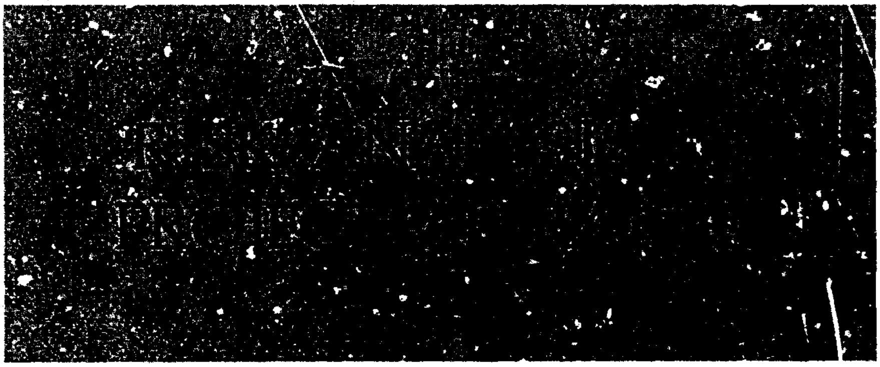
Public Disclosure Authorized

OCP 11

LOUIS Y. POULIQUEN



WORLD BANK STAFF OCCASIONAL PAPERS NUMBER ELEVEN

INTERNATIONAL BANK FOR RECONSTRUCTION AND DEVELOPMENT

World Bank Staff Occasional Papers

No. 1. *Herman G. van der Tak*, The Economic Choice between Hydroelectric and Thermal Power Developments, with a Note on Joint Costs in Multipurpose Projects, $1.00

No. 2. *Jan de Weille*, Quantification of Road User Savings, $1.00

No. 3. *Barend A. de Vries*, The Export Experience of Developing Countries, $1.50

No. 4. *Hans A. Adler*, Sector and Project Planning in Transportation, $1.50

No. 5. *A. A. Walters*, The Economics of Road User Charges, $5.00

No. 6. *Benjamin B. King*, Notes on the Mechanics of Growth and Debt, $1.50

No. 7. *Herman G. van der Tak and Jan de Weille*, Reappraisal of a Road Project in Iran, $3.00

No. 8. *Jack Baranson*, Automotive Industries in Developing Countries, $3.00

No. 9. *Ayhan Çilingiroğlu*, Manufacture of Heavy Electrical Equipment in Developing Countries, $3.00

No. 10. *Shlomo Reutlinger*, Techniques for Project Appraisal under Uncertainty, $3.00

No. 11. *Louis Y. Pouliquen*, Risk Analysis in Project Appraisal, $2.50

INTERNATIONAL BANK FOR RECONSTRUCTION AND DEVELOPMENT

WORLD BANK STAFF OCCASIONAL PAPERS NUMBER ELEVEN

LOUIS Y. POULIQUEN

RISK ANALYSIS IN PROJECT APPRAISAL

Distributed by The Johns Hopkins Press
Baltimore and London

Manufactured in the United States of America
Library of Congress Catalog Card Number 79-120739

ISBN 0-8018-1155-4

FOREWORD

I would like to explain *why* the World Bank Group does research work, and why it publishes it. We feel an obligation to look beyond the projects we help to finance toward the whole resource allocation of an economy, and the effectiveness of the use of those resources. Our major concern, in dealings with member countries, is that all scarce resources, including capital, skilled labor, enterprise and know-how, should be used to their best advantage. We want to see policies that encourage appropriate increases in the supply of savings, whether domestic or international. Finally, we are required by our Articles, as well as by inclination, to use objective economic criteria in all our judgments.

These are our preoccupations, and these, one way or another, are the subjects of most of our research work. Clearly, they are also the proper concern of anyone who is interested in promoting development, and so we seek to make our research papers widely available. In doing so, we have to take the risk of being misunderstood. Although these studies are published by the Bank, the views expressed and the methods explored should not necessarily be considered to represent the Bank's views or policies. Rather they are offered as a modest contribution to the great discussion on how to advance the economic development of the underdeveloped world.

ROBERT S. McNAMARA
President
International Bank for
Reconstruction and Development

TABLE OF CONTENTS

PART II

TABLES

FIGURES

PREFACE

This paper is part of a continuing effort in the Bank to find ways to tackle the problem of uncertainty. It relates primarily to work in the Transportation and Public Utilities Projects Departments; part of it was prepared while the author was on a temporary assignment in the Economics Department.

The reader is probably already familiar with Shlomo Reutlinger's recent paper, *Techniques for Project Appraisal under Uncertainty* (World Bank Staff Occasional Paper No. 10). A deliberate effort has been made to focus this new paper on particular problems arising in Projects Departments' work. If the attempt has been successful, it is only because of the wholehearted participation of a great many staff members of the Projects Departments. This participation has sometimes taken the form of reasoned skepticism rather than immediate acceptance, but the challenge of the former has proved at least as useful as the encouragement of the latter.

It would take too long to mention by name all those who contributed to this paper. Special mention is deserved, however, by Messrs. Aldewereld, Chadenet and Baum for their full support in this enterprise, Mr. Jaycox for taking the initiative of using risk analysis in the appraisal of three of the four projects which constitute the basis of this paper, Messrs. Higginbottom, Jones, Scoffier and Soges for their sustained assistance in solving the problems of evaluating probability distributions and correlations, Mrs. Comer for her patience and efficiency in carrying out all the computer programming work, and Miss Snell, Miss Maguire and Mr. Latimer for editing the final drafts.

ANDREW M. KAMARCK
Director
Economics Department

PART I

I

INTRODUCTION

The material in this paper is drawn from the results of about a year's experimentation with risk analysis; conclusions, therefore, can only be tentative at this stage. The purpose of the paper is threefold. First, it describes for the general reader three case studies in the use of risk analysis in project appraisal which serve to illustrate different aspects of the practical problem. Secondly, it discusses and illustrates a number of methodological problems. Thirdly, it makes some general observations on the usefulness of the approach.

The subject is introduced in general terms in Chapter II. The methodology is explained in Chapter III, a case study of the Bank Group's earliest analysis, the Port of Mogadiscio project. Chapter IV outlines the economic appraisal of a Tanzanian section of the Tanzam highway in order to show how the probability analysis fits the framework and also describes how the same analysis was used to resolve a technical problem. The benefits from disaggregation are shown in the rather special case presented by the Great East Road pre-project study in Chapter V.

In the second part of the paper some of the problems met in the analysis are explored in more detail. Chapter VI is devoted to the correlation problem. The techniques used in obtaining probability distribution judgments from technical experts are described in Chapter VII. Chapter VIII comments on the time and money costs of computer use implied by the methods described, while Chapter IX discusses questions of sample size and other statistical questions. Chapter X summarizes four ways in which risk analysis is thought to be especially useful, and the general advantages of the method, with a repeated warning about the correlation problem.

II

RISK ANALYSIS AND THE SIMULATION APPROACH

Risk analysis is essentially a method of dealing with the problem of uncertainty. Uncertainty usually affects most of the variables which we combine to obtain a cost estimate, an economic rate of return or net present value, a financial return, or any of the other indicators which may be used to evaluate a project. Sometimes we deal with this uncertainty by combining values for all input variables, chosen in such a way that they yield a conservative estimate for the result of the analysis. In other cases we may select the best estimate value, that is, the value which we think is most likely to be achieved. Both these solutions imply a decision: the first to look at the project with a conservative eye, the second, to disregard the consequences of any variation around the best estimate value. Both can lead to biased decisions. For example, if we combine only conservative estimates of our variables, our final result is likely to be "overconservative." On the other hand, by using only best estimate values we fail to take into account that other values of the variables we combine might result in substantial variations in the final estimate; thus, by basing our decision on a single value of the decision variable, we may be taking more risk than we intend.

The purpose of risk analysis is to eliminate the need for restricting one's judgment to a single optimistic, pessimistic, or "best" evaluation, by carrying throughout the analysis a complete judgment on the possible range of each variable and on the likelihood of each value within this range. At each step of the analysis these judgments are combined at the same time as the variables themselves are combined. As a result, the product of the analysis is not just a single value of the decision variable, but a judgment on the possible range of the decision variable around this value, and a judgment on the likelihood of each value within this range.

These judgments take the form of probability distributions. That is to say, each possible value of each variable is associated with a number between 0 and 1, such that for each variable the sum of all these numbers, or probabilities, is equal to 1. These probabilities, which are called subjective probabilities because they represent some degree of subjective judgment,[1] follow all the rules of traditional probability theory. From a mathematical point of view, risk analysis, therefore, consists of aggregating probabilities. Of the various ways in which this can be done, the only one we refer to in this paper, and the one which seems best fitted to risk analysis, is the Monte Carlo simulation technique.

The idea underlying the Monte Carlo technique is simple. When we say that a project has a 30 percent chance of earning a 10 percent return, we mean that if we had a great number of similar projects we would expect about 30 percent of them to earn a 10 percent return. Conversely, if we had a great number of projects and if 30 percent of them earn a 10 percent return, we could say that the probability of a 10 percent return is 30 percent. Hence the simplest application of the Monte Carlo technique is to build a great number of projects with the characteristics of the one we are interested in, and see how many of them earn 10 percent, 15 percent, 20 percent, etc. In practice, the value of each of the uncertain variables is chosen by random selection, and the rate of return or some other decision variable is computed for the project defined by these values. The process is repeated many times and the results are statistically analyzed. The only difficulty is in making sure that the distribution of the values of each of the input variables, as it emerges from the random selection, is consistent with the distribution for that variable chosen for the analysis.[2] The technique will become clearer after description of the Mogadiscio and the Tanzam highway cases.

---

[1] All "subjective" judgments that we are likely to obtain from experts are based on some sort of "objective" experience. For example, usually the past record of similar events leads the expert to attach more importance to one outcome than to another.

[2] The reader may wish to refer to James W. Butler, "Machine Sampling from Given Probability Distributions," *Symposium on Monte Carlo Methods* (John Wiley & Sons, Inc.) 1956.

III

THE PORT OF MOGADISCIO: A CASE STUDY

The risk analysis used to appraise this project was the first to be undertaken in the IBRD. Initially, a conventional cost-benefit analysis was used to appraise the project. A Bank appraisal mission consisting of an engineer, a financial analyst and an economist, in 1967 visited the existing lighterage port at Mogadiscio, Somalia, which the project would have replaced with a two-berth, deep-water port. But the conventional analysis, based on information the mission gathered and on a consultant's report, ran into serious difficulties in its effort to assess the economic justification of the project using best estimates of the variables. A sensitivity analysis, undertaken at this stage to pinpoint the most crucial elements of the project, narrowed the sources of uncertainty to seven variables. It was then decided that a risk analysis using probability distributions would be a useful tool to deal with these uncertainties, though such a risk analysis had not been undertaken before in the Bank and had not been anticipated at the time of the mission's visit to Somalia. The Bank might nowadays carry out this risk analysis slightly differently, but the general approach is thought to be correct and the later modifications would not change the decision about the economic justification.

The Project's Background

The project included the construction of a breakwater, two berths, two transit sheds, storage area and office accommodations. No dredging was necessary since natural depth existed in the approach from the sea as well as at the site of the two proposed berths.

Traffic through the existing lighterage port of Mogadiscio for the period 1964 to 1966 averaged about 125,000 tons per year. It was expected that, in addition to generating some traffic, the construction of the new port would result in the diversion of about 85,000 tons per year of bananas which were exported through the port of Merca, about 50 miles south of Mogadiscio.

Economic Justification

The cost-benefit analysis takes into consideration, on the cost side, the capital cost of the project and, on the benefit side, three types of projected savings: (a) savings in cargo handling cost, (b) savings in reduction of damages, (c) savings in ship turnaround time. These savings are applied to the projections of future traffic, broken down into the normal growth of traffic that might be expected without the port, generated traffic, and diverted traffic (see below). The result of the analysis is an internal rate of return over the average life of the assets.

Cost of the project

The cost of the project has been estimated at $14.6 million. A cost breakdown is roughly as follows:

Breakwater	43 percent
Berth and storage area	25 percent
Engineering fees	8 percent
Auxiliary works	24 percent

The major single item is the breakwater. Among the auxiliary works, the biggest single item represents only 3 percent of the total cost.

Existing composition and the uncertain trends of traffic

Traffic through the port of Mogadiscio consisted primarily of imported goods and materials. Imports constituted, on the average, about 75 percent of total traffic at Mogadiscio, and the port handled over 45 percent of the country's total imports. There was no large bulk traffic through the port. The largest single category of import traffic was cereal grains, which usually accounted for about 25 percent of total imports. Otherwise import traffic consisted of small consignments of manufactured goods, machinery and raw materials. For export, the main commodity handled during the previous six years had been charcoal, which until 1966 accounted for about 75 percent of total exports. The only other exports of individual importance were live animals, skins and hides. Traffic through the lighterage port at Merca, south of Mogadiscio, consisted almost entirely of bananas for export.

Dry cargo traffic through the port of Mogadiscio increased rapidly and fairly steadily from about 102,000 tons in 1960 to nearly 164,000 tons in 1965, at an annual rate of 10 percent. But in 1966, total traffic fell to 110,000 tons, almost to the 1960–61 level. Traffic since 1960 has been affected by a number of factors. The main reason why import tonnage doubled between 1960 and 1965 was the rapid growth of the population and of construction activities in the city of Mogadiscio during its first years as the capital of the United Republic. In 1964 and 1965, imports increased considerably due to speculation against the imposition of strict import licensing to improve the balance of payments. Another factor in 1965 was a severe drought in 1964–65 which caused food grain imports to double. In 1966 imports fell from 126,000 tons to 85,000 tons, as import restrictions were imposed and as inventories built up in 1964–65 were run down. Imports through Merca were small; they increased from about 5,500 tons to about 8,000 tons over six years, or at an average rate of 5 percent per annum.

The export tonnages handled have varied from year to year with the fluctuations of the charcoal trade, which the Government had been attempting to end in order to retard land erosion. In June 1967 trade in charcoal was made illegal. A comparison of average export tonnages (excluding charcoal) of the period 1964–1966 with the 1960–1966 average indicated a growth rate of nearly 7 percent per annum for the six years. The export of bananas through Merca increased from about 49,000 tons in 1960 to about 62,000 tons in 1964 and remained at about this level during 1965 and 1966. The growth of banana export tonnage over the six years had been about 4.5 percent per annum, but the tonnage growth rate understated trade in bananas because improved packing techniques had reduced shipping by about 12 percent.

Future traffic projections

Normal Traffic. The foreseeable long-term need for restrictions on imports and a probable slowdown of economic growth in the Mogadiscio area implied moderate "best estimates" of future import traffic. Imports were projected to grow at a rate of 3.5 percent per annum from a base of 106,000 tons, the average of import tonnages over the 1964–1966 period. Exports of live animals, hides and skins should continue to grow fairly rapidly but at a much slower rate than in recent years, because these exports had already been expanded considerably and further increases were likely to be more difficult.

Export tonnages (excluding charcoal) were projected to grow at the rate of 6 percent per annum from the 1964–1966 average.

Generated Traffic. In view of the large unit savings to be achieved by constructing a deepwater port at Mogadiscio, a large proportion of which could be passed on to the Somalian consumer and producer, considerable export and import traffic should be generated by the proposed project. Taking price elasticity of demand for this traffic as .08, generated traffic was estimated at 10,000 tons per annum by the third year of operations of the new port, increasing thereafter at the average growth rate indicated above for normal traffic.

Diverted Traffic. Once the new port was in operation, the banana exports and small import tonnages passing through Merca would be diverted to Mogadiscio. The banana production potential of the Genale/Scialambot area in the Merca hinterland, estimated at 100,000 tons, was to be realized within five years, according to current plans. The plan, however, was probably overly optimistic. The banana industry appeared to have a fairly bright future, but there were marketing problems which had not been met because of protection in the Italian market. Therefore it was assumed that banana exports would build up from the 1964–1966 average level of 61,000 tons to about 85,000 tons over the ten-year period 1967-1976.

Savings

Table 1 shows a year-by-year breakdown of the benefits from 1972 to 1978 based on the best estimate of each variable.

TABLE 1: Port of Mogadiscio: Estimated Benefits on the Basis of the Best Estimate of All the Variables

(US $'000, 1967 prices)

Type of Benefits	1972	1973	1974	1975	1976	1977	1978
Reduction in general cargo handling costs	296.4	308.4	320.8	333.8	347.3	361.5	375.9
Reduction in diverted traffic handling costs	136.6	145.2	154.1	163.4	168.6	170.3	170.7
Reduction of physical damage to cargo	656.0	678.7	704.7	735.0	761.5	790.0	820.2
Reduction in ship turnaround time	532.7	552.6	572.5	591.3	609.3	629.0	641.7
Savings from generated traffic	18.5	37.0	46.2	50.7	55.1	59.6	61.0
	1,640.2	1,721.9	1,798.3	1,874.2	1,941.8	2,010.4	2,069.5

Savings in Cargo Handling Cost. Lighterage is a rather inefficient operation since it requires double handling of the cargo. In addition, the transshipment between ship and lighter is hampered by the movements of both the ship and the lighter and is, therefore, much slower than loading or unloading on a protected quay. For the existing lighterage port the cargo handling costs—divided into a fixed and a variable component—were taken at their actual value after some adjustments to eliminate costs resulting from redundant labor and faulty work methods. For the proposed port fixed and variable costs were computed from their components, which are essentially:

—for the variable costs: the number of men and the productivity of these men, i.e., the number of tons they can handle in an hour;

—for the fixed costs: the costs of maintenance (labor and materials), administrative staff and staff in warehouses and transit sheds.

Reduction in Damages. In the lighterage port, the necessity of handling the cargo twice, including one time at sea, means not only higher costs but high damages. Benefits resulting from a reduction of these damages have been computed on the basis of the proportion (P) of the forecast tonnage (T) which is expected to be saved through easier handling and of the value (V_c) of an average ton of cargo. Therefore, the resulting saving (S_D) is given by

$$S_D = P \times V_c \times T$$

Savings in Ship Turnaround Time. Savings were expected in ship turnaround time because the higher productivity of labor anticipated in the new port implied faster loading and unloading and less ship time in port per ton of cargo. The savings formula essentially compares the observed number of ship-days required to move a given tonnage (T) of cargo of each type (general cargo, bananas) in the existing port, with the ship-days estimated to be required under the improved conditions to move the same tonnage, less an allowance for waiting time. The savings in ship-days are then put into monetary terms by multiplying them by the value of a ship working day (V_s). The savings in turnaround time (S_T) for tonnage T is:

$$S_T = V_S \left[\frac{T}{P_S} - \frac{T}{P_L \times H} - W(T) \right]^1$$

where P_S is the tonnage loaded or unloaded per ship per day, observed in existing conditions, P_L is the estimated cargo handling rate per hour under the changed conditions and H is the number of hours to be worked per day. The

[1] Simplied formula.

estimate of turnaround time derived from handling capacity has to be adjusted for the expectation that some ships may have to wait. A simple Poisson queuing model gave this waiting time (W) as a function of the total tonnage. The waiting time has then been allotted to each type of traffic proportionally to the share of this traffic in the overall traffic.

Projected savings in all three categories were considered functions of the traffic handled through the port, which was estimated year by year from the projected growth rates described earlier. The traffic demand was assumed to be linear and consequently unit savings for generated traffic were taken as one half of the unit savings for normal traffic. In the case of diverted traffic the benefits have been reduced by the cost of an increase of about 25 miles in the land transport of bananas. The traffic taken for the computation of the benefits is the real traffic in the port up to the time when the economic capacity of the two berths will be reached. Thereafter all benefits stay constant except those resulting from a sudden reduction in ship turnaround time[2] followed by a progressive increase again in ship waiting time. The economic capacity of the two berths was computed separately, using the queuing model referred to above.[3]

Shortcomings of the Analysis

We did not feel much confidence in our results. To arrive at the final 12.2 percent economic rate of return, we had used best estimates for each variable, but on some occasions, we had been obliged to resort to awkward ways of finding them (combining notions of both the mean and the mode, for example). Furthermore, the rate of return was based on highly uncertain data, and was interpreted under some rather optimistic assumptions about the variables.

On the traffic side, for example, the difficulties of accurately predicting traffic growth have already been indicated, even without taking into consideration the closure of the Suez Canal. Traffic in cereals, the major import commodity, fluctuated with the success of competing internal crops and with

---

[2] It was assumed that a third berth would be brought into operation as soon as the economic capacity of the two existing ones was reached.

[3] Since a detailed description of the financial analysis of this project would not add very much to the case, this part of our work will be described only briefly. On the basis of projected traffic, the operating revenues resulting from a given system of port charges (different for imports and exports) were estimated. By comparing these revenues with investment cost and operating expenses the rate of return was computed over the 40 year life of the project. This gave, with linear depreciation, the financial return on net fixed assets for the 5 years following the construction of the project. In addition, the converse problem was solved to determine a system of charges which would yield a 6 percent financial return over the life of the assets.

the weather, and therefore was difficult to forecast. Developments in other import traffic depended on regional development, which had shown no clear trend. Traffic in charcoal, the major export commodity, was ruled out by law, and it was hard to guess to what extent it would be replaced by increased livestock trade. Bananas, which could constitute about one-third of the traffic of the new port, might not be diverted, or they might not be able to compete when Italy's preferential treatment was abolished under the rules of the Treaty of Rome and thus might disappear as an export altogether.

Nor could the cost of the project be precisely estimated. The costing of the most expensive single item, the breakwater, was the most uncertain. The usual uncertainty about the quantity of construction material required was increased by uncertainty about contingency allowance for storm losses (the quantity of materials lost during construction because of rough seas). On the price side, the lack of detailed analysis on rock availability reduced the estimate of the unit cost of rock to a guess.

The number and nature of these and countless other project uncertainties are not new to anyone acquainted with project appraisal. It is therefore only necessary to add that the productivity of labor in African ports varies from about 5 to 12 tons per gang-hour, that we were not very sure of the value of an average ton of cargo nor of the value of a ship working day, and that our hypothesis on the reduction in cargo damage was made without much reliable data support.

Sensitivity Analysis

The conventional analysis had failed to give a satisfactory result using single best estimates. The most natural way to deal with this situation was to make a sensitivity analysis, in other words, to see what would happen if other values of the input data were substituted. Using the most unfavorable estimate for each variable, we obtained a 2 percent rate of return, which confirmed our suspicion that the project was risky. But how risky? Again, a natural approach to this question was to try to find out which variables were principally responsible for the variations of the rate of return.

For this purpose we examined each one of the 27 uncertain variables which appeared in our rate of return computations. We varied them, one at a time, holding all other variables at their best estimate value. We found the variation of the rate of return as the best estimate of each variable was replaced by the maximum value, the minimum value, and a value 10 percent above the best estimate. Table 2 shows the results we obtained for the economic rate of return. On the basis of this table, and a similar one for the financial rate of return, it appeared that the performance of the project was essentially ex-

TABLE 2: Port of Mogadiscio: Sensitivity Analysis

Parameter Varied (1)	Value Assigned (Somali shillings except where stated otherwise) Highest (2)	Best Estimate (3)	Lowest (4)	Internal Rate of Return, % Based on Highest Value (2) (5)	Lowest Value (4) (6)	Percentage Response to 10 percent Variation of Best Estimate[b] (7)
1 Cost of project ('000 sh.)	119,255	103,700	98,000	10.8	12.9	8.20
2 Gang productivity, tons per hour	12	10	5	14.2	7.3	4.10
3 Reduced damages, percentage	2	1.5	1	13.6	10.8	3.28
4 Ton cargo, average value	2,500	2,000	1,500	13.3	11.2	3.28
5 Unnecessary staff[a]	650	400	150	10.7	11.9	1.64
6 Road transport costs per ton, bananas	37.5	25.4	12.5	11.5	13.0	0.82
7 Ship working day value, general cargo	7,868	6,428	4,988	12.5	11.9	1.64
8 Life of assets, years	60	40	20	12.3	10.9	0.82
9 Growth rate percent, imports	8	3	2	13	11.8	0.82
10 Avoided investment costs ('000 sh.)	4,285	2,800	0	12.4	11.9	0.82
11 Ship working day value, banana	12,780	10,650	8,520	12.3	12.1	0.82
12 Maximum banana traffic p.a. ('000 tons)	100		0	12.1	12.1	0.82
13 Growth rate percent, exports	14	6	3	12.5	12.1	0
14 Growth rate percent, bananas	8	4	0	12.2	12.6	0
15 Maintenance cost p.a. materials ('000 sh.)	900	780	600	12.1	12.4	0
16 Men in gang, general cargo	35	30	25	12.1	12.3	0
17 Men in gang, bananas	28	24	20	12.2	12.3	0
18 Men in transit sheds	115	102	85	12.2	12.3	0
19 Men doing maintenance	115	100	85	12.2	12.3	0
20 Miscellaneous cost p.a. ('000 sh.)	360	300	270	12.2	12.2	0
21 Banana charge per ton[c]	27	18	9	12.2	12.2	0
22 Port charges per ton[c]	65	54	45	12.2	12.2	0
23 Ratio export/import charges[c]	1.5	1.19	1	12.2	12.2	0
24 Number of men in warehouses	36	32	29	12.2	12.2	0
25 Number of men in administration	95	82	65	12.2	12.2	0
26 Elasticity of traffic demand	0.2	0.08	0	12.2	12.2	0
27 Ship charges per call[c]	150	100	50	12.2	12.2	0

Note: Based on best estimates of parameters varied, the economic rate of return is 12.2 percent.

[a] This line assumes that this staff receives 3,000 shillings a year; in the rest of the analysis it was assumed that it received nothing.

[b] See page 10. It was assumed that the variations were reasonably well-behaved and that therefore the percentage response to increasing the best estimate by 10 percent was a good representation of a 10 percent variation in either direction around the best estimate.

[c] These refer to the financial rate of return only.

plained by seven variables: (1) cost of the project, (2) productivity of labor, (3) value of an average ton of cargo, (4) percentage of the tonnage which would be saved through reduction in damages, (5) rate of growth of imports, (6) value of a ship working day, and (7) the life of the assets.

Above, it was explained that in this analysis we varied only one variable at a time. However, we made exceptions to this rule in the case of variables whose variation would, in the real world, very probably be correlated. For example, when we varied the productivity of a general cargo gang, we varied at the same time the productivity of a banana gang. These two variables both depend on the efficiency of the organization of the new port, and on how efficient port operators the Somalis will turn out to be. Therefore, they are likely to be correlated. There is no reason, however, why their variations should be completely interdependent since, for example, the productivity of the general cargo gang also depends on the degree of unitization of the cargo, a factor which is unlikely to affect the productivity of the banana gang. In assuming, as we did, that the variations of these two variables were fully correlated we may have somewhat overestimated the sensitivity of the final result to the productivity of labor.

We used a more rigorous way of handling correlation in estimates of the number of persons required in the various operations of the port. The estimates of the number of men in gangs, transit sheds, warehouses, etc., are subject to uncertainty not only as to the exact numbers of men required to operate the port in the most efficient way, but also as to the Port Authority's efficiency in eliminating the redundant labor presently employed in the port. The first uncertainty is likely to affect the variables independently of one another since overestimation of the number of men needed in a banana gang needs not necessarily imply overestimation of the number of men needed in a transit shed. The second uncertainty, on the contrary, is likely to affect all the variables in the same direction. If the Port Authority does not manage to eliminate redundant personnel, it is likely that this personnel will be distributed among the various services of the port and so increase the cost of all of them.

We resolved this difficulty by creating an artificial variable, which we called "unnecessary staff," and which represents all redundant labor in the port. Then, rather than testing the sensitivity of the actual number of persons presently employed in each of the various services of the port, we tested separately the sensitivity of the theoretically most efficient number of persons required in each one of these services and the sensitivity of all unnecessary staff for the whole port. In sum, we tried to test the importance not so much of a variable *per se*, but rather of various sources of uncertainty.[4]

[4] The problem of correlation is discussed further in Chapter VI.

If, instead of isolating seven major sources of uncertainty, we had isolated only one, or eventually two, our task would have been completed. We could have concluded that if this determining variable were, say, less than a given value *a*, the project was very likely to be justified, and that if it were more than *a*, the project was very likely not to be justified. A simple evaluation of the likelihood that this variable was less than *a* would have been enough to give us an idea of the riskiness of the project. With two variables, our judgment would have been more difficult to put in words. We might, however, have been able to illustrate it with the help of a graph showing the limits within which the project would be justified.

But with seven variables, such a task is impossible. One can find an infinity of combinations of the variables for which the project is justified, and an infinity of combinations for which it is not. Ironically, the more combinations of variables one tries, the less clear the picture of the project becomes. The only way to obtain an overall, synthetic picture of the project is to proceed with a probability analysis.

Probability Analysis

The first step of this risk analysis is to assign to each variable a probability distribution. Since we had found out that the variation of the rate of return was essentially explained by the variations of seven variables, we limited our analysis to these seven variables. The distributions were based essentially on subjective judgment. This did not raise any difficulty in practice and did not take much time, since we had limited ourselves to a small number of variables. The distributions we obtained are shown in Figure 1. They were obtained in essentially two ways.

The three normal distributions that we adopted for the value of a ship working day, the value of an average ton of cargo, and the percentage reduction in damages, and the chi-square distribution that we used for the cost of the project are the result of an approach which could be compared to the portrait method used to identify suspects. On the basis of limited information a portrait is drawn and subsequently modified until the informant is satisfied with it. Similarly, on the basis of limited information obtained from the appraiser, we chose among classical probability distributions one which seemed to fit the case. We drew it, indicated the corresponding probabilities for various intervals, and went back to the appraiser. He decided whether it was too skewed or whether an interval had too high a probability, and on the basis of this new information we modified it. We repeated this process until the appraiser was satisfied with the distribution.

The distributions we used for the life of the assets, the growth rate of im-

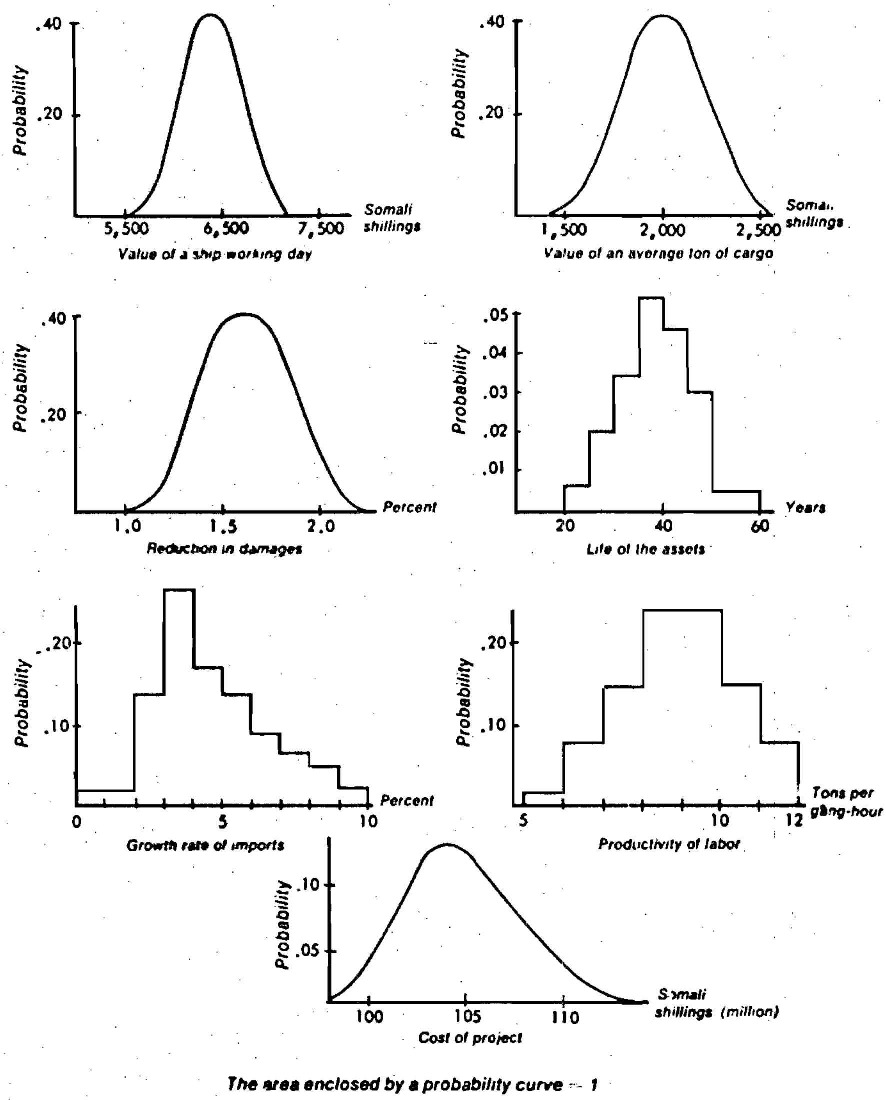


Figure 1. Probability Distributions Used in the Simulation of Mogadiscio Port Project

ports and the productivity of labor, which we have called step rectangular distributions, were obtained with the somewhat more active participation of the appraiser. Let us take for example the case of the productivity of labor. The steps by which the appraisal team set up its distribution are illustrated in Figure 2. We first divided the total range of variation we had delineated in the sensitivity analysis (5-12 tons per gang-hour) into two intervals:

5-10 and 10-12, and tried to assign a probability to each one of them. For this we used a trial and error approach based on the engineer's experience: 50%-50% gives too high a probability to the 10-12 interval, as does 60%-40% and 70%-30%, but 80%-20% does not seem to give enough; therefore, we tried 75%-25%. In other words, the appraisal team's judgment was best expressed quantitatively by saying that the probability of exceeding 10 tons per gang-hour is only one-third of the probability of getting a lower productivity of labor.

In the second step, we chose to subdivide the 5-10 interval into 5-8 and 8-10. Then following the same trial and error process we allocated a 30 percent probability to the 5-8 interval and a 45 percent probability to the 8-10. The sum of these two probabilities is, of course, equal to the 75 percent probability of the entire 5-10 range. In a third step we pushed this subdivision further and obtained the following distribution:

from	5 to 6	a 5 percent	probability
"	6 to 7	a 10 percent	"
"	7 to 8	a 15 percent	"
"	8 to 9	a 22.5 percent	"
"	9 to 10	a 22.5 percent	"
"	10 to 11	a 15 percent	"
"	11 to 12	a 10 percent	"

Finally, in a fourth step we made some minor adjustments to give the distribution a final polish. For example, we found that compared to the probability of the 6-7 range, the probability of the 8-9 and 9-10 ranges was too low. We therefore raised them to 25 percent and decreased the 6-7 range probability to 8 percent, which in turn led us to decrease the probability of the 5-6 range to 3 percent. Similar considerations for the 10-12 range led us to the final distribution in Figure 2.

This approach and the portrait method for choosing the other distributions include an iterative interaction between quantitative and qualitative judgment. On the basis of a qualitative judgment one attempts to produce tentative figures. These figures in turn are translated back into qualitative judgment which is compared to the initial qualitative one. The figures are modified in light of the discrepancy, and the procedure is repeated until the qualitative judgment derived from the quantitative one fully agrees with the initial judgment.

The simulation

The simulation is by far the fastest and the easiest operation of the entire analysis. When, as in our case, computer help is used, the computer can be

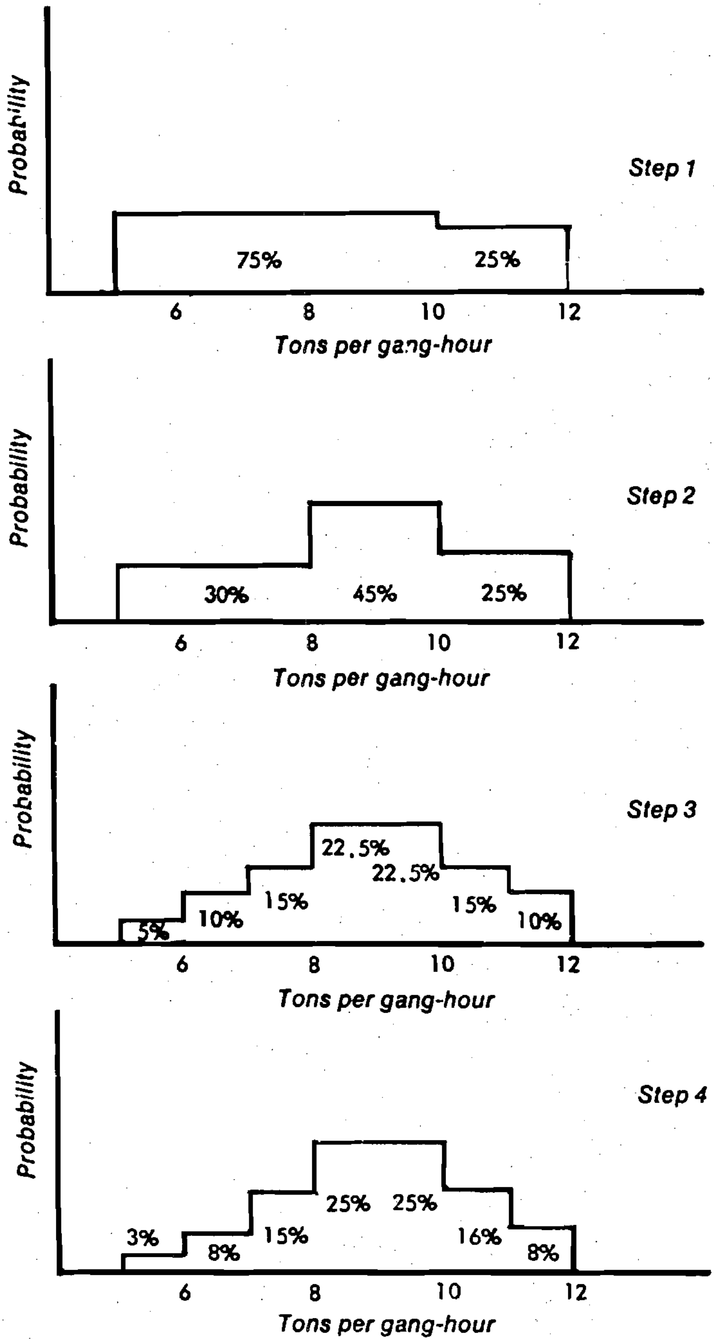


Figure 2. Steps in Establishing the Probability Distribution of the Productivity of Labor

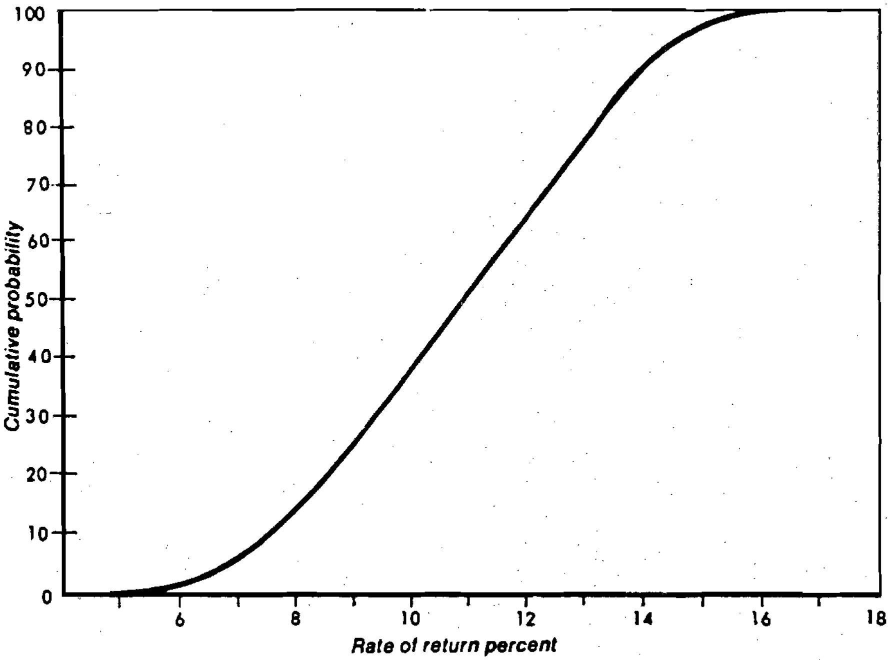


Figure 3. Mogadiscio Port Project: Cumulative Probability Distribution of the Economic Rate of Return

instructed to generate random values for each of the parameters varied in the analysis, to compute the rates of return, to repeat the process until enough values are obtained (300 times in the present case) and then to give the observed distribution of the result. In this case the computer was also used to draw the curve in Figure 3. The simulation is, therefore, an operation which requires no outside intervention and takes only a few minutes. Programming the computer to calculate rates of return is very simple; for random number generation, it is slightly more involved but still easily derived from the basic random number generators which exist in all computer libraries.

The results

The results for the economic rate of return are summarized by the cumulative probability distribution in Figure 3. It has a mean of 10.6 and standard deviation of 2.5. Along the x-axis are the rates of return and along the y-axis the probability that these rates of return will not be exceeded. For example, we find that there is a 99 percent probability that the rate of return will exceed 5 percent, a 94 percent chance of it exceeding 7 percent, and so on along the curve until we reach a 2 percent chance of exceeding a 15 percent rate of return. The curve can also be used to determine the probability that the rate of return will fall within a given range: we take the difference along the ordinate of the two extreme points of the range. For example, we find there is about a 40 percent chance that the rate of return will be between 10 percent and 13 percent. The figure also shows that the probability of getting a return inferior to 12.2 percent, the rate of return we obtained in the conventional analysis using best estimates for each variable, is 70 percent, but the probability of getting more is only 30 percent. So at first glance, the results of this risk analysis seem to indicate that doubts about the likelihood of the 12.2 rate of return were fully justified.

But, unlike the sensitivity analysis, the probability analysis gives us a complete picture of the project and enables quantification of project risk—not, of course, the "true" risk, but the risk as it appeared to the appraisal mission. The probability distribution of the rate of return summarizes this risk; one could say that it represents the complete judgment of the appraisal mission.

How to use this probability distribution in a scientific way could constitute an entire study on its own. In the absence of any such scientific criteria, we used the distribution in a very pragmatic way. We first saw that, while the sensitivity analysis had told us that the minimum rate of return was 2 percent, the chances of ever getting below 5 percent were so slim that it could be considered the minimum for all practical purposes. We then looked at the probability of getting less than 8 percent, since we thought 8 percent was a low but probably still acceptable value of the opportunity cost of capital in Somalia, and found it to be about 15 percent. We thought that this was acceptable, when combined with the information that the project had a better than even chance of earning more than 10 percent and nearly a 20 percent chance of earning more than 13 percent.

Our judgment was therefore arrived at by combining considerations of what the project could turn out to be at the extremes and the probabilities that this would happen, with a weighted estimation of how any unfavorable outcomes might be compensated by favorable ones. In this respect, the mean rate of return was particularly helpful. It indicated to us that on balance, we could expect the project to yield an 11 percent rate of return; we thought this was acceptable, especially since we did not have to fear any large variations around this value. On the basis of this simple analysis, we decided to recommend the project for financing.

We felt particularly free to make this recommendation because in presenting it, we were not just presenting our own difficult decision; we were also presenting to management all the information necessary to check this recommendation, and possibly to overrule it. If we had indicated in our appraisal report only the

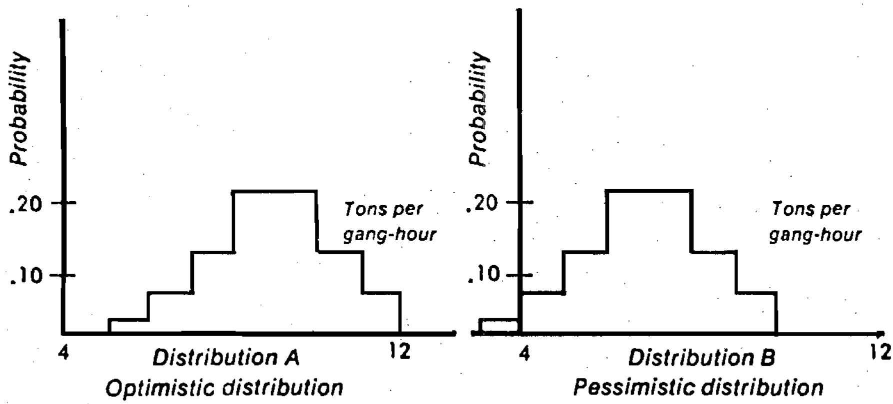


Figure 4. Mogadiscio Port Project: Shift of the Productivity of Labor Distribution

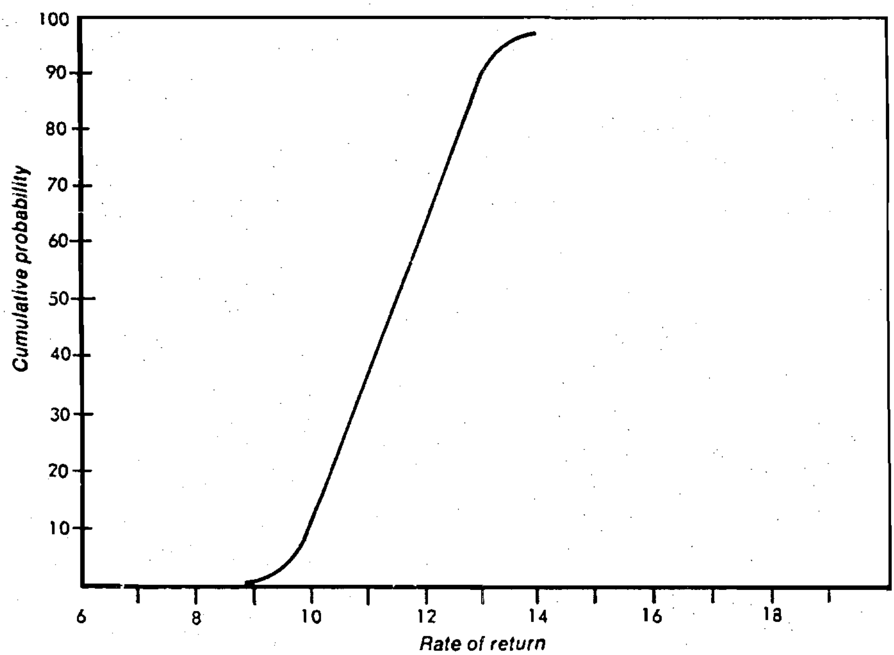


Figure 5. Mogadiscio Port Project: Cumulative Distribution of the Economic Rate of Return

12.2 percent rate of return found in our best estimates calculation, the situation would have been quite different. The decision-maker would have been acting in the dark. In fact, it would not have been possible for anybody but the team of appraisers to evaluate the risk of the project. We would have been recommending the financing of a project earning a 12.2 percent rate of return after having already decided that the risk of the project was acceptable —a dangerous mixing of analysis and decision-making.

Another important outcome of the risk analysis was that we found a way to reduce project risk. We were led to this finding by the results of a second sensitivity analysis which we carried out, this time on the entire probability distribution of each variable. We found that the productivity of labor had a much higher sensitivity in the risk analysis than in the first sensitivity analysis because we were considering its entire variation rather than its value at a single point. This higher sensitivity means that if our judgment about the probability distribution of the productivity of labor is too optimistic (i.e. in Figure 4, if the true distribution is B and not A), then the probability of having a rate of return inferior to 8 percent is no longer 15 but 30 percent (see also Figure 21). On the other hand, this assumption that the productivity of labor could be confined to the 9-10 tons per gang-hour range yields the distribution in Figure 5; the risk of the project has practically been eliminated. We therefore suggested that a consultant be engaged at an appropriate time to help organize cargo handling operations and that this be made part of the loan agreement. In this case, we were not only able to quantify the risk attached to the project, but also to find a feasible way to reduce it.

IV

THE TANZAM HIGHWAY: A CASE STUDY

The Project

The Tanzam highway is a 1170-mile-long highway from the Copper belt in Zambia to the port of Dar es Salaam in Tanzania. This route, most of which was a poor gravel road, became important after Zambia achieved independence in 1964 and after the Rhodesian unilateral declaration of independence in 1965 led to a UN embargo on traffic with Rhodesia. Compliance with the embargo requires Zambia to divert its seagoing import-export traffic from the existing main route across Rhodesia and Mozambique to a more expensive route via the Tanzam highway. Reconstruction to two-lane bituminous paved standard of the gravel or earth sections of the road from Kapiri Mposhi in Zambia to Morogoro in Tanzania (965 miles) was planned under various financing arrangements. The Bank group was asked to finance two sections in Zambia, of 122 miles and 235 miles respectively, and one in central Tanzania of 311 miles. It is this last project of 311 miles which is briefly presented here.

The road was not the only transport mode proposed to meet the Zambian demand for a new access to the sea: among other projects, a railway, to come into operation at an unknown date, was also proposed, and an oil pipeline was under construction. Uncertainty about the economic benefits therefore existed in the project from the start. Before describing the probability analysis, however, it is useful to outline the framework in which it fitted. The economic

TABLE 3: Tanzania, Highway Project: Estimated Vehicle Operating Costs on: I. Engineered Bitumen/Asphalt, II. Engineered Gravel, and III. Unimproved Earth. (1967–1968)

(US cents per vehicle mile)

Vehicle Category Fuel Type Road Type	Average Car Gasoline I	II	III	Pick-up Truck Gasoline I	II	III	Average Truck 7-Ton Capacity Diesel I	II	III	Average Bus 50 Passengers Diesel I	II	III	Truck-trailer 30-Ton Capacity Diesel I	II	III
A. Flat to Rolling Terrain															
Fuel	1.11	1.24	1.38	1.62	1.68	1.81	1.86	2.26	2.68	1.82	2.28	2.70	4.15	5.18	6.15
Lubricants	.06	.08	.10	.08	.11	.15	.14	.17	.26	.15	.19	.26	.42	.50	.63
Maintenance															
Labor	.22	.29	.39	.26	.32	.55	.88	1.39	2.32	.77	1.31	2.29	6.30	10.47	22.30
Parts	.44	.56	.82	.48	.69	1.10	1.48	2.41	4.00	2.16	3.71	6.40			
Tires	.20	.40	.68	.32	.62	1.20	1.18	2.68	6.05	.64	1.80	4.48	4.30	5.88	11.30
Crew Wages	1.12	1.40	1.85	1.24	1.56	2.08	1.40	1.77	2.34	1.77	2.34	2.80	1.95	3.55	6.30
Interest	.78	.97	1.30	.63	.79	1.05	.40	.49	.66	.87	1.15	1.40	1.10	1.60	2.52
Insurance	.37	.46	.61	.32	.40	.53	.34	.43	.57	.79	1.05	1.26	.40	.60	.90
Depreciation	2.12	2.79	4.14	2.02	2.74	4.54	1.52	2.32	3.97	4.17	6.28	10.32	6.80	12.40	20.05
Total	6.42	8.19	11.27	6.97	8.91	13.01	9.20	13.95	22.85	13.14	20.11	31.91	25.42	40.18	70.15
B. Rolling to Hilly Terrain															
Total	6.65	8.58	11.80	7.41	9.11	13.90	9.75	15.05	25.02	14.49	21.21	34.43	26.81	43.49	76.65

Note: Costs are net of taxes, license fees and other transfer payments. Costs of terminal operation, waybills, company administration, which are not directly affected by road improvement, are not included.

Sources: Jan de Weille, *Quantification of Road User Savings*, Occasional Paper No. 2. (East African prices); United Research Incorporated, Consultant to the Government of Tanzania; interviews with trucking firms in Zambia and Tanzania; mission estimates.

analysis was that normally employed in the Bank in the assessment of a road project.[1]

Estimates of costs were first established, with their probable years of occurrence. Shadow prices, sometimes used, were not required in this case, and the allowance generally made for contingencies was also unnecessary since a probability analysis was to be made. A date was set, to which the present value calculations relate. Separate analyses were made of each section of the project: in this case eight sections of the road were separately costed. Benefits and costs were assessed from the national angle. In this case separate calculations were made for Tanzania alone and for Tanzania and Zambia together.

The main quantifiable benefits were four:

a) Savings on operating costs because the road would be upgraded for existing traffic and for the normal increase of traffic which would have taken place even without the road. These were estimated by obtaining costs per mile per type of vehicle on the new road and subtracting similar costs on the old road. General comparative data were used for this purpose, particularly those in Occasional Paper No. 2, as amended by information obtained from the government of Tanzania and its consultants, and from interviews with trucking firms (Table 3). It was necessary to know for each road section the average daily traffic per vehicle type at the relevant date, and to assess the rate of normal traffic growth of each (Table 4).

TABLE 4: Tanzania, Highway Project: Estimated Tanzanian[a] Traffic on Project Road Sections

(Average Daily Traffic 1967)

Project Road Section	Cars	Trucks and Buses	Truck-Trailers	Total
Morogoro +20	120	165	20	305
Morogoro +20 – +38	80	130	20	230
Morogoro +38 – Mikumi	54	118	14	186
Mikumi-Mahenge	36	79	9	124
Mahenge-Kitonga-Iringa	58	120	10	188
Iringa-Sao Hill	62	89	6	157
Sao Hill-Makumbako	36	74	6	116
Makumbako-Iyayi	24	54	5	83
Annual Growth Rate	6%	8%	8%	

[a] Not including Zambian transit traffic. The best estimate for this was an ADT of 85.

b) Benefits from a shortening of the distance may be related directly to the annual trucking costs per vehicle-mile (as in Iran, in Occasional Paper No. 7) or as here, to the operating costs of vehicles on the road (Table 5).

[1] The methodology is described in Occasional Papers Nos. 2, 4 and 7 of this series.

TABLE 5: Tanzania, Highway Project: Estimated Vehicle Operating Costs on Project Road Sections[a]

(US cents per vehicle mile)

Road Section	Old Miles	New Miles	Miles Saved	Passenger Cars			7-Ton Truck			50-Passenger Bus			30-Ton Truck Trailers		
				Old Road	New Road	Savings per mile	Old Road	New Road	Savings per mile	Old Road	New Road	Savings per mile	Old Road	New Road	Savings per mile
Morogoro + 38	38	38	0	8.2	6.4	1.8	14.0	9.2	4.8	20.1	13.1	7.0	40.0	25.4	14.6
Morogoro +38-Mikumi	37	37	0	9.4	6.4	3.0	17.6	9.2	8.[illegible]	24.8	13.1	11.7	52.0	25.4	26.6
Mikumi-Manenge	67	65	2	10.2	6.7	3.5	19.7	9.8	9.9	27.7	14.5	13.2	59.5	26.8	32.7
Mahenge-Kitonga	7	7	0	8.2	6.4	1.8	14.0	9.2	4.8	20.1	13.1	7.0	40.0	25.4	14.6
Kitonga Gorge	5	5	0	18.6	10.1	8.5	29.3	16.9	12.4	39.2	21.6	17.6	79.8	44.9	34.9
Gorge-Iringa	33	33	0	8.2	6.4	1.8	14.0	9.2	4.8	20.1	13.1	7.0	40.0	25.4	14.6
Iringa-Sao Hill	58	49	9	9.1	6.4	2.7	16.7	9.2	7.5	23.6	13.1	10.5	49.0	25.4	23.6
Sao Hill-Makumbako	58	49	9	9.7	6.4	3.3	18.5	9.2	9.3	25.9	13.1	12.8	55.0	25.4	29.6
Makumbako-Iyayi	30	27	3	10.7	6.4	4.3	21.1	9.2	11.9	29.4	13.1	16.3	64.0	25.4	38.6
Total	333	310	23												

[a] Costs derived from Table 3 after adjustment to reflect road alignment and road surface condition.

c) Half the value of savings of traffic *induced* by the road improvement was calculated as a benefit. For this it was necessary to estimate the price elasticity of demand for traffic within Tanzania. In a feasibility study of the Tanzam highway financed by USAID in 1966, the Stanford Research Institute found the elasticity to be relatively high. Elasticity is taken at 1; that is to say, generated traffic as a percentage of normal traffic (present traffic plus normal growth) is equal to the percentage saving in vehicle operating costs.

d) The reduction in costs of road maintenance, obtained from the consultant's estimate and from the Ministry of Communications, Labour and Works, was another economic benefit. Against this were set the costs of future resurfacing or repair of the new road.

The benefits and costs were appropriated to the years in which they are expected to occur and the internal rate of interest was found which equates the stream of benefits and that of costs discounted at that rate. The rate was 19.6 percent with and 13.9 percent without the Zambian traffic (see Tables 6 and 7). The expected life of the project was put in this instance at 20 years,

TABLE 6: Tanzania, Highway Project: Streams of Costs and Benefits, Best Estimates

(US$ million)

		Benefits	
Year	Costs	With savings due to probable Zambian traffic	Without savings due to probable Zambian traffic
1969	14.235		
1970	14.235		
1971	—	5.957	2.810
1972	—	6.314	3.163
1973	—	6.685	3.531
1974	—	6.948	3.791
1975	—	7.232	4.072
1976	—	5.806	4.226
1977	.513	5.268	4.478
1978	—	5.187	4.792
1979	1.107	5.349	5.151
1980	.878	5.650	5.551
1981	—	6.036	5.986
1982	1.323	6.483	6.458
1983	—	6.981	6.969
1984	.365	7.526	7.520
1985	—	8.118	8.115
1986	—	8.758	8.756
1987	—	9.449	9.448
1988	—	10.194	10.194
1989	—	10.998	10.998
1990	−.819[a]	11.866	11.866

[a] Residual value of the pavement.

TABLE 7: Tanzania, Highway Project: Expected Economic Rate of Return (*percent*)

	Section	A Local traffic plus full savings due to probable Zambian traffic	B Local traffic only
1	Iyayi-Makumbako	29.4	16.7
2	Makumbako-Sao Hill	36.4	22.7
3	Sao Hill-Iringa	15.1	11.4
4	Iringa-Mahenge (less Kitonga Gorge)	12.1	9.5
5	Kitonga Gorge	11.0	8.8
6	Mahenge-Mikumi	20.8	14.2
7	Mikumi-pavement end	11.4	9.3
8	Pavement end-Morogoro	16.5	13.8
	Iyayi-Morogoro (overall)	19.6	13.9

which determined the point beyond which no further calculations were made. But it would have made little difference to the return if the expected life had been taken at 30 or 40 years.

Finally the internal rate of return so calculated has to be compared with the estimated cost of capital, representing conceptually the opportunity cost of other alternative investments foregone by this decision to build a road. Certain refinements, such as identification of congestion costs or accident reduction[2] were not used in this study. Nor were shadow prices necessary.

The Probability Analysis

The problem

Due to uncertainty surrounding major variables of the economic evaluation, the selection of single values for these variables and use of them in the rate of return calculation could give a misleading impression. Among these uncertainties, one was outstanding.

The Governments of Zambia and Tanzania have agreed to construct a 980-mile rail link, with the financial and engineering assistance of mainland China. The date of completion of this railway is uncertain. If and when it is built, it is probable that most, if not all, Zambian seagoing import-export traffic will be diverted to the railway. For the probability analysis, it was assumed that from the date of the railway's completion, Zambian traffic would fall off by 50 percent per annum: but it was uncertain whether this would be as early

[2] They might have been, for there had already been over 100 deaths on the existing road since 1966.

as mid-1974 or in the 1980's. Within wide bounds, any hypothesis about the duration of the Rhodesian crisis and/or the likelihood or timing of the railway would be as good as any other. Under these circumstances any reliable traffic forecast was virtually impossible. Finally, though less importantly, the uncertainty about vehicle savings, maintenance savings and traffic counts could not be neglected. A probability analysis appeared an appropriate answer to the problems, and the cost-benefit analysis was accordingly put into a probability framework.

Since a probability analysis was to be made, a computer program was written for the calculation of the rate of return of the project. The inputs of the program as variables were the Bank staff's best estimates of project costs, the volume and composition of future traffic, vehicle operating and road maintenance costs with and without the project, the price elasticity of demand for transport, and the economic life of the assets; the output was a single rate of return for each of the road sections and for the entire road (Table 7).

The analysis

In the second step a sensitivity analysis was made of the rate of return. For this purpose, each of the parameters was varied, one at a time, and the corresponding values of the rate of return computed. Each parameter was tested by a 10 percent variation around the value which had been chosen in the determination of the initial rate of return, and for those parameters which had ranges of uncertainty estimated to be larger than ±10 percent of the initial estimated value, the rates of return for the two extreme values of the range were also computed. The results of this sensitivity analysis indicated which parameters had an important effect on the value of the rate of return, and allowed the number of variables to be considered in the probability analysis to be limited to these. The methodology of such a sensitivity analysis has been more fully described in the preceding chapter on the Mogadiscio case.

The third and final step in the analysis was the probability analysis proper, which was done by way of a simulation (Monte Carlo technique). For this purpose a probability distribution was estimated of the values for each of the parameters selected in the sensitivity analysis. These distributions reflected the Bank staff's judgments of the uncertainties with respect to data, and their judgment as to probable future developments. The rate of return was then computed repeatedly (300 times), using each time, for each of the selected parameters, a value drawn at random from the range of its probability distribution. The frequency of a particular value's selection is governed of course by its probability weighting. Care was taken to correlate interdependent variables. Finally, the distribution of the rate of return was statistically deduced from the sample of values obtained.

The hypotheses employed in the analysis are shown in Table 8. There is also a graphic presentation of the results in terms of the cumulative probability of earning *more* than a specific rate of return with and without Zambian transit traffic (Figure 6).

Looking in detail at Table 8, the reader will first notice that the engineers' normal contingency allowances have been removed from the best estimate

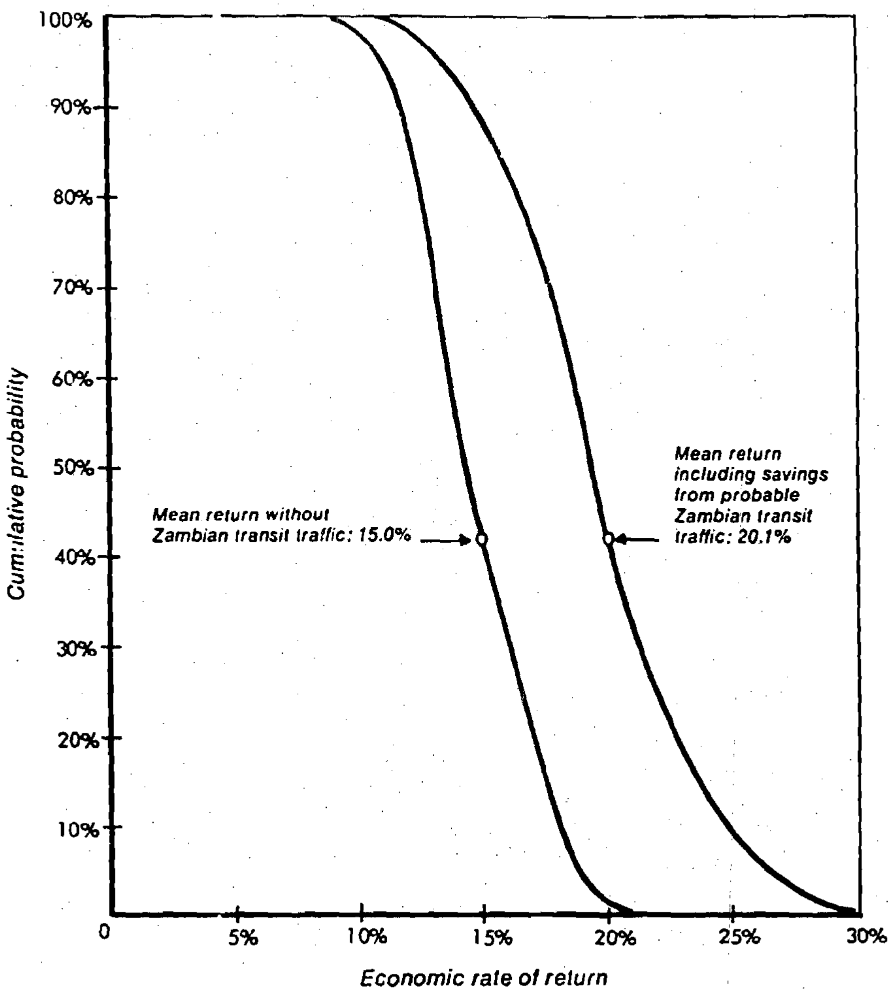


Figure 6. Tanzania, Second Highway Project: Probability Analysis of Economic Rate of Return

calculations and interpreted as a probability distribution. They are disaggregated into quantity overrun contingency allowances and bid overrun allowances. (Item A.)

In general, Tanzanian traffic on the Tanzam highway (Item B) falls off the further the distance from Dar es Salaam, except in the vicinity of major towns such as Morogoro and Iringa. The proportion of local truck and bus movements is very high at 60 percent and 75 percent, with the higher proportion on the less trafficked sections. The probability distribution allowed for doubts about the accuracy of the traffic count.

Local truck traffic was estimated to be growing at the rate of 6 percent to 10 percent on this road with automobile traffic growing more slowly at between 4 percent to 8 percent. (Item C.) Without probability analysis best estimates of 8 percent and 6 percent would have been used.

Transit traffic to and from Zambia in the period after completion of the road improvements, i.e. post-1971, was more difficult to forecast. (Item B.4.) Current transit traffic volumes and political conditions did not provide a clear guide for forecasting the needs of the future. The oil traffic which formed the bulk of Zambian import traffic would be diverted to the new pipeline. Copper export traffic over the road may be affected by the unbalanced loads that will result from diversion of the oil traffic. The buildup of goods import traffic and a balancing load of copper will depend both on the speed at which extra capacity now being provided at the port of Dar es Salaam comes into use and the future need to maintain transit traffic in this direction. Zambian import-export traffic levels will be affected by many factors, including the competitiveness and capacity of existing alternatives, the efficiency of the operation of the entire system to Dar es Salaam including trucking operations and commercial arrangements, and the actual growth of Tanzania's own port requirements as compared with the planned port expansion. These and related factors were assessed, insofar as possible, and translated into judgments as to the probable levels of annual traffic over the road during the period after 1971.

The period for which Zambian traffic is likely to utilize the improved road (Item C.4) was also uncertain, especially in view of the proposed parallel development of a rail link. This rail link would supersede the highway in the role of ensuring a reliable access to the sea for Zambia, and would divert most of this transit traffic from the road. If the Governments of Zambia and Tanzania pursued their present intention to build the railway, the earliest time when it could become operational would be mid-1974, i.e. after 18 months of engineering and about 5 years of construction. There was, however, the possibility that the railway might not be built prior to the 1980's, should the Governments of Zambia and Tanzania decide to postpone construction or find that the proposed improvement to the Tanzam highway would serve as an adequate

TABLE 8: Tanzania, Second Highway Project: Hypothesis for Probability Analysis

Item	Best Estimates	Nature of Uncertainty	Probability Distribution
A. *Engineering and Construction Costs* *Road Section*	(US$)*		
1. Iyayi-Makumbako	1,962,300	Quantity	Uniform between 0% and +15% on each item
2. Makumbako Sao Hill	3,614,700		
3. Sao Hill-Iringa	4,503,500	plus	plus
4. Iringa-Mahenge (less barge)	3,308,200		
5. Kitonga Gorge	1,162,000	Bid Uncertainty	50% probability 0% to 8% } overrun on total cost
6. Mahenge-Mikumi	7,[illegible]30,000		50% probability 8% to 15% } overrun on total cost[b]
7. Mikumi-End of Pavement	3,608,000		
8. End of Pavement-Morogoro	3,062,000		
B. *Traffic Count in 1967*			
1. Cars per section	See Table 4	Accuracy of traffic counts	Rectangular −20% and +10%
2. Trucks per section	See Table 4	Accuracy of traffic counts	Rectangular −20% and +10%
3. Truck trailers per section	See Table 4	Accuracy of traffic counts	Rectangular −20% and +10%
			Step Rectangular
			Probability / *ADT-range*
			5% / 0
4. Zambian import/export average daily traffic 1971	85	Capacity of Dar es Salaam Port	10% / 1–33
		Distribution of traffic over alternative routes.	20% / 33–66
			40% / 66–105
		Central African political situation.	20% / 105–120
			5% / 120–155
C. *Traffic Growth Rate*			
1. Cars	6%	Forecasting Uncertainty	Cars—uniform between 4% and 8%
2. Trucks	8%	Forecasting Uncertainty	Trucks and truck trailers—uniform between 6% and 10%
3. Truck trailers	8%	Forecasting Uncertainty	Trucks and truck trailers—uniform between 6% and 10%
			Discrete
			Probability / *Year in which traffic starts to decline*
			10% / 1974
4. Zambian import/export traffic			30% / 1975
Period 1968–1975	0%	Duration of traffic	10% / 1976
Period after 1975	−50%	Probability of Railway Development or other solution	7.5% / 1977
			7.5% / 1978
			7.5% / 1979
			7.5% / 1980
			20% / post 1980

D. *Elasticity of Traffic Demand*				
1. Cars	1.0		[c]	
2. Trucks	1.0			
3. Truck trailers	1.0			
4. Zambian import/export traffic	0			
E. *Savings on Vehicle Operating Costs* (*US$ per mile*)				
1. Cars per section	See Tables 3 and 4		Lack of data and uncertainty as to size of trucks in use.	*Old road* Uniform, complete correlation −12% to 15%.[d] In addition savings on operation costs of trucks are varied uniformly on a −5% + 10% range to account for uncertainty regarding truck size.
2. Trucks per section				*New road* Ignored, because variation in savings fully taken care of by variation in estimated operating costs on old road.
3. Truck trailers per section				
4. Zambian traffic per section				
F. *Road Maintenance* (*US$ per mile*)	Formulas: a + b (Traffic unit equivalent per day) = Maintenance cost per mile/p.a.			Uncertainty accounted for by varying variable term only
Old Road	*a* (*fixed term*)	*b* (*variable term*)		
Sections 1, 2, 3, 6 and 7	538.8	5.04	Divergence and lack of data	Uniform between $4.20 and $7.00
Sections 4 and 8	1,300.0	3.15		Uniform between $1.60 and $4.50
Section 5	1,685.0	4.00		Uniform between $2.50 and $5.50
New Road				
Sections 1, 2, 3, 4, 6, 7, and 8	1,225.0	1.20		Uniform between $0.50 and $1.50
Section 5	1,550.0	1.80		Uniform between $1.00 and $2.70
G. *Traffic Unit Equivalents for Maintenance Cost Calculation*	*Old Road*	*New Road*		
1. Cars	1	0	[c]	
2. Trucks	2	1		
3. Truck trailers	3	3		
4. Zambian traffic	3	3		
H. *Time and Cost of Required Pavement Strengthening*	*Cost* (*US$ per mile*)			
Overlay of 1.25 inch asphalt concrete	13,500		Unit costs	Rectangular ±25%
	Timing After 6 million passes of equivalent 18 kip axle loads in design lane		Traffic in future	According to aggregated distribution of traffic volume/duration/growth with each vehicle pass translated into equivalent 18 kip axle passes.

[a] Consultants' estimates, net of transfer payments and contingency allowances, and reflecting some adjustments by the mission.

[b] A very detailed breakdown of cost components and associated quantity and bid uncertainties for sections 6, 7, and 8, derived from discussion with the consultants was used for these three sections.

[c] A variable of low sensitivity or low range of variance. No variation according to probability within range considered necessary.

[d] The distribution is artificial in that it aims to reflect judgment as to unit savings rather than level of operating costs.

alternative for the foreseeable future. It was assumed that Zambian seagoing traffic would not increase beyond its peak in 1971 and that it would decline by 50 percent annually from the completion of the railway. The distribution shows best guesses of the commencement of this decline.

Unit savings in vehicle operating costs due to improvement will be large. The distances between the major origin and destinations on this road are also large. Thus, a large absolute fall in the cost of most vehicle trips should stimulate significant economic activity in the high potential agricultural areas south and west of Iringa. High transport costs caused by poor roads and long distances have been a significant barrier to the development of this fertile, temperate highland.

Except for trucks, the operating costs (and thence the savings from an old to a new road) will be completely correlated; i.e., it is estimated that if the costs for cars are 15 percent higher than estimated from the general data (see above), they will be 15 percent higher for buses, truck trailers or special traffic. In the case of trucks, lack of accurate data on the average truck size contributed additional uncertainty.

Road maintenance cost savings (Item F) were computed according to the formula

$$\text{Maintenance cost per mile p.a.} = a + b\,x$$

where x represents the projected average daily traffic, expressed in traffic units, and a and b are the parameters of a linear approximation of the road maintenance cost *within the range of traffic covered by this project.*[3] This is computed for the road with and without the road improvement, and the cost savings constitute a benefit. The variable term G only was varied in the probability analysis. It will also be noticed that the traffic unit, which in the case of the existing road was defined as a car, was changed to a truck in the case of the new road. In effect this implies that the passage of cars over the new (asphalt) road is estimated to have a negligible effect on the cost of the road's maintenance.

The final item H relates to the cost of pavement strengthening, estimated at $13,500 per mile (± 25 percent) and its timing. These are discussed below. Note that this cost is directly related to the corresponding benefits (the traffic use of the roads).

Results of the analysis

The overall results of the best estimates are shown in Table 7, and of the probability analysis in Figure 6. Table 7 sets out the rates of return based on the best estimate of each variable. In column A the expected benefits to the Zambian through traffic are included; in column B they are excluded. Com-

[3] See Figure 9.

parison of columns A and B shows that the rate of return is only moderately sensitive to future Zambian traffic level and duration. The overall rate of return to Tanzania is thus about 14 percent. In addition, Tanzania would probably receive a share of the savings likely to accrue in respect of the Zambian transit traffic. The rates of return for each section individually are generally satisfactory, even though from the Tanzanian point of view the return becomes marginal in the neighborhood of the Zambian border.

The probability analysis (Figure 6) indicates that while there is a high degree of uncertainty with respect to some of the major variables, there is a relatively small risk that the project will not yield a satisfactory return. The analysis showed that there is less than 5 percent probability that the project will earn less than a 10 percent return for Tanzania—i.e. more than 95 percent probability that the rate of return will be over 10 percent. The mean returns under probability analysis, 20.1 percent with the Zambian traffic and 15 percent without, are marginally higher than those found by a best estimate calculation, 19.6 percent and 13.9 percent, in Table 7.

The Problem of the Road Surface Thickness

The uncertainty about the traffic level for which the road was to be designed made the choice of an asphalt surface thickness difficult. Since the cost-benefit analysis was already in probability form, it was only necessary to feed in the alternative thicknesses of surface and their corresponding expected lives to the probability analysis, to evaluate which alternative was best.

The problem of optimization is one which is met frequently in Bank projects, and stems from the different perspectives of those who tend to seek the best, i.e. "safest" engineering solutions, and those who seek to optimize the use of scarce resources. In the Tanzanian case, two issues were at stake. The first was the strength of the road, i.e. the number of vehicle passes (of an equivalent standard 18 kip axle) for which the road should be designed. The lower the strength, the sooner the pavement was likely to wear out and to need replacement. The second issue was the manner in which the design strength should be achieved, through the base or through the pavement surface. An asphalt concrete surface of minimum (¾") thickness and a thick base were theoretically equivalent to a 2" asphalt concrete surface and a less thick base, but there was large uncertainty as to what the cost of laying a pavement of ¾" would be. The second alternative was being pressed by the consultant though its initial cost was admittedly higher.[4]

[4] A third alternative, a double-seal bituminous surface course, as originally specified, was not considered because its costs appeared very slightly higher than the minimum thickness ¾" asphalt concrete alternative.

A cost analysis was therefore conducted for various strengths and for the two extreme alternatives, with the relevant maintenance and replacement costs discounted to present value. The results are shown in the following table:

Expected (Mean) Present Value of Cost Per Mile at 10 Percent Discount Rate (US dollars)

Number of vehicle passes	125,000	150,000	175,000	200,000	250,000	300,000
Structural strength:						
¾" asphalt concrete combination	27,335	27,744	28,434	29,109	30,317	31,557
2" asphalt concrete combination	30,499	30,736	31,292	31,845	32,726	33,878

Note: Costs include initial construction cost for sub-base and surface course, maintenance costs and strengthening costs, less salvage value.

This exercise showed first that at all strengths, the ¾" asphalt concrete combination was economically preferable to the 2" one. The difference between the two was greatest at the minimum design strength of 125,000 passes, but this was rejected because the probability of the need for a strengthening as early as 1978/79 (regardless of surface course chosen) was felt to be too high. The small cost difference between this alternative and the 150,000 pass alternative appeared a reasonable cost to pay for the additional security on that score. It was estimated that the minimum cost acceptable solution thus identified at 150,000 passes strength would be on the average $1.3 million less expensive to build than the equivalent 2" pavement, and about $750,000 less expensive on a present value basis at a 10 percent rate of discount.

However, quite properly, engineering consultants have a professionally high aversion to risk, and this was heightened in this case by their relative lack of experience with lesser thicknesses of asphalt concrete. Thus the engineer's judgment was that additional risk was attached to the ¾" alternative. Even though the *mean* expected present value of cost had been shown to be less, this would not therefore be convincing if there were a *fair chance* of the cost of a less thick surface proving more expensive in the long run or of the cost difference being so small as not to be worth the trouble of optimizing.

Fortunately the probability distribution of Figure 7 shows that there was a nil probability, according to the assumptions that had been technically agreed,

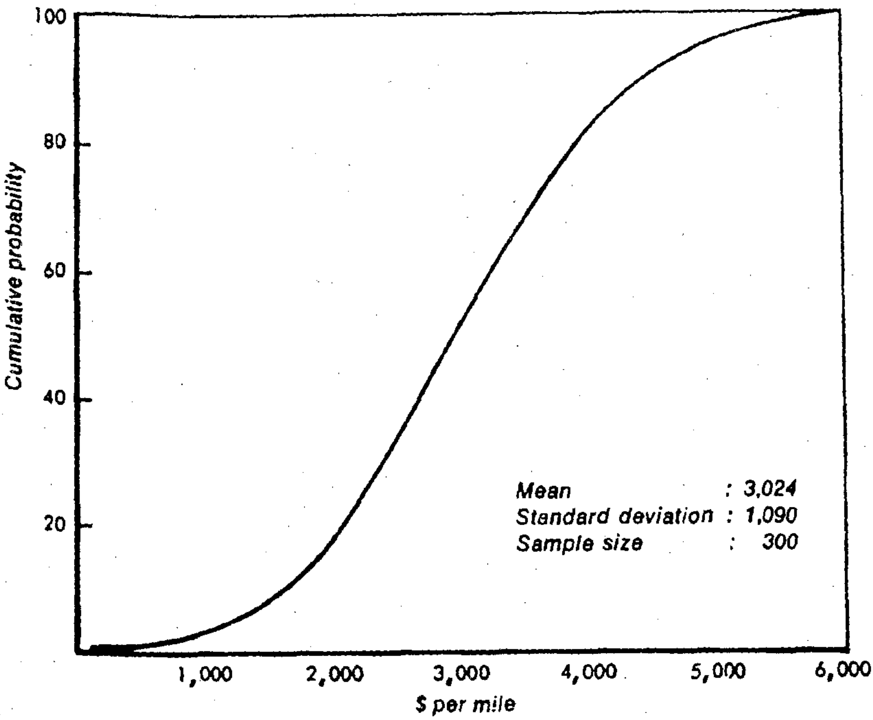


Figure 7. Probability of Cost Difference between ¾″ and 2″ Asphalt Concrete Surface Course Solutions

that the 2″ alternative would be cheaper, and over an 80 percent chance that the extra cost would be only $2,000 per mile in present value. The mean expected extra cost was $3,024 per mile. The Bank suggestion therefore appeared worth considering.

The final outcome was a compromise. The Bank financed section as well as the other sections of the road were designed to a 175,000 passes strength and a 1½″ asphalt concrete thickness.

V

DISAGGREGATION

This chapter and the next discuss the problems arising from the choice of a level of aggregation and the evaluation of correlation, which create the major difficulties of risk analysis and whose solution is a critical condition for the validity of the results.

Aggregation

By level of aggregation we mean the degree of detail which the analysis encompasses. For example, in the cost of a road, the costs of land clearance, earthwork, base, sub-base, and pavement can be distinguished. The cost of the base can be further subdivided into the costs of extracting stones, crushing them, transporting them, and laying them, and each of these stages can also be broken down. Where to stop subdividing in order to make the best risk analysis is the aggregation problem.

Our experience indicates that risk analysis calls for more disaggregation than usually is used in Bank project appraisal and that the smaller the component, the easier it is to formulate a judgment, though there is clearly a limit to this rule. To illustrate we cite the case of the Great East Road.

Great East Road Project

The proposal to pave the 64-mile Luangwa-Nyimba section of the Great East Road, which leads from the line-of-rail in Zambia to Malawi, was presented to the Bank at the same time as that of upgrading the Tanzam highway to Tanzania. The existing traffic level was known to be low, and the traffic growth rate only normal. There was a 25 percent possibility that the road might be needed as a spare carrier for Zambian export-import traffic. In that case an additional ADT of about 35 was expected.

However the dominant uncertainty in this case was the cost of the project. The detailed engineering study had not been completed at the time of appraisal and estimates ranged widely from £1.5 million to £2.8 million. The higher cost would have entailed a negative return. But a decision had to be made whether to go ahead with the project, to postpone it or even to discard it.

Accordingly the opportunity was taken to employ risk analysis, in which the chief uncertainties revolved around the technical elements of cost. The fact-finding advantages of a probability analysis in such a case soon became apparent. When the consultant was asked to give a single cost estimate, he was unwilling to do so before completing more detailed engineering work, since the figure might have been quoted against him later. Also, not being familiar with this method, he was at first unable to give variations around an estimate.

The discussion then backtracked to subjects like cost of cement, thickness of base, amount of earthworks, etc., and the attitude of the consultant changed. He had designed many other roads and, on the basis of the preliminary soil survey, it was easy for him to guess what base would be needed on this particular road. He knew that economic considerations would permit only a minimum realignment of the existing road and consequently was able to form a judgment as to how much earthmoving work could be dispensed with—and so on. By dealing with the components separately—by disaggregating—it was possible to obtain a range of cost estimates and a full probability distribution over this range—which at first sight had appeared impossible. The full details of the information we obtained is given in Table 9, and there is further discussion of the analysis in Reutlinger's "*Techniques for Project Appraisal under Uncertainty*," which is Occasional Paper No. 10 in this series.

The result of the analysis on very preliminary cost estimates indicated that if the proposed construction were to begin after the completion of engineering preparation, the mean expected rate of return would be less than 5 percent, and there was only a very low probability that the project would earn more than a 10 percent return. If construction started in 1974, the result was only slightly better. As a result it was agreed that the Bank should not at that time consider financing that part of the road, though it would be prepared to reconsider the project on receipt of further study.

Advantages of Disaggregation

Incomplete or inaccurate judgment often results from a lack of disaggregation. Vehicle operating costs and road maintenance costs in the risk analyses on

TABLE 9: Zambia, Great East Road: Hypotheses Made in the Probability Analysis of the Economic Rate of Return of the Proposed Bituminous Paving of the Luangwa-Nyimba Section

	Item	Best Estimates	Nature of Uncertainty	Probability Distribution
	Construction Costs	*(kwachas)*		
1	Pavement base	466,000	Price and quantity	Discrete: *Probability (percent)* — *Total Cost (kwachas)* 48 — 579,000 (6″ base) 52 — 466,000 (5″ base)
2	Sub-base and shoulders	311,150	Price, quantity and thickness of base required	(A) *Basic Cost:* Step rectangular distribution reflecting the uncertainty about the salvage value of the existing road *Probability (percent)* — *Cost within Range (kwachas)* 30 — 219,200-311,150 50 — 311,150-342,000 20 — 342,000-561,165 (B) *Final Cost:* Correlated to cost of pavement base (Item 1 above) (i) If cost of base is K 579,000 then: *Probability* — *Final Cost sub-base & shoulders* 42% — Equal to Basic Cost (2A) 51% — 80% of Basic Cost 7% — 60% Basic Cost (ii) If cost of base is K 477,000 then: 88% — Equal to Basic Cost 12% — 80% of Basic Cost

Table 9. Cont.

	Item	Best Estimates	Nature of Uncertainty	Probability Distribution
	Traffic Growth	*(annual rate) (percent)*		
11	Cars	6	Forecasting Error	Uniform on range 4%–3%
12	Buses	6	" "	" " " 4%–8%
13	Trucks and truck trailers	8	" "	Uniform on range 6%–10%. Growth of truck and truck trailer traffic is fully correlated
14	Special Traffic:			
	Period 1968–1972:	0	—	—
	Period 1972 on:	−20	—	—
	Elasticity of Traffic Deman..!			
15	Cars	.75		
16	Trucks	.75		
17	Buses	0		
18	Truck trailers	.75		
19	Special traffic	0		
	Vehicle Operating Costs (per vehicle mile)	*(kwachas)*		
20	*Old Road*			
	Cars	0.0613	Lack of data	Uniform on a −12% + 15% range; all fully correlated.[a] In addition operating cost of trucks is varied uniformly on a −5% +10% range to account for uncertainty on size of trucks.
	Trucks	0.1076	Lack of data and size of trucks	
	Buses	0.1516	Lack of data	
	Truck trailers	0.215	" " "	
	Special traffic	0.215	" " "	
21	*New Road*			
	Cars	0.0479		
	Trucks	0.0670		
	Buses	0.1034	—	b
	Truck trailers	0.140		
	Special traffic	0.140		

	Maintenance (Formula: $a + b$ (Traffic units per day) = Maintenance Costs per annum)			
22	*Old Road*	(*kwachas*)		
	Fixed term Variable term	417.0 = a 3.6 = b	Divergence of existing data	Uncertainty accounted for by taking variable term uniformly distributed between K 3 and K 5
23	*New Road*			
	Fixed term Variable term	600.0 = a 2.25 = b	Divergence of existing data	Uncertainty accounted for by taking variable term uniformly distributed between K 1 and K 2.5
24	***Traffic units for maintenance calculations***			
	Cars	1		
	Trucks	2		
	Buses	2		
	Truck trailers	3		
	Special traffic	3		
	Others			
25	Life of road (years)	20	Incomplete analysis	Triangular on range 12–25 years
26	Length of road	64 miles		
27	Year construction starts	1969		
28	Construction time	2 years		
29	Distribution of cost over construction period:			
	1st year	50 percent		
	2nd year	50 percent		

[a] This distribution is artificial and is only geared at getting a correct distribution of the savings from the improvement of the road.

[b] Ignored because variation of savings is fully taken care of by variation of operating costs on old road.

road projects provide examples. Lack of time prevented us from disaggregating and consequently we had great difficulty in choosing probability distributions which reflected our uncertainties. Clearly, if we went into the details of the components of vehicle operating costs, the problem would become simpler because we have had experience with fuel consumption, tire wear, depreciation, etc. The same applies to road maintenance costs because we have an idea of the amount of gravel necessary to maintain a gravel road, the number of gradings required per year, etc., and we can also find ways to express the uncertainties resulting from the weather, the geometry of the road and other such factors which are difficult to introduce into a maintenance formula. The choice of an appropriate level of aggregation, therefore, appears to be an essential condition to the expression of a clear judgment.

PART II

VI

CORRELATION

The Importance of Correlations

The example given in the previous chapter seems to indicate that, in many cases, the more disaggregation the better. Unfortunately there is a limit to disaggregation because of the problem raised by correlation. The problem was touched on in the sensitivity analysis of the Mogadiscio port project (Chapter III), where there was expected correlation between the productivity of a banana gang and the productivity of a general cargo gang, and correlation between the number of men employed in a transit shed and that in a warehouse. Correlated variables, or in simple terms, variables which are likely to vary together in a systematic way, appear in every project. However, an experienced professional may feel he is familiar with two separate variables and knows how they are related, without being able to describe their correlation—how their variations are related. Correlations are difficult to detect, and even more difficult to measure, but overlooking them may lead to a completely wrong interpretation in the analysis.

The Mogadiscio case gives an idea of the importance of correlations. We initially neglected the correlation between productivity of labor and port capacity. The probability that the project would earn less than 10 percent in this case was 15 percent. After we introduced this correlation, the probability rose to 40 percent, i.e. it almost tripled. In another project (a telecommunications project in Malaya) the standard deviation of the rate-of-return distribution without allowing for correlation between two variables was about .35. With the appropriate allowance for correlation it rose to about 1.1. Since in this case the rate-of-return distribution was practically a normal distribution, this change meant that the probability that the rate of return would fall outside of a ±1 percent range in absolute value around the best estimate rose from 5 percent to about 40 percent. The consequences of mishandling the correlations are so serious that they can eventually lead to the wrong decision.

It is easy to understand the way correlation works. When independent variables are aggregated, the effect of the variation of one may be compensated by the variation of another one in an opposite direction. If they are positively correlated, the effect of the variation of one will always be aggravated by the variation of the others. If productivity of labor and capacity of the Mogadiscio port are independent of each other, a low value of the productivity of labor may be compensated by a high value of the capacity of the port. If they are positively correlated, as is likely to be the case, the effect of a low productivity of labor on the economic rate of return is heightened by the low capacity of the port, and the probability of getting a low rate of return will be higher than if they were independent. Correlations can also be negative, that is, the variables may systematically compensate each other. However, in the type of projects we have worked with this occurs less frequently than positive correlation.

Correlations are difficult to detect. The first reason is that they do not have to be taken into consideration in the single point estimate method and are therefore not familiar to most people. Let us take the port of Mogadiscio case. An engineer generally knows that in a port which operates normally well he can expect that the productivity of labor will be around 10 tons per gang-hour and the capacity around 700 tons per linear yard of berth. He also knows from his experience how much productivity can vary from port to port and under different situations, and likewise for port capacity. But he tends to think of the port as an organic whole, rather than to analyze its functions; to ask him how the variations of capacity relate to the variations of productivity is to ask the engineer a question he does not usually ask himself. Therefore it is usually difficult to get an answer, and if the question is not asked, there is a good chance the engineer will not notice and the correlation will be overlooked.

The second reason that correlations are difficult to detect is that they are often hidden. Some correlations, particularly those which relate to engineering specifications, are not too difficult to identify. For example, the strength of a road is given by the thickness of sub-base, base and pavement. The engineer's uncertainty about each of these three parameters is tied to his uncertainty about the other two, and he will therefore recognize their correlation. Items 1 and 2 of Table 9 illustrate this point in the case of the Great East Road in Zambia, and show the hypotheses which grew out of discussion with the consulting engineer.

Other correlations are more difficult to spot and to assess. For example, vehicle operating costs are an important element in road projects. We usually make the distinction between operating costs for different types of vehicles and, *a priori*, if we have underestimated the operating cost of cars, there is no reason why we should also have underestimated the operating cost of trucks. However, if we have underestimated the operating costs of cars, it may be because we have overestimated the quality of the road. It is, therefore, likely that this will affect operating costs of all vehicles in the same direction. Item 20 of Table 9 shows that we decided to treat as fully correlated the operating costs for all vehicles except trucks (we were not sure of the composition of the truck fleet). Another example is the uncertainty about the amount of work required to build a road. If the road has been designed by one person and if this person has overestimated the amount of earthwork, is it because he has a systematic tendency to overestimate—in which case he will also probably have overestimated all the other elements of the road—or is it just by chance? We have assumed that the latter is more likely than the former, or at least will explain a greater part of the variations in the amount of earthwork, but the question is open for discussion.

The problem of correlations should therefore be approached with great care. However, its solution is not impossible and, while we may not yet have mastered it perfectly, the following points can serve as a guide.

Suggestions of Ways to Meet the Problem

Limitation of the disaggregation

To limit disaggregation is to solve the problem of correlation by eliminating it. If we work with the total cost of a road, we do not have to worry about the correlation between the cost of the base and the cost of the sub-base. The distribution we shall use for the cost will implicitly include this relation. However, as we have indicated, there is a limit to the level of aggregation which is feasible in obtaining probability distributions. Therefore the choice of the level of aggregation requires a trade-off between the advantages of clarity of judgment and of avoiding the hazards of disaggregation. It is a difficult choice and one often guided by the availability of time. Because we believe that the influence of correlations on the outcome of the analysis is more important than the influence of the shape of any particular distribution, we have usually opted for as little disaggregation as possible. The distribution of vehicle operating costs referred to earlier is a case in point. For the same reason, all the variables used in the case of the port of Mogadiscio project are rather highly aggregated.

Isolation of the sources of uncertainties

Limiting disaggregation can be considered only as an emergency measure in dealing with correlation. The advantage of risk analysis, after all, is that it permits disaggregation, and we want to retain this advantage. We found that it helps to think not so much in terms of disaggregating the technological components of the project, but in terms of disaggregating the sources of uncertainty. Let us again refer to the case of vehicle operating costs for the Great East Road in Zambia. To compute the rate of return we have to distinguish between the operating cost of cars, trucks, and truck trailers; this is a technological disaggregation. For the purpose of the risk analysis, we may ignore the technological distinctions *per se* and think in terms of the sources of uncertainty on these vehicle operating costs. We may distinguish three essential and independent sources of uncertainty: errors in the general data on which we based our estimation, errors in the way we have extrapolated these data to the particular case of the Great East Road,[1] and uncertainty about the average truck size. By so defining the uncertainties, it is easier to assess the correlations. We have treated as fully correlated for all types of vehicles the uncertainties resulting from the first two sources, because the data for all vehicle types originated from the same sources, and also because we thought the condition of the road would affect operating costs of all vehicles in the same way. But the third source of uncertainty affects only the operating cost of trucks. Variations of the vehicle operating costs resulting from this particular source of uncertainty should not be treated as correlated with others.

Applying the same line of reasoning to the analysis of the uncertainty on the cost of a project, it may be useful to distinguish between quantity uncertainty, unit cost uncertainty, and bidding uncertainty. In making this distinction in the case of the Tanzam highway, we considered all the unit cost uncertainties as independent and, except for the technological correlations, also all the quantity uncertainties. We then allowed for uncertainty about bidding on the total cost of the project, i.e. we assumed that this uncertainty would affect the cost of all the components of the project in the same direction. As a result, our

[1] Tables usually give vehicle operating cost for a typical earth road or a typical gravel road. The roads the Bank considers for financing, before they are improved, are always something of a cross between the two.

uncertainty about the actual unit price of each component of the project is explained to a very small extent by uncertainty about the economic unit cost, and to a much larger extent by the uncertainty about the outcome of the bidding process. When possible, this isolation of independent sources of uncertainty seems to be the easiest and most rigorous way of handling correlations, but in many cases we have also had to rely on the following approach.

The pessimistic-optimistic approach

Suppose we suspect some correlation between two variables but we cannot quantify its exact effect on the distribution of rates of return to a project which has tended to look favorable in the analysis to date. We can often reinforce our confidence about the project by examining a pessimistic view of the suspected correlation. If we find the project still acceptable, even in this light, then we can feel our confidence in the project has been justified. Conversely, if to date a project has tended not to look good we may be able to reassure ourselves that we are justified in rejecting it by examining an optimistic view of the effect of the suspected correlation. If the project still looks unfavorable, even with all the benefit of the doubt, our doubts are confirmed. Of course, if the reverse comes to light and the favorable project looks bad under pessimistic correlation assumptions or the unfavorable project looks good under optimistic assumptions, we must try some other route.

For each case, we have an example where this approach worked for us: in the first case, the Mogadiscio port project, and in the second, the Great East Road. In the Mogadiscio case, we were concerned with the correlation between productivity of labor and port capacity. The project had looked good to date, so we looked at the pessimistic case: the complete dependence of the two variables. Curve 1 of Figure 8 illustrates the result we obtained. Also indicated in the figure is the curve if the two variables were completely independent (Curve 2).[2] We know that the true curve (3) lies somewhere in between.

The decision to accept a project is based—among other things—on the probability of having more than a 10 percent return, if 10 percent represents the opportunity cost of capital. If we are sure in the Mogadiscio case that the project is acceptable under the assumption of complete dependence of one variable on another, it will be even more acceptable under the true assumptions. In the Mogadiscio case, 10 percent fell at A; this is still a favorable result. In the case of the Great East Road, a project which did not look good,

[2] The figure as here drawn and described applies to the correlation of the variables mentioned in the port of Mogadiscio case and the Great East Road case; it is not a universal diagram. In fact, the positions of the curves of complete dependence and independence as here drawn may be reversed in different cases or using different variables.

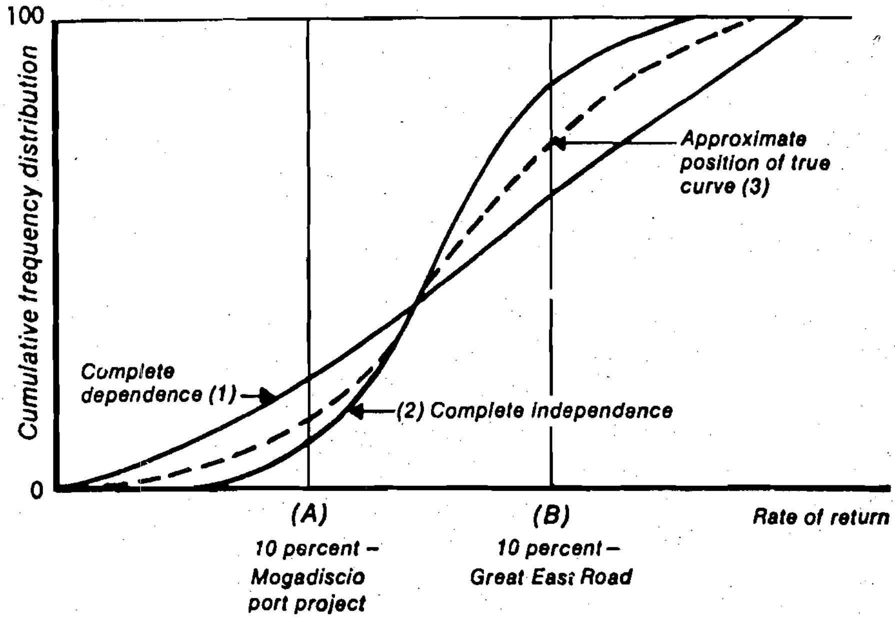


Figure 8. Cumulative Distribution Functions with and without Correlation of Variables

we were concerned with the correlation of vehicle operating costs. We made the optimistic assumption and got the same picture as in Figure 8, except that 10 percent fell at B, still a poor result. Thus, introducing a pessimistic correlation hypothesis has not damaged the Mogadiscio case; and introducing an optimistic correlation hypothesis has not helped the Great East Road case.

We mention this approach for what it may be worth. We have used it rather extensively because we did not have time to handle the problem more accurately and because it was the best method we could think of. It turned out to be practical because the models we were using were simple and it was easy to anticipate the consequences of various correlation assumptions. In the future, however, our models will undoubtedly become more and more complicated, making it difficult to say whether an assumption is more or less optimistic. Let us hope that at the same time we will improve our understanding of the correlation problem as it affects working practice and be able to handle it more rigorously.

Collection of more data

An essential step toward a better understanding of correlation is to make a serious effort to collect more data. We have noticed the great difficulty, and sometimes impossibility, of making subjective judgments about correlations. We have imputed this to a lack of experience on a problem which is new to us,

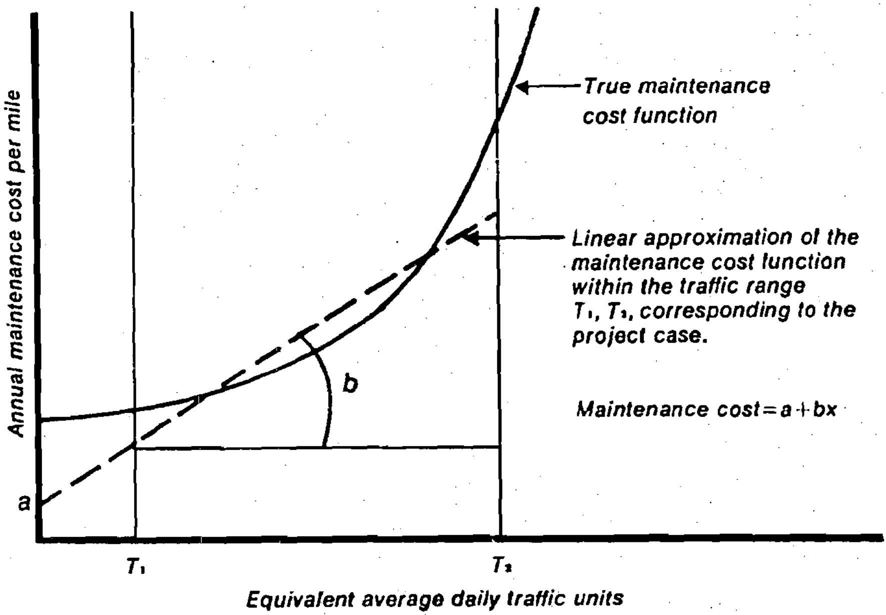


Figure 9. True Maintenance Function and Linear Approximation

and for which we must somehow develop a feeling. Whether we do it by way of statistical analysis or more empirically, we need data to acquire this feeling. In the case of the Mogadiscio project the availability of port data enables us to check a number of possible correlations. Besides the correlation between imports and exports, which we neglected because it was not sensitive, we suspected some correlation between the productivity of labor and GNP, or anything which would reflect improvement of living conditions in Somalia. Fortunately, at least for the sake of the mathematical analysis, available data indicated that there is none. In the case of roads we have not been so lucky. As an example, consider the case of the road maintenance cost (M), which we usually approximate by a linear function of the type:

$$M = a + bx$$

where x represents the projected average daily traffic (see figure 9). This formula is surrounded by uncertainties which affect both the constant coefficient a, and the variable coefficient b. The uncertainty about these two coefficients is likely to be correlated: if a is overestimated, b is probably underestimated, and conversely. In the absence of more reliable data on road maintenance cost this correlation has proven to be very difficult to assess. The only way to do better is to collect and analyze data, a relatively easy assignment in this case. In other cases it will be more difficult.

VII

THE CHOICE OF A PROBABILITY DISTRIBUTION

Choosing probability distributions for the variables is what seems to worry most people about risk analysis, possibly because they think it requires finding the true distribution of each variable. This is indeed quite impossible, though ways to improve the drawing-up of distributions through team evaluation and Bayesian approaches have been studied.[1] The aim of risk analysis is more modest. Risk analysis does not aim to give the exact true distribution of the rate of return, that is, the distribution we would obtain if we were omniscient rather than human beings, but rather the one which best represents the judgment of an appraisal team. Therefore, it is not a question of finding the true distributions of the input variables, but for each variable the distribution which best expresses the judgment of the appraiser. The distribution corresponding to a vague judgment will be as appropriate and useful as the one corresponding to a detailed judgment.

We have already mentioned in the Mogadiscio case two ways of obtaining probability distributions, one which we have called the portrait approach and the other which leads to the step rectangular distributions. Bank appraisal missions have now practically abandoned the portrait approach. When participating in this approach the appraiser tends to accept any smooth distribution. Possibly he is aesthetically influenced by the deceptively attractive appearance of the smooth curve, and impressed by the complicated formulas. His judgment

[1] See Robert L. Winkler, "The Consensus of Subjective Probability Distributions," in *Management Science*, Vol. 15, No. 2, October 1968, and its bibliography (19 articles).

seems to lose its sharpness, and in the end the approach means more work for fewer results. Therefore, as often as possible the second approach is used which leads to step rectangular distributions. Attempts have also been made to supplement it by using distributions which would fit cases in which not enough information is available to obtain a good step rectangular distribution, but in which information would be wasted by using a distribution which failed to discriminate between the likelihoods of any two values on a given range. The need felt by the appraisers for such distributions is a good illustration of what seems to be the main objective of the probability distribution choice, namely, to make use of all information available but not to require more information than is, in fact, available. The various distributions we shall now review, all of which have been used, are precisely geared to making the maximum use of available information.

The Step Rectangular Distribution

We have already described in detail in the Mogadiscio case how to obtain the distribution shown in Figure 10. This distribution is an attractive one for a number of reasons. In the first place it takes explicit advantage of the fact that the quantification of subjective probability judgments, in both theory and practice, is based on preference ranking. It also has the advantage that it can be drawn up by the appraiser himself. He has the freedom to choose whatever intervals he wants and to divide them into as many sub-intervals as he wants. This complete freedom of initiative, which he lacked in the case of the portrait approach, seems to help him considerably in the expression of his judgment.

In use, this distribution has proven astonishingly reliable: when the data generation process has been repeated for several distributions after a period of

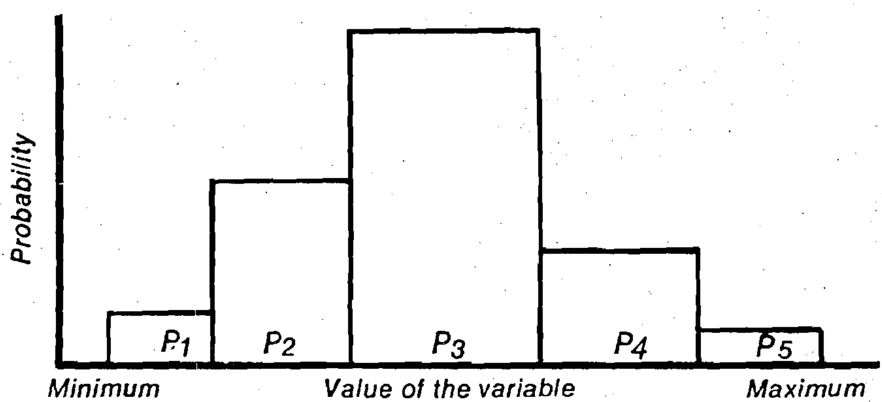


Figure 10. The Step Rectangular Distribution

time, it has usually come up with the same or a very similar result. It is also a distribution which fits well with the rule of using all the information available but not requiring more. If the appraiser thinks that he can express more accurately a judgment he has just made, he can sub-divide intervals one step further and create a more detailed distribution. If, on the contrary, he thinks that he will be guessing to say that one value in an interval is more probable than another, he may stop sub-dividing. Furthermore, this distribution lends itself well to the final review and polishing, described as the fourth and last step in the process described in Figure 2.

We have asked ourselves whether it is useful to smooth out this distribution before beginning the risk analysis. In many cases a continuous distribution would appear better fitted to the type of judgment we wish to express. But there are difficulties. The final outcome of our smoothing has to be a distribution from which random numbers can be easily generated. But by trying to improve the presentation of the judgment we want to simulate, we may end up with a less useful distribution. First, it may be difficult to find a smooth distribution which is close to the step distribution. Even if we do find a continuous distribution approximating well the one we start with, the improvement we gain may not be worth the trouble we may run into in the generation of random numbers. All computer random number generators start from uniformly distributed numbers; for the type of distribution we described in Figure 10 this generation is extremely simple. On the other hand, most other types of distribution require mathematical transformations which are often difficult and usually time-consuming. Therefore, the extra accuracy which can be obtained through smoothing is usually not worth the supplementary work which it requires.

The Discrete Distribution

The discrete distribution is very similar to the step rectangular distribution. The only difference is that the probabilities P_1, P_2.... etc., of Figure 10, instead of being assigned to a range, are assigned to one value only. This distribution is obtained in the same way and has the same properties as the step rectangular distribution. We have used it when the variables we were considering were, by nature, discrete variables—for example, in the Tanzam highway case, the year in which the Tanzam railway might come into operation.

The Uniform Distribution

From the point of view of information availability, the uniform distribution covers the opposite case from the step rectangular distribution. It is used where judgment is very vague and the appraiser is not able to differentiate between

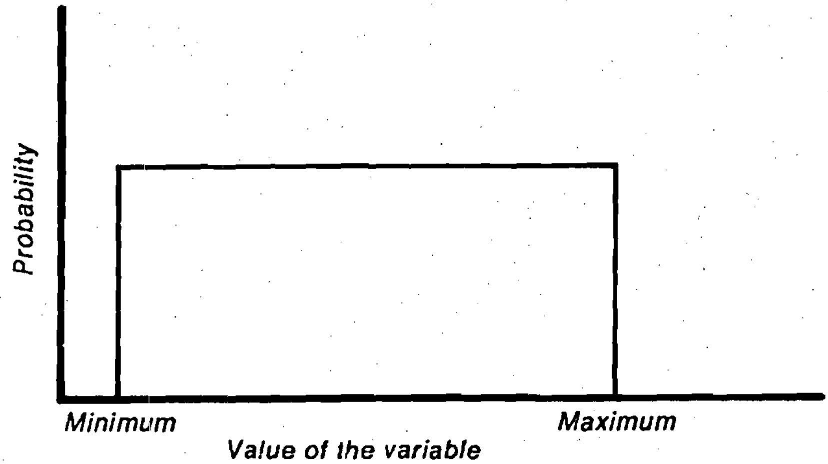


Figure 11. Uniform Distribution

any two values within the range of the variable. It is shown in Figure 11 and can be viewed as a particular case of the step rectangular distribution, with only one sub-range. As far as is possible, this distribution should be avoided. It is almost contradictory to suppose that a project with an equal chance of costing anywhere from $10 million to $15 million will under no circumstances cost $9.9 million or $15.5 million, which is what we assume when we say that the cost of a project is uniformly distributed between $10 million and $15 million. We have therefore used the uniform distribution only in the case of low sensitivity variables or whenever we wanted, to be on the safe side, to overestimate the probability of the extremes of the variables' range.

The Beta Distribution

The Beta distribution is the first distribution we tested to fill the gap between the step rectangular distribution, for which detailed information is needed, and the uniform distribution, for which minimal information is needed. Figure 12 shows its appearance. Use of the Beta distribution was suggested by the wide use made of the Beta distribution in the PERT system (Program Evaluation and Review Technique). The Beta distribution is entirely defined, if in addition to its range, one fixes two parameters. The literature on PERT[2] suggests use of the mode and a standard deviation

[2] See for example D. G. Malcom, J. H. Roseboom, C. E. Clark and W. F. Fazar, "Application of a Technique for Research and Development Program Evaluation," *Operations Research*, Vol. 7, pp. 646–669, 1959.

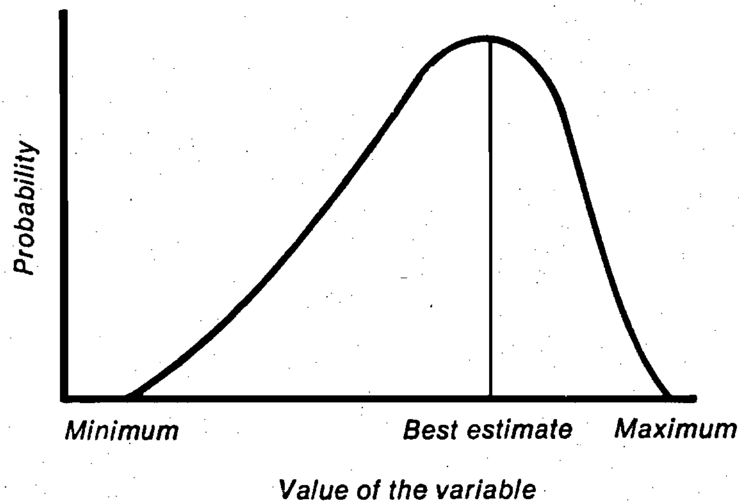


Figure 12. Beta Distribution

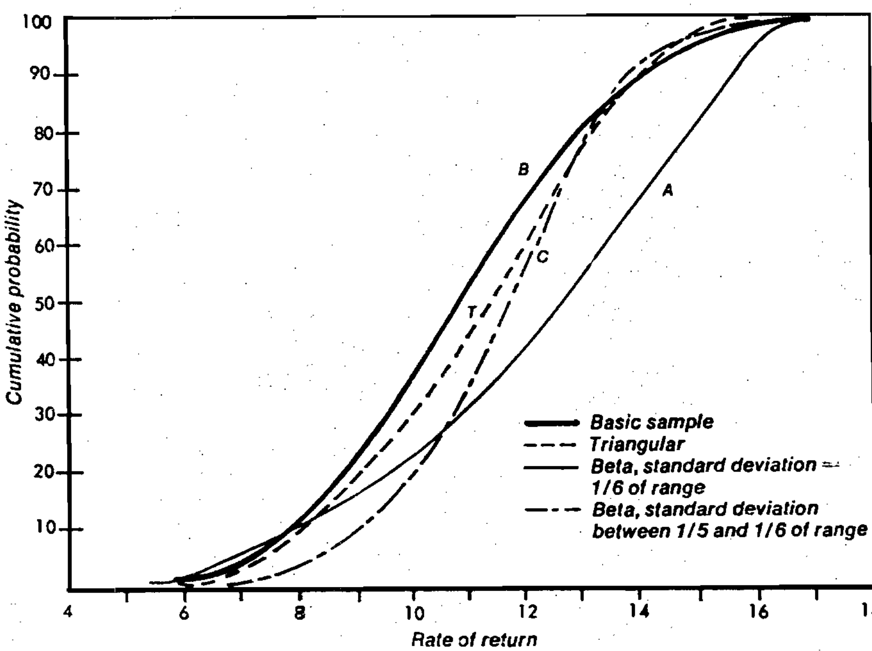


Figure 13. Mogadiscio Port Project: Substitution of Triangular and Two Different Beta Distributions for the Original Distribution of the Productivity of Labor

equal to ⅙ of the range. We have compared the results of substituting this distribution for the step rectangular distribution of the productivity of labor in the Mogadiscio case. The result, shown in Figure 13 (Curve A), indicates that this particular Beta distribution is a bad choice; it is far from being close to the basic sample distribution. We have not investigated further the use of the Beta distribution defined in this way because we think that it relies too much on the value assigned to the best estimate. In our experience the best estimate is not a reliable datum, and in practice it often is an imprecise mixture of the value with the highest probability and the mean.

To find a way to limit the influence of the best estimate, let us reconsider the case of the port of Mogadiscio. The step distribution in Figure 13 shows that all values between 8 tons per gang-hour and 10 tons per gang-hour have the same probability. The best estimate (m) could therefore be anything

TABLE 10: Degrees of Freedom of Selected Beta Distributions of Range 0–1

P = probability of getting less than the best estimate

P	Degrees of Freedom	m (best estimate) .55	.60	.65	.70	.75	.80	.85
	a	3.0	4.5	1.5	4.0			
.50	b	2.5	3.0	2.5	2.0	—		
	a	4.0	4.0	4.0				
.55	b	3.5	3.0	2.5				
	a	4.0	3.0	3.5	4.5	4.0	4.0	
.60	b	4.0	2.5	2.5	2.5	2.0	1.5	—
	a		4.0	4.0	4.0	3.5	3.5	
.65	b	—	3.5	3.0	2.5	2.0	1.5	—
	a		3.0	4.5	4.5	4.5	4.5	4.0
.70	b		3.0	3.5	3.0	2.5	2.0	1.5
	a				4.0	4.0	4.0	3.5
.75	b		—	—	3.0	2.5	2.0	1.5
	a						4.5	4.5
.80	b				—	—	2.5	2.0
	a							4.0
.85	b						-	2.0

The Beta distribution given in this table is such that:

1. $P = \frac{1}{B(a,b)} \int_0^m t^{a-1} (1-t)^{b-1}\, dt$

   with $B(a,b) = \int_0^1 t^{a-1} (1-t)^{b-1}\, dt$; accuracy on $P = \pm 0.02$.
2. The standard deviation is between ⅕ and ⅙.
3. The mode falls within the range $m - 0.10$ to $m + 0.05$.
4. The degrees of freedom a and b are such that $2a$ and $2b$ are integers.

between the value we originally chose (10 tons per gang-hour) and a value inferior to this by almost 30 percent of the range. From Table 10 it will now be seen that if we try to keep the standard deviation of the Beta distribution between ⅕ and ⅙ of the range, a small shift in the mode will result in a substantial modification of the shape of the distribution. For example, let us look at the column $m = 0.70$. It indicates that there exist five different Beta distributions of range 0-1 with a standard deviation between ⅕ and ⅙ and a mode between .60 and .75. The first column shows that for these five distributions the probability of exceeding .70 (or $1 - P$) varies from 50 percent to 25 percent. These five distributions are, therefore, quite different. If we had decided to use a Beta distribution defined in this way for the productivity of labor in the Mogadiscio case, this distribution would have given a probability for exceeding 10 tons per gang-hour of anywhere from 25 to 50 percent, depending on where we had decided to place the mode.

This led to the idea of making the choice of the degrees of freedom of the Beta distribution depend not only on estimates of the standard deviation and the mode of the distribution, but also on an estimate of *the probability that this mode will be exceeded.* Table 10 is designed for this purpose. Given a best estimate m and the probability $(1 - P)$ of exceeding m, it gives the degrees of freedom of a Beta distribution over the 0-1 range such that its mode will be between $m + 0.05$ and $m - 0.10$ and the standard deviation between ⅕ and ⅙. In introducing this form of estimation, the appraisal team's idea was not only to limit the importance of the best estimate but also to make use of information on the probability of exceeding the best estimate, which is often available. Figure 13, curve C, shows the result of introducing this information; it is clearly closer to the basic sample than the previous Beta distribution (curve A).

The Trapezoidal Distribution

This distribution is shown in Figure 14 below. It owes its appearance here to the experience of the lack of reliability of the best estimate, referred to in the previous section, and to the observation that it is often helpful to distinguish a smaller range around the best estimate within the total range. This smaller range will often correspond to what may happen under normal circumstances as opposed to what may happen under extreme circumstances. For example under normal circumstances a project may cost between −5 percent and +10 percent of the estimated cost. Under unusual circumstances it may cost between −25 percent and +100 percent of the estimated cost. While the appraiser may not be able to say that it is more likely that the project will exceed its cost by 10 percent than by 5 percent, he probably knows that +100 percent is less likely than +25 percent. Therefore, while within the inner range

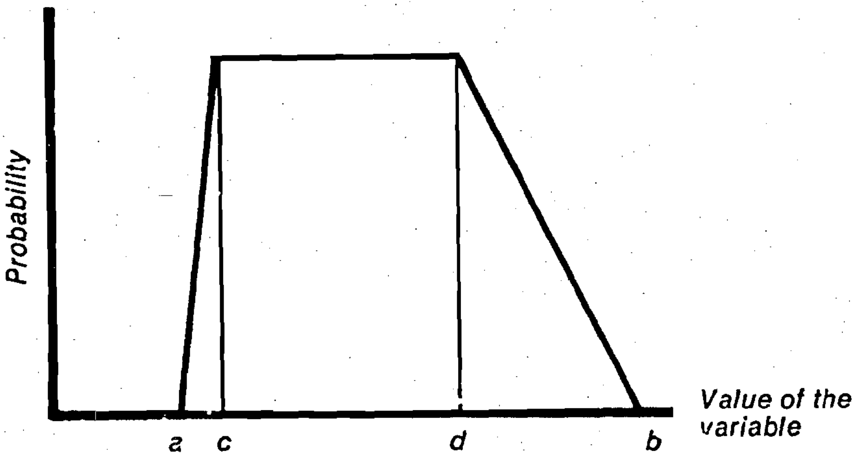


Figure 14. Trapezoidal Distribution

all values have the same probability, outside of this range and toward the outer limit these probabilities decrease. Use of this distribution indicates that it fits well a large class of subjective judgments.

The Triangular Distribution

This distribution, illustrated in Figure 15 below, is a particular case of the trapezoidal distribution and requires little comment. It simply reflects the fact that one is tempted to assign to a value close to the extreme of a range a lower probability than to a value close to the best estimate. It is only a convenient guess that this probability varies linearly from the value of the best estimate to the extreme value of the range—a guess which makes random number generation very easy. Surprisingly, especially if we consider the high sensitivity of this variable in the Mogadiscio analysis, curve T in Figure 13 shows that if we had substituted a triangular distribution for the basic sample distribution of labor productivity, we would have obtained a result remarkably close to our original result.

The Normal Distribution

In our admittedly limited experience, the normal distribution (see Figure 16) seems to be of little use with risk analysis variables.[3] Outside of its appearance in the Mogadiscio case as the result of a portrait approach, we have now abandoned it. We have used it only on one occasion when the availability of an

[3] However, normality may be a good assumption about the final rate of return distribution.

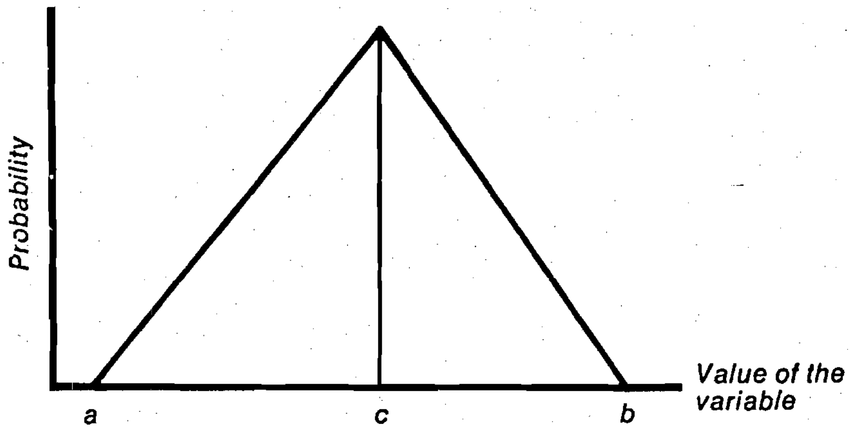


Figure 15. Triangular Distribution

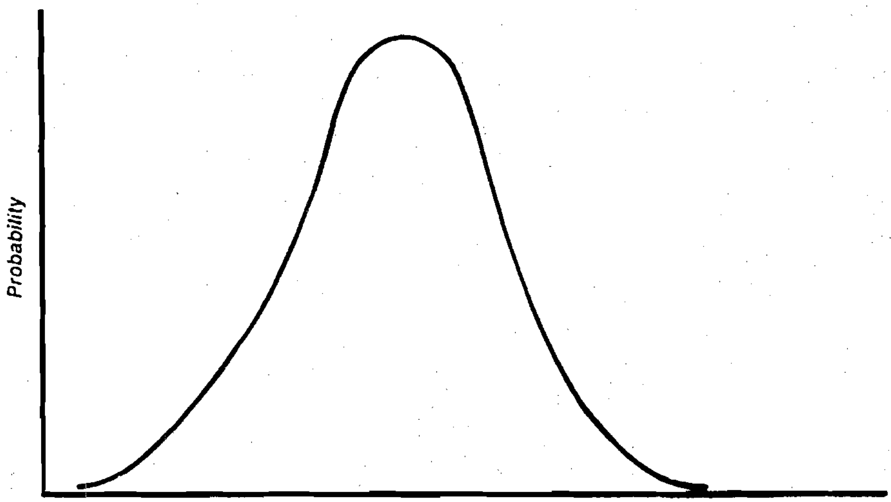


Figure 16. Normal Distribution

exceptional amount of data permitted a statistical analysis, and it turned out in the analysis that a normal distribution was an appropriate choice. But, except in rare cases, there is probably no justification for expressing a subjective judgment by a normal distribution. The variations we are trying to anticipate are the result of neither statistical errors nor random disturbances. Take the distribution of the value of an average ton of cargo in the port of Mogadiscio. Only a minor part of the uncertainty on this variable originated from the specificity of the sample of merchandise from which the mean was computed.

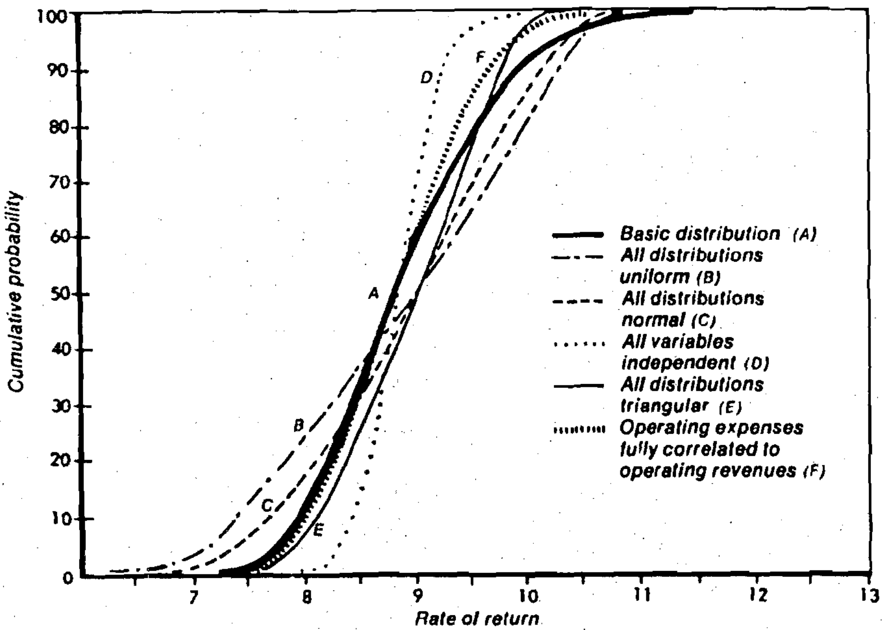


Figure 17. Telecommunications Project: Cumulative Distribution of the Rate of Return under Different Distribution and Correlation Hypotheses

The bulk of the uncertainty we wanted to represent resulted from three facts: (a) in 1966, Somalia imposed restrictions on its imports; thus their composition was likely to change over time; (b) by the time the port was constructed, livestock exports were likely to represent a greater share of the traffic; (c) the data on which the analysis was based were taken from customs statistics and were likely to be biased. There was nothing about these uncertainties which pointed to a normal distribution. They represented subjective doubts about the course of events, which may happen to be best represented by a skewed distribution, a bimodal one, or anything else.

The Comparative Importance of Correlations and Probability Distributions

We have emphasized the importance of correlation and pointed out that it constitutes a much more serious problem than the choice of the probability distributions. As an example, Figure 17 shows what would have happened in the case of another project[4] if we had:

a) replaced the distributions of uncertain variables with uniform distributions (curve B);

[4] The telecommunications project in Malaysia.

b) replaced all the distributions with normal distributions of the same mode[5] and same standard deviation as the original distributions (curve C);
c) replaced all the distributions by triangular distributions of the same mode[5] and same range as the original distributions (curve E);
d) kept the same distributions but considered all the variables independent (curve D); and
e) kept the same distributions for revenues and assumed that the operating expenses were fully correlated with the operating revenues (curve F).

The effect of a different assumption about the correlations (curve D) is as great as the greatest effect from changing the shape of the variable distributions (curve B, all distributions uniform). That curve B is so much less steep than the basic distribution is to be expected, because the uniform distribution exaggerates the probabilities of the extremes. Using normal distributions for all variables yields a curve less close to the basic distribution than using triangular distributions. This result confirmed experience in the Mogadiscio case, when the [illegible] distribution was used for the profile approach; the use of the normal distribution does not appear appropriate when subjective judgment is involved. Curve F (operating expenses fully correlated to operating revenues) shows that adding a full correlation assumption need not flatten the distribution curve, as it did for vehicle operating costs in the Great East Road case, and in the case for the correlation of port capacity and gang productivity in the port of Mogadiscio, as shown in Figure 4 (curve 2). Adding a correlation assumption in the case we are now looking at has resulted in a curve steeper than the original or true distribution, because the two variables assumed to be correlated —operating expenses and operating revenues—pull on the rate of return in opposite directions. Correlation of their variation logically tends to increase the likelihood that the internal rate of return will fall close to its mode and to decrease the likelihood it will fall in the tails of the distribution, thus tipping the cumulative distribution curve to the vertical. In the case of port capacity and gang productivity in Mogadiscio, the two correlated variables pull the same way on the rate of return; correlation tends to increase the weight of the more extreme probabilities, thus flattening the curve. However, even this kind of reasoning about correlations should be applied with caution, because often the effect of correlation between two variables will be more complex and not at all obvious before the data are analyzed. This is especially true when more than two variables are correlated.

---

[5] Whenever the mode was undetermined, we chose the value originally given as best estimate.

VIII

MEANS AND COST OF RISK ANALYSIS

The financial cost and the time constraints involved in a risk analysis are important elements in the decision whether to undertake a risk analysis. There follow a few comments, which lead us to the conclusion that these constraints should not be allowed to limit the use of the method.

All the simulations we made have been carried out on a computer. An important step of our work has, therefore, consisted in the development of computer programs. In the case of Mogadiscio this took us about six months and in the case of the first road project about two months. The lower time cost in the latter case results partly from the fact that the model used for roads is conceptually simpler and more standard than the one developed for ports. It also results from the fact that we have been able to use for the roads some of the programs we had developed in the Mogadiscio case. Development of computer programs, therefore, appears to be an important capital investment which can speed up a risk analysis considerably. For example, we can now carry out a road risk analysis in a maximum of three calendar days, the actual programming time spent by the programmer being from one-half day to one day. Even in the case of the pavement cost analysis made for the Tanzam highway, which was a completely new exercise, we were able to use parts of the existing programs and the whole programming work did not take more than a calendar week.

It should be pointed out, however, that while we have tried to standardize our programs to the maximum, we found it best to rewrite each time the part dealing with probabilities. The appraiser can, therefore, eliminate or introduce in the probability analysis any variables he wants. He can also use for each variable any probability distribution, and have the variables correlated in any way he wishes. This, of course, requires more programming work, but we think that this is necessary to ensure a good risk analysis. Even though a road is a road the world around, all road projects are different because the judgments attached to each of the elements of the project are different. Furthermore, these judgments may take different forms in the minds of different persons. If, as we think, a good risk analysis depends essentially on how well one is able to capture the appraiser's judgment, a very flexible framework is needed which can adapt to any judgment and retain its integrity, no matter what form it may take. There is, therefore, a limit to standardization. The major drawback to a lack of standardization does not seem to be the delay which may follow in obtaining the results—three days or even a week is still an acceptable time for a risk analysis—but the greater possibility of errors.

A computer program is a delicate tool which, once it is tested, should be modified as little as possible. While the computer will not make any error in the computation, it will not be able to detect any error in the logic of the program unless it is instructed to do so. It is difficult enough to detect an error in a rate of return and nearly impossible to detect an error in the probability distribution of this rate of return. The best way found so far to overcome this danger seems to be to present the program in a form which makes its checking as easy as possible. However, this does not seem good enough and we are now thinking of introducing into our models built-in tests which will detect possible anomalies in the results.

Outside of the programming work, which requires a programmer's time as well as computer time—and the former may be very expensive, about $5,000 in the Mogadiscio case—the risk analysis proper is inexpensive. To give an idea of the order of magnitude of the cost, the Mogadiscio simulation takes about 7 minutes on an IBM 7090 computer; on the same computer, the road simulation takes from 4 to 10 minutes depending upon the number of sections into which the road is divided. At commercial rates the machine costs about $6 a minute. Exclusive of the program preparation, a risk analysis will cost from $50 to $100 of computer time. It may become more expensive when the models become more sophisticated but, on the other hand, the cost of the program preparation should become cheaper as a result of more experience, better organization and the existence of a program library, which is already beginning to accumulate.[1]

[1] Robert Schlaifer, *Analysis of Decisions Under Uncertainty* (New York: McGraw-Hill, 1969); and "Computer Programs for a First Course Decision Under Uncertainty," Boston, Division of Research, Harvard Business School.

A last observation is that risk analysis does not seem to require any particular mathematical skill. Points which require some knowledge of mathematics or statistics, such as the sampling problem, once resolved for one project, are resolved for all. Difficulties might arise in getting a good feeling for the importance of the problem raised by correlations. But, here again, the mathematical treatment of correlations does not raise any problem, and as far as finding and quantifying correlation goes, practical knowledge of a project is more important to those in charge of the appraisal than is theoretical knowledge of the properties of correlations.

IX

MISCELLANEOUS TECHNICAL PROBLEMS

Sample Size

The choice of an appropriate sample size relates to the decision as to how many times the computer should repeat the computation of the rate of return, based each time on randomly generated values for the variables. This is a problem of statistics. The solution essentially focuses on:

a) the mean and standard deviation of the distribution; or
b) the probability of achieving a minimum return; or
c) the shape of the entire distribution.

More details on the first two approaches can be found in most statistics textbooks, and on the third one in the article on the subject by Feller.[1] Although, in the cases we have dealt with, the rate of return is nearly normally distributed and thus use of the first approach may be justified,[2] we were initially interested in the entire distribution and therefore based the choice of our sample size on the third approach, using samples of size 300. Kolmogorov's theorem then indicates that there is 95 percent probability that the maximum vertical distance between the true distribution and the distribution we obtain with this sample will be inferior to 8 percent. This sample size gives more

[1] W. Feller, on the Kolmogorov-Smirnov limit theorems for empirical distributions, *The Annals of Mathematics*, volume 19, no. 2, June 1948.
[2] See following section.

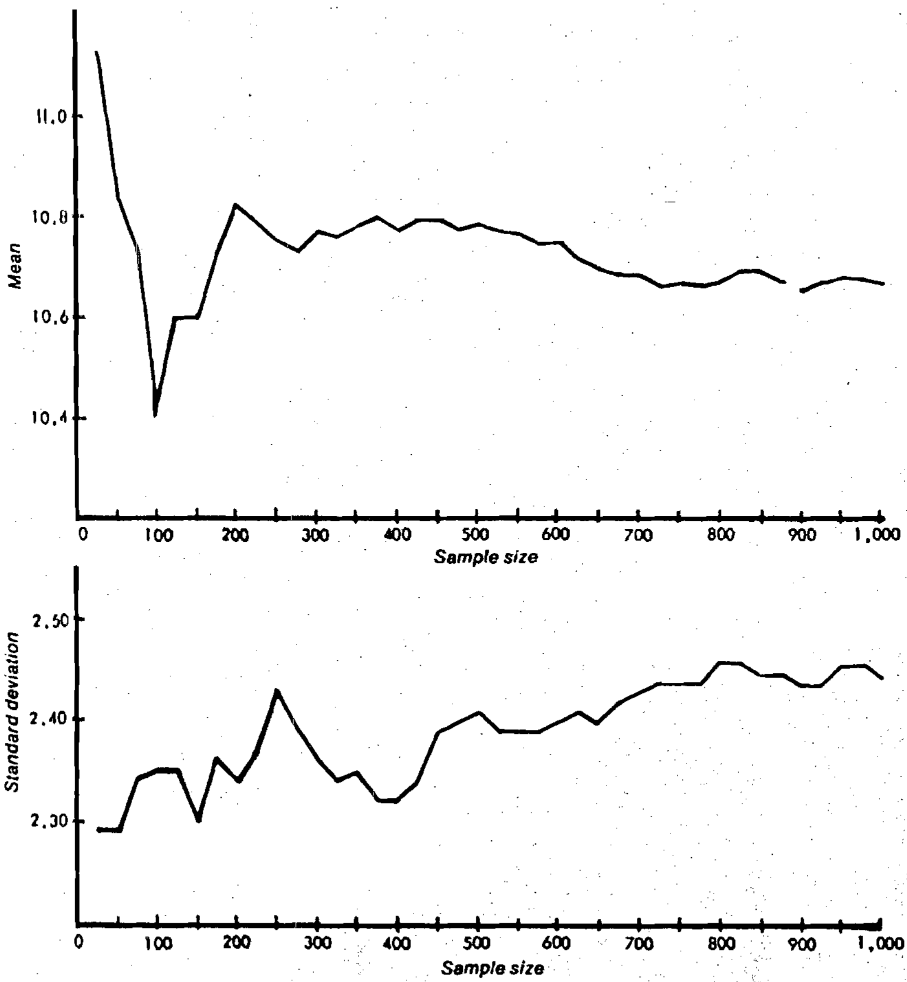


Figure 18. Mogadiscio Port Project: Variations of Mean and Standard Deviation According to Sample Size

accuracy than we need for the mean of the distribution and an acceptable accuracy for the standard deviation.

To illustrate the effect of sample sizes, we experimented with different sample sizes in the Mogadiscio case. Figure 18 shows what may happen to the mean and standard deviation when the sample size is increased to 1,000. In this exercise we increased the sample size by 25 at a time and computed, each time, the mean and standard deviation of the total sample. Even the smallest sample (25) gives a mean which differs from the mean for the largest samples by not more than 0.4 units or 4 percent, which is accurate enough for project appraisal. Therefore, a sample of size 50 would have yielded an acceptable estimate of the mean. From the results for the standard deviation, it is more difficult to make a similar judgment, but again, a sample of size 25 or 50 yields a standard deviation less than 7 percent different from the standard deviation for the largest samples. This is an acceptable result. Figure 19 shows the type of dispersion which can be expected with samples of size 100. With the sample of this size we have an 80 percent probability that the Kolmogorov's distance between sample distribution and the true distribution will not exceed 11 percent. This result seems to be confirmed by Figure 19 which—assuming that the distribution we obtained with a sample of size 1,000 is very close to the true distribution—indicates that only two out of the ten observed distributions of sample size 100 differ from the true distribution by more than 10 percent. The reader will of course realize that this is only a simplistic illustration of a complex statistical problem.

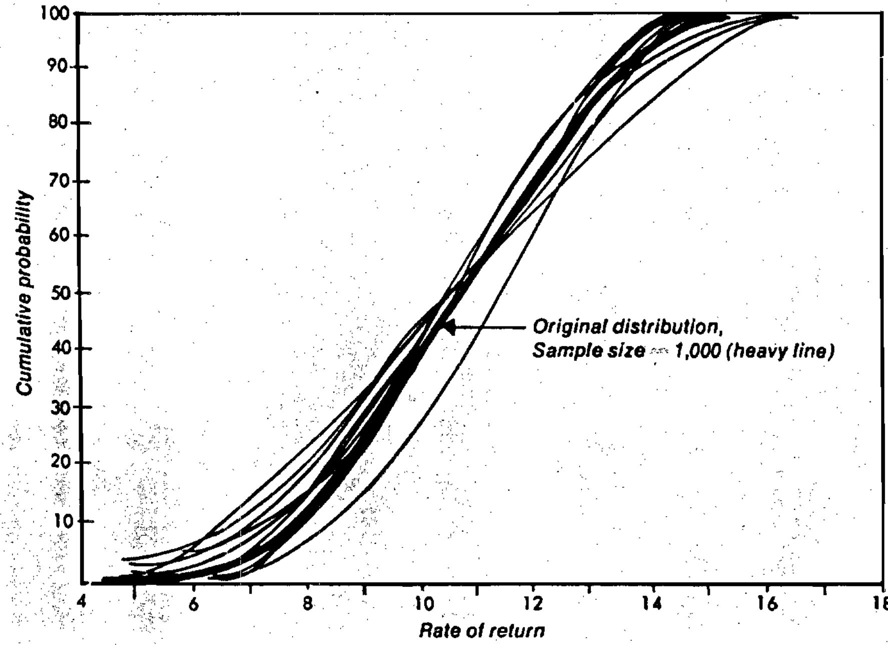


Figure 19. Mogadiscio Port Project: Ten Cumulative Distributions of the Rate of Return

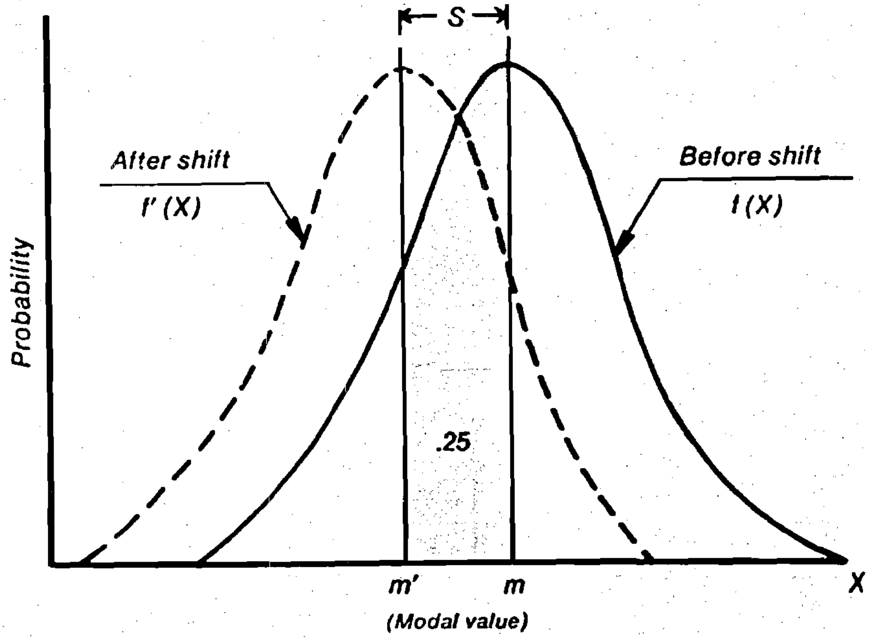


Figure 20. Quartile Shift of a Distribution

$$f'(X) = f(X + S)$$

where S is such that

$$\int_{m-S}^{m} f(X)\,dX = .25$$

and m is such that: $\dfrac{df(X)}{dX} = 0$

Normality of Rate of Return

Theoretical considerations indicate that, under certain conditions, the rate of return should follow a normal distribution.[3] If this were always the case, the distribution would be entirely defined by its mean and standard deviation, and our work would be somewhat simplified. In practice, we did obtain normal distributions in the cases we undertook. However, the conditions of applicability of the central limits theorem (sufficient in proving a distribution normal) were only partially fulfilled. The following table shows for the Mogadiscio case that the distribution obtained was very close to a normal distribution, with the mean standard deviation equal to the sample mean and standard deviation:

[3] F. S. Hillier, "The Derivation of Probabilistic Information for the Evaluation of Risky Investments," *Management Science*, April 1963.

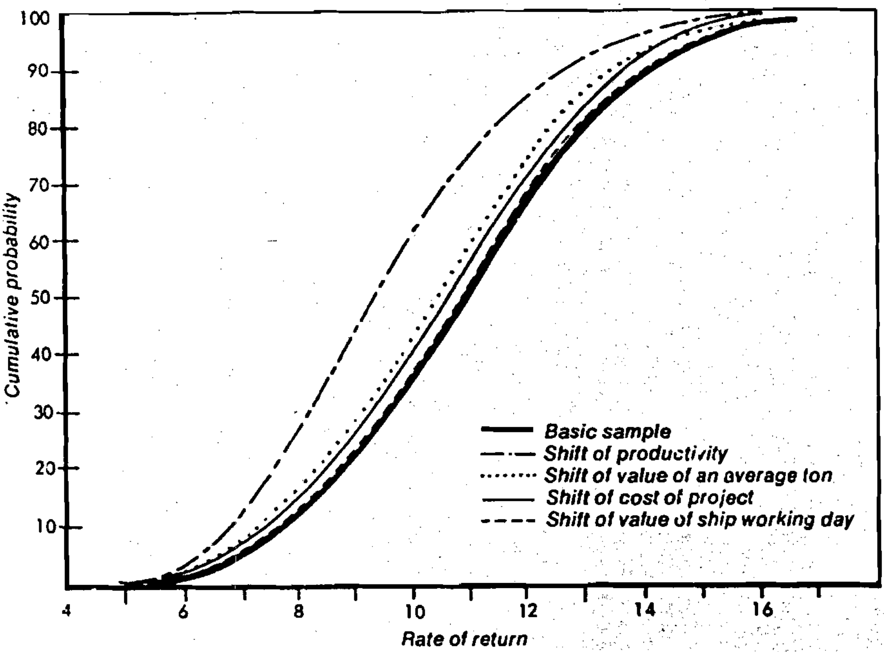


Figure 21. Mogadiscio Port Project: Effect of a Quartile Shift of Variable Distributions on Distribution of the Rate of Return

Cumulative probability of rate of return R being less or more than the rate shown	*Normal Distribution*	*Actual Results*
Prob. (R ≤ 6%)	3.5%	3%
Prob. (R ≤ 8%)	15%	15%
Prob. (R ≤ 10%)	41%	41%
Prob. (R ≥ 12%)	29%	30%
Prob. (R ≥ 14%)	9%	9%

We have made the same observation in the other cases of risk analysis and, though the divergences were greater than in the Mogadiscio case, they were never great enough to change the conclusion of the analysis. Though it cannot be demonstrated by the central limits theorem, normality may be a good assumption after all.[4]

[4] This applies, however, to the rate of return, not to the individual variables, as we explained earlier in Chapter VII.

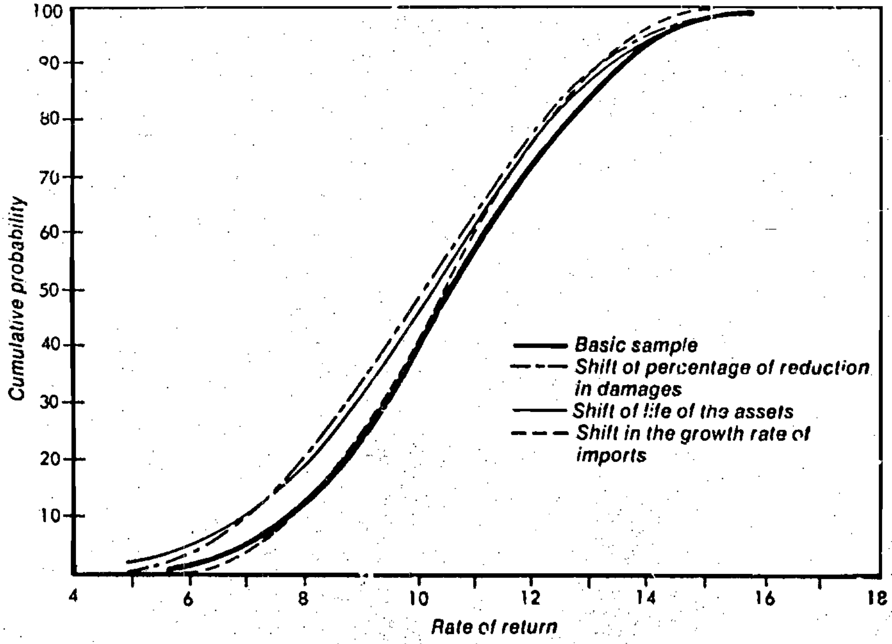


Figure 22. Mogadiscio Port Project: Effect of a Quartile Shift of Variable Distributions on Distribution of the Rate of Return

Sensitivity to Quartile Shifts

In order to discover the effect on the distribution of the rate of return of any possible variations in the distribution of the variables, we performed a quartile shift as defined in Figure 20 on each of the variable distributions used in the Mogadiscio port project (keeping the other variables unchanged) and compared the resulting rate of return distributions with the original.[5] The results are shown in Figures 21 and 22.

The additional information does not justify the expense of repeating the simulation for each variable. However, we believe that it may be possible to carry out this shift sensitivity analysis without actually repeating the simulation, just by extracting from the original sample a sub-sample in which one of the variables is distributed according to the modified distribution. This may lead to using somewhat bigger samples (say 500), but would make it possible to make the shift sensitivity analysis without using any extra computer time.

[5] This exercise was suggested by Mr. David Herz of McKinsey Inc.

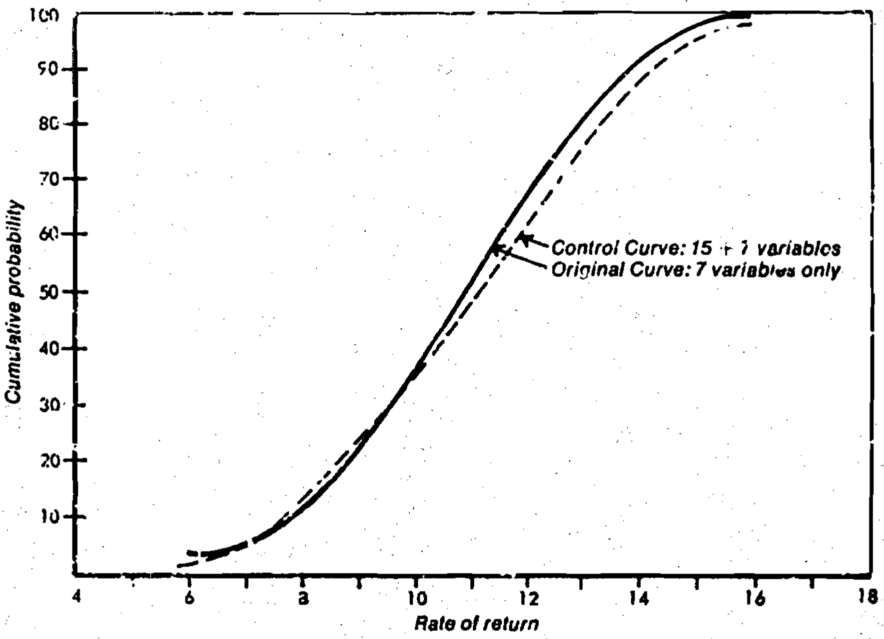


Figure 23. Mogadiscio Port Project: Effect of Including Low Sensitivity Variables on Distribution of the Rate of Return

Elimination of Low Sensitivity Variables

To simplify the simulations it was [illegible] a general rule that low sensitivity variables should be kept constant, as in the Mogadiscio case. The computer timesaving is negligible, but we think that this rule helped the appraiser focus on the important distributions of the analysis. In order to test the legitimacy of this approach, in the Mogadiscio case we made a simulation in which fifteen of the variables which were originally kept constant were varied randomly according to specified distributions as were the seven basic variables. Figure 23 shows that the result is very similar to our original result and indicates that this simplification is probably justified.

X

THE USEFULNESS OF RISK ANALYSIS

The major advantage of risk analysis is that it enables us to attack problems that we would otherwise avoid and to make decisions we would not otherwise feel competent to make. In cases like the Tanzam highway and the Mogadiscio port project, where uncertainty is high, the appraisers would usually follow the procedure of calculating several rates of return under different assumptions, basing their overall decision about the project on these few calculations and their best judgment, and presenting that unique and final rate of return which most accurately reflects the sum total of their knowledge of the project. Without probability analysis, this is the best they can do: the best estimate technique confines them to packing all the complexities of their understanding into a single number and then defending it as well as they can. With probability analysis, not only are the conclusions presented by the appraisers less limited, but the supporting material has all been quantified in easily comprehensible, standardized form. This means that, whereas previously it might have been recognized that some further information was needed in a particular area, with this kind of presentation it is usually possible to specify what kind of information is needed and how much difference it will make—that is, problems can be attacked which might otherwise have had to be passed over.

Special Advantages: Four Cases in which Risk is a Major Factor

Among the projects to which we have applied risk analysis there seem to be four distinct kinds of problems in which uncertainty plays an important role: whether to undertake a marginal project, how to handle a project with unusual uncertainties, how to settle on the best combination of specifications in a single project, and how to identify a project with only minimal information.

1. Marginal Projects. For some projects, like Mogadiscio port or the telecommunications project, the rate of return computed on the basis of the best estimate for each variable is very close to the estimated opportunity cost of capital. Then normal kinds of uncertainties about the value of the input variables are enough to turn a satisfactory rate of return into an unsatisfactory one. The decision to accept such a project implies judgments on the likelihood that the project will earn a satisfactory rate of return nonetheless and on the extreme ranges of possible results.

2. Unusual Uncertainties. For other projects like the Tanzam highway, despite a satisfactory rate of return based on the best estimate of each variable (say 13 percent to 18 percent), the uncertainty on some of the variables is so great that there is a distinct possibility that the project may not earn a satisfactory return. This kind of uncertainty is built into the project and cannot be eliminated or even reduced by any amount of additional study.

3. Optimization of Project Specifications. In many cases, the overall justification of the project has already been established at the identification or preappraisal stage. But the analysis of design standards, project timing, project phasing, and project size can only be done at the appraisal stage, and such analyses may lead to saving millions of dollars in project cost, as the Tanzam highway case has suggested. Specification analysis is basic to most Bank project work. In both the Great East Road case and the Mogadiscio case there were problems of timing and scale, though we have not described them explicitly here. Choice among alternatives on such specific issues is made particularly difficult by uncertainty.

In the choice between alternatives A and B shown in Figure 24, for example, uncertainty about data is not critical to the decision whether to go ahead with the project, for either alternative will return an adequate yield. But it may still be critical to the choice between alternatives.

In the neighborhood of S*, alternative A should be preferred to alternative B because it would yield a much higher return. On the other hand, in the neighborhood of S**, B should be preferred to A. The choice between A and B will therefore involve some estimation of the probability distribution of S, and if this choice also involves other uncertain variables it will very likely require a probability analysis.

4. Project Identification. The best example of project identification among our four examples is the Great East Road (Chapter V). Here no detailed

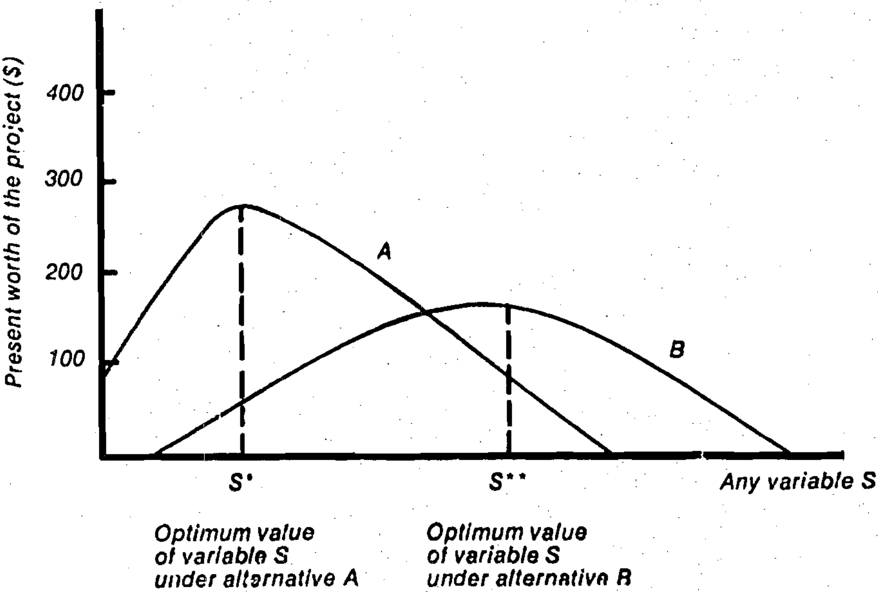


Figure 24. Choice between Two Alternatives, Both with Adequate Rates of Return

study had been made; only rough data estimates were available. Yet a decision had to be made to go ahead with the project, to postpone it, or discard it. The uncertainty is less elusive than that associated with marginal projects, since most of it could be eliminated through study. However, studies are expensive, take time, and even the decision to undertake a study requires careful analysis and involves a judgment on the possible outcome of the study.

The frequency of such cases

It is for these four classes of problems that risk analysis seems best designed. They seem also to represent the essential cases in which uncertainty has to be dealt with in one way or another. The first two classes, marginal projects and projects with unusual uncertainties, while they are not frequent, particularly among transportation projects, often raise critical issues because they involve important decisions. The third case (optimization) is much more frequent but less critical because in practice it is easy to bypass the issue by falling back on rules-of-thumb. It is hard enough to conceive a workable project under the usual uncertainties without also trying to optimize. But if, as our experiences suggest, risk analysis provides an efficient tool to handle the difficulties of optimization under uncertainty, we may be able to do it more often. Finally, in project identification, risk analysis can increase the scope for action. The more sophisticated project planning becomes, the longer in advance decisions must be made and the more elements—all uncertain—intervene in these decisions. The present need to improve decision-making processes at the identification stage may be reflected in the present overwhelming number of project studies. In turn, more and better decisions at the project identification stage should have the effect of decreasing the number of studies and focusing them on the most important issues.

Risk analysis might well find its major application in the optimization field. Consequently it may become a tool for the consultants even more than for the Bank. Once a project has been fully designed, it is often too late to optimize, but if, at the Bank's request, the optimization were to be made by the consultants or by the project designer, many improvements could probably be achieved at that stage.

General Advantages

Risk analysis requires only one set of computations, either by mental calculation or by computer, to obtain a complete picture of the project. To obtain a similarly adequate picture of the project using the conventional method and the identical computational aids,[1] one must repeat the entire computational process at least once. Therefore, even though a single rate of return can be calculated more quickly by the conventional method, in practice appraisers at the Bank seldom stop at the first rate of return obtained by conventional analysis. The final range of alternatives from which a decision is made can be calculated more quickly by using risk analysis. In the case of the Great East Road in Zambia, we obtained a result from the risk analysis no later than three days after the return of the mission from the field. By the usual Bank method, it would probably have taken two full working days to figure out the traffic, the cost of the project, the savings in vehicle operating cost, the savings in maintenance cost and, finally, the rate of return. Then we would have found that this rate of return was too low and did not correspond to the opinion we had of the project. We would then have repeated the operation, changing the value used for, say, the traffic level, savings in vehicle operating cost, or the cost of the project. In this particular case, it would have been easy to pick second and third values just as good as our first, because of the great uncertainty about most of the variables. This unsystematic sensitivity analysis might have required another two to five days. In the same time, using risk analysis, the final report on the project was already finished. In addition, it took only about one programmer-hour to rerun the program six months after

[1] Naturally, developing computer programs takes time and using them speeds up calculations, but these facts are true for any method and do not affect the validity of the argument here. See also Chapter VIII

the original decision, when we received new information on the cost estimate. With faster computer turnaround time, the time to get the computer results could be reduced from three days to one day, which is the time it takes the programmer to prepare the probability part of the computer program.

Clarity of presentation

Another benefit of risk analysis is that it results in greater report clarity and thereby permits more people to make useful contributions to project appraisal. Our appraisal reports give the values of the elements used in the evaluation of a project. However, they very seldom give the judgment lying behind these values and when they do, it is always in qualitative form. As a consequence it is very difficult for anyone to discuss these judgments, and comments often focus more on the presentation than on the substance of a report. In some cases, a high degree of technicality creates a natural barrier to wide discussion. But often, as with the Tanzam highway, a discussion of the assumptions is both possible and desirable.

This transparency of the analysis, while serving to make discussion more effective, also seems to facilitate the adoption of recommendations. However, our experience is limited and it would be interesting to investigate this point further, in particular in connection with negotiations relating to technical questions.

Convenience

Because of its compactness, a probability distribution not only communicates information well, but also is very convenient to work with. It was surprising to find that after even a very little experience with risk analysis, it became easier and more natural to express a judgment in probabilistic terms than in terms of a best estimate or, indeed, of any other kind of estimate. We found that an expert consultant may be unwilling to commit himself to a single cost estimate before the completion of his study, but he may quite readily proffer a range of cost estimates and a full probability distribution over that range.

Rigor in analysis

Risk analysis both demands and permits the use of greater rigor in analysis. Risk analysis demands more rigor simply because it is a more systematic method. It permits more rigor than the single-estimate approach because even in a simple analysis it allows for more than one course of action. Paradoxically, it is always easier to be rigorous than to approximate.

In concluding this chapter it may be suggested that, for the purposes of risk analysis, in many cases present worth may be a better criterion than the internal rate of return. One difficulty in decision-making is the estimation of the opportunity cost of capital. This is often cited in the Bank as an argument in favor of using an internal rate of return rather than a present value. We need not know the opportunity cost of capital to compute the internal rate of return; it is used only at the last stage to decide whether the calculated rate of return is acceptable or not. Since in practice the rates of return obtained are often higher than the highest likely value of the opportunity cost of capital, the need to calculate the latter in detail does not then arise. However, if risk analysis is to be applied to marginal projects or to marginal components of a project, the decision as to the acceptability of the project will no longer be so obvious. Mogadiscio's port project, for example, implies the comparison of an internal rate of return, varying over a wide range, to an estimated opportunity cost of capital, also varying over a wide range. This is not easy. Under these circumstances it seems possible to consider the opportunity cost of capital as an uncertain variable similar in all respects to the other variables of our analysis, and to use a present worth approach. The probability of failure of the project would thus simply appear as the probability that the project has negative net present worth, and the decision as to its acceptability would be made very simple.

Summary and Conclusions

The overall conclusions are numerous, and many are, of course, still tentative. They can be summarized in the following four points:

(a) Risk analysis is a powerful technique which permits the use of a great deal of information which would otherwise be lost. It enables us to handle uncertainty not only about the viability of a marginal project, but also about the most appropriate design or phasing or size of a clearly acceptable project.

(b) Perhaps even more importantly, the entire framework of risk analysis provides a highly efficient medium of communication, a focus for evaluation and discussion, whether between one person and his superior, among the various members of a team, or possibly (looking toward the future) between consultants and the Bank or a borrower and the Bank.

(c) Risk analysis is in no sense a technique which replaces skilled judgment. On the contrary, it often requires the use of far more judgment than the tradirional analysis. The technique cannot provide correct answers on the basis of false assumptions.

(d) Despite the method's value, the treatment of correlations between variables remains a major problem. It is clear that results can be completely misleading if these correlations are not properly handled. This danger is not merely theoretical; there is apparently a systematic tendency to overlook correlations. It follows that risk analysis should be undertaken only with great caution.

[illegible] REPRESENTATIVES

[illegible]

[illegible] CHA[illegible]

[illegible] The Johns Hopkins [illegible] with the [illegible]

[illegible] Book Co. Pty [illegible]

[illegible] Ltd. [illegible]

INTERNATIONAL BANK FOR RECONSTRUCTION AND DEVELOPMENT

WORLD BANK

1818 H STREET, N.W., WASHINGTON, D.C. 20433 U.S.A.

Telephone number: EXecutive 3-6360 Cable address: INTBAFRAD