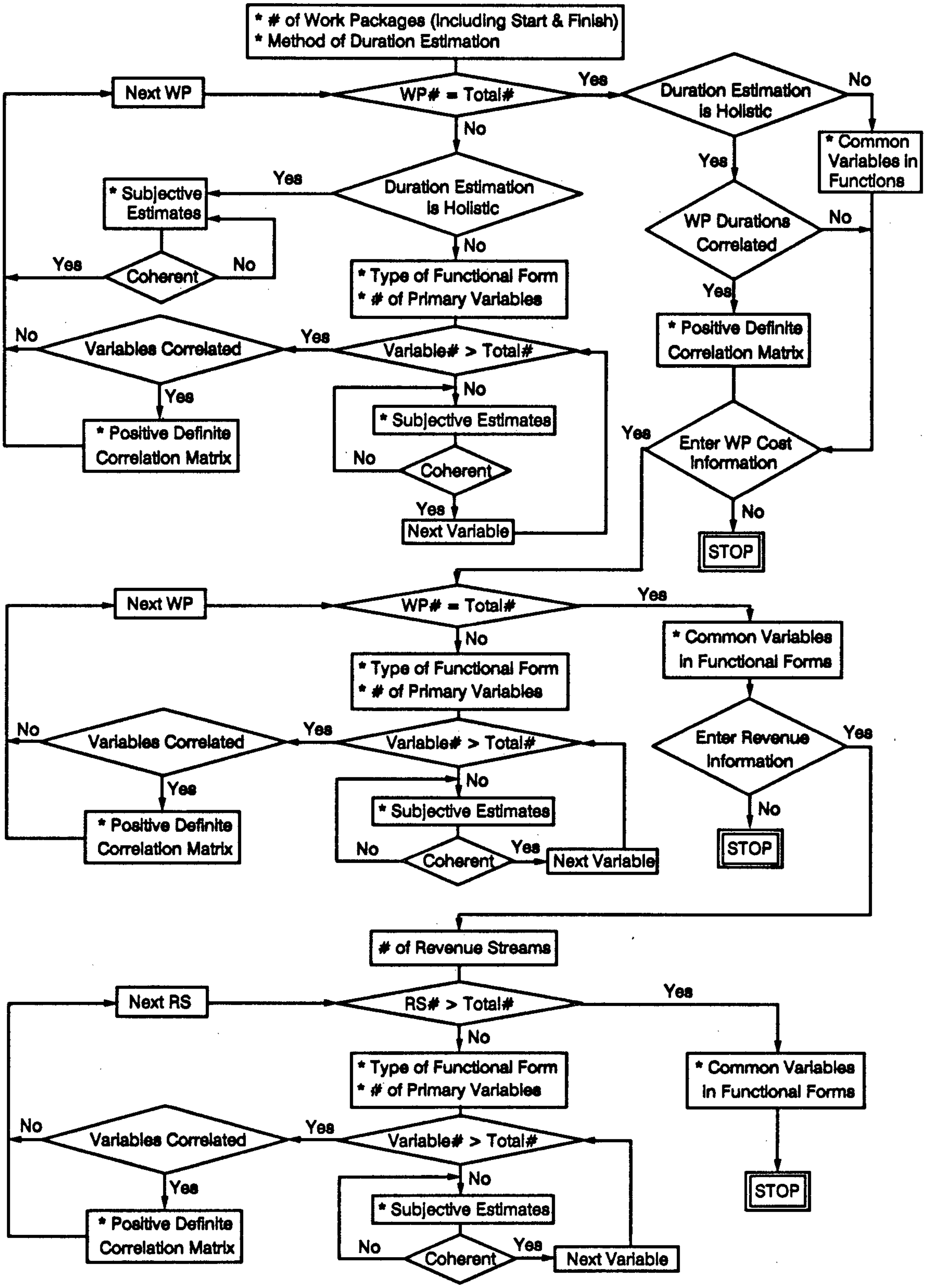

Figure D.1: Flowchart for ELICIT

11, 12 and 13 which are the input files for TIERA.

To ensure that the subjective percentile estimates for a primary variable are coherent ELICIT approximates them to a Pearson type distribution. The flowchart of this process is depicted in Figure (D.2). When the subjective estimates are coherent, ELICIT displays the expected value, standard deviation, skewness and kurtosis of the approximated Pearson distribution as verification and proceeds to the next variable. When the estimates are not coherent the analyst/expert are given an opportunity to re-estimate percentiles as suggested in section 3.7.

To ensure that a correlation matrix is positive definite ELICIT follows the theoretical development described in section 4.2.2. When the vector of correlation coefficient values - $\mathbf{b}$ is elicited, the program checks for the condition given by equation (4.14). If the condition is satisfied the program accepts the $\mathbf{b}$ vector as valid correlation coefficients between the primary variables. When the condition is not satisfied the program informs the user that the theoretical requirement for a valid $\mathbf{R_n}$ has been violated, and requests the user to identify a previous variable in the ordered list whose correlation coefficient with the current variable that should be changed. Once the user has identified a variable, ELICIT calculates the bounds for that correlation coefficient (if they exist) from equation (4.15), thereby giving guidance for the user to conform to the theoretical requirement.

Thirdly, the common (shared) variables between the functions for work package durations, the functions for work package costs and the functions for revenue streams are elicited. Utilizing this information, TIERA develops positive definite correlation matrices for work package durations, work package costs and revenue streams. These correlation matrices are then used in the evaluation of moments for path durations (hence project duration), project cost and project revenue respectively.

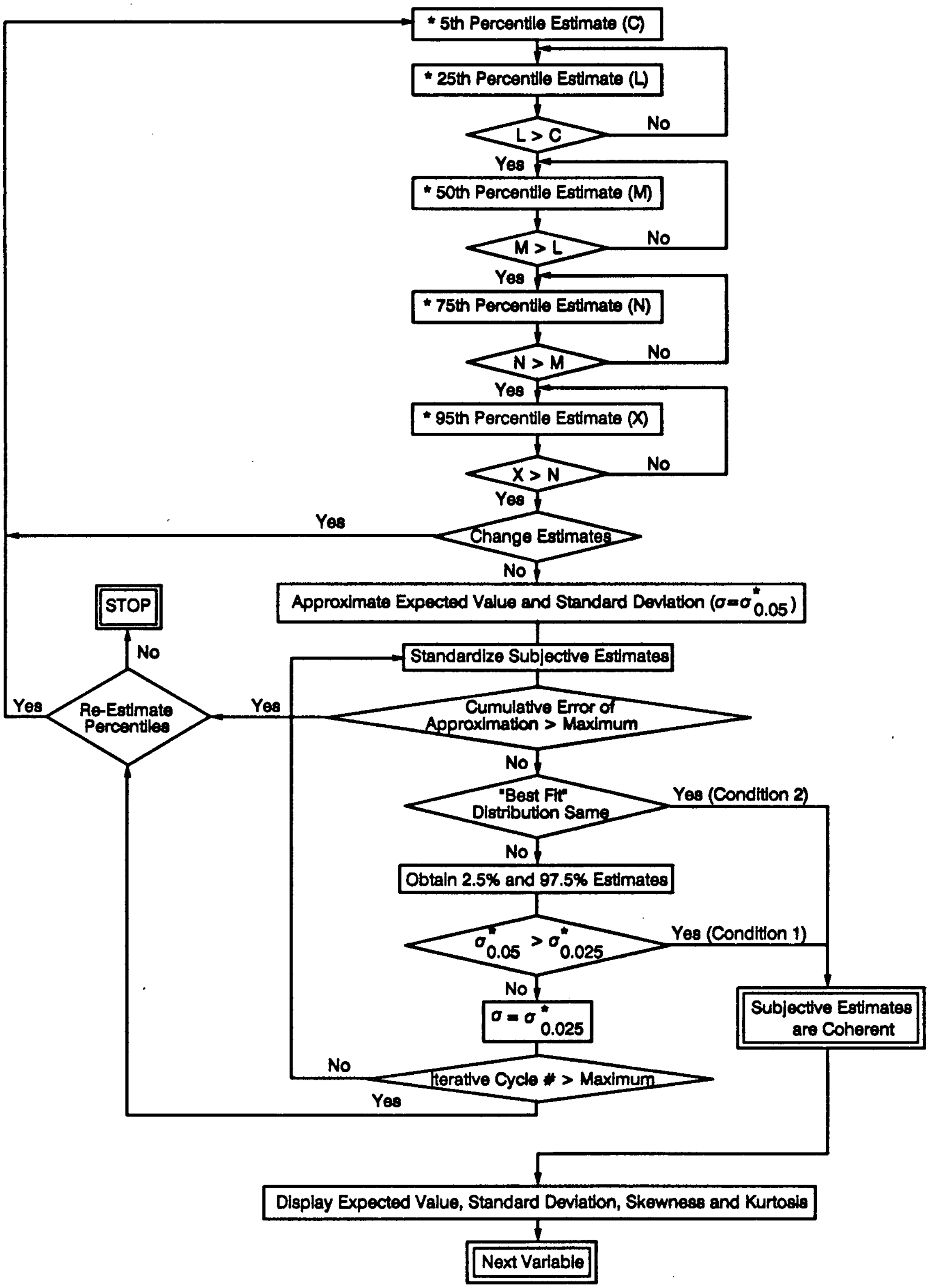

Figure D.2: Flowchart to Ensure Coherence of Subjective Estimates

D.3 TIERA - Program for Risk Quantification

TIERA is the computer program of the analytical method developed in this thesis for time and economic risk quantification. It is developed in two modules. The main module follows Figure (6.3) and consists of all the analytical derivations described in chapter 6. Except for the reverse arrow in Figure (6.3), where the first four moments for work package start times are evaluated from the modified PNET algorithm, all of the other arrows use the moments of the primary variables at the lower level to evaluate the first four moments of the derived variables at the higher level.

The flowchart for the modified PNET algorithm is depicted in Figure (D.3). When the transitional correlation, $\rho = 0$, the modified PNET algorithm defaults to the longest path approach because there is only one representative path. Then the process will always stop at the third decision node. Figure (D.4) depicts the flowchart for the process to trace all the paths to a work package (or milestone) from the start work package of the precedence network. The algorithm to trace all the paths to a work package was based on the "stack" concept.

The second module for TIERA is an external subroutine consisting of functions for work package durations, work package costs and revenue streams that are specified by the analyst. At present, the analyst can specify five functions for work package durations, and ten each for work package costs and revenue streams. If more functions are needed for an analysis, the number can be increased with a small modification to the main program. The main program of TIERA should always be executed in combination with a compiled version of the external subroutine consisting of the functions.

To execute TIERA the main program looks for data from Units 1, 10, 11, 12 and 13. The data file containing the table of Pearson distributions should always be specified at Unit 1. At present the data file contains 2665 distributions. When the table

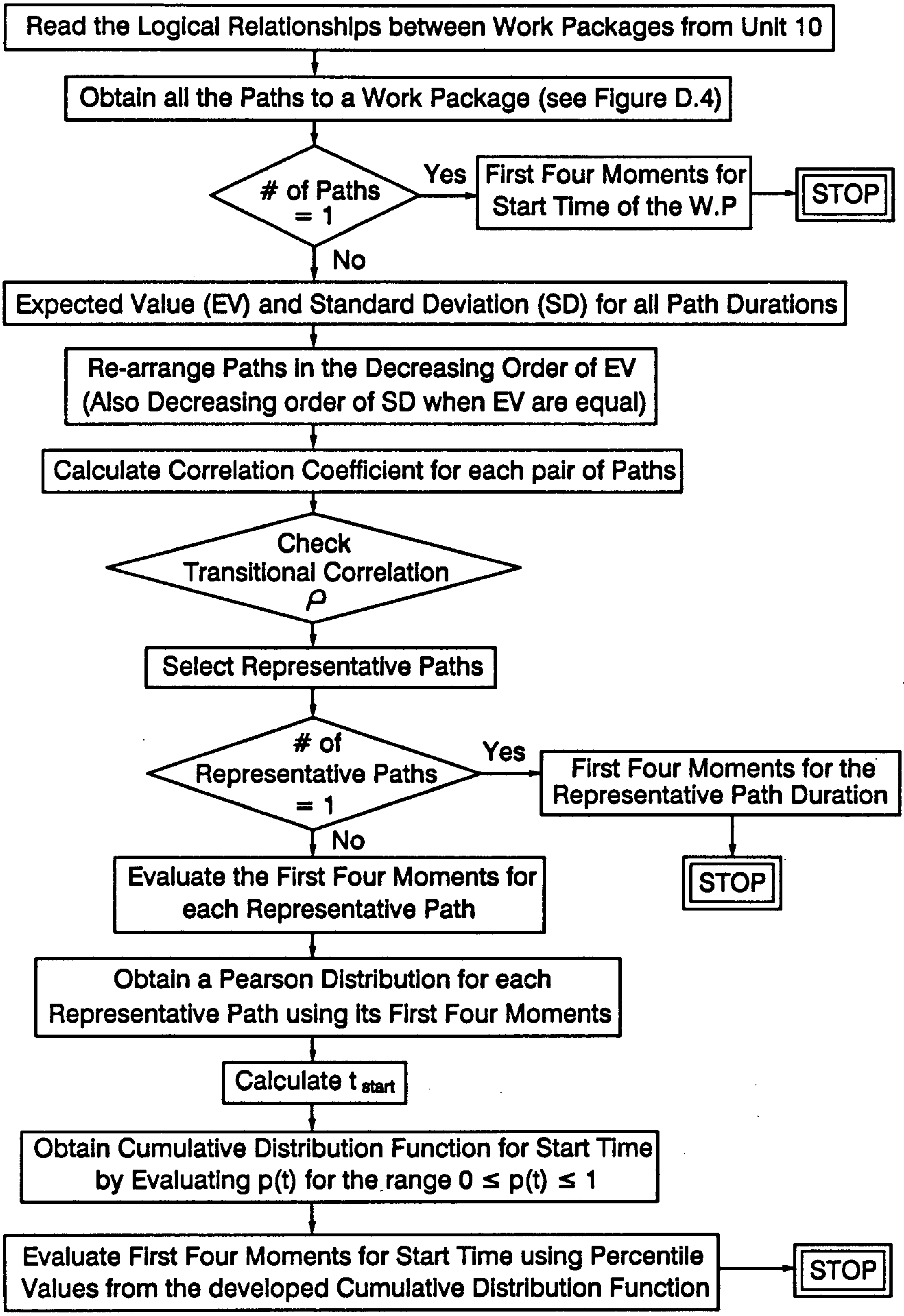

Figure D.3: Flowchart of the Modified PNET Algorithm

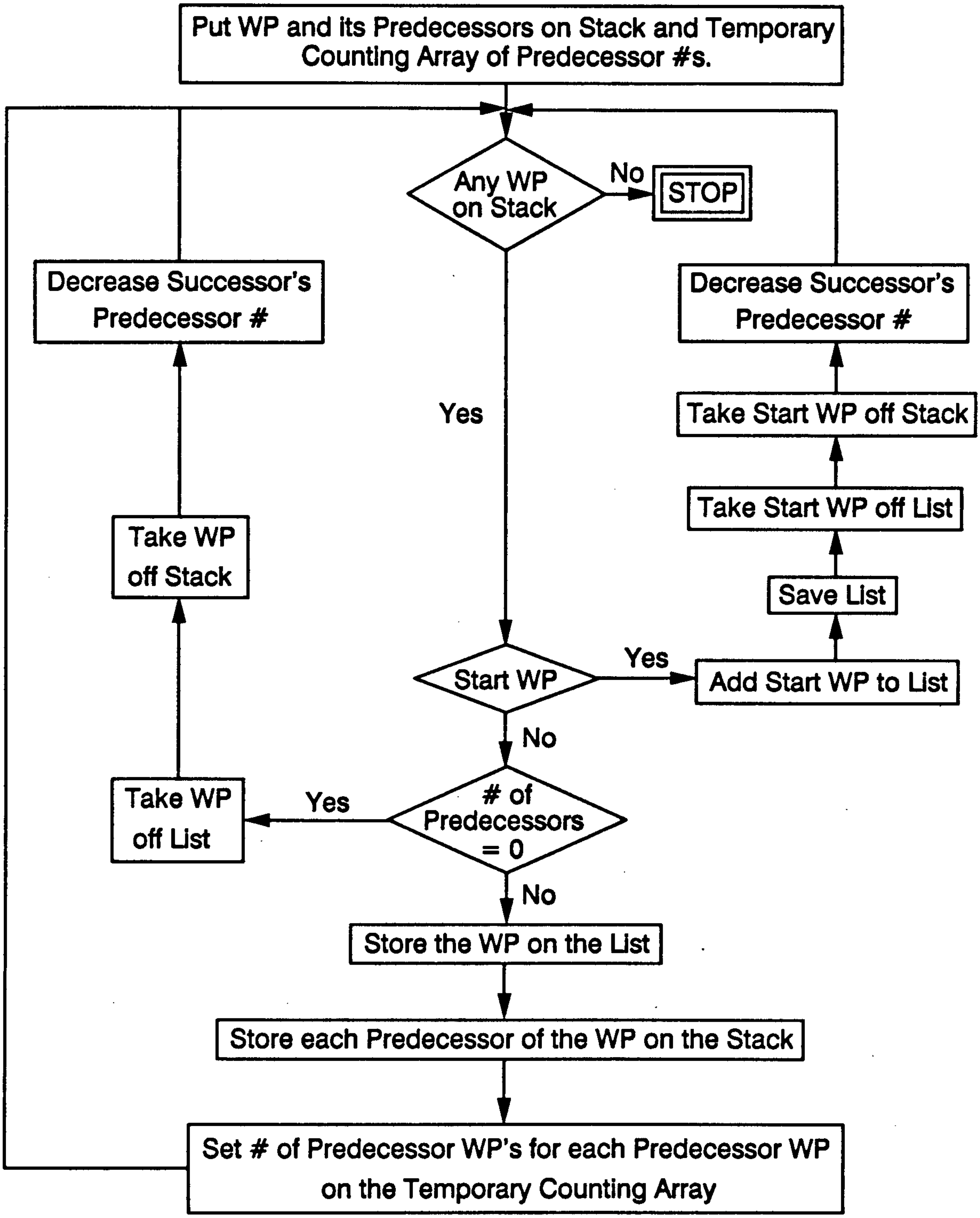

Figure D.4: Flowchart to Trace all the Paths to a Work Package developed by Amos and Daniel (1971) is included the file will contain approximately 12,100 distributions. The data file containing the logical relationships between work packages should be specified at Unit 10. At present only *finish to start* = 0 relationships between work packages are permitted. The data files for work package durations, work package costs and revenue streams should be specified at Units 11, 12 and 13 respectively. These data files are the output files from ELICIT.

The output from TIERA is written to Unit 7. A typical output from TIERA is illustrated in Figure (D.5). The output is for the third case of the second example given in chapter 7 (*i.e* correlations between primary variables and between derived variables are treated). Units 5 and 6, the reading and writing units for FORTRAN are left free to permit the joint execution of ELICIT and TIERA. When the two programs are executed together the reading and writing for user - computer interaction are from Units 5 and 6 respectively.

***** WORK PACKAGE DURATIONS : THE TIME UNIT IS MONTHS *****

W.P #	*** EXPECTED VALUE ***	*** STANDARD DEVIATION ***	*** SKEWNESS ***	*** KURTOSIS ***
1	0.00	0.00	0.000	0.000
2	7.68	2.63	0.487	2.339
3	4.95	1.27	1.954	8.580
4	6.92	2.37	0.484	2.282
5	3.36	1.05	0.846	2.859
6	3.36	0.90	3.189	14.203
7	1.69	0.56	1.900	6.330
8	6.44	1.70	4.348	24.684
9	5.80	2.00	0.514	2.364
10	2.69	0.72	3.106	13.575
11	4.55	1.11	0.478	2.275
12	6.96	2.40	0.511	2.314
13	6.79	2.44	1.469	4.589
14	6.35	2.44	0.900	2.971
15	180.00	0.00	0.000	0.000

***** WORK PACKAGE EARLY START TIMES FOR A TRANSITIONAL CORRELATION OF 0.50 : THE TIME UNIT IS MONTHS *****

W.P #	*** EXPECTED CALENDAR MONTH ***		*** STANDARD DEVIATION ***	*** SKEWNESS ***	*** KURTOSIS ***
1	JUN / 1988	0.00	0.00	0.00	0.00
2	JUN / 1988	0.00	0.00	0.00	0.00
3	FEB / 1989	7.68	2.63	0.49	2.34
4	FEB / 1989	7.68	2.63	0.49	2.34
5	JUL / 1989	12.63	2.91	0.52	2.59
6	JUL / 1989	12.63	2.91	0.52	2.59
7	AUG / 1989	14.60	4.02	0.36	2.70
8	AUG / 1989	14.60	4.02	0.36	2.70

Figure D.5: Typical Output from TIERA

Figure (D.5) contd.

9	OCT / 1989	15.99	3.10	0.48	2.67
10	OCT / 1989	16.00	3.24	0.71	2.86
11	OCT / 1989	16.00	3.24	0.71	2.86
12	OCT / 1989	16.00	3.24	0.71	2.86
13	MAY / 1990	23.00	4.24	0.40	2.90
14	MAY / 1990	23.39	4.27	0.50	3.10
15	FEB / 1991	32.57	4.75	0.40	2.90

****** THE PROJECT DURATION FOR A TRANSITIONAL CORRELATION OF 0.50 : THE TIME UNIT IS MONTHS ******

** CALENDAR MONTH **	** EXPECTED VALUE **	** STANDARD DEVIATION **	*** SKEWNESS ***	*** KURTOSIS ***
FEB / 1991	32.57	4.75	0.40	2.90

*** WORK PACKAGE COSTS DISCOUNTED AT A RATE OF RETURN OF 0.090 ***

W.P #	*** EXPECTED VALUE ***	*** STANDARD DEVIATION ***	*** SKEWNESS ***	*** KURTOSIS ***
1	0.	0.	0.000	0.000
2	3697632.	1523500.	0.398	2.190
3	5578746.	2056956.	0.831	2.830
4	7125176.	2954487.	0.401	2.193
5	1524454.	625166.	0.511	2.314
6	810978.	315346.	1.324	4.105
7	1882968.	845204.	1.007	3.218
8	3170606.	1307522.	1.591	5.038
9	3505625.	1459957.	0.419	2.211
10	1497887.	608433.	1.261	3.909
11	2042131.	739623.	0.343	2.141
12	7037942.	2931777.	0.420	2.211
13	1960845.	865159.	0.900	2.971

Figure (D.5) contd.

14	6824271.	3225952.	0.592	2.421
15	0.	0.	0.000	0.000

**** THE PROJECT COST DISCOUNTED AT A RATE OF RETURN OF 0.090 ****

*** EXPECTED VALUE ***	*** STANDARD DEVIATION ***	*** SKEWNESS ***	*** KURTOSIS ***
46659262.	11195712.	0.265	2.789

*** NET REVENUE STREAMS DISCOUNTED AT A RATE OF RETURN OF 0.090 ***

W.P #	*** EXPECTED VALUE ***	*** STANDARD DEVIATION ***	*** SKEWNESS ***	*** KURTOSIS ***
1	32676410.	11681997.	-0.677	6.146
2	18432211.	4951469.	-0.230	2.260
3	18389084.	4885079.	0.292	2.604

**** THE PROJECT REVENUE DISCOUNTED AT A RATE OF RETURN OF 0.090 ****

*** EXPECTED VALUE ***	*** STANDARD DEVIATION ***	*** SKEWNESS ***	*** KURTOSIS ***
69497704.	13595959.	-0.427	4.695

**** THE PROJECT NET PRESENT VALUE AT A DISCOUNT RATE OF 0.090 ****

*** EXPECTED VALUE ***	*** STANDARD DEVIATION ***	*** SKEWNESS ***	*** KURTOSIS ***
22838442.	17612327.	-0.265	3.567

Figure (D.5) contd.

```
**** THE INTERNAL RATE OF RETURN FOR THE PROJECT (%) ****

**** THE EXPECTED VALUE ****          **** STANDARD DEVIATION ****
            16.923                               6.616

***** 2.5% VALUE *****      ***** 5.0% VALUE *****      ***** 10.0% VALUE *****
                  4.40                        6.69                         9.07

***** 25.0% VALUE *****     ***** 50.0% VALUE *****     ***** 75.0% VALUE *****
                  12.55                       16.56                        20.91

***** 90.0% VALUE *****     ***** 95.0% VALUE *****     ***** 97.5% VALUE *****
                  25.46                       28.40                        30.99
```

Appendix E

The Correction Factor α

The proof given in this appendix is an extension to that from Van Tetterode (1971). Let Q and P be two independent variables from $U(0,1)$. Then,

$$E[Q] = E[P] = \frac{1}{2} \; ; \; \sigma_Q^2 = \sigma_P^2 = \frac{1}{12} \; ; \; cov(Q,P) = 0$$

Let R be a new random variable formed as,

$$R = P + \alpha\,(Q - P) \tag{E.1}$$

Then from equation (E.1),

$$\sigma_R^2 = (1-\alpha)^2\,\sigma_P^2 + \alpha^2\,\sigma_Q^2 \tag{E.2}$$

From the definition of covariance of R and Q,

$$cov(R,Q) = cov(P + \alpha(Q-P),\; Q) \tag{E.3}$$

$$= cov(P,Q) + cov(\alpha(Q-P),Q) \tag{E.4}$$

$Since\ cov(P,Q) = 0$

$$= \alpha\; cov(Q,Q) - \alpha\; cov(P,Q) \tag{E.5}$$

$$cov(R,Q) = \alpha\,\sigma_Q^2 \tag{E.6}$$

Let the correlation coefficient between R and Q be ρ. From the definition,

$$\rho = \frac{cov(R,Q)}{\sigma_R\,\sigma_Q} \tag{E.7}$$

Substituting equation (E.2) and (E.6) in (E.7),

$$\rho = \frac{\alpha\, \sigma_Q^2}{(\sqrt{(1-\alpha)^2\, \sigma_P^2 + \alpha^2\, \sigma_Q^2}\ )\, \sigma_Q} \tag{E.8}$$

Since $\sigma_Q^2 = \sigma_P^2 = \dfrac{1}{12}$

$$\rho = \frac{\alpha}{\sqrt{1 - 2\,\alpha + 2\,\alpha^2}} \tag{E.9}$$

Therefore, from equation (E.9),

$$(2\rho^2 - 1)\,\alpha^2 - 2\rho^2\,\alpha + \rho^2 = 0 \tag{E.10}$$

The correction factor α as the solution of equation (E.10) is given by,

$$\alpha = \frac{\rho^2 \pm \rho\,\sqrt{1 - \rho^2}}{2\rho^2 - 1} \tag{E.11}$$

Therefore, the random number correction is as follows (Van Tetterode, 1971).

$$RN_{ij} = RN_j + \alpha_{ij}\,(RN_i - RN_j) \tag{E.12}$$

where RN_i and RN_j are the random numbers generated for variables X_i and X_j respectively, and RN_{ij} is the random number corrected for the linear correlation ρ_{ij} between X_i and X_j relative to X_i. When

$$\rho_{ij} = 0; \quad \alpha_{ij} = 0; \quad RN_{ij} = RN_j$$

$$\rho_{ij} = +1; \quad \alpha_{ij} = 1; \quad RN_{ij} = RN_i$$

$$\rho_{ij} = -1; \quad \alpha_{ij} = 1; \quad RN_{ij} = RN_i$$

When $\rho_{ij} = 0$ the corrected random number is given by the j^{th} random number demonstrating independence. When the correlation is either perfect positive or perfect negative the corrected random number is same as the i^{th}, demonstrating perfect correlation. Therefore, for all values of ρ_{ij},

$$0 \leq \alpha_{ij} \leq 1.$$

Appendix F

Input Data for Numerical Examples

This appendix contains the input data used for the numerical examples in chapter seven for the validation studies and applications of the analytical method for time and economic risk quantification.

Road Pavement Project

Table F.1 is Table 1 from Ang et al. (1975). It describes the various activities of the project involving the paving of 2.2 miles of roadway pavement and the construction of appurtenant drainage structures, excavation to grade, placement of macadam shoulders, erection of guardrails, and landscaping. The respective mean durations and corresponding standard deviations are also listed.

Industrial Building Project

Table F.2 is Table 3 from Ang et al. (1975). It describes the various activities of the project involving the construction of a single-story industrial building. The building is comprised of reinforced concrete piers, frost walls, structural steel columns, and a precast roof deck. The respective mean durations and corresponding standard deviations are also listed.

Table F.1: Activities and Estimated Durations (Pavement Project)

#	Description of Activities	$E[D]$ days	σ_D days
01	Dummy	0	0
02	Set-up batch plant	2	0.5
03	Order and deliver paving mesh	5	1.0
04	Deliver rebars for double barrel culvert	6	1.5
05	Move in equipment	3	0.5
06	Deliver rebars for small box culvert	7	4.0
07	Build double barrel culvert	10	2.0
08	Clear and grub from station 42 - station 100	3	1.0
09	Clear and grub from station 100 - station 158	7	1.5
10	Build box culvert at station 127	5	2.0
11	Build box culvert at station 138	3	1.5
12	Cure double barrel culvert	9	2.0
13	Move dirt between station 42 - station 100	5	1.5
14	Start moving dirt between station 100 - station 158	3	0.5
15	Cure box culvert at station 127	9	4.5
16	Cure box culvert at station 138	6	2.0
17	Order and stockpile paving material	2	0.5
18	Place subbase from station 42 - station 100	7	1.73
19	Finish moving dirt between station 100 - station 158	5	2.0
20	Pave from station 42 - station 100	10	2.0
21	Place subbase from station 100 - station 158	7	3.31
22	Cure pavement from station 42 - station 100	6	1.5
23	Pave from station 100 - station 158	10	4.5
24	Cure pavement from station 100 - station 158	6	1.5
25	Place shoulders from station 42 - station 100	3	1.0
26	Place shoulders from station 100 - station 158	3	1.0
27	Place guardrail and landscape	5	1.5
28	Dummy	0	0

Table F.2: Activities and Estimated Durations (Industrial Building Project)

#	Description of Activities	$E[D]$ days	σ_D days
01	Mobilization	32	3.2
02	Move in	2	0.5
03	Initial layout	2	0.5
04	Dummy	0	0
05	Site rough grading	2	0.5
06	Layout of piers	1	0.5
07	Excavate piers	2	1.0
08	Dummy	0	0
09	Order and deliver rebars	40	12.0
10	Form and rebars piers	2	0.5
11	Pour piers	2	0.5
12	Cure piers	4	0.8
13	Strip piers	1	0.1
14	Dummy	0	0
15	Dummy	0	0
16	Excavate frost walls	1	0.5
17	Order and deliver structurl steel columns	60	12.0
18	Erect structural steel columns	5	1.0
19	Order and deliver precast roof deck	30	6.0
20	Form and mesh frost walls	3	0.9
21	Pour frost walls	1	0.3
22	Cure frost walls	4	0.4
23	Strip frost walls	1	0.1
24	Backfill	2	0.5
25	Grade and compact gravel for floor	2	0.2
26	Rebar floor and set screeds	2	0.5
27	Pour and finish floor	2	0.5
28	Dummy	0	0
29	Excavate and grade parking	2	0.2
30	Stone base for parking	1	0.2
31	Dummy	0	0
32	Set roof deck	5	1.5
33	Hang siding and waterproof roof	6	1.2
34	Hang doors	4	1.2
35	Clean up	2	0.5
36	Bituminous surface in parking	3	0.3

First Example

The first example is an actual deterministic feasibility analysis conducted for a mineral project in South America.

Deterministic Estimates

Table F.3 contains the description and deterministic estimates for duration cost of work packages. The work package durations were developed to correspond to the modified construction schedule. The work package costs were estimated such that the sum of the work package costs is equivalent to the constant dollar cost estimate of the deterministic feasibility analysis.

The Statistics

The deterministic values are assumed as the median values of probability distributions for work package durations and costs. Table F.4 contains the expected value, standard deviation, skewness and kurtosis for work package durations and costs used in this example.

Revised Durations

Table F.5 contains the statistics for the revised work package durations. The coefficients of variation for work package durations are approximately 40% instead of the 3% to 13% used in the previous case.

Table F.3: Deterministic Values for Work Package Durations and Costs

WP#	Work Package Description	Dura	Cost
		mths	$
01	Start Work Package	–	–
02	Engineering & Mobilization	4	2800000
03	Construction of a temporary fuel tank	3	200000
04	Road & Rail for equipment transfer	3	2520900
05	Camp expansion	3	2620000
06	Roads for construction requirements	8	2400000
07	Water supply scheme	11	2501100
08	Mine auxiliary building	11	4233800
09	Town-site - Phase 1	8	3552200
10	Power house construction	5	865800
11	Rainy season : Downtime	3	–
12	Office, changehouse & lab for plant	10	2497200
13	Road/rail/port transfer facilities	10	4198000
14	Construction of process plant	8	4996300
15	Tailings Dam	8	3980000
16	Town-site - Phase 2	8	4000000
17	Power plant - supply & distribution	13	6958300
18	Roads for operational requirements - Phase 1	9	3500000
19	Construction of permanent fuel system	9	743900
20	Tailings Pipeline - Phase 1	4	550000
21	Plant shop & warehouse	13	1513600
22	Pre-production	21	33047700
23	Rainy season - Downtime	3	–
24	Tailings thickner - Phase 1	5	440000
25	Town-site - Phase 3	6	2000000
26	Tailings pipeline - Phase 2	5	682500
27	Tailings thickner - Phase 2	4	346000
28	Equipment & installation of process plant	6	11853700
29	Roads for operational requirements - Phase 2	6	1475000
30	Reclaim water system	9	1356100
31	Start up	3	600000
32	Project mgmt., org. expenses, import tax	33	18018000
33	Finish Work Package (Revenue Period)	180	–
	Total Base Estimate	36	124450100

Table F.4: Statistics for Work Package Durations and Costs

WP#	Duration (months)				Constant Dollar Cost ($)			
	$E[WPD]$	σ_{WPD}	$\sqrt{\beta_1}$	β_2	$E[C_0]$	σ_{C_0}	$\sqrt{\beta_1}$	β_2
01	–	–	–	–	–	–	–	–
02	3.98	0.51	0.4	9.0	2836999	913186	0.1	2.1
03	3.02	0.27	0.7	9.0	203700	67051	0.2	2.2
04	2.98	0.33	-0.3	3.5	2550166	912688	0.1	2.1
05	2.98	0.33	-0.3	3.5	2649599	912707	0.1	2.1
06	7.99	0.92	0.2	2.1	2418499	881804	0.1	2.1
07	11.04	0.84	0.1	2.1	2537692	913156	0.1	2.1
08	11.02	0.45	0.5	5.5	4258293	1337784	0.1	2.0
09	7.99	0.92	0.2	2.1	3579135	1140382	0.1	2.0
10	5.01	0.44	0.1	2.6	860250	273656	0.0	2.0
11	3.02	0.27	0.7	9.0	–	–	–	–
12	10.03	0.58	0.2	2.3	2535235	913261	0.1	2.1
13	10.05	0.55	0.3	2.4	4198739	1276596	0.0	2.0
14	7.93	0.27	0.4	3.7	5034668	1581374	0.1	2.0
15	7.94	0.33	0.3	3.5	4042899	1370345	0.1	2.0
16	7.93	0.27	0.4	3.7	4073999	1341012	0.1	2.0
17	13.01	1.06	0.0	2.0	7029228	2220857	0.1	1.9
18	9.02	0.33	0.3	3.5	3536999	1095334	0.1	2.0
19	9.02	0.33	0.3	3.5	746157	237101	0.0	1.9
20	3.98	0.51	0.4	9.0	555550	191632	0.1	2.0
21	13.01	0.55	0.0	2.3	1517817	471158	0.1	2.1
22	21.02	1.06	0.1	2.1	33400048	10952316	0.1	2.0
23	3.02	0.27	0.7	9.0	–	–	–	–
24	4.99	0.40	0.0	2.8	445550	149120	0.1	2.0
25	6.02	0.40	0.2	2.8	2018499	638774	0.2	2.2
26	5.01	0.30	0.4	5.6	688975	219015	0.1	1.9
27	3.96	0.51	0.0	2.4	349330	118624	0.1	2.0
28	6.02	0.55	0.1	2.2	12074330	3992590	0.2	2.0
29	6.02	0.48	0.2	2.6	1502749	518038	0.2	2.1
30	9.02	0.33	0.3	3.5	1390842	458266	0.3	2.3
31	3.02	0.33	0.3	3.5	607400	194779	0.1	1.9
32	33.45	0.91	0.7	6.4	18751328	6156604	0.5	2.6
33	180.00	–	–	–	–	–	–	–

Table F.5: Statistics for Revised Work Package Durations

	Duration (months)			
WP#	$E[WPD]$	σ_{WPD}	$\sqrt{\beta_1}$	β_2
02	3.99	1.59	0.3	6.0
03	3.04	1.21	0.4	5.4
04	3.04	1.21	0.4	5.4
05	3.04	1.21	0.4	5.4
06	7.99	3.19	0.1	2.9
07	11.15	4.41	0.2	2.9
08	11.15	4.41	0.2	2.9
09	7.99	3.19	0.1	2.9
10	5.01	2.00	0.1	3.1
11	3.04	1.21	0.4	5.4
12	10.09	4.01	0.2	3.8
13	10.09	4.01	0.2	3.8
14	7.99	3.19	0.1	2.9
15	7.99	3.19	0.1	2.9
16	7.99	3.19	0.1	2.9
17	13.06	5.22	0.2	5.9
18	9.12	3.59	0.3	3.9
19	9.12	3.59	0.3	3.9
20	3.99	1.59	0.3	6.0
21	13.06	5.22	0.2	5.9
22	21.08	8.41	0.1	3.1
23	3.04	1.21	0.4	5.4
24	5.01	2.00	0.1	3.1
25	6.05	2.40	0.1	2.9
26	5.01	2.00	0.1	3.1
27	3.99	1.59	0.3	6.0
28	6.05	2.40	0.1	2.9
29	6.05	2.40	0.1	2.9
30	9.12	3.59	0.3	3.9
31	3.04	1.21	0.4	5.4
32	33.91	5.65	0.8	8.0

Revenue Streams

The expected value, standard deviation, skewness and kurtosis for annual revenue and operating cost for the revenue streams are given in Table F.6.

Table F.6: Statistics for Annual Revenue and Operating Costs

RS#	Annual Revenue ($)				Annual Operating Cost ($)			
	$E[R]$	σ_R	$\sqrt{\beta_1}$	β_2	$E[O\&M]$	$\sigma_{O\&M}$	$\sqrt{\beta_1}$	β_2
01	57771120	7925719	-0.2	2.5	18714624	2917630	0.5	2.9
02	66244656	5804341	-0.2	2.2	18951120	3503251	0.1	2.1
03	68975456	8889606	-0.3	2.1	21454432	4331357	0.7	2.5
04	77449584	6990992	0.0	2.1	21453872	4331492	0.7	2.5
05	61242768	7014973	-0.2	2.1	21638864	4058441	0.9	2.8
06	60687760	8011651	-0.3	2.1	19176192	3222784	0.4	2.5
07	61242768	4596042	-0.4	2.7	13533594	1982918	0.2	2.5
08	32325072	3656270	-0.2	2.6	10425628	2472484	0.7	3.5

Second Example

The second example is a hypothetical engineering project of thirteen work packages and three revenue streams.

Work Package Duration

The statistics for primary variables for work package duration model are given in Tables F.7, F.8 and F.9. The positive definite correlation matrix for primary variables that was used for all the work package durations is given by $\mathbf{R_{WPD}}$.

$$\mathbf{R_{WPD}} = \begin{bmatrix} 1.00 & -0.30 & 0.40 \\ -0.30 & 1.00 & -0.35 \\ 0.40 & -0.35 & 1.00 \end{bmatrix}$$

Table F.7: Statistics for Quantity Descriptor Q_i (ft^3)

WP#	$E[Q_i]$	σ_{Q_i}	$\sqrt{\beta_1}$	β_2	Common
02	38397.3	12186.1	0.5	3.3	*
03	60555.0	8829.3	0.9	9.0	
04	76850.0	24440.5	0.5	3.2	**
05	16185.0	3527.4	0.8	7.8	
06	8092.5	1373.3	0.4	3.6	
07	20370.0	5802.8	0.8	9.0	
08	32429.2	7030.8	0.8	7.8	
09	38397.3	12186.1	0.5	3.3	*
10	16160.8	2820.8	0.3	2.6	
11	21998.0	2621.4	0.2	2.4	
12	76850.0	24440.5	0.5	3.2	**
13	20413.0	5782.4	0.7	8.5	
14	76850.0	24440.5	0.5	3.2	*

Table F.8: Statistics for Labour Productivity Rate P_{L_i} ($ft^3/m.d$)

WP#	$E[P_{Li}]$	$\sigma_{P_{Li}}$	$\sqrt{\beta_1}$	β_2	Common
02	9.0	1.25	0.0	5.6	*
03	10.2	2.23	0.8	8.0	**
04	9.0	1.25	0.0	5.6	*
05	10.1	2.28	0.1	2.2	***
06	10.2	2.23	0.8	8.0	**
07	10.2	2.23	0.8	8.0	
08	8.4	1.28	0.1	8.8	
09	9.0	1.25	0.0	5.6	*
10	10.2	2.23	0.8	8.0	**
11	10.1	2.28	0.1	2.2	***
12	9.0	1.25	0.0	5.6	*
13	9.9	2.22	0.9	9.0	
14	10.2	2.23	0.8	8.0	**

Table F.9: Statistics for Labour Usage L_i ($m.d/year$)

WP#	$E[L_i]$	σ_{L_i}	$\sqrt{\beta_1}$	β_2	Common
02	6833.2	692.7	0.4	2.4	
03	15185.0	1539.5	0.4	2.3	*
04	15185.0	1539.5	0.4	2.3	*
05	6074.0	615.8	0.4	2.4	**
06	3074.0	761.0	1.0	7.2	
07	15370.0	3805.3	1.0	7.2	
08	7777.5	2339.8	1.1	5.7	
09	9055.5	832.9	0.4	4.3	
10	7685.0	1902.6	1.0	7.2	
11	6074.0	615.8	0.4	2.4	**
12	15092.5	1388.1	0.4	4.3	***
13	3850.8	393.4	0.4	2.3	
14	15092.5	1388.1	0.4	4.3	***

Work Package Cost

The statistics for primary variables for work package cost model are given in Tables F.8, F.9, F.10, F.11 and F.12. The statistics for primary variables in Table F.12 are common for all the work package costs. Therefore, when the primary variables are assumed to be correlated, from the definition all of the work package costs are correlated. The positive definite correlation matrix for primary variables that was used for all the work package costs is given by $\mathbf{R}_{\mathbf{WPC}}$.

Table F.10: Statistics for Equipment Usage E_i ($e.d/year$)

WP#	$E[E_i]$	σ_{E_i}	$\sqrt{\beta_1}$	β_2	Common
02	512.0	126.8	0.8	5.0	
03	600.0	60.8	0.0	3.2	
04	851.0	136.6	0.7	9.0	
05	300.0	30.4	0.0	3.2	*
06	256.1	63.4	0.8	5.0	
07	303.7	30.8	0.4	2.4	**
08	425.5	68.3	0.7	9.0	
09	305.5	31.8	0.5	2.3	
10	230.5	57.1	1.0	7.2	
11	303.7	30.8	0.4	2.4	**
12	1063.8	170.7	0.7	9.0	
13	461.0	114.2	1.0	7.2	
14	300.0	30.4	0.0	3.2	*

Table F.11: Statistics for Subcontractor Cost S_i (\$)

WP#	$E[S_i]$	σ_{S_i}	$\sqrt{\beta_1}$	β_2	Common
02	20370.0	5802.8	0.8	9.0	*
03	10185.0	2901.4	0.8	9.0	**
04	38425.0	12220.2	0.5	3.2	
05	10185.0	2901.4	0.8	9.0	**
06	21966.5	2593.1	0.3	2.6	
07	32370.0	7054.8	0.8	7.8	
08	40462.5	6866.5	0.4	3.6	***
09	20555.0	4918.8	0.6	3.4	
10	8092.5	1373.3	0.4	3.6	
11	6464.3	1128.3	0.3	2.6	
12	40462.5	6866.5	0.4	3.6	***
13	20370.0	5802.8	0.8	9.0	*
14	10185.0	2901.4	0.8	9.0	**

Table F.12: Statistics for Common Primary Variables

Primary Variable	$E[X]$	σ_X	$\sqrt{\beta_1}$	β_2
$C_{L_i}(\$/m.d)$	141.85	22.76	0.7	9.0
$C_{M_i}(\$/ft^3)$	76.85	19.03	1.0	7.2
$C_{E_i}(\$/e.d)$	301.85	27.76	0.4	4.3
$I_{C_i}(\$/year)$	161850.0	35274.18	0.8	7.8
θ_{L_i} (%)	6.07	0.64	1.0	7.2
θ_{M_i} (%)	5.04	0.58	0.8	9.0
θ_{E_i} (%)	5.04	0.58	0.8	9.0
θ_{S_i} (%)	6.07	0.64	1.0	7.2
θ_{I_i} (%)	6.07	0.64	1.0	7.2
r (%)	7.54	0.85	0.2	2.5

$$\left[\begin{array}{cccccccccccccc}
1.0 & -.56 & 0 & 0 & .15 & .65 & 0 & 0 & 0 & .25 & 0 & 0 & 0 & 0 \\
-.56 & 1.0 & 0 & 0 & .34 & -.7 & 0 & 0 & 0 & -.4 & -.2 & 0 & 0 & 0 \\
0 & 0 & 1.0 & 0 & .20 & 0 & 0 & -.56 & 0 & 0 & 0 & -.2 & 0 & 0 \\
0 & 0 & 0 & 1.0 & .30 & .15 & 0 & .15 & 0 & 0 & 0 & 0 & .7 & 0 \\
.15 & .34 & .20 & .30 & 1.0 & .20 & .15 & .20 & 0 & 0 & 0 & 0 & 0 & -.4 \\
.65 & -.7 & 0 & .15 & .20 & 1.0 & 0 & 0 & 0 & .20 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & .15 & 0 & 1.0 & 0 & 0 & 0 & .50 & 0 & 0 & 0 \\
0 & 0 & -.56 & .15 & .20 & 0 & 0 & 1.0 & 0 & 0 & 0 & .30 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1.0 & .60 & .30 & .25 & .3 & .30 \\
.25 & -.4 & 0 & 0 & 0 & .20 & 0 & 0 & .60 & 1.0 & .20 & 0 & 0 & 0 \\
0 & -.2 & 0 & 0 & 0 & 0 & .50 & 0 & .30 & .20 & 1.0 & 0 & 0 & 0 \\
0 & 0 & -.2 & 0 & 0 & 0 & 0 & .30 & .25 & 0 & 0 & 1.0 & 0 & 0 \\
0 & 0 & 0 & .7 & 0 & 0 & 0 & 0 & .30 & 0 & 0 & 0 & 1. & 0 \\
0 & 0 & 0 & 0 & -.4 & 0 & 0 & 0 & .30 & 0 & 0 & 0 & 0 & 1.0
\end{array}\right]$$

Revenue Streams

The expected value, standard deviation, skewness and kurtosis for annual revenue and operating cost for the revenue streams are given in Table F.13.

Table F.13: Statistics for Annual Revenue and Operating Costs

RS#	Annual Revenue ($)				Annual Operating Cost ($)			
	$E[R]$	σ_R	$\sqrt{\beta_1}$	β_2	$E[O\&M]$	$\sigma_{O\&M}$	$\sqrt{\beta_1}$	β_2
01	5907500	1763709	-0.8	7.8	590750	176371	-0.8	7.8
02	3453750	694077	-0.3	3.5	323670	70548	0.8	7.8
03	3027749	441466	0.9	9.0	509249	145070	0.8	9.0