ROBUSTNESS AS A DECISION CRITERION IN CAPITAL INVESTMENT ANALYSIS

Robert H. Berry,

School of Information Systems,

University of East Anglia,

Norwich NR4 7TJ United Kingdom

Phone: +44 (603) 561 61, Fax: +44 (603) 50 77 20

E-Mail: rhb@sys.uea.ac.uk

Duarte Trigueiros,

ISCTE,

Avenida das Forças Armadas,

1600 Lisbon, Portugal

Phone: +351 (1) 793 50 00, Fax: +351 (1) 796 47 10,

E-Mail: dmt@sara.inesc.pt

Goksan Ergodan,

School of Information Systems,

University of East Anglia,

Norwich NR4 7TJ United Kingdom

Phone: +44 (603) 561 61, Fax: +44 (603) 50 77 20

ROBUSTNESS AS A DECISION CRITERION IN CAPITAL INVESTMENT ANALYSIS

Abstract

In this study we show that Quinlan's ID3 algorithm (1979), originally intended for extracting rules from sequential data, is very effective in interpreting complex decision trees like those used in Investment Appraisal. As a result of using the ID3 algorithm for post-processing decision trees, a tree of rules is obtained, in which sequential events are hierarchically placed according to their causal relationship to outcomes. We show that this post-processing is capable of reducing redundant information, and leads to a new way of assessing the robustness of decisions.

We firstly introduce and discuss the previous literature on robustness. Then, we explain the algorithm ID3 and the meaning of the rules obtained using it. We point out that the data structure this algorithm requires is the hierarchical one, and that other characteristics of decision trees used in Investment Appraisal also match its requirements. Next, we show how the ID3 algorithm can be used for post-processing decision trees. Finally, we present a well-known example, illustrating this new technique, and we discuss the results in the light of the assessment of the robustness of decisions.

ROBUSTNESS AS A DECISION CRITERION IN CAPITAL INVESTMENT ANALYSIS

This study tackles the problem of how to interpret complex decision trees of the kind used in Investment Appraisal. We focus on robustness as a decision criterion in Capital Budgeting. We expect to show the potential interest of this criterion for managers, and a new quantification of the robustness of decisions in the case of sequential models.

1   Robustness as a Decision Criterion

The models used in Investment Appraisal only maximize expected earnings or other equivalent function. This reduces complex multiple-criteria problems to optimization exercises.

The intuitive meaning attached to the robustness of a decision is the thickness of the optimum. If an attribute can generate several possible outcomes, the manager will be interested in knowing which of the values of the attribute leads to an optimal outcome. However, the knowledge about this optimum isn't enough: In practice, managers know that attributes suffer deviations during the execution of projects. An outcome is robust when its optimality isn't significantly affected by such deviations. An outcome is fragile when even small deviations from the optimum in the attributes deteriorate or invert its optimality.

Rosenhead *et al.* (1972) [9] elaborated the intuitive idea of robustness presented above. After stressing that robustness only makes sense after establishing the difference between plan and decision they suggest a definition of robustness near the idea of flexibility of paths or operative:

> "Little of what is known about the system being planned for is known with certainty. Much of what is not known cannot be expressed in terms of probabilities. The situation is one of uncertainty. As these uncontrollable and often unpredictable external events unfold, more information becomes available on the desirable future state and how to achieve it. In the light of this information it is natural and appropriate to reconsider and perhaps modify the as yet unimplemented stages of the plan. But if the possibility of making revisions has played no role in the specification of the earlier, implemented decisions, there may no longer be adequate residual flexibility.
>
> All decisions limit the future by commiting the present. A plan whose initial decisions limit the future as little as possible has an evolutionary advantage in an uncertain world.

Consider a planning problem in which one decision must be chosen from a set $D\{\equiv (d_i)\}$ of short-term decisions; and in which one of a set S of alternative plans (or solutions) will be realized in the long run. Any initial decision d_i will restrict the attainable plans to a subset S_i of S.

Suppose that some subset $\hat{S}$ of S is curently considered "good" or acceptable according to some combination of satisaficing criteria. A subset $\hat{S}_i$ of $\hat{S}$ will be attainable after an initial decision d_i. Then the *robustness* of d_i (see also [3] e [2]) is defined as:

$$r_i = \frac{n(\hat{S}_i)}{n(\hat{S})}$$

where $n(S)$ is the number of elements in set S.

Robustness, a measure of the useful flexibility maintained by a decision, has characteristics which make it a suitable criterion for sequential decision making under conditions of uncertainty. It handles the uncertainty of the environment, not by imposing a probabilistic structure, but by stressing the importance of flexibility. (...)"

The line of thought sketched above has been object of further exploring by one of the authors, Rosenhead, in several papers (for example, [8]).

The definition of robustness introduced by Rosenhead *et al.*, despite the useful insights it provides, isn't suited for financial budgetting. Next we examine the main causes of this.

No place for risk: The first aspect of such definition is the explicit denying of the use of risk measures such as the spread of similar cases, or probability assessments.

When outcomes are uncertain — in the sense used by Rosenhead *et al.* — the information about them is not zero. It is nonexistent. It may be that the likelihood of a given outcome is very high while the one of another is low. This definition ignores such a difference. When Rosenhead *et al.* refer an uncertain world, they mean a world in which managers would make plans whithout any knowkedge about the likelihood of future events.

We think that, despite Rosenhead's remarks, the cases of planning under total uncertainity are rare in Investment Appraisal. By ignoring all the a-priori available information, the suggested criterion throws away data that could be important concerning the decision. As an example, if, in a collection of outcomes, one of them has a likelihood of 99 in 100, this strong expectation would be ignored by the criterion suggested by Rosenhead *et al.*.

The causal weight of the decisions: A consequence of rennouncing to any a-priori information is the fact that the criterion above can't avoid considering all the set of possible

decision as if they had the same effectiveness in obtaining the desired outcome. The only thing that matters for establishing Rosenhead's definition is the total number of paths in each set. The causal weight of each path is ignored. However, it's clear that all these paths should have equal effectiveness in causing the desired outcome in order for the definition to be appropriate.

We shall see later on that decisions aren't equally effective in causing an outcome: The causal weight or effectiveness in causing a desired outcome can vary widely from decision to decision. Some decisions are able to bring about an outcome and some others contribute very little to its happening. We'll show examples of projects in which some decisions, apparently important, are completely indifferent regarding the desired outcome.

We think that the fact that the causal *nexus* linking outcomes to decisions can be strong or weak, should be in the basis of the concept of robustness.

Causal weight and commitment of the future: Rorenhead basis his definition of robustness in the assumption that

> All decisions limit the future by commiting the present. A plan whose initial decisions limit the future as little as possible has an evolutionary advantage in an uncertain world.

This is indeed an interesting statement in negotiation processes but not in Finance. When negotiating, the goal is to achieve an agreement between the parts involved no matter which. In Capital Budgeting the goal is a positive NPV. In most of the negotiation processes, the particular shape of the solution found is not important when compared with the fact that the parts were brought to agree. This strongly contrasts with the importance of the outcome in financial projects.

Moreover, the second statemant of the above paragraph is eventually against the common experience of managers: A plan designed so that its decisions will commit the future as little as possible, will not have any special advantage even in an uncertain world.

The idea of marching to the desired outcome through non-commiting decisions seems somehow misplaced in Management. To walk through decisions carefully chosen so as to avoid any commitment, is generally considered as the best way of not attaining any goals.

The extent to which a decision commits the future isn't, in general, independent of its causal weight. A strong commitment often means the perception by the manager that the decision is important for causing an outcome. In the limit, if a decision doesn't commit the future at all, its capacity for causing the outcome is probably zero.

Therefore, the definition of robustness suggested by Rosenhead *et al.*, based as it is in the strict independence between the degree of commiting the future and the power to bring outcomes about, is strange to Capital Budgeting.

Robustness and struture: Finally, the flexibility referred to by Rosenhead *et al.* depends on the complexity of the model. It is achieved by means of a growing sophistication. Only after introducing many alternatives is it possible to select the flexible ones. Therefore, only sophisticated, very complete models can be checked according to this criterion. Simplified models will not allow such an assessment.

However, it is well known that simplicity and generality are linked. Analysts often try to prune out the non-significant branches or attributes of a model so that the resulting one reflects general trends and avoids very particular ones. In this sense, simplicity means also a sort of robustness: With large numbers of unimportant attributes, it becomes more likely that the model will fail to recognize the desired goal based on its attributes.

A New Definition of Robustness We suggest that the robustness of attributes should be based on their causal weight. Robust decisions should be those which, in a project, are more able to cause the desired outcome than the attributes the manager can't command. A project is itself robust when the robust attributes are decisions. It is a fragile one when the most causal attributes are out of reach of the manager.

A decision with causal weight means that the manager has in his hands a real capacity of command. Before its execution he can shape it so that the desired outcome is the most likely. We call such decisions robust.

As defined here, the robustness is not dependent on particular structures of decision or on the degree of sophistication of the model. The same decision tree can have all sorts of robustness associated with each attribute, depending on the probabilities present in the model. It is also a measure able to use all the information available about future events. Notice that it is possible to introduce in probability models like decision trees the ignorance about the future (uncertainity) along with the lack of information (risk) if necessary. Therefore, our notion of robustness will not be bounded by the need to make available all the probabilities of future events.

How to implement this robustness? We'll see that the algorithm ID3 can do it in the particular case of sequential decision structures.

2   An Introduction to ID3

This section introduces the algorithm ID3. The application of this algorithm in Finance is discussed. Finally, an example is given illustrating its use in sequential decision problems having nominal outcomes.

Originally, algorithms for rule induction were intended to a convert complex deterministic experience into logical structures. For example, the ID3 algorithm was presented as a tool able to extract rules from Chess ends. The ID3 algorithm is a hierarchical search of attributes which are the most informative for explaining outcomes. The criterion for selecting attributes is the gain in information about the outcomes they apportion.

Probabilities are one amongst several ways of assessing expectations. When using them, knowledge about the number of outcomes under question is also required. For example, a probability of 1/2, in a game having two outcomes (the toss of a coin) means something different from the same probability of 1/2 in a six outcomes one. In the first game it expresses no expectations or the absense of prior information available about the outcome. In the second, it expresses a strong trend towards a given outcome and few chances associated with the other ones. In general, if N is the number of outcomes in a game, a probability of $1/N$ means the absense of prior knowledge or expectation about the outcome.

It seems useful to gather in a unique measure of uncertainty both probabilities and the number of outcomes. This would allow the comparing of games in terms of the uncertainty involved. Given a game having N possible outcomes, the number of digits needed to signal one of them is proportional to the logarithm of the N. $logN$ is known as the logical variety of a game and it measures the missing amount of information required to uniquely identify an outcome.

When outcomes show some regularity, the missing amount of information about them is no longer $logN$. There is a gain in information by knowing that, say, K_i of the N possible outcomes share the same attribute and hence can be lumped into one class known in advance. The knowledge of this attribute brings about some amount of information, which must be subtracted from the missing information in order to obtain the uncertainty about outcomes.

After a multiple classification, the remaining missing information needed to correctly identify each case is just the difference between the missing information before the classification and the average information that such a classification carries with it:

$$H = logN - \sum_i \frac{K_i}{N} logK_i$$

This difference is known as "Entropy". It measures the missing information in a game: When $H = 0$ there is no missing information. Outcomes can be predicted by knowing its attributes.

When $H = log N$ there is total ignorance about outcomes. Between those extreme cases, any expectation can be assessed.

The use of the Entropy of games simplifies the process of decision making. It allows the direct comparing of games in terms of uncertainty. Instead of a collection of probabilities and N, only a unique measure is now required. Also, Entropy is often more precise. For example, in a two-outcomes game the difference between a $p = 1/2$ uncertainty and a $p = 1/3$ one is neglectible in terms of information gain and need not to be considered. This is clearly displayed by using Entropy instead of probability.

The ID3: Hierarchically scaling attributes. The gain in information due to the existence of prior knowledge about outcomes, $G = \sum_i \frac{K_i}{N} log((K_i)$, is what the ID3 maximizes through a hierarchical search. It begins by selecting the attribute which best predicts outcomes when G is used as criterion. Next, the sample is devided into as many sub-samples as the classes of the selected attribute and a similar search is performed in each of them. This search selects, for each sub-sample, the remaining attribute able to explain in the best way outcomes according to G. By repeating this process, a hierarchical structure is obtained, having the most informative attribute at the root and the least informative ones at the leaves.

The ID3 gathers together probabilities, number of classes and structure into one unique information measure. Sequential structures involving probabilities are converted into an hierarchical display of the relative importance of each attribute to the outcome.

The ID3 is intended for sequential, deterministic data: The problem of extracting rules from observations is part of the more general problem of acquiring knowledge from particular cases by induction. An amount in knowledge by induction can be achieved using many different ways. Here we are interested in comparing two of them: Generalization, which applies to non-deterministic data, and re-arranging or ordering, which applies to complex deterministic data.

When some regularity is found in non-deterministic cases it can be possible to consider it as a common feature of the whole population. It is an information reduction process, aimed at neglecting what is considered particular and keeping only the general description. The aim of the second way of extracting knowledge is to eliminate redundancy or emphasize some quality hidden by the complexity. It is just a transformation: No information is lost. Only redundancy is "lost".

Statistical modelling techniques often represent a prior opinion about the population. Linear Regression is used when the assumption of linear co-variance seems plausible. Anal-
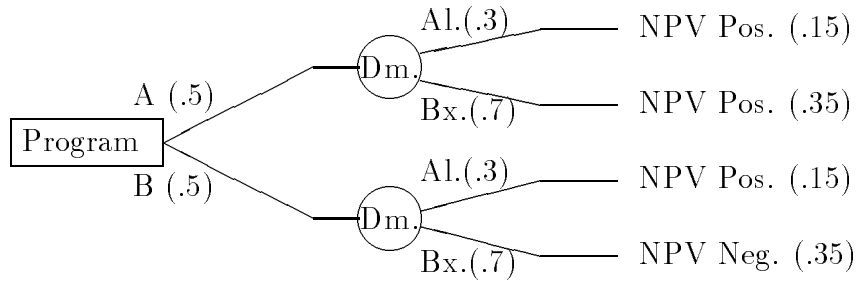
Figure 1: A simple decision tree. *Dm.* is the attribute "Demand". The likelihood of each class is also shown.

ysis of Variance assumes linear effects, Principal Components looks for common sources of variability and so on. The experience has shown that prior assumptions about general qualities are of great practical importance.

When a modelling tool is unlimited in its fitting capacity, the problem is to discover when to stop adding details to it. Cluster Analysis, Log-linear modelling, and indeed, Rule Induction, all suffer from this lack of self-moderation to fit data. Using them, it is easy to *over-fit* the data, building detailed and useless models. And it's difficult to see real underlying features of the data between so many possibilities.

Rule Induction produces rules. Can linear relations be modelled by rules if needed? This seems a more interesting problem. The answer is straightforward: When working with nominal observations, linearity means pure additivity or no interaction between attributes. The class variable is explained only by main effects. Being so, the ID3 or other similar algorithm will work out in first place the attribute with more explanatory power, and then all the others in a decreasing scale of explanatory capacity. However, the way the model is built (effects nested within previous effects) is certainly not the natural way for describing additivity. Hierarchical structures doesn't seem adequate as linear descriptors.

Rule induction, if not stopped, induces automatically a ternary relation, just because of its intrinsic hierarchical and step-wise way of modelling.

Again, we find it useful to distinguish statistical modelling from ordering, or a search for regularity from the search for *simplicity*. Chess ends are not simple but they are regular. More simplicity is required. The distribution of death penalties in Florida is quite simple and much less regular then a chess end. A model is required. And when a model is required Rule Induction fails to do it.

Methodology: The likelihood of each outcome were calculated multiplying the probabilities along its path. This likelihood is the expected frequency of that particular outcome.

All the outcomes with associated expected frequencies and attributes are the collection

Net Present Value:	Demand	Program	Frequence
Positive (+)	High	A	15
Positive (+)	Low	A	35
Positive (+)	High	B	15
Negative (−)	Low	B	35

Table 1: The set of Outcomes, their attributes and expected frequencies.

of observations Rule Induction transform. The result is a rule tree. It is important, in this case, to examine results as a tree, not only as a set of rules.

A simplified example will show the entire process. Figure 1 is a decision tree. The attribute "Project" is a decision one and the attribute "Demand" is an uncertain one. At the top of the tree the outcomes (NPV) are displayed. The steps are:

Consider only nominal outcomes: NPV are considered as just "positive" and "negative".

Consider any prior trend associated with the decision attributes: If no prior trend exists, the decision attributes are equally alike. So, in this case, a probability of 1/2 is assigned to them.

Construct the set of observations: the above table shows the outcomes associated with its attributes, and with its likelihood expressed as a relative frequency.

Apply Rule Induction to the set of observations. See [6] for a description of ID3. Simple experiments can be carried out using the selective "query" facilities of a data-base software associated with contingency table facilities.

In this simple case, the first step of ID3 would generate two contingency tables: Outcome by Program and Outcome by Demand.

Demand	Outcome	
	$NPV > 0$	$NPV < 0$
High	30	0
Low	35	35
Qui-Square: 23.1		

Program	Outcome	
	$NPV > 0$	$NPV < 0$
A	50	0
B	15	35
Qui-Square: 53.8		

Knowledge about the demand attribute discriminates between outcomes better then a corresponding knowledge about the program to pursue. So, the first attribute to be considered is the demand.

Next, we would consider separately low and high demand observations, and would repeat the process for each of them. The result would be a logical structure like this:

If	Program	is A then:		the outcome is NPV Positive.
If	Program	is B then:		
	If	Demand is High then:		the outcome is NPV Positive.
	If	Demand is Low then:		the outcome is NPV Negative.

3   ID3 and Decision Trees in Capital Investment

What is the future of Rule Induction in financial modelling? Cross-sectional studies or those who use systems of equations, don't seem to fit well in a so specific structure. If, in a financial model, a variable influences itself, it is likely to generate circular structures after discretization.

On the other hand, nested designs are important in investment appraisal. Sequential decisions lead to nested designs: Quinlan original work on extraction of rules from chess ends is a typical example of sequential data: The former movements will condition all the next ones. In financial modelling, sequential data appears as a result of using *decision trees* for sensitivity analysis or discrete simulation [4].

But before exploring this capability let us point out two potentially dangerous limitations of Rule Induction when applied to financial data. These limitations have not been displayed in previous work.

Measurement   Information can be available in different ways and, for statistical modelling purposes, in different scales of measurement. It is usual, following Stevens [10], to consider four basic scales according to measurement power. Nominal measurements, which are simply sets of unordered categories, that is, *labels*. They classify mapping disjounct sets as *the same* or as *different* things, without quantifying the differences.

Rule Induction deals with the poorest type of measurements, simple non-ranked categories. When applying it to ratio scaled data it is important to realize how much information is lost, and how can this loss influence the results.

Using ratio observations as simple labels involves three damaging steps in the information content of observations:

9

To ignore the existence of a real zero is equivalent to consider the zero as an arbitrary position. With financial data, for instance, it could mean that the difference between profits and losses is flattened away.

To ignore the measure of intervals is equivalent to limit the information about a set of companies to a simple ranking ("first", "second", and "third") even when they are as distant as GM, IBM, and the local store.

Finally, to ignore even the ranking, as Rule Induction does when applied to ratio-scaled data, is equivalent to limit the information about the three companies above to the point of saying that they are different, as indeed they are, and nothing else.

Clearly, the loose of information is severe and this will be reflected in the model's behaviour. Only a very special reason cal lead to such simplification.

The damage in the informational content of data by using unadequate modelling techniques is sometimes difficult to recognize intuitively. One possible reason is that labels also contain information, although not accessible to the model. The label "second" doesn't carry on itself any information about ranking when treated by Rule Induction. But, as it exists an external correspondence between "second" as a label and "second" as a rank, the ranking information is not lost. However, when interpreting results, such a control over the information the model left behind can be very difficult or impossible.

Rule Induction, when used with financial "discretized" data, will not produce wrong rules. Nevertheless, the results can be very misleading due to this *lack of control over lost information.* Two very similar situations can generate rules who seem to establish very different alternatives; and two really different positions will easily be flattened into a single, bolt, common rule.

Controlled experiments will not display all the damage. But when a set of rules is generated for expert systems to apply to data other then the original one, it is quite sure that this misleading capacity will become more visible.

When modelling, there are reasons to use tools based on less powerful scales. Analytical statistical tools are very sensitive to deviations from underlying assumptions like normality or homogeneity of variance. If the data being analyzed shows strong deviations from assumptions, the solution is to use less sensitive tools. Generally, these tools, though very general, are unable to use all the existing information. But there was a good reason to loose it.

4 An Experiment: Prism Paints Inc

The power of Rule Induction as a post-processor for interpreting decision trees has been assessed using a version of the well-known problem described by Magee [5]. Let us quote some periods of Magee's introduction.

> "Prism Paints Inc. must decide what to do with one of its manufacturing plants which is rather small and unable to supply the quality of products required in the current market. There is considerable managerial controversy over the proper course of action — whether to modernize the operation by construction of better facilities at the location, or to scrap the existing plant and supply the area involved from the company's facilities elsewhere. (...)"
>
> "There are three basic patterns of of operation offering promise:

Program A: To modernize the plant in question and also expand elsewhere. This program is less expensive when annual demand is less then a known threshold.

Program B: To close the plant in question and expand elsewhere. This program is less expensive when annual demand is between two known values.

Program C: To modernize and expand the plant in question. This program is less expensive when demand is above a known threshold."

> "The possible alternatives involve significantly different capital expenditures and appear to lead to substantial differences in operating economy as well. Underlying the controversy over what to do, is a concern for future product demand in the area served by the plant. (...)"
>
> "Evidence on market demand and the data underlying the demand forecasts lead to an estimation of how likely it is that the demand will fall in the low, medium, or high range in each of the three stages of the project. These estimations must be made in relation to the demand in the preceding stage. That is, *demands are conditional on previous stage's demand*."
>
> "The cash flows are estimated from the types of marketing, operations, engineering, and financial analysis mentioned elsewhere. In our Prism Paints case, distribution analysis show that the relative operating profitability of the alternatives facing management can be expressed by net annual cash flows in terms of demand levels. (...)"

Level of Demand	Stage 1	Stage 2 If Stage 1 is:			Stage 3 If Stage 2 is:		
		Low	Medium	High	Low	Medium	High
Low	.50	.35	.15	0	.20	.05	0
Medium	.43	.50	.45	.40	.60	.35	.20
High	.07	.15	.40	.60	.20	.60	.80

Table 2: Probabilities that demand will be low, medium and high

> "The management of Prism Paints establishes a desired return on investment
> or cost of capital of 14% per year. This is the rate to be used to reduce future
> cash flows to a present value for comparative purposes."

Knowing the initial investment each alternative involves, and establishing a required return on investment, it is possible to express outcomes by its Net Present Value (see [1] or a text on financial management like [7] for further development). In our case, NPV were not taken from the original model. Simulated values were created and checked to avoid dominance or other intuitive or obvious decisions.

Table 3 identifies the 81 NPV by its attributes. Under normal conditions, it is unusual to have only high positive NPV in a project. Such a situation would lead to sharp competition and profits would come down quickly. The real alternatives are between a positive NPV project and a negative one. NPV seems to be a very appropriate class attribute for Rule Induction, since only two situations are really important. In our case, however, we allowed NPV to have three possible outcomes ("Negative", "Positive" and "Extra Profits") in order to make it more complex. The majority of outcomes (60%) were negative.

In our reduced version the three choices under discussion, Program A, B or C, remain unchanged. Also similar to the original are the three stages, each with an uncertain demand level (low, medium and high). Thus the resulting decision tree has $3^4 = 81$ leaves. All other refinements, like the allowances for program shifts, have not been considered.

No prior trends about the final decision were introduced in the model: The three programs are equally considered if no other information is available.

4.1   Discussion of Results

Rule Induction has been applied to this set of 81 observations. The likelihood of each of them were interpreted as a proportion of observed frequency. Appendix 2 displays the used methodology.

DEMAND			OUTCOME (NET PRESENT VALUE)		
Year 1	Year 2	Year 3	Program A	Program B	Program C
High	High	High	Positive	Very Positive	Very Positive
High	High	Medium	Positive	Positive	Positive
High	High	Low	Positive	Negative	Negative (*)
High	Medium	High	Positive	Positive	Positive
High	Medium	Medium	Positive	Positive	Positive
High	Medium	Low	Negative	Negative	Negative
High	Low	High	Positive	Negative	Negative (*)
High	Low	Medium	Negative	Negative	Negative (*)
High	Low	Low	Negative	Negative	Negative (*)
Medium	High	High	Positive	Positive	Positive
Medium	High	Medium	Positive	Positive	Positive
Medium	High	Low	Negative	Negative	Negative (*)
Medium	Medium	High	Positive	Positive	Positive
Medium	Medium	Medium	Positive	Positive	Positive
Medium	Medium	Low	Negative	Negative	Negative
Medium	Low	High	Negative	Negative	Negative
Medium	Low	Medium	Negative	Negative	Negative
Medium	Low	Low	Negative	Negative	Negative
Low	High	High	Positive	Negative	Negative
Low	High	Medium	Negative	Negative	Negative
Low	High	Low	Negative	Negative	Negative (*)
Low	Medium	High	Negative	Negative	Negative
Low	Medium	Medium	Negative	Negative	Negative
Low	Medium	Low	Negative	Negative	Negative
Low	Low	High	Negative	Negative	Negative
Low	Low	Medium	Negative	Negative	Negative
Low	Low	Low	Negative	Negative	Negative

Table 3: Prism Paints Inc: Original values of the NPV to each program for different demands during the three years of the project.

How can managers profit from the resulting rule tree? Clearly, in three ways: Rule Induction performed a *flexibility quantification*, a *complexity quantification* and, as Race and Thomas already pointed out, a *sensitivity analysis*. These features are consequence of a *complexity reduction* capability (see figure 2).

To understand why these features can be useful to managers it is convenient to remember what managers actually get from a decision tree. Managers get an optimal path, a figure which is an optimum and nothing else. No information is available about the robustness of that particular solution; nothing can be said about flexibility during the process. An optimum is a static parameter. The problem, for managers, is also dynamic. They must calculate the consequences of changing plans during the process, or the robustness of the outcome to changes in conditions.

Rosenhead *et alt.* describe these limitations of decision analysis [9]. Optimality is not the only important decision criteria: Flexibility and Robustness, between other qualities, are also considered by managers. Rule Induction performs a transformation that seems to be able to display dynamic qualities of the underlying rule tree.

Basically, we classify the useful features of Rule Induction in five items.

Complexity Reduction: The original 81 leaves have been reduced to 20 rules which bear the same information. Lots of redundancy was removed. A more simple interpretation of the whole problem is thus expected.

Flexibility Quantification: In a decision tree there are *decision attributes*, which correspond to choices managers can make, and *chance attributes* corresponding to uncertain events. ID3 hierarchically scales attributes according to its importance to the outcome. So, the relative position of decision and chance attributes is related with the degree of control managers have over the developing sequential process.

In our example there is only one decision attribute, the choice of the program. Since the question is whatever to choose program A, B, or C, branches in which this attribute appears *far away* from the root or even doesn't appear at all, show that manager's decision is irrelevant to the outcome and therefore they have little control over it.

An irrelevant decision means a *not* flexible path: Managers, whatever they do, are likely to have to face a pre-determined outcome. Their intervening power over the running process is little. Conversely, a relevant decision means flexibility, for the outcome is mostly determined by decisions managers can take during the running process, not by events they don't control.

In our case, a low demand in the first stage leads to important differences between

programs, while for a medium or a high demand the chosed program doesn't seems to influence the outcome very much. The decision attribute doesn't appear in first place and this means that whatever managers decide, first year demand is what really will condition the results. According to this, managers could emphasize investment related with promotion and advertissment instead of trying to force the chance investing in productive assets. Of course, our example is a somehow pessimistic scenario, with lots of negative NPV and only one decision attribute.

Quantification of the complexity: In our rule tree, some branches lead to "bushy" zones and others to simple ones. The former developed many leaves and several attributes before identify unique NPV; the later, on the contrary, are quite simple leading to a unique NPV after considering a few attributes.

Rule induction can separate paths leading to complex sets of possibilities from others leading to more simple ones. Until now, little attention has been given to the *complexity of uncertainty* in managerial decision theory. We think that simple branches can be viewed as bearing an important quality for managers. In fact, the uncertainty in branches leading to simple outcomes is caused by few attributes. It is *less dimensional* then the one found in paths leading to bushy zones. There are less degrees of freedom to consider when going along simple paths.

In our case, a high or medium demand in the first stage leads to bushy branches. Positive NPV is reached after complicated paths, while negative outcomes are very simple. This is a useful warning for managers.

Sensitivity Analysis: Rule Induction also performs a sort of sensitivity analysis over the decision tree. Managers would be interested in the order by which attributes enter the model. The first attribute is the most influential and so on. For Prism Paints project, the demand in the first year is the most influential attribute. More then new plants, what Prism really seems to need is promotion. The program choice only appears to be influential for a low first stage demand, which is a good piece of information for managers to think about: The investment doesn't seem to fit in the demand levels.

Rules: On a first thinking, rules are more attractive for direct interpretation then decision trees. But twenty of them, all together, don't seem to be interpretable by managers. However, they can be fed into an expert system which, along with other information, would provide support to decisions both in the beginning and during the development. Concepts like those pointed out above could be quantified by this system. Expertise could work out a very useful set of suggestions.

```
                                      ⌐  = B, NPV = 1
                                      │
                 ⌐ = B,   P 2         │  = M, NPV = 1                      ⌐ = M, NPV = 1
                 │                    │                    ⌐ = "A",  D 3  │
                 │                    └  = A,  Prog.        │              └ = A, NPV = 2
                 │                                          │  = "B", NPV = 1
                 │                                          └  = "C", NPV = 1
                 │
                 │                    ⌐ = B, NPV = 1
                 │                    │
    P 1          │ = M,   P 2         │  = M,    P 3        ⌐ = B, NPV = 1
                 │                    │                    │  = M, NPV = 2
                 │                    └ = A, NPV = 2        └ = A, NPV = 2
                 │
                 │                    ⌐ = B, NPV = 1
                 │                    │
                 └ = A,   P 3         │  = M, NPV = 2
                                      │
                                      └ = A,    P 2        ⌐ = M, NPV = 2     ⌐ = "A", NPV = 2
                                                           └ = A,   Proj.    │ = "B", NPV = 3
                                                                             └ = "C", NPV = 3
```
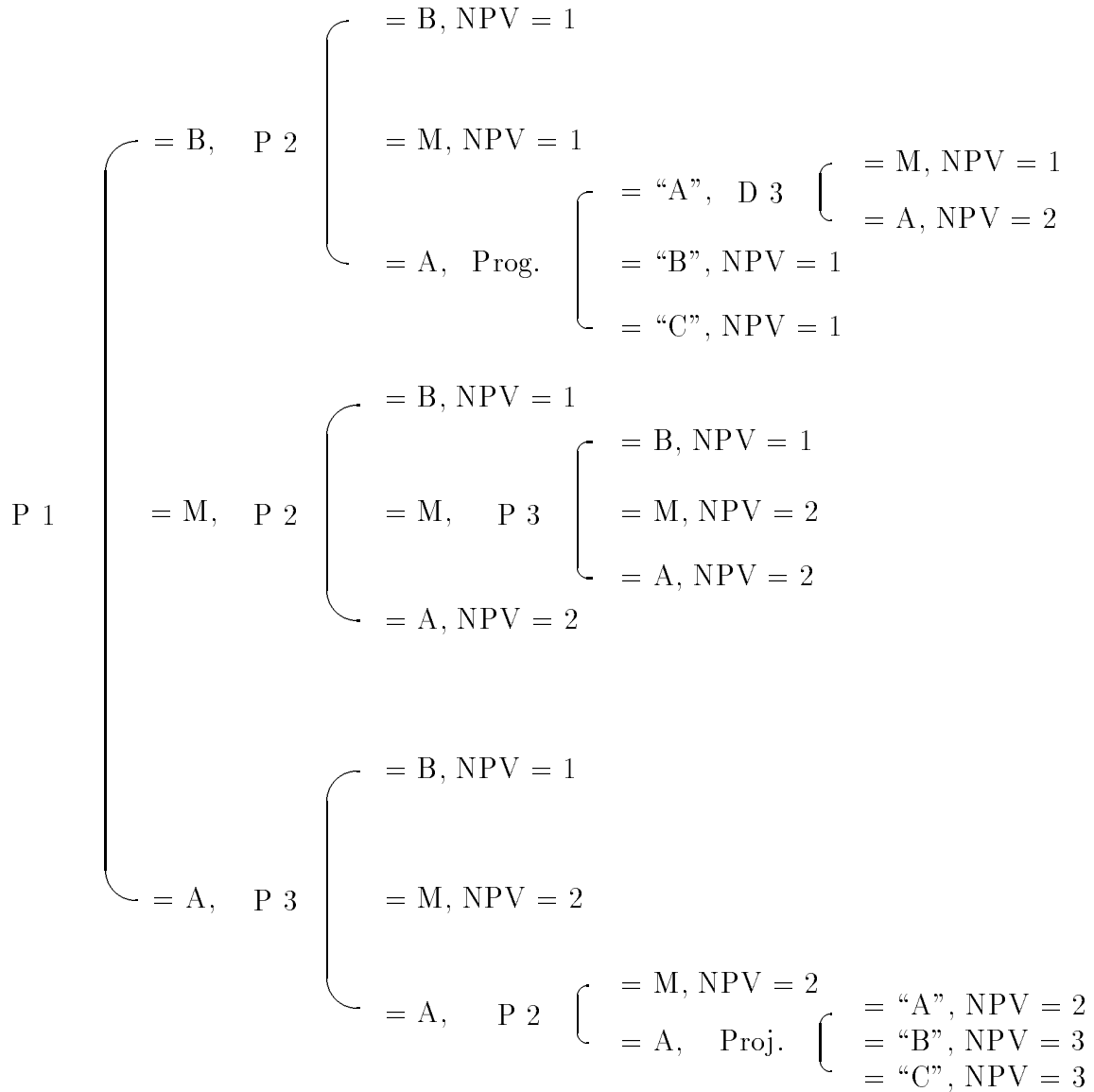
Figure 2: Prism: Resulting Rule Tree

This simple experiment seems to show that Rule Induction can be useful as a *Post Processor* for sequential decision models. Its complexity reduction capacity is welcome. And the way ID3 scales attributes can provide managers with new important measures of a project's dynamic behaviour. On the other hand, there is no danger, in this sort of problems, of inducing misleading relations: No generalization took place.

References

[1] R. Ashford, R. Berry, and R. Dyson. Operational research and financial management. *European Journal of Operational Research*, 36:143–152, 1988.

[2] J. Friend and W. Jessop. *Local Government and Stratergic Choice*. Tavistock, London, 1969.

[3] S. Gupta and J. Rosenhead. Robustness in sequential investment decisions. *Management Sciences*, 15:B–18, 1968.

[4] F. Hespos and P. Strassman. Stochastic decision trees for the analysis of investment decisions. *Management Science*, 11:224–259, 1965.

[5] J. Magee. How to use decision trees in capital investment. *Harvard Business Review*, pages 79–96, September 1964.

[6] J. Quinlan. Discovering rules from large collections of samples — a case study. In D. Michie, editor, *Expert Systems in the Micro Electronic Age*. Edimburgh University Press, 1979.

[7] Brealey R. and Myers S. *Principles of Corporate Finance*. McGraw-Hill Book Co., fourth edition, 1981.

[8] J. Rosenhead. Planning under uncertainty. *Journal of the Operational Research Society*, 31:209–216, 1980.

[9] J. Rosenhead, M. Elton, and S. Gupta. Robustness and optimality as criteria for stratergic decisions. *Operational Research Quarterly*, 23(4):413–431, 1972.

[10] S. Stevens. On the theory of scales of measurement. *Science*, 103:677–680, 1946.