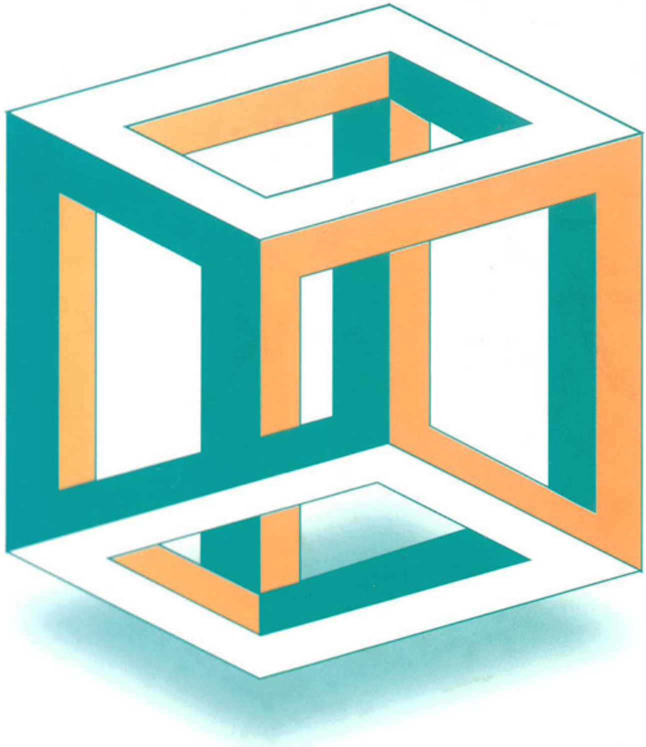
Fuzzy Sets in Management, Economics and Marketing



Editors

Constantin Zopounidis
Panos M. Pardalos
George Baourakis

World Scientific

Fuzzy Sets in Management, Economics and Marketing

This page is intentionally left blank

Fuzzy Sets in Management, Economics and Marketing

Editors

Constantin Zopounidis
Technical University of Crete, Greece

Panos M. Pardalos
University of Florida, USA

George Baourakis
Mediterranean Agronomic Institute of Chania, Greece

World Scientific
New Jersey • London • Singapore • Hong Kong

Published by

World Scientific Publishing Co. Pte. Ltd.

P O Box 128, Farrer Road, Singapore 912805

USA office: Suite 1B, 1060 Main Street, River Edge, NJ 07661

UK office: 57 Shelton Street, Covent Garden, London WC2H 9HE

Library of Congress Cataloging-in-Publication Data

Fuzzy sets in management, economics, and marketing / edited by Constantin Zopounidis,
Panos M. Pardalos, George Baourakis.

p. cm.

"This book is prepared on the occasion of the 7th international congress of SIGEF"--Editorial

Includes bibliographical references and index.

ISBN 9810247532 (alk. paper)

1. Decision making--Mathematical models. 2. Marketing--Decision making--Mathematical models. 3. Investments--Decision making--Econometric models. 4. Fuzzy sets. I. Zopounidis, Constantin, 1958– . II. Pardalos, P. M. (Panos M.), 1954– . III. Baourakis, George.

HD30.25 .F89 2001
658.4'033--dc21 2001045590

British Library Cataloguing-in-Publication Data

A catalogue record for this book is available from the British Library.

Printed in Singapore by Mainland Press

To our children:

Dimitrios Zopounidis
Miltiadis Pardalos
Niovi and Manos Baourakis

This page is intentionally left blank

There is only one good, the right knowledge and only one evil, the illiteracy

SOCRATES
(Diog, Laert. Vioi Fil. II31)

This page is intentionally left blank

CONTENTS

EDITORIAL

The most salient feature of the late 20th century is the scale of changes affecting social, economic and corporate life. Our environment is changing at a rate that would once have been inconceivable. Nowadays, the speed of events has become astonishing. The problems posed by these new situations are increasingly complex and changeable and traditional models based on determinism and chance are no longer able to cope with this reality. The fuzzy sets theory constitutes an innovative approach to handle this increasing complexity of corporate management problems. The continuous advances made within the context of the fuzzy sets theory, raise the need to bring together all those individuals, companies and institutions which, from an educational, research or professional point of view, wish to collaborate in the development and promotion of the new techniques with academic researchers, students as well as executives.

Based on the above, the INTERNATIONAL ASSOCIATION FOR FUZZY-SET MANAGEMENT AND ECONOMY (SIGEF) has been set up to encourage research and study relating to all aspects of the economy in general, and corporate management in particular. In fulfilling its goals, SIGEF organizes an international congress on an annual basis. Since the first congress held in Reus, Spain, in 1994, six congresses have been held in Santiago de Compostela (Spain, 1995), Buenos Aires (Argentina, 1996), Santiago de Cuba (Cuba, 1997), Lausanne (Switzerland, 1998) and Morelia (Mexico, 1999).

This book is prepared on the occasion of the 7th international congress of SIGEF, held in Chania, Greece. The objective of the book is to present the recent advances in the modeling of the fuzziness and uncertainty that prevail in the modern economic and management environments. In pursuing this objective the paper included in the book propose a wide range of diversified methodological approaches, including the fuzzy sets theory, neural networks, genetic algorithms, machine learning, optimization, data analysis, multivariate statistics and econometrics. Special focus is given on the real-world applications of these advances in addressing economic, marketing, financial and management problems.

Overall, the book is organized into three major parts. The first part involves decision making, management and marketing, the second part deals with stock markets and portfolio management, whereas the third part of the book is devoted to corporate finance and banking management (credit scoring and asset/liability management).

The first part of the book includes eight papers. The first paper by Gil-Aluja presents algorithms for generalising the latticed schemes by means of configurations that do not require the conditions of a lattice, within the context of financial decision making. The second paper by Arenas, Bilbao, Jiménez and Rodríguez Uría involves the hospital management problem. The authors propose the use of a fuzzy goal programming model for the management of the clinical and surgical operations of a hospital in terms of the goals of providing quality service, minimizing costs and maintaining job allocation. The paper of García-Lapresta, Martínez-Panero and Lazzari introduces a Borda-type group decision making procedure. The authors employ a fuzzy approach to preference modelling, based on the use of triangular fuzzy numbers. The fourth paper by Doumpos and Zopounidis involves the development of sorting models using the preference disaggregation of multicriteria decision aid. The authors present the results of a simulation study that investigates some important issues regarding the use of preference disaggregation analysis for decision making in sorting problems. The paper of Koutsoupias and Papadimitriou presents the application of a statistical multidimensional data analysis clustering method for processing and managing life insurance contract portfolios, in order to reveal the trends, formations and groupings that emerge directly from life insurance contract, taking into account all life contract parameters. The paper of Dimara, Baourakis and Kalogeras presents the results of an econometric analysis to identify the factors that influence consumers to highly value extrinsic (packaging, certification) as opposed to intrinsic characteristics (aroma), for quality wine, a very important Greek product. The results of the analysis provide useful guidelines to firms in deciding whether they can use extrinsic cue differentiation to target market segments. The paper of Baourakis, Drakos and Apostolakis uses a similar data analysis methodology to identify the market segments of juices and the factors that influence the buying behavior of juice consumers in Greece. The study deals with consumer preferences, attitudes and perceptions with regard to the special characteristics pertaining to juices such as packaging, color, price, taste, and advertisement. Finally, the paper by Baourakis presents the results of a market survey to provide a thorough insight into the behavior, attitudes and knowledge of tourism enterprises concerning Cretan agricultural products.

The second part of the book, which involves stock markets and portfolio management, includes five papers. The first paper by De Andrés, Glòria Barbera and Terceno deals with the construction of fixed income portfolios. The methodology proposed by the authors is based on the use of fuzzy mathematical programming. Gupta, Chevalier and Sayekt investigate the existence of a causality relationship between interest rates, exchange rates and stock prices within the context of an emerging market, namely the Jakarta stock exchange. The third paper by Koulouriotis, Diakoulakis and Emiris addresses the stock market forecasting problem proposing a new modelling approach based on the use of fuzzy cognitive maps. The paper by Kanas compares the forecasting performance of linear and nonlinear neural network models of monthly aggregate stock returns in order to examine whether forecasts from a nonlinear stock returns model are preferable to forecasts from a linear stock returns model. Finally, the paper of Tsakonas, Dounias and Merikas proposes the use of genetic algorithms for constructing fuzzy rule bases, as part of an hybrid decision support architecture, involving neural networks for wavelet-filtered daily stock rates of change.

The third and last part of the book is devoted to corporate finance and banking management problems. This part includes four papers. The first paper by Couturier and Fioleau deals with the corporate performance assessment problem and in particular with the evaluation of the adequacy of the decisions taken by companies with regard to the corporate objectives. The methodology proposed by the authors is based on the fuzzy sets theory and the theory of expertons. The second paper by González, Flores R.J., Flores R.B. and Mendoza presents a new fuzzy methodology to determine multiple IRRs in investment decision problems. The third paper by Michalopoulos, Dounias, Hatas and Zopounidis present the application of a machine learning methodology in developing a credit scoring model for corporate credit risk assessment purposes. Finally, the paper of Kosmidou and Zopounidis presents an overview of the bank asset-liability management techniques that have been developed and used over the last 20 years. The review involves both deterministic and stochastic approaches to bank asset-liability management.

Sincere thanks must be expressed to the authors whose contributions have been essential in creating this volume. We are also grateful to the many referees of all papers, whose thoughtful critical reviews played a decisive role in improving the quality of this book.

Finally, we would like to thank Dr. Michael Doumpos for his administrative assistance in communicating with the authors and coordinating the material presentation of this volume.

June 2001

Prof. Constantin Zopounidis
Technical University of Crete, Dept. of Production Engineering and Management,
Financial Engineering Laboratory, University Campus,
73100 Chania, GREECE

Prof. Panos M. Pardalos
University of Florida, Dept. of Industrial and Systems Engineering, 303 Weil Hall,
Gainesville, FL 32611-6595, U.S.A.

Dr. George Baourakis
Mediterranean Agronomic Institute of Chania, Department of Economic and
Management Sciences, PO. Box: 85, 73100Chania, GREECE

DECISION MAKING, MANAGEMENT AND MARKETING

This page is intentionally left blank

ALGORITHMS FOR ORDERLY STRUCTURING OF FINANCIAL "OBJECTS"

JAIME GIL-ALUJA
University of Barcelona, Barcelona, Spain

Uncertainty, which is all the time becoming more obvious within social life, economy and management, makes it progressively more difficult to use numerical mathematics for arrivng at a solution to the multiple problems of decision appearing in financial activities. The attempt to incorporate non numerical mathematics in order to be able to establish a new theory of decision in uncertainty [1] has opened up for research in the different areas of knowledge, such as, banking knowledge. The reformulation of concepts such as relation, assignment, grouping and order have been the starting out point for new lines of research which we feel will be most fruitful. In conjunction with Professor Kaufmann [4] we have worked, for many years, in order to construct a body, as homogenious as possible, able to allow optimum groupings of physical or mental objects. For this we drew up what, over the years, has become to be known as the theory of affinities. In this theory, and as a culmination, we introduced the Galois lattices, a simple and elegant means of presenting, in a wholly structured manner, those groups which had been previously found. On this occasion what we intend to do is to present certain algorithms that are capable of generalising the latticed schemes by means of configurations that do not require the conditions of a lattice. In this way we feel that financial decisions will be made easier by the presence of structured and ordered information in a manner that is adequate for the requirements of the decision subject.

1 Presentation of the problem

Often arising within the sphere of banking is the need to select a placement of financial means, from among several possibilities, or to decide among several clients which should be the beneficiary or beneficiaries of a credit or loan. In either case, it would appear to be recommendable to establish an order of priority that allows for providing a whole range of possibilities. But this does not only occur by means of a simple ordering but requires to be structured.

With the object of "visualising" this process we have resorted on several occasions to the use of lattice forms. Now, as is sufficiently well known, the lattice is a particular case of algebraic structure. Our objective is to generalise, wherever possible, the algorithms that have been used up to now in order for them to become of greater value.

Let us recall that every lattice can be defined either as an algebraic structure, or as an ordered set.

In the first case a set T and two laws of composition in T are considered. If:

$$\forall\ x, y, z \in T$$

the properties of commutativity, associativity, idempower and absorption are complied with and in this case T possesses the configuration of a lattice[1].

In the second we start out from an ordered set represented by T and an order relative to T.

It can be said that an ordered set is a lattice if all the sub-sets of its elements {x, y}, x, y ∈ T, possess a lower and a upper point.

$$T \text{ is a lattice} \Leftrightarrow (\forall\ x, y \in T, \exists\ (x \Delta y) \in T \text{ y } \exists\ (x \nabla y) \in T)$$

We are now going to eliminate these latter two conditions, with which the algorithms that are presented will cover a greater range of situations which may arise in daily banking activities.

2 Algorithm for arriving at the HASSE diagram

In a recent work [2] we proposed an algorithm for ordering taking as its primary support the properties of relations. For this the concept was defined of a chain of an ordered set as follows:

"If E is a set ordered by means of a relation of order R, it can be said that a sub-set A that is not void of E is a **"chain"** when it is totally ordered by R."

A chain that is not contained in any other is defined as the**"maximum chain".**

We are now in a position to resort to an element of matrix analysis, that can provide excellent results. We are referring to the "Hasse diagram".

In order to find the Hasse diagram, we start from a relation of order which can be represented by a matrix R. We then continue with the following steps:

1) List of all the chains.
2) Elimination of those chains that are contained in others. In this way we arrive at the maximum chains.
3) Vertical placing (u horizontal, if necessary) such as placing all those ordered chains containing common elements one alongside the other.
4) Formation of a single chain, joining the parts of identical chains with a single line.
5) The resulting diagram is a Hasse diagram.

With the object of illustrating this simple algorithm we are going to start from a relation of order (in this case strict) given by the following Boolean matrix:

[1] If total strictness is required in the axiomatic construction, the requirement that the axioms be independent must be complied with. Because in reality the axioms include idempower. Now, since the custom exists of listing the eight axioms, we have done it in this way also, even knowing that with six it is sufficient.

$$R = \begin{array}{c} \\ a \\ b \\ c \\ d \\ e \\ f \\ g \\ h \\ i \\ j \end{array} \begin{array}{c} \begin{array}{cccccccccc} a & b & c & d & e & f & g & h & i & j \end{array} \\ \left[\begin{array}{cccccccccc} & 1 & & 1 & & 1 & & & & \\ & & & 1 & & 1 & & & & \\ & & & 1 & 1 & & & & & 1 \\ & & & & & & & & & \\ & & & 1 & & & & & & \\ & & & & & & & & & \\ 1 & 1 & & 1 & & 1 & & & 1 & \\ & & & 1 & 1 & & & & & 1 \\ 1 & 1 & & 1 & & 1 & & & & \\ & & & 1 & 1 & & & & & \end{array}\right] \end{array}$$

1) We now list all the chains:

$a \prec b \prec d$
$a \prec b \prec f$
$a \prec d$
$a \prec f$
$b \prec d$
$b \prec f$
$c \prec d$
$c \prec e \prec d$
$c \prec j \prec d$
→ $c \prec j \prec e \prec d$
$e \prec d$
$g \prec a \prec b \prec d$
$g \prec a \prec b \prec f$
$g \prec a \prec d$

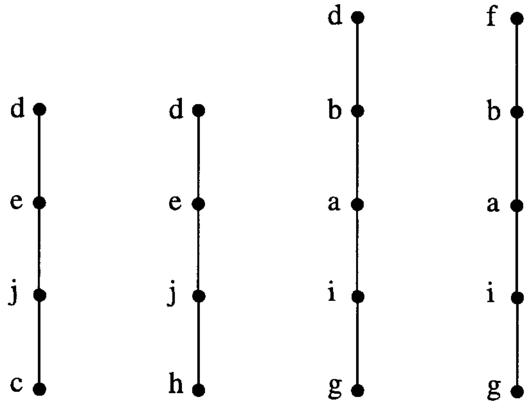
$g \prec a \prec f$
$g \prec b \prec d$
$g \prec b \prec f$
$g \prec d$
$g \prec f$
→ $g \prec i \prec a \prec b \prec d$
→ $g \prec i \prec a \prec b \prec f$
$g \prec i \prec a \prec d$
$g \prec i \prec a \prec f$
$g \prec i \prec b \prec d$
$g \prec i \prec b \prec f$
$g \prec i \prec d$
$g \prec i \prec d$
$g \prec i \prec f$
$h \prec d$

$h \prec e \prec d$
$h \prec j \prec d$
→ $h \prec j \prec e \prec d$
$i \prec a \prec b \prec d$
$i \prec a \prec b \prec f$
$i \prec a \prec d$
$i \prec a \prec f$
$i \prec b \prec d$
$i \prec b \prec f$
$i \prec d$
$i \prec f$
$j \prec d$
$j \prec e \prec d$

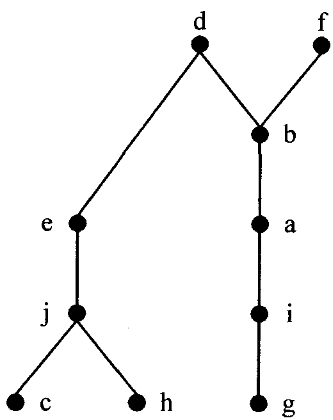
2) Those chains contained in others are now eliminated and we are left with:

$c \prec j \prec e \prec d$
$g \prec i \prec a \prec b \prec d$
$g \prec i \prec a \prec b \prec f$
$h \prec j \prec e \prec d$

3) We now place these chains vertically and put those chains possessing elements of other chains alongside each other:



4) We form a single chain:



5) This is the Hasse diagram that establishes a relation of order between the elements of the referential set {a, b, ..., j}.

3 Algorithm based on the p-latin composition

In one of our earlier works [1] we proposed an algorithm for order based on the P-latin composition. On this occasion and taking as a reference the P-latin convolution of matrices we are going to show a variation that will lead us to the same result as that arrived at by the Hasse diagram. Below we describe the steps to be followed:

1) Transformation of the bolean matrix of relations into a latin matrix $[L]^1$.
2) From the latin matrix $[L]^1$ we find the latin matrix amputated to the left [L'].
3) The latin convolution between matrix $[L]^1$ and matrix [L'] where property P = elemental path. We now arrive at latin matrix $[L]^2$.

4) Latin convolution between matrix $[L]^2$ and matrix $[L']$. In this way we find matrix $[L]^3$.
5) The convolution process is repeated until latin matrix $[L]^{r-1}$ is arrived at, r being the cardinal of the referential set, unless prior to this a void matrix were to have been found, in which event the process would have stopped at that point.
6) The **non void boxes** of matrix $[L]^{r-1}$ provide us with the relations of maximum order.
7) Comparing these maximum chains (relations) with those of the latin matrices $[L]^{r-2}$, $[L]^{r-3}$, ... by annulling those chains of the previous matrices that are contained within others.
8) The remaining relations constitute the relations of order that we are looking for.

We are now going to use this algorithm with the same relation of order [R] that we used for the previous algorithm.

1) We transform Boolean matrix [R] into latin matrix $[L]^1$. We arrive at:

		a	b	c	d	e	f	g	h	i	j
	a		ab		ad		af				
	b				bd		bf				
	c				cd	ce					cj
	d										
$[L]^1 =$	e				ed						
	f										
	g	ga	gb		gd		gf			gi	
	h				hd	he					hj
	i	ia	ib		id		if				
	j				jd	je					

2) We find the amputated latin matrix to the left $[L']$. In our case it is the following:

		a	b	c	d	e	f	g	h	i	j
	a		b		d		f				
	b				d		f				
	c				d	e					j
	d										
$[L'] =$	e				d						

f										
g	a	b		d		f			i	
h				d	e					j
i	a	b		d		f				
j				d	e					

3) On making the latin convolution $[L]^1$ or [L'] we arrive at matrix $[L]^2$ we have reproduced below:

$[L]^2 =$

	a	b	c	d	e	f	g	h	i	j
a				abd		abf				
b										
c				ced cjd	cje					
d										
e										
f										
g	gia	gab gib		gad gbd gid		gaf gbf gif				
h				hed hjd	hje					
i				iad ibd		iaf ibf				
j				jed						

4) We continue the process of convolution with $[L]^2$ or [L'] and now arrive at $[L]^3$.

$[L]^3 =$

	a	b	c	d	e	f	g	h	i	j
a										
b										
c				cjed						
d										
e										
f										
g		giab		giad		giaf				

	a	b	c	d	e	f	g	h	i	j
				gabd gibd		gabf gibf				
h				hjed						
i										
j										

5) The latin convolution $[L]^3$ or $[L']$ results in the following matrix $[L]^4$

$[L]^4 =$

	a	b	c	d	e	f	g	h	i	j
a										
b										
c										
d										
e										
f										
g		giabd				giabf				
h										
i										
j										

Taking into account the fact that the convolution of $[L]^4$ or $[L']$ provides $[L]^5 = [\varnothing]$, that is to say a void matrix, we stop the process.

6) The boxes that are not empty of matrix $[L]^4$ provide us with the maximum chains:

$$g \prec i \prec a \prec b \prec d$$
$$g \prec i \prec a \prec b \prec f$$

7) Comparing the chains from matrix $[L]^4$ with those from matrix $[L]^3$ we only find the following that are not contained in the previous two we arrived at. These are:

$$c \prec j \prec e \prec d$$
$$h \prec j \prec e \prec d$$

The comparison of $[L]^4$ and $[L]^3$ with $[L]^2$ and $[L]^1$ does not give us chains that are not contained in those four that we have retained.

8) The relations of order, we were looking for, are:

$$g \prec i \prec a \prec b \prec d$$
$$g \prec i \prec a \prec b \prec f$$
$$c \prec j \prec e \prec d$$
$$h \prec j \prec e \prec d$$

It is obvious that these chains coincide with those found by following the algorithm of the previous section thus we can immediately construct the Hasse diagram.

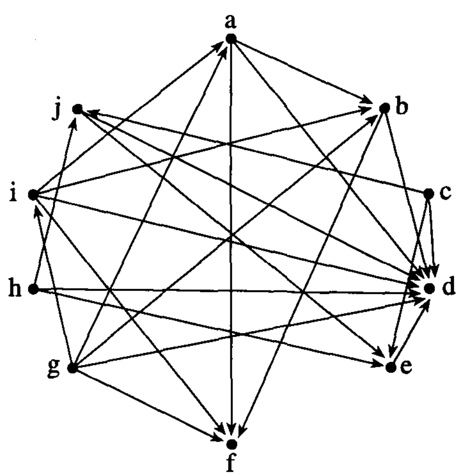
4 Algorithm based on the concept of ordinal function

We could wonder now what happens if instead of using one of the algorithms we have shown, we resorted the classical one built on the basis of the notion of ordinal function, in one of its many variations[2]. For our presentation we have selected the algorithm in its arrow form, the steps of which are reproduced below [5]:

1) Establish the graph in the arrow form.
2) Look for the vertices without predecessors (those to which no arc arrives) which form level N_0.
3) In the graph, mark all the vertices that are members of level N_0 and eliminate the arcs leaving from the same. In this way we arrive at another graph.
4) In the new graph we have arrived at, look for the vertices without predecessors. With these we form level N_1.
5) Return to point 3) but without the vertices relative to level N_1, and so on successively until emptying the graph.
6) With the disappearance of the graph we arrive at the sought after order, given by levels $N_0, N_1, N_2, ..., N_s$.

We will now move on and use this algorithm on our matrix [R] that represents the relations between the elements of referential {a, b, c, ...,j}.

1) We transform the matrix relation [R] into its corresponding arrow form. Thus arriving at:

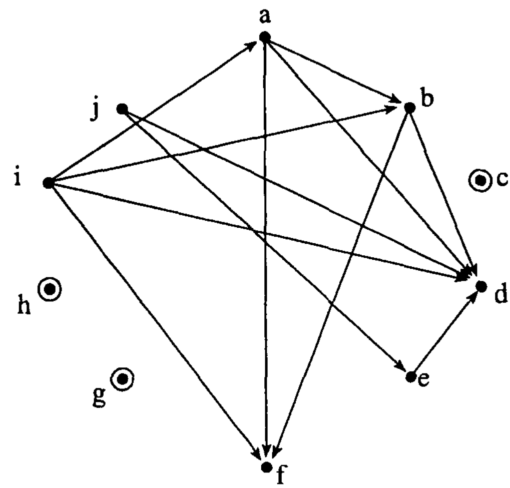


[2] We are referring to the work by Kaufmann, A. and Gil-Aluja, J., developed from an arrow form graph and to that of Demoocrom, M., shown in the matrix form.

2) The vertices without predecessors are c, g, h that form level N_o:

$$\{c, g, h\}$$

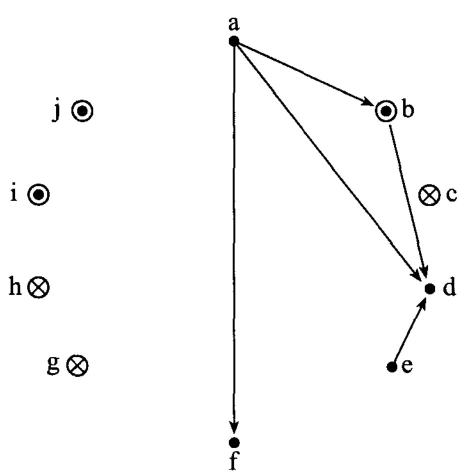
3) Vertices c, g, h are marked with a circle and we eliminate the arcs leaving from the same. We finally, arrive at the following graph:



4) The new vertices without predecessors are now i, j, which form the next level N_1:

$$\{i, j\}$$

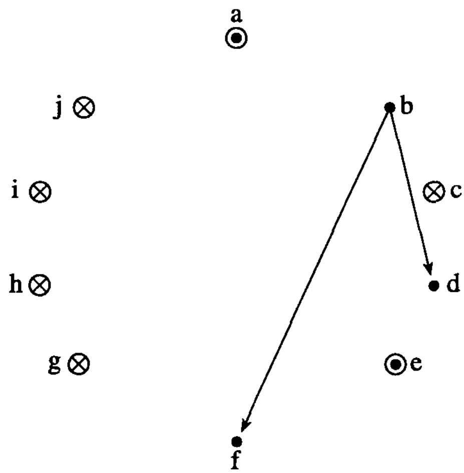
5) We now mark these with a circle. In order to distinguish these from those relative to level N_1 we cross out the latter. The result is then the following graph:



The new vertices without predecessors are a, e which form level N_2:

$$\{a, e\}$$

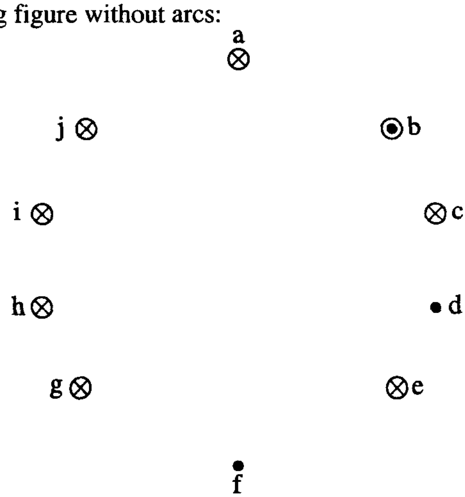
We mark these vertices and cross out the previous ones from level N_2. We now arrive at the following graph:



The new vertex without predecessor is b, with which we form the next level N_3:

$$\{b\}$$

When vertex b is marked and the arcs leaving from the same are eliminated we arrive at the following figure without arcs:



Finally, the last two vertices that have no predecessors are d, f which form the last level N_4:

$$\{d, f\}$$

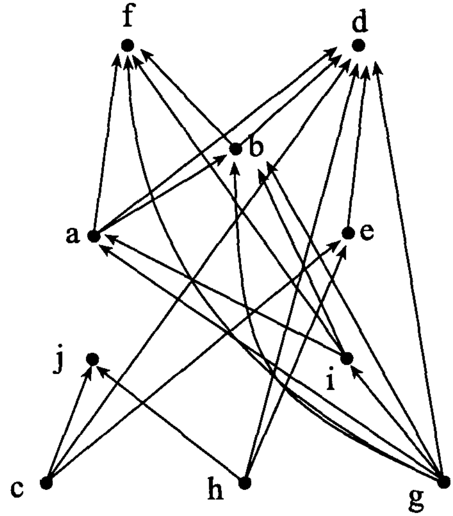
6) The graph has now vanished the levels are identified and with them the order we were looking for with the algorithm. This is as follows:

$$\{c, g, h\}\ ,\ \{i, j\}\ ,\ \{a, e\}\ ,\ \{b\}\ ,\ \{d, f\}$$

N_0 N_1 N_2 N_3 N_4

From these levels the arrow-form graph can be reconstructed corresponding to the originally considered matrix [R], but now, placing the vertices according to the ordinal function found with levels N_n, {n = 0, 1,..., 4}. For better visual comparison with the results arrived at by means of the use of the previous algorithms we will establish the order vertically from bottom to top, on the understanding that this could also be done from left to right, as has been done on may other occasions.

The graph associated to matrix [R], with the same arcs that the arrow-form possessed originally is the following, once recomposed according to the ordinal function:



At this point the algorithm normally used to find the ordinal function ends. But in order to find the maximum chains, a complementary step must be added consisting of eliminating all those arcs that join two vertices that are members of shorter paths, and only leaving the one with the most arcs. Therefore, for example, in order to go from **c** to **d** there are the following paths:

$$c \prec j \prec d$$
$$c \prec d$$
$$c \prec e \prec d$$
$$c \prec j \prec e \prec d$$

The first three are eliminated, leaving us with the maximum chain:

$$c \prec j \prec e \prec d$$

From **h** to **d** there are the following paths:

$$h \prec d$$
$$h \prec e \prec d$$
$$h \prec j \prec d$$
$$h \prec j \prec e \prec d$$

This contains the first three which, therefore must be eliminated. The maximum chain linking **h** with **d** is:

$$h \prec j \prec e \prec d$$

There are no paths linking **c** with **f** nor **h** with **f**.

The paths linking **g** with **f** are:

$$g \prec f$$
$$g \prec a \prec f$$
$$g \prec b \prec f$$
$$g \prec i \prec f$$
$$g \prec a \prec b \prec f$$
$$g \prec i \prec a \prec f$$
$$g \prec i \prec b \prec f$$
$$g \prec i \prec a \prec b \prec f$$

The first six are eliminated and we are left with the maximum chain of;

$$g \prec i \prec a \prec b \prec f$$

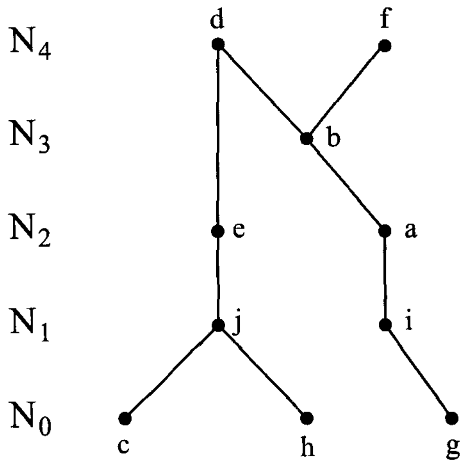
Finally the paths (chains) linking **g** with **d** are the following:

$$g \prec d$$
$$g \prec a \prec d$$
$$g \prec b \prec d$$
$$g \prec i \prec d$$
$$g \prec a \prec b \prec d$$
$$g \prec i \prec a \prec d$$
$$g \prec i \prec b \prec d$$
$$g \prec i \prec a \prec b \prec d$$

Here we are left with the maximum chain of:

$$g \prec i \prec a \prec b \prec d$$

With the eliminations we have proposed, the graph ordered by the ordinal function, in accordance with our algorithm remains as follows:



Here, once again we have the same Hasse diagram as the one found by using the other algorithms. Perhaps with the one we have just used it could be considered that the advantage relative to the other is that here the levels in which each of the vertices can be found appears very clearly. In our case this is knowing with all certainty at what level each element of the chains (vertices) should be placed. Here, the apparent problem can be found in determining at what level it is necessary to place vertex **e**: as we have done it correctly ourselves at the same level as vertex **a**, or (incorrectly) at the same level as **b**.

With the object of making the algorithm based on the ordinal function suitable for finding the Hasse diagram, it is necessary to add two new phases, which are as follows:

7) The initial arrow form graph is reconstructed by placing the initial arrow form graph according to the ordinal function.

8) Eliminate the arcs that make up the paths that join the same initial vertices with the same end vertices that are members of a shorter path, only leaving the one with the greatest number of arcs (greater length).

By linking these two stages with the six we described before, we have an algorithm that, in conjunction with those we showed earlier allow us, through different procedures (at least apparently different), to find the Hasse diagram, which is equivalent to the, basis for taking decisions within the sphere of non numerical mathematics.

5 Conclusion

Throughout our development we have started out, as we normally do, from a **relation of order**, expressed, in its matrix form. Later, we moved on to the arrow form or latin form according to the requirements of the algorithm which we intended to introduce. But, in whichever of these forms, it was always with a relation of order. We do not feel that it is necessary to insist on the fact that not all relations allow for an order, be this total or partial, strict or not strict. In a graph circuits may occur, that is to say that an object x is preferable to another object y, which at the same time is preferred to a third object z which is considered better than x. Therefore, between x, y, z it is not possible to establish any order whatsoever. In this case the algorithms we have proposed lack any use, if the corresponding relation is considered directly.

Fortunately, today we can count on sufficient theoretical and technical elements to provide us with a good solution to this problem. In fact, a graph with more than one circuit can be broken down [3] into sub-graphs that are strongly connected, or in the language of matrices, equivalency classes.

Obviously it is true to say that the objects forming each equivalency class do not admit any order among themselves (it is then said that they are indifferent or *equivalent*) *but on the other hand* ***it is*** *possible to order equivalency classes* or strongly connected graphs (circuits). And here once again appear the interesting possibilities of the algorithms we have presented.

The proposed algorithms constitute procedures that are destined to one and the same object: to structure an order, so as to provide not only information in itself, but also information included in an overall interconnected set. This allows for a complete viewing of possible financial decisions and the position of each one of them relative to the others.

In this way we have developed an idea which, revolving around the area of the concept of decision, is suitable for assisting in the perception, with far greater **clarity**, of the interest in non numerical mathematics of uncertainty for financial decisions.

References

1. Gil-Aluja, J.: Elements for a theory of decision in uncertainty. Kluwer Academic Publishers. Boston, London, Dordrecht, 1999.
2. Gil-Aluja, J.: "Entre la epistemología y la lógica. Decisión y Algoritmo de ordenación". [Between epistomology and logic. Decision and algorithm for order]. Work done in posthumous hommage to Profesor Emilio Soldevilla. Bilbao, 19 January 2001.
3. Kaufmann, A. y Gil-Aluja, J.: Nuevas técnicas para la dirección estratégica. [New techniques for strategic management]. Publications from the University of Barcelona, Barcelona 1991.
4. Kaufmann, A. y Gil-Aluja, J.: Técnicas de gestión de empresas. Previsiones, decisiones y estrategias. [Business management techniques. Estimates, decisions and strategies]. Publ. Pirámide. Madrid 1992.

5. Kaufmann, A. y Gil-Aluja, J.: Grafos neuronales para la economía y la gestión de empresas. [Neural graphs for economy and business management]. Publ. Pirámide. Madrid 1995.

This page is intentionally left blank

A FUZZY GOAL PROGRAMMING MODEL FOR EVALUATING A HOSPITAL SERVICE PERFORMANCE

M. ARENAS, A. BILBAO AND M.V. RODRÍGUEZ URÍA

Dpto. Economía Cuantitativa, Universidad de Oviedo, Spain

E-mail: vrodri@.econo.uniovi.es

M. JIMÉNEZ

Dpto. Economía Aplicada I., Universidad del País Vasco, Spain

E-mail: eupjilom@se.ehu.es

Recently we have developed a method to solve a linear goal-programming problem whose parameters are crisp and where both the constraints and the achievement of goals are flexible. In this paper we use this method to solve a model for evaluating a hospital service performance. The Decision Maker usually offers information that allows us to establish suitable values, in fuzzy terms, of the degree to which the goals and constraints are reached. Without distinction between goals and constraints, we will accept that a solution may verify each one with a certain degree of achievement, and the membership function of each fuzzy set describing them represents this degree. Then, we propose as problem solution the decision vectors x that maximize a global measure for the degree of achievement of goals and constraints. Using linear membership functions and, as aggregate function, the weighted mean, the initial fuzzy problem was converted into an equivalent goal programming problem that we solve in an interactive way. Our proposal is applied here to present a model to design the real performance of a surgical service at a local general hospital.

1 Introduction

In a previous paper [2] we presented a standard Goal Programming model to design the performance of a surgical service at a local general hospital. Later we proposed a method [4] to solve a linear goal-programming problem whose parameters are crisp and where both the constraints and the achievement of goals are flexible. That method establishes a direct relation between the standard Goal Programming [8, 13] and the fuzzy approach. Like other fuzzy approaches, our method allows the Decision Maker (DM) to establish the goals and the fulfilment of constraints in an imprecise way, i.e. the DM may understand that the objectives should be *"essentially greater than or equal to target g"*, *"approximately equal to h"* or *"essentially smaller than or equal to b"*. But what is more specific in our method is that the DM can express a different degree of satisfaction according to the deviation values, and he/she can establish a tolerance threshold for the deviation values. We think that those characteristics are very useful to model the surgical performance of a hospital.

The DM usually considers many criteria of a diverse nature in the decision process and, in general, some of them are at odds with to each others. So, in practice, many decision making problems are described by multi-objective programming models and sometimes it is necessary to formulate them with elements of imprecision or uncertainty. Usually, the simultaneous optimization of all objectives is impossible, so the DM should be willing to make flexible the attainment of objectives and the satisfaction of constraints. Therefore, it is reasonable to construct a model reflecting imprecise data or flexibility in terms of fuzzy sets.

The fuzzy programming approach to multiobjective programming problems was first introduced by Zimmerman [17] and further developed by other authors: i.e. [1, 3, 9, 15]. It is possible also to find in literature diverse formulation of GP models reflecting imprecision on problem formulation, i.e. [7 ,12].

The first step in the formulation of a general GP model is to establish a set of target values i.e., the achievement level desired for each objective considered in the decision making situation. Once the target value is fixed for each objective, this becomes a goal. Usually the DM is the one who provides those numerical values in order to set an acceptable level of achievement for any of the attributes considered. Then deviations not desired with respect to those values must be minimized. In section 2 we will present our method to solve a GP model. In section 3 we will use this method to improve the performance of a surgical service at the local hospital that provided us with data.

2 Flexible approach to a goal programming problem

A Fuzzy Multi-Objective Linear Programming can be presented as the following system [16]:

$$\begin{array}{lll} & \textit{Find } x & \\ \text{s.t.} & c_i x \tilde{\geq} g_i & i = 1,\dots,q \\ & c_i\, x \cong h_i & i = q+1,\dots,k \\ & a_i x \tilde{\leq} b_i & i = k+1,\dots,m \\ & x \geq 0 & \end{array} \tag{1}$$

where symbols $\tilde{\geq}, \tilde{\leq}, \cong$ indicate that the inequalities or equalities are flexible. Such inequalities or equalities may be described by a fuzzy set whose membership function, $\alpha_i = \mu_i(x)$, represents the degree to which each target is attained.

To solve (1) is to find a solution x^* that maximizes a global measure $F(\alpha_1, \alpha_2, \dots, \alpha_m)$ for the degree of fulfilment of goals or constraints [16]. But not all constraints may be considered of equal importance. The most simple aggregation function that considers the different importance of constraints is the weighted mean: $F(\alpha_1,\alpha_2,...,\alpha_m)=\sum_i w_i\alpha_i$, where the weighted vector w verified $w_i \geq 0$ and $\sum_i w_i = 1$.

Then the model (1) can be written as a fuzzy constraint satisfaction problem:

$$\begin{aligned} &\text{Max } \; w_1\alpha_1 + w_2\alpha_2 + ... + w_m\alpha_m \\ &\text{s.t. } \; \mu_i(x) = \alpha_i \qquad i = 1,2,...,m \\ &\qquad 0 < \alpha_i \leq 1 \;\; \sum_i w_i = 1 \\ &\qquad\quad x \geq 0 \;\; w_i \geq 0 \end{aligned} \tag{2}$$

In order to define the membership function $\mu_i(x)$ for the *i-th* relation due to goals and constraints, we will have to know the tolerance margin, with respect to the target and right-hand side values, that DM is willing to accept. The DM will not consider the solutions with deviations larger than the tolerance margin.

In any case, $\mu_i(x)$ should be 0 if the *i-th* constraint is strongly unfulfilled, and 1 if it is completely satisfied. Then, for flexible inequalities of type $\tilde{\geq}$, $\mu_i(x)$ should increase monotonously from 0 to 1 over the tolerance interval, and for inequalities of type $\tilde{\leq}$, $\mu_i(x)$ should decrease monotonously from 1 to 0 over the same interval.

Then, if we suppose that membership function $\mu_i(x)$ is linear, to approach the model (2) we consider the following results (see [4]):

- For flexible inequalities of type $c_i x \tilde{\geq} g_i$ (see fig. 1):

$$\begin{aligned} &\mu_i(x) = 1 \Leftrightarrow c_i x - p_i = g_i, \text{ where } p_i \geq 0 \\ &\mu_i(x) = \alpha_i \in (0,1] \Leftrightarrow c_i x + n_i = g_i, \text{ where } n_i = (1-\alpha_i)\, d_i \\ &\mu_i = 0 \Leftrightarrow n_i \geq d_i \Leftrightarrow c_i x + d_i + r_i = g_i, \text{ where } r_i \geq 0 \end{aligned} \tag{3}$$

where the left tolerance margin is $d_i \geq 0$, the target negative deviation is $n_i = g_i - c_i x > 0$ and $p_i = c_i x - g_i \geq 0$ is the target positive deviation. r_i denote the amount by which the constraint is strongly unfulfilled, i.e. $c_i x$ is out of the tolerance interval. From now on we specify this type of fuzzy sets as (g_i, d_i, ∞)

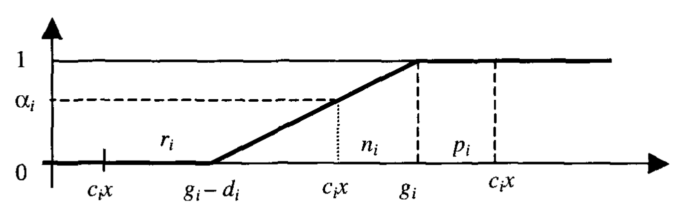


Figure 1. The fuzzy goal (g_i, d_i, ∞)

- For flexible inequalities of type $a_i x \tilde{\leq} b_i$ (see fig. 2):

$$
\begin{aligned}
&\mu_i(x) = 1 \Leftrightarrow a_i x + n_i = b_i, \text{ where } n_i \geq 0 \\
&\mu_i(x) = \alpha_i \in (0,1] \Leftrightarrow a_i x - p_i = b_i, \text{ where } p_i = (1-\alpha_i)\, t_i \qquad (4) \\
&\mu_i = 0 \Leftrightarrow p_i \geq t_i \Leftrightarrow a_i x - t_i - q_i = b_i, \text{ where } q_i \geq 0
\end{aligned}
$$

where the right tolerance margin is represented by $t_i \geq 0$ and where n_i, p_i are negative and positive deviation of targets. q_i is the amount by which the constraint is strongly violated. From now on, we specify this type of fuzzy sets as (b_i, ∞, t_i).

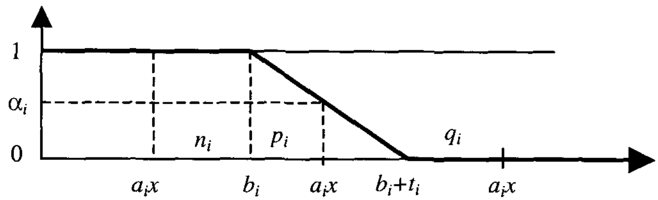


Figure 2. The fuzzy goal (b_i, ∞, t_i)

- For flexible equalities $c_i x \cong h_i$ (see fig. 3), if $d_i \geq 0$ and $t_i \geq 0$ represent both tolerance margins, we have:

$$
\mu_i(x) = \alpha_i \in (0,1] \Leftrightarrow \begin{cases} c_i x + n_i = h_i, \text{ where } n_i = (1-\alpha_i) d_i \\ \text{or} \\ c_i x - p_i = h_i, \text{ where } p_i = (1-\alpha_i) t_i \end{cases}
$$

$$
\mu_i(x) = 0 \Leftrightarrow \begin{cases} n_i \geq d_i \Leftrightarrow c_i x + d_i + r_i = h_i, \text{ where } r_i \geq 0 \\ \text{or} \\ p_i \geq t_i \Leftrightarrow c_i x - t_i - q_i = h_i, \text{ where } q_i \geq 0 \end{cases} \qquad (5)
$$

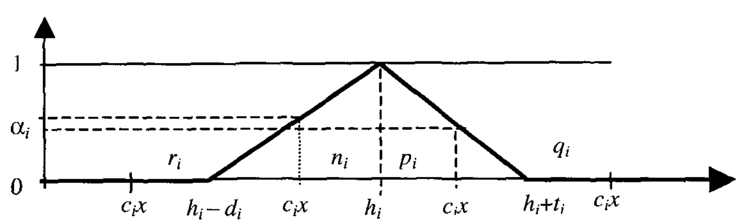


Figure 3. The fuzzy goal (h_i, d_i, t_i)

From now on we will specify this type of fuzzy sets as (h_i, d_i, t_i).

According to the above considerations, we can write the model (2) as follows:

$$\text{Max } w_1\left(1-\frac{n_1}{d_1}\right)+\ldots+w_q\left(1-\frac{n_q}{d_q}\right)+w_{q+1}\left(1-\frac{n_{q+1}}{d_{q+1}}-\frac{p_{q+1}}{t_{q+1}}\right)+\ldots+$$
$$+w_k\left(1-\frac{n_k}{d_k}-\frac{p_k}{t_k}\right)+w_{k+1}\left(1-\frac{p_{k+1}}{t_{k+1}}\right)+\ldots+w_m\left(1-\frac{p_m}{t_m}\right)$$
$$\begin{array}{lll} \text{s.t.} & c_i x+n_i-p_i=g_i & i=1,\ldots,q \\ & c_i x+n_i-p_i=h_i & i=q+1,\ldots,k \\ & a_i x+n_i-p_i=b_i & i=k+1,\ldots,m \\ & n_i \le d_i & i=1,\ldots,k \\ & p_i \le t_i & i=q+1,\ldots,m \\ & x\ge 0,\ p_i\ge 0, n_i\ge 0 & \end{array} \tag{6}$$

Which is equivalent to the next formulation:

$$\text{Min } \quad w_1\frac{n_1}{d_1}+\ldots+w_q\frac{n_q}{d_q}+w_{q+1}\left(\frac{n_{q+1}}{d_{q+1}}+\frac{p_{q+1}}{t_{q+1}}\right)+\ldots+w_k\left(\frac{n_k}{d_k}+\frac{p_k}{t_k}\right)+$$
$$+w_{k+1}\frac{p_{k+1}}{t_{k+1}}+\ldots+w_m\frac{p_m}{t_m} \tag{7}$$

s.t. *The same feasible set as model* (6).

Observe that we have transformed the fuzzy problem (1), into a goal programming one, whose objective function expresses the DM's satisfaction function [11] according to the deviation values n_i, p_i and the tolerance margins d_i, and t_i.

The feasible set of problem (6) could be the empty set. It means that DM's requirements have been excessive, i.e. DM has suggested a rigid maximum deviation of targets. For further detail about the analysis of infeasible LPs, see i.e. [6].

Using LINDO as a problem solver it is possible, from its unfeasibility report, to know what the relationships are, in the feasible set, that may have been unfulfilled. For simplicity, we suppose that they are the first and the last (i =1 and i =m); then, besides maximising the total measure of fulfilment of constraints $F(\alpha_1,\alpha_2,\ldots,\alpha_m)$, we have to minimize the excess over the tolerance interval for constraints 1 and m. Then, we rewrite model (7) as the following bi-objective program:

$$\text{Min } \quad w_1\frac{n_1}{d_1}+\ldots+w_q\frac{n_q}{d_q}+w_{q+1}\left(\frac{n_{q+1}}{d_{q+1}}+\frac{p_{q+1}}{t_{q+1}}\right)+\ldots+w_k\left(\frac{n_k}{d_k}+\frac{p_k}{t_k}\right)+$$
$$+w_{k+1}\frac{p_{k+1}}{t_{k+1}}+\ldots+w_m\frac{p_m}{t_m}$$

$$\text{Min } w_1 \frac{1}{g_1} r_1 + w_m \frac{1}{b_m} q_m \tag{8}$$

s.t. *Constraints of model* (6), *correcting the following:*

$c_1 x + n_1 + r_1 - p_1 = g_1$; $a_m x + n_m - p_m - q_m = b_m$; $n_1 = d_1$; $p_m = t_m$; $r_1 \geq 0$; $q_m \geq 0$

Where r_1 and q_m are the amounts by which deviations of constraints 1 and m are larger than the respective tolerance margins (see (3) and (4)).

The solution to this bi-objective problem maximizes a global measure for the degree of fulfilment of goals or constraints and minimizes the excess over the tolerance thresholds on unfulfilled goals or constraints. This solution is proposed as an optimal approximated solution to the initial problem (1).

3 Evaluating a hospital service performance

Hospitals, in general, function in a political, highly complex environment in which there are conflicting opinions about how best to carry out the primary commitments of that institution. There is a set of values that stresses the quality of care and a set of pressures to meet community medical service needs. At the same time, there is pressure to maintain a financially viable institution.

Several authors, i.e. [5, 10], have applied GP formulation to problems in the health care field. In [2] we had presented a crisp model to design the real performance of a surgical service at a local general hospital. Now, with this fuzzy approach, we aim to improve the performance of the local hospital that provided us with data. This data may be considered vague in the sense that small variations may be accepted.

The model described here was developed for the COT[1] service of a local state-owned hospital, to plan one year ahead, because the economic forecasts are also made on a yearly basis. To simplify the model design, we will omit a detailed description of the Hospital.

Let us now write a fuzzy linear model for clinical and surgical operations. The model allows the DM to evaluate the Hospital's operation plans in terms of the goals of providing quality service, minimizing costs and maintaining job allocation.

Priority 1.- *Assure the necessary personnel and spatial resources to give satisfactory services to the patient:* The prior objective is to ensure the personnel needed to offer the patient a satisfactory service; the decision center estimates that the current staff will be enough to satisfy demand. Table 1 shows the staff requirements:

---

[1] Traumatology and Orthopedic Surgery- Cirugía Ortopédica y Traumatología.

Table 1

Variables	Products	Physicians	Nurses
x_{11}	Scheduled operation with admission	200′	200′
x_{21}	Emergency operation with no admission	200′	200′
x_{12}	Ambulatory scheduled operation	90′	90′
x_{22}	Ambulatory operation non-scheduled	90′	90′
y_{11}	First appointments	20′	20/3′
y_{12}	Following appointments	10′	10/3′
y_{13}	Ambulatory emergency	32′	32′
Z	Hospitalization[2]	15 ‘	70.28′

Next we will show table 2 containing the global time available for Doctors, Surgery Nurses, Hospitalization Nurses, and the Supervisor Nurse.

Table 2

	Physicians	Hospitalization Nurses	Consulting Nurses	Surgical Nurses
Time resource (in hours)	23333	23920	3640	5980

Following this priority, the spatial requirements are also a constraint. Spatial requirements are shown in table 3.

Table 3

	Beds	Consulting Rooms	Operating theatres
Available	66	8420 h/annum	2100 h/annum

These are the constraints corresponding to this priority:

Human resources: For the DM the global available time for Doctors and Nurses should be used up, but it can admit a tolerance threshold of $\pm 5\%$. Then the constraints corresponding to human resources can be expressed as follows:

Physicians:

$$200x_{11} + 200x_{21} + 90x_{12} + 90x_{22} + 20y_{11} + 10y_{12} + 32y_{13} + 15z$$
$$\cong (1400000,70000,70000)$$

Surgical nurses:

$$200x_{11} + 200x_{21} + 90x_{12} + 90x_{22} \cong (358800,17940,17940)$$

Consulting nurses:

[2] Each person being in hospital use 70.28=1435200/(0.8477 x 66 x 365) minutes of hospitalization nurses.

$$\frac{20}{3}y_{11}+\frac{10}{3}y_{12}\cong(218400,10920,10920)$$

Hospitalisation nurses:

$$70.28z\cong(1435200,71760,71760)$$

Spatial resources: The available spatial resources can be increased up to 5%. Then the corresponding constraints (in minutes) can be written as follows:

Consulting Room:

$$20\,y_{11}+10\,y_{12}\ \tilde{\leq}\ (505200,\infty,25260)^3$$

Operating Theatre:

$$100\,x_{11}+100\,x_{21}+45\,x_{12}+45\,x_{22}\ \tilde{\leq}\ (126000,\infty,6240)$$

Priority 2.- *Minimize cost.* As a second priority we have taken into account the need to decrease or, at least, not surpass the actual COT service costs by more than 5%. Therefore, we will carry out the service costs analysis in order to minimize them. Our-base year has been 1994. Having made the cost evaluation, we have developed a model to allocate it into each of the variables. In order to do that, we have applied a cost assignment technique used in most the Spanish Hospitals. Table 4 shows costs corresponding to each variable.

Table 4

	Activity Level current	Activity Level Desirable	Cost per unit
x_{11}	1230	1320	19,437.57
x_{21}	66	360	19,437.57
x_{12}	27	130	9,718.78
x_{22}	8	20	9,718.78
y_{11}	14187	10000	9,719.47
y_{12}	22663	20000	5,830.5
y_{13}	16539	10000	11,663.25
z	20655	20850	38,875.15

Here is the corresponding inequality[4]:

$$19437.57\,x_{11}+19437.57\,x_{21}+9718.78\,x_{12}+9718.78\,x_{22}+9719.47\,y_{11}+5830.5\,y_{12}+11663.25\,y_{13}+38875.15\,z\ \tilde{\leq}\ (1291423,\infty,64571)$$

Priority 3. *Adjust waiting list.* This third group includes some aspects related to quality service. One of the main problems of the National Health Service is waiting

[3] Data in constraints are expressed in minutes.

[4] In thousand pesetas.

lists in some medical and surgical services, specially in surgical ones. Due to strictness and the impossibility of making any change in the short term, we have been obliged to restrict the model with the following bounds-upper and lower- which show the need for expanding activities in order to reduce waiting lists. The desired activity levels are shown in table 4. The Dm doesn't wants to decrease the actual activity level, and the desirable activity level can be surpassed by as much as 2%. That may be described by a fuzzy set similar to the one drawn in figure 4:

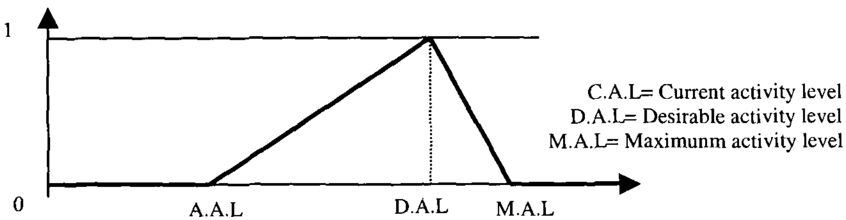


Figure 4

Thus, the corresponding constraints can be written as follows:

$$x_{11} \cong (1320,90,26.4) \qquad y_{11} \cong (10000,200,4187)$$
$$x_{21} \cong (360,294,7.2) \qquad y_{12} \cong (20000,400,2663)$$
$$x_{12} \cong (130,103,6..5) \qquad y_{13} \cong (1000,200,6539)$$
$$x_{12} \cong (20,12,0.4) \qquad z \cong (20850,417,1042.5)$$

Priority 4. *Provide a quality service.* In this priority we set constraints referring to the quality of C.O.T. service, through the most frequent activity indexes:

* Occupation Index: (Patient-days / Annual Average of beds working x 365days) x 100: in 1994, the number of patient-days was 20422, with 66 beds:

O.I. = (20422/(66x365)) x 100 = 84.77%.

* Turnover: Number of Admissions / Number of beds: in 1994, the number of admitted patients was 1779, then :T.I. = 1779 / 66 = 26.95.

* Average length of stay : Patient-days / Number of Admissions (scheduled and emergency). In 1994: A.L.S. = 20422 / 1779 = 11.47

The next table shows data and aims:

Table 5

	Current	Minimum	Maximum
Occupation Index	84.77%	75	85
Turnover	26.95	27	30.4
Average Length of stay	11.47	9	11.5

As we have seen previously (see footnote 2), the Occupation Index appears in the first goal. Average Length of Stay as well as Turnover are matched and they could be placed in the same equation. An increase in the length of stay will imply a decrease in turnover. So, this is the constraint that expresses this last priority:

$$27 \leq \frac{\dfrac{x_{11}+x_{21}}{0{,}73}}{66} \leq 30.4$$

The right hand side has been obtained from historical data obtained from the Hospital Decision Center, so Total Admissions could be separated into Surgical and Clinical Admissions, the latter being about 27% of the total amount.

Maximum and minimum degrees of satisfaction are attained respectively with maximum and minimum turnover, so the above constraint can be expressed as follows:

$$x_{11}+x_{21} \tilde{\lesssim} (1465,\ 164, 0)$$

Using (3), (4) and (5) we define the corresponding fuzzy sets that represent the flexible relationship and whose membership function $\alpha_i = \mu_i(x)$ represents the degree to which each goal is attained.

Now we ask the decision-maker to weigh up the different priorities. The DM gives us the following information:

- Human and spatial resources: 35%. The goals which are in this group have the following weight: Physicians: 20%, Surgical nurses: 15%, Consulting nurses: 10%, Hospitalization nurses: 10%, Consulting room: 20%, Operating room: 25%
- Minimize cost: 22%
- Adjust waiting list: 18%. The goals in this group have the same weight.
- Provide a quality service: 16%

Then we can write our model as follows (see eq. (2)):

$$\text{Max } F(\alpha_1,\alpha_2,\ldots,\alpha_{16}) = 0.35(0.20\alpha_1 + 0.150\alpha_2 + 0.10\alpha_3 + 0.10\alpha_4 + 0.20\alpha_5 + 0.25\alpha_6) +$$
$$0.22\alpha_7 + 0.18\left(\frac{1}{8}(\alpha_8+\alpha_9+\alpha_{10}+\alpha_{11}+\alpha_{12}+\alpha_{13}+\alpha_{14}+\alpha_{15})\right) + 0.16\alpha_{16}$$

s.t.

$$\mu_i(x) = \alpha_i$$
$$x \geq 0,\ \ 0 < \alpha_i \leq 1 \qquad i = 1,2,\ldots,16$$

Below, according to model (7), we can reformulate the above problem as follows:

$$\text{Min}\ \ 0.35\left(0.20\left(\frac{n_1}{70000}+\frac{p_1}{70000}\right)+0.15\left(\frac{n_2}{17940}+\frac{p_2}{17940}\right)+0.10\left(\frac{n_3}{10920}+\frac{p_3}{10920}\right)+\right.$$
$$\left.+0.10\left(\frac{n_4}{71760}+\frac{p_4}{71760}\right)\right)+0.35\left(0.20\frac{p_5}{25260}+0.25\frac{p_6}{6300}\right)+0.22\frac{p_7}{64571}+$$
$$+0.18\frac{1}{8}\left(\frac{n_8}{90}+\frac{p_8}{26.4}+\frac{n_9}{294}+\frac{p_9}{7.2}+\frac{n_{10}}{103}+\frac{p_{10}}{6.5}+\frac{n_{11}}{12}+\frac{p_{11}}{0.4}+\frac{n_{12}}{200}+\frac{p_{12}}{4187}+\frac{n_{13}}{400}+\right.$$
$$\left.+\frac{p_{13}}{2663}+\frac{n_{14}}{200}+\frac{p_{14}}{6539}+\frac{n_{15}}{417}+\frac{p_{15}}{1042.5}\right)+0.16\frac{n_{16}}{164}$$

s.t. (9)

$200x_{11}+200x_{21}+90x_{12}+90x_{22}+20y_{11}+10y_{12}+32y_{13}+15z+n_1-p_1=1400000$;
$200x_{11}+200x_{21}+90x_{12}+90x_{22}+n_2-p_2=358800$; $6.66y_{11}+3.33y_{12}+n_3-p_3=218400$
$70.28z+n_4-p_4=1435200$; $20y_{11}+10y_{12}+n_5-p_5=505200$;
$100x_{11}+100x_{21}+45x_{12}+45x_{22}+n_6-p_6=126000$;
$19437.57x_{11}+19437.57x_{21}+9718.78x_{22}+9719.47y_{11}+5830.5y_{12}+$
$11663.25y_{13}+38875.15z+n_7-p_7=1291423$;
$x_{11}+n_8-p_8=1320$; $x_{21}+n_9-p_9=360$; $x_{12}+n_{10}-p_{10}=130$; $x_{22}+n_{11}-p_{11}=20$; $y_{11}+n_{12}-p_{12}=10000$; $y_{12}+n_{13}-p_{13}=20000$; $y_{13}+n_{14}-p_{14}=10000$; $z+n_{15}-p_{15}=20850$; $x_{11}+x_{21}+n_{16}-p_{16}=1465$; $n_1<70000$; $p_1<70000$; $n_2<17940$; $p_2<17940$; $n_3<10920$; $p_3<10920$; $n_4<71760$; $p_4<71760$; $p_5<25260$; $p_6<6300$; $p_7<64571$; $n_8<90$; $p_8<26.4$; $n_9<294$; $p_9<7.2$; $n_{10}<103$; $p_{10}<6.5$; $n_{11}<12$; $p_{11}<0.4$; $n_{12}<200$; $p_{12}<4187$; $n_{13}<400$; $p_{13}<2663$; $n_{14}<200$; $p_{14}<6539$; $n_{15}<417$; $p_{15}<1042.5$; $n_{16}<164$
With all variables non-negative.

This problem is unfeasible. If we use the computer program LINDO as a solver, we find that it is necessary to correct one of the following constraints:

3) $n_3 - p_3 + 6.66\,y_{11} + 3.33\,y_{12} = 218400$; 12) $n_{12} - p_{12} + y_{11} = 10000$;
13) $n_{13} - p_{13} + y_{12} = 20000$; 21) $n_3 <= 10920$; 37) $p_{12} <= 4187$; 39) $p_{13} <= 2663$ (10)

According to this result, to make the problem feasible the decision-maker should increase the left tolerance margin of Consulting Nurses and the right tolerance margin of desirable activity level of Following Appointments (y_{12}) and Ambulatory Emergency (y_{13}). Then we reformulate program (9) as a bi-objective problem (see (8)) in which we assume that the aforementioned tolerance margins can be exceeded:

Min *The same objective function as* (9)

$$\text{Min}\ \ 0.35\cdot 0.10\frac{r_3}{218400}+\frac{0.18}{8}\left(\frac{q_{12}}{10000}+\frac{q_{13}}{20000}\right)$$

(11)

s.t. *Constraints of model* (9), *correcting (10) that become:*

3) $6.66y_{11}+3.33y_{12}+n_3+r_3-p_3=218400$; 12) $y_{11}+n_{12}-p_{12}-q_{12}=10000$;

13) $y_{12}+n_{13}-p_{13}-q_{13}=20000$; 21) $n_3=10920$; 37) $p_{12}=4187$; 39) $p_{13}=2663$

Where we have included three new non-desirable deviation variables: r_3, q_{12} and q_{13}, that represent the amount by which the left tolerance margin of Consulting Nurses and the right tolerance margin of desirable activity level of Following Appointments (y_{12}) and Ambulatory Emergency (y_{13}), are exceeded (see (3) and (4)). By means of the second objective function we try to minimize this excess. To solve program (11) we formulate the following mono-objective program:

$$\begin{aligned} &\text{Min } \text{“}\textit{The sum of the two objectives functions of model } (11)\text{”} \\ &\text{s.t. } \textit{The same feasible set as model } (11) \end{aligned} \tag{12}$$

This problem is still unfeasible. If we use LINDO as a solver, we find that it is sufficient to correct one of the following constraints:

$$2)\ n_2 - p_2 + 200\,x_{11} + 200\,x_{21} + 90\,x_{22} + 90\,x_{12} = 358800;\quad 19)\ n_2 <= 17940 \tag{13}$$

That means that to make the problem feasible the decision-maker should increase the left tolerance margin of Surgical Nurses. Then we reformulate program (12) as follows:

$$\begin{aligned} \text{Min }\ & 0.35\left(0.20\left(\frac{n_1}{70000}+\frac{p_1}{70000}\right)+0.15\left(\frac{n_2}{17940}+\frac{p_2}{17940}\right)+0.10\left(\frac{n_3}{10920}+\frac{p_3}{10920}\right)+\right.\\ & \left.+0.10\left(\frac{n_4}{71760}+\frac{p_4}{71760}\right)\right)+0.35\left(0.20\frac{p_5}{25260}+0.25\frac{p_6}{6300}\right)+0.22\frac{p_7}{64571}+\\ & +0.18\frac{1}{8}\left(\frac{n_8}{90}+\frac{p_8}{26.4}+\frac{n_9}{294}+\frac{p_9}{7.2}+\frac{n_{10}}{103}+\frac{p_{10}}{6.5}+\frac{n_{11}}{12}+\frac{p_{11}}{0.4}+\frac{n_{12}}{200}+\frac{p_{12}}{4187}+\frac{n_{13}}{400}+\right.\\ & \left.+\frac{p_{13}}{2663}+\frac{n_{14}}{200}+\frac{p_{14}}{6539}+\frac{n_{15}}{417}+\frac{p_{15}}{1042.5}\right)+0.16\frac{n_{16}}{164}+0.35\left(0.15\frac{r_2}{358800}+0.10\frac{r_3}{218400}\right)+\\ & +\frac{0.18}{8}\left(\frac{q_{10}}{130}+\frac{q_{11}}{20}+\frac{q_{12}}{10000}+\frac{q_{13}}{20000}\right) \\ \text{s.t. }\ & \textit{Constraints of model } (9),\ \textit{and the new expression of (13):} \end{aligned} \tag{14}$$

2) $200x_{11}+200x_{21}+90x_{12}+90x_{22}+n_2+r_2-p_2=358800$; 19) $n_2=17940$

that includes a new non-desirable deviation variable: r_2, representing the amount by which the left tolerance margin of Surgical Nurses is surpassed.

Table 6 shows the optimal solution for model (14); the solution of the same problem through Standard Weighted Goal Programming (WGP) is included in order to compare them.

Table 6

	Variable	Desirable level	Activity level	**Our approach solution**	Standard GP solution
Scheduled operation with admission	x_{11}	1320	1230	**1235**	1178
Emergency operation with no admission	x_{21}	360	66	**66**	66
Ambulatory scheduled Operation	x_{12}	130	27	**27**	27
Ambulatory operation non-scheduled	x_{22}	20	8	**20**	8
First appointments	y_{11}	10000	14187	**14187**	13128
Following Appointments	y_{12}	20000	22663	**22663**	22663
Ambulatory Emergency	y_{13}	10000	16539	**10000**	10405
Hospitalization	z	20850	20655	**20850**	20655
	Target deviation[5]	Tolerance Margin			
Physicians	p_1	70000		**7550**	0
	n_1	70000		**0**	0
Surgical nurses	$n_2(+r_2)$	17940		**94370**	106800
Consulting nurses	$n_3(+r_3)$	10920		**48446**	50168
Hospitalization nurses	p_4	71760		**30138**	16433
Consulting rooms	n_5	25260		**0**	0
	p_5	25260		**5170**	0
Operating theatres	n_6	6300		**0**	0
	p_6	6300		**6215**	0
Cost	n_7	∞		**68734**	75315
Turnover	n_{16}	164		**164**	220

Observe that in our approach there is a redistribution of activity that demands more staff occupation and a little more financial support than in a standard GP solution, but it allows improve ment of waiting list time and also the turnover index is better. Therefore, we consider that our solution is more balanced and that it allows all resources to be used more intensively, as we had supposed.

The current work time of Physicians should be lightly augmented. There are some surpluses in surgical and consulting nurses, whereas more hospitalization nurses are necessary: some surgical and consulting nurses should be moved to hospitalisation. Current spatial availability, is totally used up: the consulting room time should be lightly augmented, while the operating theatres have to be used intensively. The optimum activity level could be reached without using up all the resources available in the service. It means inefficient allocation of resources. Current level of activity is lower than the optimum one and costs are higher. Some

---

[5] n_i: target negative deviation. p_i: target positive deviation.

of the desirable levels of activity have been reached in spite of the other ones. Turnover has been reached at its minimum desirable value.

4 Conclusion

In classic goal programming, the DM has to provide crisp goals and he/she cannot express a satisfaction degree concerning observed target deviations. In this paper we have used a new approach to flexible GP: by means the fuzzy logic we have tried to lessen both drawbacks. In [11] another flexible approach to the same problem is expressed.

We proposed a problem formulation taking into account not only goal achievement but also minimizing the amount by which goals or constraints may be strongly unfulfilled. This usually expresses the DM's preferences better than the classical approach. The flexible relationships in a GP problem are described by fuzzy sets whose membership function, $\alpha_i = \mu_i(x)$, represents the degree in which each relationship is attained.

We have used the simplest shapes of membership function i.e. linear shapes that give a linear program, that in several authors′ opinions [14, 17], are valid for most practical applications. Nevertheless other shapes may be used.

We deal with the unfeasibility of the problem, due to initial DM preferences, relaxing some veto threshold, in an interactive way.

Our problem solution maximizes a global measure of the degrees of achievement of goals. We have used the weighted mean, although more complex aggregation functions may also be considered.

The solution obtained through our approach is more balanced than the classic one with regard to the achievement of the aspiration levels of objective functions because, in general, it is preferable to satisfy fewer goals in total, with only minimum deviations in those which are not satisfied.

Also in this paper we have shown that the use of the weighted mean to aggregate fuzzy sets is suited to in the Fuzzy Goal Programming approach.

We applied our solution method to evaluate a hospital service performance and the DM was very interested in it, because it permits him to carefully analyze the hospital performance observing where there are surpluses and deficits. Then, from the information given by our solution approach, DM can propose organization changes in order to improve the operational results.

References

1. Arenas, M., Bilbao, A., Rodríguez, M. V. and Jiménez, M., A Fuzzy Solution Approach to a Fuzzy Linear Goal Programming Problem. In Fandel. G. and Gal. T. (Eds.), *Multiple Criteria Decision Making. Lecture Notes in Economics and Mathematical Systems* **448**, (Springer, New York, 1996) pp.255-264.

2. Arenas, M., Lafuente, E. and Rodríguez, M.V., Goal Programming Model for Evaluating Hospital Service Performance. In Caballero, R., Ruiz, F. and Steuer, R. (Eds.) *Advances In Multiple Objective And Goal Programming. Lectures Notes in Economics and Mathematical Systems* **455** (Springer, Heidelberg, 1997) pp. 57 -65.
3. Arenas, M., Bilbao, A., Rodriguez, M. V. and Jiménez, M., Fuzzy Parameters multiobjective linear programming problem: Solution analysis. In Stewart. T. and Van den Honert, R. (Eds.), *Trends in Multicriteria Decision Making*. Lecture Notes in Economics and Mathematical Systems **465** (Springer, New York, 1998) pp. 128-137.
4. Arenas, M., Bilbao, A., Jiménez, M. and Rodríguez, M.V., On considering flexible constraints of different importance in goal programming problems. *The Fourth International Conference on Multi-Objective Programming and Goal Programming: Theory & Applications.* (Ustron, Poland. May 29-June 1, 2000)
5. Chae, Y., Suver, J. and Chou, D., Goal programming as a capital investment tool for teaching hospitals. *HCM review*, Winter (1985) pp. 27-35.
6. Chinneck J.W. and Michalowsky, W., MOLP formulation assistance using LP infeasibility analysis. In Tamiz (Ed.), (Springer, New York, 1996) pp. 87-106.
7. Hannan, E. L., Linear Programming with Multiple Fuzzy Goals. *Fuzzy Sets and Systems* **6** (1981) pp. 235-248.
8. Ignicio, J. P., *Goal programming and extension.* (Lexington Books, Massachusetts, 1976)
9. Lai, Y.J. and Hwang, C.L. Multi-Objective fuzzy Mathematical Programming: Theory and Applications. *Lecture Notes in Economics and Mathematical Systems* **404** (Springer, Berlín, 1993)
10. Lee, S., An Aggregative Resource Allocation Model for Hospital Administration. *Socio-Economic Planning Science.* Vol. **7** (1973) pp. 381-395.
11. Martel, J.M. and Aouni, B., Diverse Imprecise Goal Programming Model Formulations. *Journal of Global Optimization* **12** (1998) pp. 127-138.
12. Narashimhan, R., Goal programming in fuzzy environment. *Decision Sciences* **11** (1980) pp. 325-336.
13. Romero, C., *Handbook of critical issues in goal programming.* (Pergamon Press. Oxford, 1991)
14. Rommelfanger, H., Some problems of fuzzy optimization with t-norm based extended addition. In Delgado, M., Kacprzyk, J., Verdegay, J.L. and Vila, M.A. (Eds.) *Fuzzy Optimization*, (Physica-Verlag, Heidelberg, 1994) pp. 158-168.
15. Sakawa, M., Fuzzy sets and interactive multiobjective optimisation. *Applied Information Technology Series.* (Plenum Press, New York, 1993)
16. Torra I Reventós, V., On considering constraints of different importance in fuzzy constraints satisfaction problems. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems.* Vol. **6**, N° **5** (1998) pp. 489-481.
17. Zimmermann, H.J. Fuzzy programming and linear programming with several objective functions. *Fuzzy sets and systems* **1 (1)** (1978) pp. 45-55.

This page is intentionally left blank

A GROUP DECISION MAKING METHOD USING FUZZY TRIANGULAR NUMBERS

J. L. GARCÍA-LAPRESTA AND M. MARTÍNEZ-PANERO

Departamento de Economía Aplicada (Matemáticas). Facultad de Ciencias Económicas y Empresariales. Universidad de Valladolid. Av. Valle de Esgueva 6, 47011 Valladolid, Spain
E-mail: lapresta@cpd.uva.es and panero@eco.uva.es

L. L. LAZZARI

Departamento de Matemática. Facultad de Ciencias Económicas. Universidad de Buenos Aires. Córdoba 2122, Ciudad de Buenos Aires (1120), Argentina
E-mail: ilazzari@econ.uba.ar

In this paper we introduce a Borda-type decision procedure taking into account agents' intensities of preferences by means of linguistic labels. The advantages of the classic Borda count hold, and are even improved, because the flexibility of gradation increases.

1 Introduction

Among the great variety of methods in collective decision making[1], in this paper we begin by focusing our attention on the Borda count. This procedure will be the pattern of the one introduced in this paper. It will be shown that we are right to do so. For the present, we can mention that the Borda count has been considered "optimal" by Saari [40, p. 12]. In the same direction, Dummett [15, p. 290] also argues that it is "the best tool for reaching the decision most likely to be correct when the object is to reconcile different judgements about effective means to a common aim, and the most equitable method of determining a resultant of divergent desires". Moreover, an institution called the *De Borda Institute*[2] exists, which shows its advantages and tries to shed light on it.

Despite works presented earlier[3], this method was proposed by Jean Charles de Borda, as indicated by its name. This engineer and navy officer denounced, in a Memory read in 1770 in the French Academy of Sciences, that the usual collective decision procedures only considered the most preferred alternative for each agent, ignoring the rest. Taking into account this partial information, the final output could not faithfully reflect the agents' preferences. Borda showed examples with this fault, and then advocated the following scheme: each agent arranges all the alternatives linearly, and gives integer marks to each of them: the highest score,

[1] See Fishburn [16], for example.
[2] http://members.tripod.co.uk/deBordaInstitute.
[3] McLean and Urken [32, pp. 19 and ff.] indicated that Nicolaus Cusanus (XVth c.) foresaw the idea of the Borda count.

which coincides with the number of alternatives, to the most preferred; one point less to the next alternative; and so on, in a descent manner, till the least preferred is reached, which is given only one point. Then, each alternative is given the total score summed over all agents' individual rankings, and the collective decision is to rank the alternatives from highest to lowest total score.

This Borda count was chosen by the French Academy of Sciences to select its members, and it was both critiziced and praised. Indeed, the Marquis of Condorcet censured its manipulability and pointed out the following fact: an alternative which defeats all others one by one in the majority of the cases[4] might not be selected if the Borda count is used (see Black [7, pp. 156 and ff.]). On the other hand, the Spanish enlighted mathematician Morales [35] considered it as the most appropriate and representative procedure. However, it was said that the Borda count did not respect the agents' freedom of scoring and did not reflect the candidates' merit with accuracy. In order to refute these arguments, Morales [36, pp. 18 and ff.] showed that, if the scale of values 0,1,2..., is used instead of 1,2,3..., to score the alternatives from the least to the most preferred, then the total score of a fixed alternative is the number of alternatives evaluated worse than it. This fact had already been pointed out by Borda and commented by Condorcet to ensure that the range of values used by the Borda count was not arbitrary (see McLean and Urken [32, pp. 81–89]). More recently, the unfulfillment of the independence of irrelevant alternatives principle considered by Arrow [1] has been added to the above-mentioned drawbacks[5].

In contrast, advantages of the Borda count which promote this method are the use of information from the entire preference rankings of all the voters, and its straightforward execution. In addition, Black [7, 8], Mueller [37] and Straffin Jr. [43], among others, have noted that the Borda count chooses the alternative which stands highest on average in the agents' preference orderings. Even more, distance-based consensus arguments can be found in Cook and Seiford [11], and in Biswas [6] reasons of efficiency and consistency related to the Borda count appear.

The Borda rule has been re-considered many times, inspiring other collective decision making procedures. In 1882, Nanson (see McLean and Urken [32, pp. 321–359]) introduced a sort of Borda count with eliminations. Copeland [12] considered Borda-type scores for each alternative, which were the number of alternatives better than that marked minus the number of alternatives worse than it (see Saari and Merlin [41]). Black [7, p. 66] proposed a hybrid procedure consisting in the choice of the Condorcet winner, if it exists; otherwise, Borda count should be used. Kemeny [27] formulated a rule defined with arguments similar to those of the Borda rule (see Young [49, pp. 61 and ff.]). Later on, Sen [42] considered the so-called narrow and broad Borda counts in connection with the independence of irrelevant alternatives principle. In addition, Dummett [15] introduced the revised

[4] This alternative, if it exists, is called the *Condorcet winner.*

[5] Again, Condorcet anticipated Arrow's independence of irrelevant alternatives principle. See McLean and Urken [32, pp. 32–34] and [33].

and adjusted Borda counts to penalize the manipulability of the classic Borda count. On the other hand, in Marchant [29, 30] and García-Lapresta and Martínez-Panero [18] generalizations of the Borda count, by using fuzzy preferences, can be found.

In this paper the classic Borda count is re-considered and extended in a natural way to non-linear orderings, where the agents can declare indifference among distinct alternatives. Although Nitzan and Rubinstein [38] consider the Borda count when agents have not necessarily transitive preferences, in this paper we justify that transitivity is a necessary coherence assumption for ensuring reasonable Borda scores. According to Zadeh [50], individuals tend to express their preferences in a linguistic manner rather than in precise numerical values. Thus, in this paper we also generalize the Borda count by considering linguistic labels for the agents to compare the alternatives. In each case, special attention will be paid to the underlying rationality assumptions.

The paper is organized as follows. In Section 2 we extend and formalize the classic Borda count. In Section 3 we define linguistic preference relations as a class of fuzzy binary relations based on linguistic labels. We shall focus our attention on fuzzy triangular numbers as labels, and so, in Section 4, the Borda-type procedure based on these labels is designed. Finally, in Section 5 we present some concluding remarks about the considered methods.

2 The Borda count

At first, the Borda count was designed assuming the agents' preferences to be linear orderings. If so, the marks assigned to the alternatives ought to be arranged, from the best to the worst, in a sequential manner. However, ties may appear in agents' preferences, and then, the previous scheme could not be used, although the original idea can be extended in several ways to be applied in this situation, as pointed out by Gärdenfors [19]. Our proposal is one of these possibilities that generalizes the original Borda count in a natural way, assigning to each alternative the number of alternatives worse than that to be scored. This can be formalized as follows. Let P^1, P^2, ..., P^m be the preference relations (asymmetric binary relations) of m agents over n alternatives $x_1, x_2, \dots, x_n$. Each agent gives a mark to each alternative, according to the number of alternatives worse than it: $r_k(x_i) = \#\{x_j \mid x_i P^k x_j\}$ is the score given by agent k to the alternative x_i. The same result is obtained by considering, for each agent k, its preference matrix:

$$\begin{pmatrix} r_{11}^k & r_{12}^k & \cdots & r_{1n}^k \\ r_{21}^k & r_{22}^k & \cdots & r_{2n}^k \\ \cdots & \cdots & \cdots & \cdots \\ r_{n1}^k & r_{n2}^k & \cdots & r_{nn}^k \end{pmatrix},$$

where

$$r_{ij}^{k} = \begin{cases} 1, & \text{if } x_i P^k x_j, \\ 0, & \text{otherwise.} \end{cases}$$

In this way, the agent k gives the alternative x_i the following mark:

$$r_k(x_i) = \sum_{x_i P^k x_j} r_{ij}^{k}.$$

We note that the possible score range is contained in the set of values $\{0,1,\ldots,n-1\}$. Some of the upper values might not be reached if there is indifference (absence of preference) among distinct alternatives. However, if the orderings were linear, the above-mentioned set would exactly ranged.

With these individual counts a collective one is obtained:

$$r(x_i) = \sum_{k=1}^{m} r_k(x_i).$$

So, the collective preference relation P^B is defined by:

$$x_i P^B x_j \Leftrightarrow r(x_i) > r(x_j),$$

which is negatively transitive[6]. Thus, the highest scored alternatives (maximals for P^B) will be chosen.

Nevertheless, it is reasonable for the Borda count to require the fulfillment of the following property of monotonicity: when two alternatives are compared by an agent, the highest scored must be the preferred one.

Definition 1. The count r_k is *monotonic* if and only if

$$x_i P^k x_j \quad \Rightarrow \quad r_k(x_i) > r_k(x_j),$$

for all pairs of alternatives $x_i, x_j \in X$.

As we shall show, this demand is verified by assuming transitivity in individual preferences.

Proposition 1. If the preference relation P^k is transitive, then the count r_k is monotonic.

Proof: Suppose $x_i P^k x_j$. Being P^k transitive, if $x_j P^k x_p$, then $x_i P^k x_p$ and, consequently, $\{x_p \mid x_j P^k x_p\} \subset \{x_p \mid x_i P^k x_p\}$. The inclusion is strict because x_j belongs to the second set and not to the first one. Finally,

$$r_k(x_i) = \#\{x_p \mid x_i P^k x_p\} > r_k(x_j) = \#\{x_p \mid x_j P^k x_p\}.$$

---

[6] An ordinary preference relation P is negatively transitive if, not being verified $x_i\ P\ x_j$ nor $x_j\ P\ x_k$, then $x_i\ P\ x_k$ is not verified either. It is easy to prove that if P is negatively transitive, then it is also transitive.

Remark 1. If P^k is not transitive, then the count r_k is not necessarily monotonic, as it is showed in the following example. Suppose a preference relation P^k over 4 alternatives, explicitly given by: $x_1 P^k x_2$, $x_2 P^k x_3$, $x_3 P^k x_4$, $x_2 P^k x_4$. Then $r_k(x_1) = 1 < r_k(x_2) = 2$, although, as pointed out, $x_1 P^k x_2$.

3 Linguistic preference relations

As in the previous section, suppose m agents who have to show their preferences over each pair of alternatives of $X = \{x_1, \ldots, x_n\}$. Now, let $L=\{l_0, l_1, \ldots, l_s\}$ be a set of linguistic labels[7], ranked by a linear order: $l_0 < l_1 < \cdots < l_s$. We consider that individuals use these labels for declaring their preferences over the pairs of X. Suppose that there is an intermediate label which represents indifference, and the rest of labels are around it in a symmetric way; so, the number of labels, s+1, will be odd and, consequently, $l_{s/2}$ is the central label. Arguments for using no more than 13 labels can be found in Miller [34] and Yager [46], among others.

Definition 2. We say that R is a *linguistic preference relation over X based on L* if and only if R is a fuzzy binary relation defined by[8] $\mu_R : X \times X \to L$, with $\mu_R(x_i, x_j) = r_{ij}$ satisfying $r_{ij} = l_h \Leftrightarrow r_{ji} = l_{s-h}$, for every $i, j \in \{1, \ldots, n\}$ and every $h \in \{0, 1, \ldots, s\}$.

In what follows, r_{ij} will denote the level of preference of x_i over x_j, according to the following scheme:

1. $r_{ij} = l_s$, if x_i is definitely preferred to x_j,
2. $l_{s/2} < r_{ij} < l_s$, if x_i is slightly preferred to x_j,
3. $r_{ij} = l_{s/2}$, if x_i is indifferent to x_j, i.e., if x_i is not preferred to x_j and x_j is not preferred to x_i,
4. $l_0 < r_{ij} < l_{s/2}$, if x_j is slightly preferred to x_i,
5. $r_{ij} = l_0$, if x_j is definitely preferred to x_i.

---

[7] See Zadeh [50], Bonissone and Decker [10] and Delgado, Verdegay and Vila [14], among others.

[8] Generally, fuzzy sets and fuzzy preferences are evaluated on L = [0,1]. Nevertheless, some authors have used lattices and other kind of sets. See Goguen [20] and Barrett, Pattanaik and Salles [3], among others.

This scheme is similar to others used in the literature[9], but in [0,1] instead of L. We note that the condition $r_{ij} = l_h \Leftrightarrow r_{ji} = l_{s-h}$ appearing in Definition 2 is related to the reciprocity axiom sometimes required in the modelization of preferences by means of fuzzy binary relations[10].

Definition 3. For each agent $k \in \{1,\ldots,m\}$, the linguistic preference relation over X based on L, R^k, induces an ordinary binary relation over X defined by:

$$x_i \succ_k x_j \Leftrightarrow r_{ij}^k > l_{s/2},$$

for all $i, j \in \{1,\ldots,n\}$.

Remark 2. Taking into account Definition 2, it is easy to see that $\succ_k$ is asymmetric, hence an ordinary preference relation over X. Consequently, the indifference relation associated with $\succ_k$ is reflexive and symmetric, i.e., $r_{ii}^k = l_{s/2}$, and $r_{ij}^k = l_{s/2}$ implies $r_{ji}^k = l_{s/2}$, for every $i, j \in \{1,\ldots,n\}$. These facts will be consider in Section 4 to represent the preference intensities of an agent k in the matrix associated with R^k:

$$\begin{pmatrix} r_{11}^k & r_{12}^k & \cdots & r_{1n}^k \\ r_{21}^k & r_{22}^k & \cdots & r_{2n}^k \\ \cdots & \cdots & \cdots & \cdots \\ r_{n1}^k & r_{n2}^k & \cdots & r_{nn}^k \end{pmatrix}.$$

In the following section, a class of linguistic preference relations will be needed in order to ensure a monotonicity property of the individual Borda counts based on linguistic labels. Now we introduce this class.

Definition 4. A linguistic preference relation R^k over X based on L is *weak*[11] *max-max transitive* if and only if it is verified:

$$(x_i \succ_k x_j \quad \text{and} \quad x_j \succ_k x_p) \quad \Rightarrow \quad r_{ip}^k \geq \max\{r_{ij}^k, r_{jp}^k\},$$

for every $i, j, p \in \{1,\ldots,n\}$. With $T_L(X)$ we denote the set of weak max-max transitive linguistic preference relations over X based on L.

---

9 See Bezdek, Spillman and Spillman [5], Kacprzyk [24], Kacprzyk, Fedrizzi and Nurmi [25] and Marimin, Umano, Hatono and Tamura [31], among others.

10 See Bezdek, Spillman and Spillman [4], Nurmi [39] and García-Lapresta and Llamazares [17], among others.

11 Max-max transitivity for R^k was initially defined by $r_{ip}^k \geq \max\{r_{ij}^k, r_{jp}^k\}$. Weak (or restricted) conditions have been considered by Tanino [44] and Dasgupta and Deb [13], among others, when certain additional hypotheses are required. Concerning Definition 4 we consider preference intensities stronger than $l_{s/2}$.

We note that Bana e Costa and Vansnick [2, p. 109] consider a condition of rationality, similar to the above Definition, in the assessment of alternatives by means of linguistic labels.

Although linguistic labels can be represented by real numbers, intervals or fuzzy sets, according to Zadeh [50] fuzzy sets are the most appropriate tool for this purpose. In this paper we represent each element of L by means of a triangular fuzzy number (TFN). On this, see Zadeh [50], Delgado, Verdegay and Vila [14], Yager and Filev [47, pp. 23-24 and 203-214], Klir and Yuan [28, p. 102], Herrera, Herrera-Viedma and Verdegay [22], Bojadziev and Bojadziev [9, pp. 44-50], Von Altrock [45, pp. 324-332], Marimin, Umano, Hatono and Tamura [31], Herrera, Herrera-Viedma and Martínez [21] and Herrera and Martínez [23], among others.

Definition 5. Given a triplet of real numbers (a_1, a_2, a_3) such that $a_1 \leq a_2 \leq a_3$, the *triangular fuzzy number (TFN)* l associated with (a_1, a_2, a_3) is defined by its membership function $\mu_l : \nabla \to [0,1]$, where

$$\mu_l(x) = \begin{cases} 0, \text{if } x \leq a_1 \text{ or } x \geq a_3, \\ \dfrac{x - a_1}{a_2 - a_1}, \text{if } a_1 \leq x \leq a_2, \\ \dfrac{a_3 - x}{a_3 - a_2}, \text{if } a_2 \leq x \leq a_3. \end{cases}$$

We denote the previous TFN by $l = (a_1, a_2, a_3)$.

Now addition and subtraction of TFNs, necessary in the next section, are presented. On this, see for example Kaufmann and Gupta [26].

Definition 6. If $l = (a_1, a_2, a_3)$ and $l' = (a_1', a_2', a_3')$ are two TFNs, then their sum and subtraction are the TFNs $l + l' = (a_1 + a_1', a_2 + a_2', a_3 + a_3')$ and $l - l' = (a_1 - a_3', a_2 - a_2', a_3 - a_1')$, respectively.

In the following section it will be necessary to compare the TFNs corresponding to the new Borda count based on linguistic labels. In order to do this, we present a method for ranking TFNs which is related to one appearing in Yao and Wu [48].

Definition 7. Given a TFN $l = (a_1, a_2, a_3)$, the *real number associated with* l is $\bar{l} = \dfrac{a_1 + 2a_2 + a_3}{4}$. Then, the binary relation $>$ over the set of TFNs is defined by

$$l > l' \iff \begin{cases} \bar{l} > \bar{l}' \\ \text{or} \\ \bar{l} = \bar{l}' \quad \text{and} \quad a_3 > a_3' \\ \text{or} \\ \bar{l} = \bar{l}', \quad a_3 = a_3' \quad \text{and} \quad a_1 > a_1', \end{cases}$$

where $l = (a_1, a_2, a_3)$ and $l' = (a_1', a_2', a_3')$ are two TFNs.

It is easy to check that $>$ is a linear order. We note that, being L a set of TFNs, this ranking will be used to order its elements.

The above mentioned ranking system for fuzzy numbers considered by Yao and Wu [48] is based on the decomposition principle and a signed distance which is defined, specifically for TFNs, as

$$d(l,0) = \frac{1}{2}\int_0^1 [a_1 + (a_2 - a_1)\,\alpha + a_3 - (a_3 - a_2)\,\alpha]\,d\alpha = \frac{1}{4}(2a_2 + a_1 + a_3),$$

where $l = (a_1, a_2, a_3)$ and $0 \le \alpha \le 1$. And then, a non-complete ordering is defined by $l > l' \Leftrightarrow d(l,0) > d(l',0)$. We note that, taking into account $d(l,0) = \bar{l}$, the ordering introduced in Definition 5 becomes complete.

Table 1: Linguistic labels.

Label	Meaning	TFN
$r_{ij}^k = l_6$	agent k prefers x_i to x_j in definite degree	(1.0, 1.0, 1.0)
$r_{ij}^k = l_5$	agent k prefers x_i to x_j in high degree	(0.8, 0.9, 1.0)
$r_{ij}^k = l_4$	agent k prefers x_i to x_j in low degree	(0.5, 0.7, 0.9)
$r_{ij}^k = l_3$	agent k is indifferent between x_i and x_j	(0.3, 0.5, 0.7)
$r_{ij}^k = l_2$	agent k prefers x_j to x_i in low degree	(0.1, 0.3, 0.5)
$r_{ij}^k = l_1$	agent k prefers x_j to x_i in high degree	(0.0, 0.1, 0.2)
$r_{ij}^k = l_0$	agent k prefers x_j to x_i in definite degree	(0.0, 0.0, 0.0)

Remark 3. If $l = (a_1, a_2, a_3)$ is a TFN such that $a_2 - a_1 = a_3 - a_2$, then $\bar{l} = a_2$. Consequently, if $l = (a_1, a_2, a_3)$ and $l' = (a_1', a_2', a_3')$ are two TFNs such that $a_2 - a_1 = a_3 - a_2$ and $a_2' - a_1' = a_3' - a_2'$, then

$$l > l' \iff \begin{cases} a_2 > a_2' \\ \text{or} \\ a_2 = a_2' \quad \text{and} \quad a_3 > a_3'. \end{cases}$$

Example 1. We consider a concrete set $L=\{l_0,l_1,l_2,l_3,l_4,l_5,l_6\}$ of linguistic labels. Table 1 contains the meaning of the linguistic preference relation over X based on L of the agent k.

4 A Borda count based on linguistic labels

For each agent $k\in\{1,...,m\}$, we shall define a new count to evaluate each alternative according to intensities of preference in pairwise comparisons. In the classic Borda count, the value assigned by an individual count is the number of alternatives considered worse than it for an ordinary preference relation P^k. Now, analogously, this role will be played by the ordinary preference relation $\succ_k$ induced by the linguistic preference relation over X based on L, R^k.

In what follows, we shall suppose that each linguistic label of L is related to a TFN, and $l_0=(0,0,0)$. Consequently, $l_h < l_h + l_{h'}$, for all $h,h'\in\{1,\ldots,s\}$. Next, the introduced individual count will add up the preference intensities among a fixed alternative and those considered worse than it according to $\succ_k$. Taking into account individual counts considered in Section 2 and Definition 3, agent k gives the alternative x_i the value:

$$r_k(x_i)=\begin{cases}\sum\limits_{x_i\succ_k x_j} r_{ij}^k, & \text{if } \{j\in\{1,\ldots,n\}\mid r_{ij}^k > l_{s/2}\}\neq\varnothing \\ l_0, & \text{otherwise.}\end{cases}$$

Again, with these individual counts a collective one is obtained:

$$\boldsymbol{r}(x_i)=\sum_{k=1}^{m} r_k(x_i).$$

The collective preference relation P^{LB} is defined by:

$$x_i P^{LB} x_j \Leftrightarrow \boldsymbol{r}(x_i) > \boldsymbol{r}(x_j),$$

which is negatively transitive. Thus, the highest scored alternatives (maximals for P^{LB}) will be chosen.

We have to point out that the classic Borda count gives individual scores belonging to the set {0, 1,..., $n-1$ }, while the fuzzy Borda count takes its values in the set of TFNs.

In this linguistic framework we also consider a monotoniciy property. As in Section 2, when two alternatives are compared by an agent, the highest assessment must be given to the preferred alternative according to $\succ_k$.

Definition 8. The count r_k is *monotonic* if and only if

$$x_i \succ_k x_j \quad\Rightarrow\quad r_k(x_i) > r_k(x_j),$$

for all pairs of alternatives $x_i, x_j \in X$.

According to the classic Borda count, one can believe that the count r_k is monotonic if the ordinary preference relation $\succ_k$ is transitive. However, as we shall show, this assumption is not sufficient. However, weak max-max transitivity in the agents' individual preferences implies the monotonicity of the count r_k.

Proposition 2. If $R^k \in T_L(X)$, then the count r_k is monotonic.

Proof: For each $i \in \{1,...,n\}$ consider the set $P(i) = \{p \in \{1,\dots,n\} \mid r_{ip}^k > l_{s/2}\}$. Suppose $x_i \succ_k x_j$, i. e., $r_{ij}^k > l_{s/2}$. Since $P(i) \neq \varnothing$, we have $r_k(x_i) = \sum_{p \in P(i)} r_{ip}^k$. If $P(j) = \varnothing$, then $r_k(x_j) = l_0$ and, consequently, $r_k(x_i) > r_k(x_j)$. If $P(j) \neq \varnothing$, then $r_k(x_j) = \sum_{p \in P(j)} r_{jp}^k$. Now we justify that $P(j) \subset P(i)$ is verified. Suppose $p \in P(j)$, i.e., $r_{jp}^k > l_{s/2}$. By hypothesis, $r_{ij}^k > l_{s/2}$; therefore, $r_{ip}^k \geq \max\{r_{ip}^k, r_{jp}^k\} > l_{s/2}$ and so $p \in P(i)$. What is more, this inclusion is strict, because $r_{ij}^k > l_{s/2}$ implies $j \in P(i)$, while $j \notin P(j)$, since by Remark 2 $r_{jj}^k = l_{s/2}$. Now, if $l \in P(j) \subset P(i)$, then $r_{ip}^k \geq \max\{r_{ij}^k, r_{jp}^k\} \geq r_{jp}^k$ and, consequently, $r_k(x_i) > r_k(x_j)$.

Remark 4. If R^k is a linguistic preference relation over X based on L, such that $\succ_k$ is transitive, but does not satisfies weak max-max transitivity, then r_k is not necessarily monotonic. This is shown in the following example. Let R^k be the linguistic preference relation over $X = \{x_1, x_2, x_3, x_4, x_5\}$ based in the set $L = \{l_0, l_1, l_2, l_3, l_4, l_5, l_6\}$ of Example 1, whose matrix is:

$$\begin{pmatrix} l_3 & l_4 & l_4 & l_4 & l_4 \\ l_2 & l_3 & l_6 & l_6 & l_6 \\ l_2 & l_0 & l_3 & l_3 & l_3 \\ l_2 & l_0 & l_3 & l_3 & l_3 \\ l_2 & l_0 & l_3 & l_3 & l_3 \end{pmatrix}.$$

It is easy to check that $\succ_k$ is transitive and $R^k \notin T_L(X)$. We note that $x_1 \succ_k x_2$, because $r_{12}^k = l_4 > l_3$. Nevertheless,

$$r_k(x_1) = l_4 + l_4 + l_4 + l_4 = (2, 2.8, 3.6) < (3, 3, 3) = l_6 + l_6 + l_6 = r_k(x_2).$$

Now we justify that if an agent ranks alternatives in a linear manner, then different intensities of preference among alternatives ensure a kind of monotonicity in differences in scorings between pairs of alternatives.

Definition 9. Let R^k be a linguistic preference relation over X based on L. R^k is *linear* if and only if it is satisfied $x_i \neq x_j \Rightarrow r_{ij}^k \neq l_{s/2}$, for all $x_i, x_j \in X$.

Proposition 3. If $R^k \in T_L(X)$ is linear, then for all $x_i, x_j, x_l \in X$ it is satisfied:

1. $r_{il}^k > r_{ij}^k > l_{s/2} \quad \Rightarrow \quad r_k(x_i) - r_k(x_l) > r_k(x_i) - r_k(x_j)$.
2. $r_{il}^k > r_{jl}^k > l_{s/2} \quad \Rightarrow \quad r_k(x_i) - r_k(x_l) > r_k(x_j) - r_k(x_l)$.

Proof.

1. Firstly, we will see $r_{jl}^k > l_{s/2}$. If $r_{jl}^k < l_{s/2}$, i.e., $r_{lj}^k > l_{s/2}$, then we would have $r_{ij}^k \geq \max\ \{r_{il}^k, r_{lj}^k\}$, hence $r_{ij}^k \geq r_{il}^k$, contrary to the hypothesis; on the other hand, if $r_{jl}^k = l_{s/2}$, then $x_j = x_l$, in contradiction with the hypothesis. Then, by Proposition 2, we have $r_k(x_j) > r_k(x_l)$ and, consequently,
$$r_k(x_i) - r_k(x_l) > r_k(x_i) - r_k(x_j).$$
2. Analogously. □

Remark 5. The previous results are false if $R^k \in T_L(X)$ is not linear. Let R^1, R^2 be the linguistic preference relations over $X = \{x_1, x_2, x_3, x_4\}$ based on the set L of the Example 1, whose matrices are

$$\begin{pmatrix} l_3 & l_5 & l_4 & l_6 \\ l_1 & l_3 & l_3 & l_5 \\ l_2 & l_3 & l_3 & l_4 \\ l_0 & l_1 & l_2 & l_3 \end{pmatrix}, \quad \begin{pmatrix} l_3 & l_3 & l_5 & l_4 \\ l_3 & l_3 & l_4 & l_6 \\ l_1 & l_2 & l_3 & l_3 \\ l_2 & l_0 & l_3 & l_3 \end{pmatrix},$$

respectively. We note that $R^1, R^2 \in T_L(X)$ and they are not linear ($r_{23}^1 = r_{12}^2 = l_3$). While $r_{12}^1 = l_5 > l_4 = r_{13}^1$, we have
$$r_1(x_1) - r_1(x_2) = (1.3, 1.7, 2.1) < (1.4, 1.9, 2.4) = r_1(x_1) - r_1(x_3).$$
Consequently, the first item of Proposition 3 is not verified. On the other hand, $r_{13}^2 = l_5 > l_4 = r_{23}^2$, but
$$r_2(x_1) - r_2(x_3) = (1.3, 1.6, 1.9) < (1.5, 1.7, 1.9) = r_2(x_2) - r_2(x_3).$$
Therefore, the second item of Proposition 3 is not verified either.

In the following examples we note that the new Borda count based on linguistic labels shares the same drawbacks as the classic one: unfulfillment of the independence of irrelevant alternatives principle and manipulability.

Example 2. Consider $X = \{x_1, x_2, x_3\}$ and L the set of linguistic labels introduced in Example 1. Let $R^1, R^2, R^3 \in T_L(X)$ whose matrices are

$$\begin{pmatrix} l_3 & l_5 & l_6 \\ l_1 & l_3 & l_6 \\ l_0 & l_0 & l_3 \end{pmatrix}, \begin{pmatrix} l_3 & l_0 & l_1 \\ l_6 & l_3 & l_6 \\ l_5 & l_0 & l_3 \end{pmatrix}, \begin{pmatrix} l_3 & l_5 & l_2 \\ l_1 & l_3 & l_1 \\ l_4 & l_5 & l_3 \end{pmatrix},$$

respectively. Then, we have

$$r(x_1) = l_5 + l_6 + l_5 = (2.6, 2.8, 3) < (3, 3, 3) = l_6 + l_6 + l_6 = r(x_2)$$

and, consequently, $x_2 P^{LB} x_1$.

Now we consider another profile of individual preferences, $\overline{R}^1, \overline{R}^2, \overline{R}^3 \in T_L(X)$ whose matrices are

$$\begin{pmatrix} l_3 & l_5 & l_6 \\ l_1 & l_3 & l_6 \\ l_0 & l_0 & l_3 \end{pmatrix}, \begin{pmatrix} l_3 & l_0 & l_6 \\ l_6 & l_3 & l_6 \\ l_0 & l_0 & l_3 \end{pmatrix}, \begin{pmatrix} l_3 & l_5 & l_6 \\ l_1 & l_3 & l_6 \\ l_0 & l_0 & l_3 \end{pmatrix},$$

respectively. Then, we have

$$\overline{r}(x_1) = l_5 + l_6 + l_6 + l_5 + l_6 = (4.6, 4.8, 5) > (4, 4, 4) = l_6 + l_6 + l_6 + l_6 = \overline{r}(x_2)$$

and, consequently, $x_1 P^{LB} x_2$.

We note that the preferences of the three agents over the alternatives x_1 and x_2 are the same in the two profiles. However, x_2 defeats x_1 in the first profile, and x_1 defeats x_2 in the second profile. Therefore, x_3 has been relevant in the comparison between x_1 and x_2.

Example 3. Again consider $X = \{x_1, x_2, x_3\}$ and L the set of linguistic labels introduced in Example 1. Let $R^1, R^2, R^3 \in T_L(X)$ whose matrices are

$$\begin{pmatrix} l_3 & l_6 & l_6 \\ l_0 & l_3 & l_5 \\ l_0 & l_1 & l_3 \end{pmatrix}, \begin{pmatrix} l_3 & l_0 & l_1 \\ l_6 & l_3 & l_6 \\ l_5 & l_0 & l_3 \end{pmatrix}, \begin{pmatrix} l_3 & l_4 & l_2 \\ l_2 & l_3 & l_1 \\ l_4 & l_5 & l_3 \end{pmatrix},$$

respectively. Since

$$r(x_1) = l_6 + l_6 + l_4 = (2.5, 2.7, 2.9),$$
$$r(x_2) = l_5 + l_6 + l_6 = (2.8, 2.9, 3),$$
$$r(x_3) = l_5 + l_4 + l_5 = (2.1, 2.5, 2.9),$$

we have $r(x_2) > r(x_1) > r(x_3)$, and x_2 is the winner alternative.

The matrix of the third agent shows that $x_3 \succ_3 x_1 \succ_3 x_2$. Therefore, the worst alternative for this agent is the winner. Now, we suppose that the third agent declares insincerely its preferences by $\overline{R}^3 \in T_L(X)$, whose matrix is

$$\begin{pmatrix} l_3 & l_6 & l_2 \\ l_0 & l_3 & l_0 \\ l_4 & l_6 & l_3 \end{pmatrix}.$$

Then, the new results are

$$\begin{aligned} \bar{r}(x_1) &= l_6 + l_6 + l_6 = (3, 3, 3), \\ \bar{r}(x_2) &= l_5 + l_6 + l_6 = (2.8, 2.9, 3), \\ \bar{r}(x_3) &= l_5 + l_4 + l_6 = (2.3, 2.6, 2.9). \end{aligned}$$

Thus, $\bar{r}(x_1) > \bar{r}(x_2) > \bar{r}(x_3)$, and now x_1 is the winner alternative, the second best alternative for the third agent. Consequently, the new Borda count based on linguistic labels can be manipulated.

5 Concluding remarks

The classic Borda count requires a very strict assumption: agents have to order linearly the alternatives. The generalization of the classic Borda count presented in Section 2 allows the agents to show indifference among alternatives, and only demands transitivity in their preferences. In both cases, agents only can show which alternatives are preferred in pairwise comparisons. The use of fuzzy preferences allows the agents to express intensities of preference among alternatives. The most extended tradition in the fuzzy approach of the preference modeling consists in assessing the intensities of preference in the unit interval. However, this approach does not take into account the way in which agents feel their preferences: the natural language used in real life. This is the reason why we have considered a Borda-type decision procedure based on linguistic labels. These labels have been represented by fuzzy triangular numbers and an example has been shown. Nevertheless, other linguistic labels can be given, according to the characteristics of the alternatives and the agents, in each concrete decision problem.

One of the advantages of the Borda count in connection with non-ranked methods in collective decision making, such as approval voting, is its flexibility of gradation. This aspect is even improved in our approach, allowing the agents to express their preferences more faithfully than in the classic Borda count, through different linguistic labels. Moreover, ties are less likely to appear than in the classic case by using TFNs to represent linguistic labels:

- In the classic Borda count total scores belong to $\{0, 1, \ldots, m(n-1)\}$.

- In the Borda count based on linguistic labels total scores are the addition of at most $m(n-1)$ TFNs, each one of them belonging to $\{l_{s/2+1},\ldots,l_s\}$.

Consequently, the probability of ties appearance in total scores of the Borda count based on linguistic labels is lower than in the classic Borda count.

6 Acknowledgements

We acknowledge an anonymous referee for valuable comments. The financial support of the Junta de Castilla y León (Consejería de Educación y Cultura, Proyecto VA09/98) is gratefully acknowledged.

References

1. Arrow K. J., *Social Choice and Individual Values,* 2nd edition (Wiley, New York, 1963).
2. Bana e Costa A. and Vansnick J. C., Applications of the MACBETH approach in the framework of an additive aggregation model, *Journal of Multi-Criteria Decision Analysis* **6** (1997) pp. 10–14.
3. Barrett C. R., Pattanaik P. K. and Salles, M., Rationality and aggregation of preferences in an ordinally fuzzy framework, *Fuzzy Sets and Systems* **49** (1992) pp. 9–14.
4. Bezdek J. C., Spillman B. and Spillman R., A fuzzy relation space for group decision theory, *Fuzzy Sets and Systems* **1** (1978) pp. 255–268.
5. Bezdek J. C., Spillman B. and Spillman R., Fuzzy relation spaces for group decision theory: An application, *Fuzzy Sets and Systems* **4** (1979) pp. 5–14.
6. Biswas T., Efficiency and consistency in group decisions, *Public Choice* **80** (1994) pp. 23–34.
7. Black D., *The Theory of Committees and Elections* (Cambridge University Press, London, 1958).
8. Black D., Partial justification of the Borda count, *Public Choice* **28** (1976) pp. 1–15.
9. Bojadziev G. and Bojadziev M., *Fuzzy Logic for Business, Finance, and Management* (World Scientific, Singapore, 1997).
10. Bonissone P. P. and Decker K. S., Selecting uncertainty calculi and granularity: an experiment in trading-off precision and complexity. In *Uncertainty in Artificial Intelligence,* ed. by L. H. Kanal and J. F. Lemmer (North-Holland, Amsterdam, 1986) pp. 217–247.
11. Cook W. D. and Seiford L. M., On the Borda – Kendall consensus method for priority ranking problems, *Management Science* **28** (1982) pp. 621–637.

12. Copeland A. H., A 'reasonable' social welfare function, Mimeo, Notes from a seminar on applications of mathematics to the social sciences, University of Michigan, 1951.
13. Dasgupta M. and Deb R., Transitivity and fuzzy preferences, *Social Choice and Welfare* **13** (1996) pp. 305–318.
14. Delgado M., Verdegay J. L. and Vila M. A., Linguistic decision making models, *International Journal of Intelligent Systems* **7** (1993) pp. 479–492.
15. Dummett M., The Borda count and agenda manipulation, *Social Choice and Welfare* **15** (1998) pp. 289–296.
16. Fishburn P. C., Multiperson decision making: a selective review. In *Multiperson Decision Making Using Fuzzy Sets and Possibility Theory,* ed. by J. Kacprzyk and M. Fedrizzi (Kluwer Academic Publishers, Dordrecht, 1990) pp. 3–27.
17. García-Lapresta J. L. and Llamazares B., Aggregation of fuzzy preferences: Some rules of the mean, *Social Choice and Welfare* **17** (2000) pp. 673–690.
18. García-Lapresta J. L. and Martínez-Panero M., Borda count versus approval voting: A fuzzy approach, *Public Choice,* forthcoming.
19. Gärdenfors, P., Positionalist voting functions, *Theory and Decision* **4** (1973) pp. 1–24.
20. Goguen J. A., L-fuzzy sets, *Journal of Mathematical Analysis and Applications* **18** (1967) pp. 157–168.
21. Herrera F., Herrera-Viedma E. and Martínez L., A fusion approach for managing multi-granularity linguistic term sets in decision making, *Fuzzy Sets and Systems* **114** (2000) pp. 43–58.
22. Herrera F., Herrera-Viedma E. and Verdegay J. L., A linguistic decision process in group decision making, *Group Decision and Negotiation* **5** (1996) pp. 165–176.
23. Herrera F. and Martínez L., A 2-tuple fuzzy linguistic representation model for computing with words, *IEE Transactions of Fuzzy Systems* **8** (2000) pp. 746–752.
24. Kacprzyk J., Group decision making with a fuzzy linguistic majority, *Fuzzy Sets and Systems* **11** (1986) pp. 105–118.
25. Kacprzyk J., Fedrizzi M. and Nurmi H., Group decision making and consensus under fuzzy preferences and fuzzy majority, *Fuzzy Sets and Systems* **49** (1992) pp. 21–32.
26. Kaufmann A. and Gupta M., *Introduction to Fuzzy Arithmetic* (International Thomson Computer Press, Boston, 1991).
27. Kemeny J., Mathematics without numbers, *Daedalus* **88** (1959) pp. 571–591.
28. Klir G. and Yuan B., *Fuzzy Sets and Fuzzy Logic. Theory and Applications* (Prentice-Hall, New Jersey, 1995).
29. Marchant T. , *Agrégation de relations valuées par la méthode de Borda en vue d'un rangement. Considérations axiomatiques*, Ph. D. Thesis, Université Libre de Bruxelles, 1996.

30. Marchant T., Does the Borda Rule provide more than a ranking?, *Social Choice and Welfare* **17** (2000) pp. 381–391.
31. Marimin, Umano M., Hatono I. and Tamura H., Linguistic labels for expressing fuzzy preference relations in fuzzy group decision making, *IEEE Transactions on Systems, Man, and Cybernetics. Part B: Cybernetics* **28** (1998) pp. 205–218.
32. McLean I. and Urken A. B. (eds.), *Classics of Social Choice* (Ann Arbor – The University of Michigan Press, 1995).
33. McLean I. and Urken A. B., Independence of irrelevant alternatives before Arrow, *Mathematical Social Sciences* **30** (1995) pp. 107–126.
34. Miller G. A., The organization of lexical memory. In *The Pathology of Memory*, ed. by G. A. Talland and N. C. Waugh (Academic, New York, 1969).
35. Morales J. I., *Memoria Matemática sobre el Cálculo de la Opinion en las Elecciones* (Imprenta Real, Madrid, 1797). English version in [32, pp. 197–235].
36. Morales J. I., *Apéndice á la Memoria Matemática sobre el Cálculo de la Opinion en las Elecciones* (Imprenta de Sancha, Madrid, 1805).
37. Mueller D. C., *Public Choice* (Cambridge University Press, London, 1979).
38. Nitzan S. and Rubinstein A., A further characterization of Borda ranking method, *Public Choice* **3** (1981) pp. 153–158.
39. Nurmi H., Approaches to collective decision making with fuzzy preference relations, *Fuzzy Sets and Systems* **6** (1981) pp. 249–259.
40. Saari D. G., *Basic Geometry of Voting* (Springer – Verlag, Berlin, 1995).
41. Saari D. G. and Merlin V. R., The Copeland method I: relationships and the dictionary, *Economic Theory* **8** (1996) pp. 51–76.
42. Sen A. K., Social Choice Theory: a re-examination, *Econometrica* **45** (1977) pp. 53–89.
43. Straffin Jr. P. D., *Topics in the Theory of Voting* (Birkhäuser, Boston, 1980).
44. Tanino T., Fuzzy preference orderings in group decision making, *Fuzzy Sets and Systems* **12** (1984) pp. 117–131.
45. Von Altrock C., *Fuzzy Logic & Neuro Fuzzy Applications in Business & Finance* (Prentice Hall, New Jersey, 1997).
46. Yager R. R., Families of OWA operators, *Fuzzy Sets and Systems* **59** (1993) pp. 125–148.
47. Yager R. and Filev D., *Essentials of Fuzzy Modeling and Control* (John Wiley & Sons, New York, 1994).
48. Yao J. S. and Wu K., Ranking fuzzy numbers based on decomposition principle and signed distance, *Fuzzy Sets and Systems* **116** (2000) pp. 275—288.
49. Young H. P., Optimal voting rules, *Journal of Economic Perspectives* **9** (1995) pp. 51–64.
50. Zadeh L. A., The concept of a linguistic variable and its application to approximate reasoning, *Information Sciences.* Part I: **8** pp. 199–249. Part II: **8** pp. 301–357. Part III: **9** pp. 43–80 (1975).

DEVELOPING SORTING MODELS USING PREFERENCE DISAGGREGATION ANALYSIS: AN EXPERIMENTAL INVESTIGATION

MICHAEL DOUMPOS AND CONSTANTIN ZOPOUNIDIS

Technical University of Crete, Dept. of Production Engineering and Management, Financial Engineering Laboratory, University Campus, 73100 Chania, Greece, kostas@ergasya.tuc.gr

Within the field of multicriteria decision aid (MCDA), sorting refers to the assignment of a set of alternatives into predefined homogenous groups defined in an ordinal way. The real–world applications of this type of problem extend to a wide range of decision–making fields. Preference disaggregation analysis provides the framework for developing sorting models through the analysis of the global judgment of the decision–maker using mathematical programming techniques. However, the automatic elicitation of preferential information through the preference disaggregation analysis raises several issues regarding the impact of the parameters involved in the model development process on the performance and the stability of the developed models. The objective of this paper is to shed light on this issue. For this purpose the UTADIS preference disaggregation sorting method (UTilités Additives DIScriminantes) is considered. The conducted analysis is based on an extensive Monte Carlo simulation and useful findings are obtained on the aforementioned issues.

1 Introduction

A wide range of real–world decision making problems require the assignment of a discrete set of alternatives described over a set of attributes into predefined homogenous groups. Such problems are usually referred to as "discrimination" or "classification" problems. In case the groups are defined in an ordinal way, MCDA researchers have introduced the term "sorting". In sorting problems the attributes that describe the alternatives have the form of evaluation criteria, while the groups are defined in a preferentially ordered way from the most to the least preferred ones.

MCDA provides a variety of techniques to address sorting problems. The development of sorting models through MCDA techniques often requires the decision maker to define specific information on the parameters of the developed model. This is the case in the outranking relations approach and the corresponding sorting techniques [2, 10, 14, 15, 16]. The required information includes technical and non–technical parameters such as the weights of the evaluation criteria, preference, indifference and veto thresholds, etc. The direct specification of these parameters by the decision–maker ensures that the developed sorting model fits his/her judgment policy. Nevertheless, the procedure required for the decision analyst to elicit this information from the decision maker is often time–consuming. This can be an inhibitory factor for its real–world application.

Preference disaggregation analysis [6, 7, 8] provides the framework and the means to overcome this problem. The preference disaggregation approach refers to the analysis (disaggregation) of the global preferences (judgement policy) of the decision maker in order to identify the criteria aggregation model that underlies the preference result (sorting). Similarly to multiattribute utility theory (MAUT), preference disaggregation analysis uses common utility decomposition forms to model the decision maker's preferences. Instead of employing a direct procedure for estimating the global utility model (MAUT), preference disaggregation analysis uses regression–based techniques (indirect estimation procedure). More specifically, in preference disaggregation analysis the parameters of the utility decomposition model are estimated through the analysis of the decision maker's overall preference on some reference alternatives. The problem is then to estimate a utility function (usually additive) that is as consistent as possible with the known subjective preferences of the decision maker. Of course, this methodological framework can be used to develop alternative criteria aggregation models, other than utility functions. For instance, Mousseau and Slowinski [11] used the philosophy of preference disaggregation analysis to develop an outranking relation for sorting purposes on the basis of the ELECTRE TRI method [16]. The authors also developed a multicriteria decision support system for this purpose [12].

Overall, the general scheme used in preference disaggregation analysis is similar to the use of a training sample for model development in statistics/econometrics, neural networks, and machine learning. However, using such an indirect procedure to estimate a sorting model and to elicit preferential information regarding the decision maker's judgment policy, requires a careful investigation of the parameters of the procedure employed. Such an investigation can provide insight on the impact of these parameters to the performance of the developed model (classification accuracy), as well as on the stability of the structural parameters of the model. The former issue is of major importance in terms of the consistency of the recommendations that the developed model provides as opposed to the observed outcome (e.g., decisions taken by the decision maker), whereas the latter issue is crucial in terms of the interpretation of the model in providing meaningful decision support.

The objective of this paper is to address the above issues for the case of the UTADIS method [18]. The UTADIS method employs the framework of preference disaggregation analysis in developing an additive utility model for sorting purposes. The analysis conducted in this paper is based on an experimental investigation of all the major parameters of the method that affect the outcome of the model development procedure. On the basis of the results, useful conclusions are obtained with regard to the parameters affecting the performance (classification accuracy) of the developed sorting models as well as to the parameters affecting the stability of the models.

The rest of the paper is organized as follows. Section 2 outlines the UTADIS method and its use for developing sorting models. Section 3 discusses the major issues considered during the model development process in the context of the UTADIS method and presents the methodology that is used to investigate these issues, while section 4 discusses the obtained results. Finally, section 5 summarizes the major findings of this study and proposes interesting future research directions.

2 The UTADIS method

Formally stated, the sorting problem involves the assignment of a set A of alternatives evaluated along a vector of n criteria $\mathbf{g} = (g_1, g_2, \ldots, g_n)$ into q groups C_1, C_2, …, C_q. The groups are defined in an ordinal way such that C_1 includes the most preferred alternatives and C_q the least preferred ones. Based on the methodological framework of preference disaggregation analysis, the UTADIS method performs the assignment (sorting) of the alternatives through the development of an additive utility function of the following form:

$$U(\mathbf{g}) = \sum_{i=1}^{m} u_i(g_i) \in [0,1]$$

which can be equivalently written as $U'(\mathbf{g}) = \sum_{i=1}^{m} p_i u_i'(g_i) \in [0,1]$, where p_i are scaling constants representing the significance of the evaluation criteria defined such that they sum up to 1, and $u_i'(g_i) = u_i(g_i)/p_i \in [0,1]$. The marginal utilities functions $u_i(g_i)$ have a piece–wise linear form (Figure 1). For each evaluation criterion g_i the range $[g_{i*}, g_i^*]$ of its values is defined. The values g_{i*} and g_i^* of criterion g_i correspond to the least and the most preferred values of the criterion respectively. In defining the piece–wise linear form of the marginal utility functions, each criterion's range is divided into b_i–1 subintervals $[g_i^p, g_i^{p+1}]$, p=1, 2, ..., b_i–1 (Figure 1). The piece–wise linear marginal utility function estimated through this process approximates the actual function.

Using the piece–wise linear marginal utility functions of the evaluation criteria, the global utility $U(\mathbf{g}_j)$ of every alternative $a_j=(g_{j1}, g_{j2}, \ldots, g_{jn})$ can be estimated through simple linear interpolation in the appropriate linear segments of the criteria marginal utility functions as follows:

$$U(\mathbf{g}_j) = \sum_{i=1}^{n} \left(\sum_{p=1}^{r_{ji}-1} w_{ip} + \frac{g_{ji} - g_i^{r_{ji}}}{g_i^{r_{ji}+1} - g_i^{r_{ji}}} w_{ir_{ji}} \right) \tag{1}$$

where, r_{ji} $(1 \le r_{ji} \le b_i - 1)$ denotes the subinterval $[g_i^{r_{ji}}, g_i^{r_{ji}+1}]$ in which the performance g_{ji} of alternative a_j on criterion g_i belongs, and $w_{ih} = u_i(g_i^{h+1}) - u_i(g_i^h) \ge 0$ $(1 \le h \le b_i - 1)$.

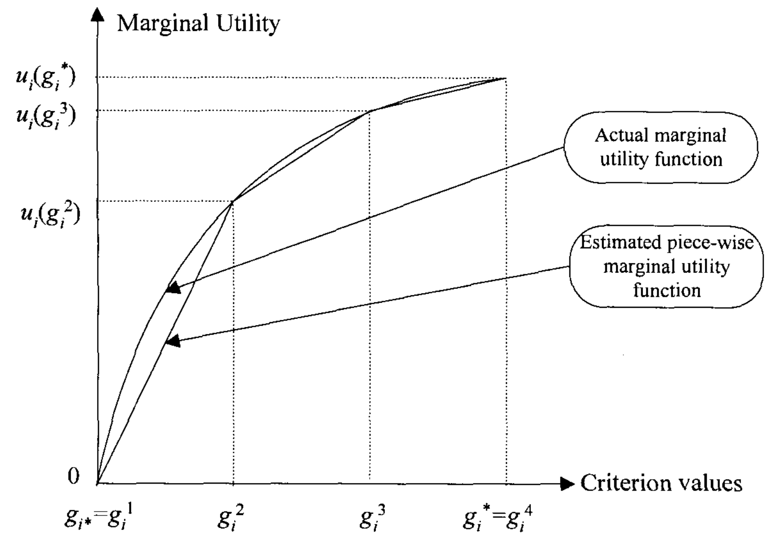


Figure 1. Piece–wise linear form of marginal utility functions.

The global utility serves as an index used to decide upon the sorting of the alternatives into the predefined groups. The sorting is performed through the comparison of the global utilities of the alternatives to some utility thresholds that define the lower bound of each group, as follows:

$$\left.\begin{array}{ll} U(\mathbf{g}_j) \geq u_1 & \Rightarrow a \in C_1 \\ u_2 \leq U(\mathbf{g}_j) < u_1 & \Rightarrow a \in C_2 \\ \dots\dots\dots\dots & \\ u_k \leq U(\mathbf{g}_j) < u_{k-1} & \Rightarrow a \in C_k \\ \dots\dots\dots\dots & \\ U(\mathbf{g}_j) < u_{q-1} & \Rightarrow a \in C_q \end{array}\right\} \qquad (2)$$

Thus, the development of the sorting model through the UTADIS method requires the determination of the marginal utility functions to obtain the specific form of the global utility function, as well as the selection of the utility thresholds u_k. The general scheme of preference disaggregation analysis is employed for this purpose. In particular, a reference set A' consisting of m alternatives $a_1, a_2, \ldots, a_m$ is used for model development (training sample). The reference set may include [8]:

1) a set of past decision alternatives,
2) a subset of the alternatives under consideration, such that $A' \subset A$, especially when A is large (this is analogous to statistical analysis where a sample is used for model development instead of the whole population).

3) a set of fictitious alternatives, consisting of performances on the criteria which can be easily judged by the decision maker to express his/her global judgment (sorting).

Using the sorting of the alternatives included in the reference set, the objective of the UTADIS method is to develop the additive utility model so that the sorting rules (2) can reproduce the predetermined sorting of the alternatives as accurately as possible. The classification errors that may occur in this process are the following:

$\sigma_j^+ = \max\{0, u_k - U(\mathbf{g}_j)\}$, $\forall a_j \in C_k$, $k = 1, 2, \ldots, q-1$

$\sigma_j^- = \max\{0, U(\mathbf{g}_j) - u_{k-1}\}$, $\forall a_j \in C_k$, $k = 2, 3, \ldots, q$

The errors σ^+ occur when the lower bound of group C_k is violated (utility threshold u_k), whereas the errors σ^- occur when the upper bound of group C_k is violated (utility threshold u_{k-1}). The development of the additive utility model that minimizes these errors is performed though the solution of the linear program[1]:

$$\text{Min } f = \left\{ \sum_{k=1}^{q} \left[\frac{\sum\limits_{\forall a_j \in C_k} \left(\sigma_j^+ + \sigma_j^-\right)}{m_k} \right] \right\} \tag{3}$$

s.t:

$$\sum_{i=1}^{n} \left(\sum_{p=1}^{r_{ji}-1} w_{ip} + \frac{g_{ji} - g_i^{r_{ji}}}{g_i^{r_{ji}+1} - g_i^{r_{ji}}} w_{ir_{ji}} \right) - u_1 + \sigma_j^+ \geq \delta_1, \qquad \forall a_j \in C_1 \tag{4}$$

$$\left.\begin{array}{l} \displaystyle\sum_{i=1}^{n} \left(\sum_{p=1}^{r_{ji}-1} w_{ip} + \frac{g_{ji} - g_i^{r_{ji}}}{g_i^{r_{ji}+1} - g_i^{r_{ji}}} w_{ir_{ji}} \right) - u_k + \sigma_j^+ \geq \delta_1 \\ \displaystyle\sum_{i=1}^{n} \left(\sum_{p=1}^{r_{ji}-1} w_{ip} + \frac{g_{ji} - g_i^{r_{ji}}}{g_i^{r_{ji}+1} - g_i^{r_{ji}}} w_{ir_{ji}} \right) - u_{k-1} - \sigma_j^- \leq -\delta_2 \end{array}\right\}, \qquad \forall a_j \in C_k \ (k = 2, 3, \ldots, q-1) \tag{5}$$

$$\sum_{i=1}^{n} \left(\sum_{p=1}^{r_{ji}-1} w_{ip} + \frac{g_{ji} - g_i^{r_{ji}}}{g_i^{r_{ji}+1} - g_i^{r_{ji}}} w_{ir_{ji}} \right) - u_{q-1} - \sigma_j^- \leq -\delta_2, \qquad \forall a_j \in C_q \tag{6}$$

$$\sum_{i=1}^{n} \sum_{p=1}^{b_i-1} w_{ip} = 1 \tag{7}$$

$$u_k - u_{k+1} \geq s, \qquad \forall k = 1, 2, \ldots, q-2 \tag{8}$$

---

[1] m_k denotes the number of alternatives in A' that belong into group C_k, whereas $\mathbf{g}_*$ and $\mathbf{g}^*$ are the vectors of the least and the most preferred values of the criteria, respectively.

$$w_{ip} \geq 0,\ \sigma_j^+ \geq 0,\ \sigma_j^- \geq 0, \quad \forall\ i=1, 2, ..., n, p=1, 2, ..., b_i-1, j=1, 2, ..., m \qquad (9)$$

$$\delta_1, \delta_2, s : \text{user–defined small positive constants } (s > \delta_1, \delta_2 \geq 0) \qquad (10)$$

After the optimal solution f^* of this linear program is obtained, a post optimality stage is carried out to investigate the stability of the optimal solution (generally, the above linear program has multiple optimal solutions). During the post optimality stage, other optimal or near optimal solutions are identified such that each satisfies the additional constraint $f' \leq (1+z)f^*$, where f' is the value of the objective function (3) for each new solution found during the post optimality analysis and z is a user–defined constant representing a small proportion of f^*. During the post optimality stage $n+q-1$ alternate optimal or near optimal solutions are identified. The first n solutions maximize the significance of each of the n criteria, whereas the remaining $q-1$ solutions maximize the corresponding utility thresholds. The final additive utility function is constructed from the average of all the solutions obtained through this process.

3 Experimental analysis

3.1 Issues to be considered

The use of the linear program (3)–(10) for the development of the sorting model in the UTADIS method provides increased flexibility to the model development process in terms of the measurement of the classification error [4] and the form of the marginal utility functions (additional constraints can be imposed to develop concave or convex functions). However, the resulting additive utility model is affected both by the technical parameters involved in the solution process (i.e., the user–defined constants δ_1, δ_2, s and z), as well as by the way that the piece–wise linear form of the marginal utility functions is considered (i.e., the way that each criterion's range is divided into subintervals).

The technical parameters define the solution space of the linear program (3)–(10) as well as of the linear programs solved during the post optimality stage. Generally, one should expect that a wider solution space of the linear programming formulation used in the UTADIS method is associated with increased instability in the developed additive utility model. In such case the developed model will be difficult to interpret. Nevertheless, there is no reason for considering an a–priori relationship between instability and classification performance. Consequently, the investigation of the impact of the technical parameters involved in the solution of the linear program (3)–(10) to the stability and the classification performance of the developed additive utility model is a significant issue.

The number of subintervals defined on the range of each criterion is also a crucial factor that affects the outcome of the model development process, especially in the case of quantitative criteria. From the brief discussion of the piece–wise linear form of the marginal utility functions in the previous section, it may seem that the definition of a larger number of subintervals provides the ability to perform a better approximation of the actual marginal utility function. However, this is rarely the case. In fact the specification of an arbitrary large number of subintervals increases the degrees of freedom of the additive utility function, resulting to its over–fitting on the data of the reference set (training sample).

In addition to the over–fitting phenomenon, it is also worth considering the stability of the developed additive utility model. In traditional statistical regression, to derive statistically significant conclusions on the estimates of n parameters of a model, one should use at least n+1 observations. Horsky and Rao [5] emphasize that this remark also holds in the case where a mathematical programming formulation is used for model development. In the case of the UTADIS method and the linear program (3)–(10) every basic solution includes as many variables as the number of constraints (i.e., $m_1 + 2\sum_{k=2}^{q-1} m_k + m_q + q-1$; [18]). Furthermore, the optimal solution will include the q–1 utility thresholds. Therefore, the optimal solution will include only $t = m_1 + 2\sum_{k=2}^{q-1} m_k + m_q$ incremental variables w. The remaining incremental variables will not be included in the optimal basis of the linear program (3)–(10). This indicates that if the number of subintervals is not defined properly, there can be cases where the linear program will have redundant incremental variables w. This case is likely to increase the instability of the model in terms of the significance of the evaluation criteria and the form of the marginal utility functions.

The traditional approach used to define the number of subintervals is to divide each quantitative criterion's range into a number of equal subintervals such that there is at least one alternative belonging into each one. Henceforth this naïve heuristic will be referred to as HEUR1. However, this heuristic does not consider the distribution of the alternatives belonging into different groups over each criterion's range. To accommodate this valuable information, this study proposes a new simple heuristic, which will be referred to as HEUR2. This heuristic is performed for all quantitative criteria in five steps as follows:

Step 1: Rank–order all alternatives a_j in A' according to their performances g_{ji} on each quantitative criterion g_i, from the least to the most preferred ones. Set the minimum acceptable number of alternatives belonging into a subinterval equal to zero ($\beta = 0$).

Step 2: Form all subintervals $[g_i^h, g_i^{h+1}]$, such that the alternative, whose performance is equal to g_i^h, belongs to a different group from the alternative whose performance is equal to g_i^{h+1}.

Step 3: Check the number of alternatives belonging into each subinterval formed after step 2. If the number of alternatives in a subinterval is less than β, then merge this subinterval with the precedent one (this check is skipped when $\beta = 0$).

Step 4: Check the consistency of the total number of subintervals formed after step 3 for all criteria, as opposed to the size of the linear program (3)–(10), i.e. the number of constraints. If the number of subintervals leads to the specification of more than $m_1 + 2\sum_{k=2}^{q-1} m_k + m_q$, incremental variables w, then set $\beta := \beta + 1$, and repeat the process from step 3; otherwise the procedure ends.

All the above issues are directly related to the development of an additive utility model for sorting purposes on the basis of the UTADIS method. The investigation of their impact on the performance and the stability of the developed sorting models is performed through an extensive simulation based on an experimental design, whose characteristics are outlined in the subsequent subsection.

3.2 Experimental design

The investigation of all issues raised in the previous subsection is performed in this study through a Monte Carlo simulation. The structural factors of this simulation are outlined in Table 1.

Table 1. Factors considered in the simulation.

Factors		Levels
F_1:	Statistical distribution of the data	Multivariate normal, Multivariate log–normal
F_2:	Number of groups	Two, Three
F_3:	Ratio of the number of alternatives in the reference set to the number of alternatives in the validation sample	20%, 30%, 40%, 50%, 60%, 70%

The simulation is based on the generation of random data that follow two kinds of distributions: the multivariate normal distribution and the multivariate log–normal distribution. The former is a symmetric distribution, whereas the latter is asymmetric (the methodology of Vale and Maurelli [17] is used to generate the data from the multivariate log–normal distribution). The generated data involve a set of alternatives described over five criteria. The alternatives are assigned into two or three groups, thus covering a wide range of sorting problems encountered in real–world practice. Each group C_k is characterized by the vector $\boldsymbol{\mu}_k$ of the mean performance on the evaluation criteria of the alternatives that belong to the group and the corresponding vector of standard deviations $\boldsymbol{\sigma}_k$, as follows:

$$(\boldsymbol{\mu}_1 = 4,\ \boldsymbol{\sigma}_1 = 4),\ (\boldsymbol{\mu}_2 = 2,\ \boldsymbol{\sigma}_2 = 2),\ (\boldsymbol{\mu}_3 = 1,\ \boldsymbol{\sigma}_3 = 1)$$

According to this specification the variation coefficient for the performance of the alternatives is equal to one for all criteria and groups. This ensures that in the case where three groups are considered, the overlapping between groups C_1 and C_2 is the same with the one between groups C_2 and C_3. Thus, the results obtained for the three–groups case are comparable to the results for the two–groups case, in the sense that the former are not affected by the differences in the overlapping of each pair of consecutive groups. The off–diagonal elements of the variance–covariance matrix are specified such that the correlation coefficient is 0.25 for all pairs of criteria and for all groups. This provides a reasonable degree of correlation while ensuring that the groups are not linearly separable, even in the multivariate normal case, since the corresponding variance–covariance matrices differ ($\mathbf{\Sigma}_k = 0.25\mathbf{\Sigma}_{k-1}$).

To ensure the ordering of the groups and the consistency of the generated data, the alternatives are produced such that there is no alternative of a less preferred group that dominates an alternative of a more preferred group. This is realized by imposing the following constraint during data generation:
$\mathbf{g}_j > \mathbf{g}_t, \forall a_j \in C_k, a_t \in C_{k+1}, k = 1, 2$.

On the basis of the above specifications, two samples are generated. The first one forms the reference set used for the development of the additive utility sorting model, whereas the second one forms the validation sample. The validation sample is used to investigate the generalizing ability of the sorting model developed on the basis of the reference set. In the two–groups case the number of alternatives included in the reference set ranges between 40–140 (40, 60, 80, 100, 120, 140), while the number of alternatives in the validation sample is fixed to 200 alternatives. In the three–groups case the number of alternatives included in the reference set ranges between 48–168 (48, 72, 96, 120, 144, 168), while the number of alternatives in the validation sample is fixed to 240 alternatives. Thus, in both cases the ratio of the number of alternatives in the reference set to the number of alternatives in the validation sample ranges between 20%–70% (Table 1). This enables the investigation of the stability and performance of the developed additive utility sorting models when additional information is included in the reference set (larger number of alternatives). In both the reference set and the validation sample the size of the groups (number of alternatives per group) is balanced such that each group includes m/q alternatives, where m is the number of alternatives in the reference set (or in the validation sample) and q is the number of groups (two or three).

For each combination of the three factors presented in Table 1 (24 combinations), the above experiment is replicated 10 times. Thus, during the above simulation 240 reference sets are examined. Each reference set corresponds to a validation sample having the same properties defined by the three factors of Table 1. In each replication of the experiment the UTADIS method is applied with different specifications of its main parameters that were outlined in the previous subsection. In particular the parameters that are considered in the simulation and their corresponding values tested are the following:

1. The user-defined constants δ_1 and δ_2 used in the right hand side of constraints (4)–(6): $\delta_1=\delta_2=\delta$=0.005, 0.015, 0.03.
2. The proportion z of the optimal classification error f^* considered during the post-optimality stage: z=0.5%, 2%, 5%, 10%.
3. The way that the subintervals are formed for the realization of the piece-wise linear form of the marginal utility functions. In this simulation four approaches are used to specify these subintervals: the heuristics HEUR1 and HEUR2, as well as the specification of an arbitrary limited number of subintervals for each criterion, namely 4 and 8 subintervals.

Therefore, the UTADIS method is applied 48 times in each replication of the experiment, whereas overall the method is applied 11,520 times throughout the whole simulation (48 combinations of the parameters of the method in 240 reference sets).

4 Presentation of results

The results obtained through the simulation described in the previous section are subject to a six-way analysis of variance (ANOVA). The six factors considered in this analysis include the three structural factors of the simulation (cf. Table 1) and the three factors related to the parameters of the UTADIS method that are considered (F_4: δ, F_5: z, F_6: subintervals). The ANOVA is performed for the sorting error rates in the validation sample and the stability of the additive utility models. In performing the ANOVA for the error rates, the transformation $2\arcsin\sqrt{\text{error rate}}$ is used in order to stabilize the variance of their variance [1, 9]. The stability of the additive utility models is measured as the variance of the significance of the evaluation criteria, over all the solutions obtained during the post optimality stage. Each of these solutions corresponds to a different additive utility model. Higher variance of the significance of the evaluation criteria in the developed additive utility models indicates increased instability and consequently a difficulty in deriving a secure interpretation of the model.

The ANOVA results for both the error rates and the stability of the sorting models are summarized in Tables 2 and 3 respectively. The effects presented in these tables are all significant at the 1% level, and each of them explains at least 0.3% of the total variance of the corresponding results (Hays ω^2). Other effects that did not meet these two requirements are not reported in order to retain the complexity of the analysis within manageable levels.

Table 2. ANOVA results for the sorting error rates in the validation sample.

	df	Sum of squares	Mean squares	F	Hays ω^2
F_2	1	222.782	222.782	28951.465	40.56%
F_1	1	128.339	128.339	16678.195	23.37%
F_6	3	42.112	14.037	1824.190	7.66%
$F_1 \times F_6$	3	40.335	13.445	1747.247	7.34%
$F_1 \times F_2 \times F_6$	3	7.764	2.588	336.335	1.41%
F_3	5	7.590	1.518	197.262	1.37%
$F_1 \times F_2$	1	5.142	5.142	668.276	0.93%
$F_2 \times F_6$	3	4.223	1.408	182.924	0.76%
$F_2 \times F_3$	5	2.393	0.479	62.203	0.43%
$F_3 \times F_6$	15	1.858	0.124	16.096	0.32%

Notes: *F_1: statistical distribution, F_2: number of groups, F_3: Ratio of the number of alternatives in the reference set to the number of alternatives in the validation sample, F_4: δ, F_5: z, F_6: subintervals*

Table 3. ANOVA results for the stability of the additive utility models

	df	Sum of squares	Mean squares	F	Hays ω^2
F_2	1	11.153	11.153	27187.938	49.14%
F_4	2	1.740	0.870	2121.341	7.66%
$F_2 \times F_4$	2	1.730	0.865	2108.148	7.62%
F_6	3	0.866	0.289	703.886	3.81%
F_5	3	0.798	0.266	648.238	3.51%
$F_2 \times F_6$	3	0.651	0.217	528.930	2.86%
F_1	1	0.259	0.259	631.180	1.14%
$F_2 \times F_5$	3	0.154	0.051	125.462	0.67%
F_3	5	0.146	0.029	71.112	0.63%
$F_2 \times F_3$	5	0.103	0.021	50.086	0.44%
$F_1 \times F_2 \times F_3$	5	0.082	0.016	40.122	0.35%

Notes: *F_1: statistical distribution, F_2: number of groups, F_3: Ratio of the number of alternatives in the reference set to the number of alternatives in the validation sample, F_4: δ, F_5: z, F_6: subintervals*

The above results indicate that all three structural factors considered in the simulation (F_1, F_2 and F_3) have a significant impact both on the sorting error rate as well as on the stability of the developed additive utility models. From these three factors the decision maker controls only the third one. Both the statistical distribution of the data (F_1) as well as the number of groups (F_2) are specified according to the nature of the problem itself, while the decision maker can decide upon the number of alternatives to be included in the reference set (F_3). Therefore, the subsequent analysis will focus only on the interactions of factors F_1 and F_2 with the rest of the factors considered in the experiment. From the remaining factors regarding the parameters of the model development process (F_4, F_5 and F_6), it is interesting to observe that neither the value of δ (F_4) nor the value of z (F_5) have a significant impact on the error rates of the additive utility sorting models developed through the UTADIS method. As expected (cf. section 3.1), their impact is significant only for the stability of the models. On the other hand, the way that the subintervals are specified (F_6) is a significant factor affecting both the error rate as well as the stability of the models.

Except for the one–way effect corresponding to the subintervals factor, its interaction with factors F_1, F_2 and F_3 is also significant in considering its impact on the error rates, whereas the only interaction that is significant for the stability of the models is the one with factor F_2. The corresponding results are summarized in Tables 4 and 5.

Table 4. Error rates and stability of the sorting models for the interactions of statistical distribution (F_1) by subinterval's specification (F_6), number of groups (F_2) by subinterval's specification (F_6) and reference set size (F_3) by subintervals' specification (F_6).

		Error rates					Stability				
		Subintervals (F_6)					Subintervals (F_6)				
Factors	Levels	4	8	HEUR1	HEUR2	Overall[1]	4	8	HEUR1	HEUR2	Overall[1]
F_1	Normal	21.62%	20.02%	20.46%	20.57%	20.67%	*	*	*	*	0.0332
	Log–normal	33.15%	32.73%	32.61%	21.72%	30.05%	*	*	*	*	0.0426
F_2	Two	21.24%	20.98%	21.25%	14.48%	19.49%	0.0717	0.0812	0.0813	0.0418	0.0690
	Three	33.53%	31.77%	31.82%	27.81%	31.23%	0.0074	0.0077	0.0073	0.0047	0.0068
F_3	20%	27.55%	27.08%	27.46%	23.66%	26.44%	*	*	*	*	0.0313
	30%	28.34%	27.46%	27.82%	22.90%	26.63%	*	*	*	*	0.0360
	40%	27.31%	26.40%	26.65%	21.29%	25.41%	*	*	*	*	0.0394
	50%	27.69%	26.49%	26.63%	20.56%	25.34%	*	*	*	*	0.0378
	60%	27.20%	25.85%	26.00%	19.72%	24.69%	*	*	*	*	0.0407
	70%	26.25%	24.97%	24.66%	18.75%	23.66%	*	*	*	*	0.0422
Overall[2]		27.39%	26.38%	26.54%	21.15%		0.0395	0.0445	0.0443	0.0233	

Notes: * *The blank entries for the interactions $F_1 \times F_6$ and $F_3 \times F_6$ indicate that these interactions are not significant in describing the stability of the additive utility models*

[1] *Overall results for factors F_1, F_2, F_3 irrespective of the subintervals specification (F_6).*

[2] *Overall results for factor F_6 (subintervals) irrespective of factors F_1, F_2, F_3.*

Table 5. Sorting error rates for the three–way interaction between the statistical distribution (F_1), the number of groups (F_2) and the subintervals' specification (F_6).

Subintervals	Normal distribution		Log–normal distribution	
	Two groups	Three groups	Two groups	Three groups
4	15.58%	27.66%	26.91%	39.40%
8	15.52%	24.53%	26.44%	39.01%
HEUR1	16.36%	24.56%	26.14%	39.08%
HEUR2	16.31%	24.83%	12.65%	30.80%

The above results indicate that the heuristic procedure HEUR2 for specifying the subintervals during the piece–wise linear formulation of the marginal utility functions, is a quite efficient approach to overcome the subintervals specification problem when quantitative criteria are considered. Overall, this procedure yields significantly lower error rates (cf. last line of Table 4) compared to the definition of an arbitrary small number of equal subintervals (4 or 8) or to the use of the commonly used heuristic HEUR1. This improvement is mainly due to the clear superiority of HEUR2 over the other approaches in the case of asymmetric data (log–normal distribution). In addition, using HEUR2 leads to a significant increase in the stability of the model, despite the fact that throughout the simulation the heuristic HEUR2 led to the definition of the larger number of subintervals compared to the other approaches. These results support the finding that this new approach proposed in this paper constitutes a useful, and yet efficient tool to confront an issue that is of major importance for the development of an additive utility sorting model through the preference disaggregation framework.

Another issue that is of major importance during model development involves the number of alternatives used in the reference set. The overall results presented in Table 4, show that larger reference sets are negatively related both to the error rate of the developed model as well as to its stability. In that regard, the use of a larger reference set (in terms of the number of alternatives) leads to lower error rate, but at the same time the instability of the model is increased. It is also worth noting that the decrease in the error rate is more significant when the HEUR2 procedure is employed. A further investigation taking also into consideration the number of groups ($F_2 \times F_3$ interaction; cf. Table 6) shows that using larger reference sets is most beneficial (in terms of the error rate) in the two–groups cases, while the improvement of the error rate is rather marginal in the three–groups case. On the other hand, the increase in the instability of the developed model is higher in the three–group case compare to the two–group one. Therefore, during model development the analyst should consider the trade–off between lower error rate and higher model instability when larger reference sets are used. The performance of a careful selection of any new alternative included in an existing reference set can eliminate this problem. In particular, if none of the existing alternatives dominates the new one, then the consideration of the new alternative adds new information to the reference set. This is translated into a new non–redundant constraint to the linear program (3)–(10), thus reducing its solution space. Asymptotically (for large reference sets) such an approach will eliminate the instability problem during model development.

Table 6. The impact of the reference set size (number of alternatives) on the error rate and the stability of the sorting models.

F_3	Two groups		Three groups	
	Error rate	Stability	Error rate	Stability
20%	21.26%	0.0569	31.61%	0.0057
30%	21.11%	0.0651	32.15%	0.0069
40%	19.45%	0.0728	31.37%	0.0061
50%	19.14%	0.0690	31.54%	0.0067
60%	19.05%	0.0745	30.34%	0.0068
70%	16.92%	0.0759	30.40%	0.0085

The impact of the reference set size to the stability of the sorting model can also be realized considering its interaction with the statistical distribution of the data and the number of groups ($F_1 \times F_2 \times F_3$). The results of Table 7 show that an increased reference set size leads to increased instability in the following two cases: (1) when the data are symmetric (multivariate normal) and two groups are considered, and (2) when the data are asymmetric (log–normal distribution) and three groups are considered. In the remaining two cases (normal with three groups, or log–normal with two groups) an increase in the training sample size leads to marginal changes in the stability of the developed sorting model.

Table 7. The impact of the interaction between the reference set size (number of alternatives), the statistical distribution and the number of groups on the stability of the sorting models.

F_3	Normal distribution		Log–normal distribution	
	Two groups	Three groups	Two groups	Three groups
20%	0.0408	0.0041	0.0729	0.0073
30%	0.0590	0.0031	0.0713	0.0107
40%	0.0698	0.0039	0.0757	0.0082
50%	0.0617	0.0045	0.0762	0.0088
60%	0.0725	0.0037	0.0765	0.0100
70%	0.0715	0.0034	0.0803	0.0137

The final two issues that are of interest with regard to the stability of the developed sorting models include the values of the technical parameters δ (F_4) and z (F_5). Both these parameters do not have a significant impact on the error rates of the developed models, but their impact is quite significant in terms of model stability (cf. Table 3). The results of Table 8 provide some insight on the way that these parameters should be defined. In general, increasing the right hand side of constraints (4)–(6), that is increasing the parameter δ, reduces the instability of the model. This is to be expected: an increase in δ (F_4) reduces the solution space of the linear program (3)–(10), thus leading to more stable solutions. The stability improvement that can be achieved by increasing δ is most significant in the two–groups case, whereas in the three–groups case there is no noticeable change. In contrast to the parameter δ, an increase in the z parameter increases the model's instability. This is also no surprise: z defines the range of the solution space explored during the post–optimality analysis. The larger the z, the wider is this range and vice versa. Generally, the results show that the value of this parameter can be more crucial for the stability of the model, since both in two–groups and three–groups sorting problems the increase of the instability caused by the selection of higher values for z is quite significant.

Table 8. The impact of the δ and z parameters on the stability of the sorting models.

Factor	Levels	Overall	Two groups	Three groups
	0.005	0.0538	0.1007	0.0068
F_4 (δ)	0.015	0.0361	0.0654	0.0067
	0.030	0.0238	0.0409	0.0068
	0.5%	0.0275	0.0534	0.0015
F_5 (z)	2%	0.0335	0.0632	0.0038
	5%	0.0410	0.0742	0.0078
	10%	0.0497	0.0853	0.0140

5 Concluding remarks and future perspective

The objective of this paper was to shed light on some critical issues to be considered during sorting model developed through preference disaggregation analysis. For this purpose the UTADIS method was considered. An extensive experimental investigation was conducted to examine the impact of the major parameters of the method on the performance and stability of the developed sorting models. The major finding of the analysis involves the significant effect of the way that the piece–wise linear marginal utility functions are formulated. This effect involves both the error rates of the developed models as well as their stability. The new heuristic approach proposed in this paper to address this issue (HEUR2), showed high efficiency in improving the stability of the models as well as their sorting performance, mainly when the data are derived from a non–symmetrical distribution. This is a quite encouraging result towards the use of this approach to specify the piece–wise linear form of the marginal utility functions of the criteria in real–world sorting problems. Along with this significant result, additional useful conclusions where drawn upon the effect of other technical parameters of the method as well as upon the effect of the reference set size, which is an issue of major importance to every regression–based sorting methodology.

Of course, this kind of analysis is not restricted to the specific MCDA method used in this study. Recently, new techniques based on preference disaggregation analysis, have been proposed to infer the parameters of other sorting methods based on utility function models similar to the UTADIS method (e.g., the M.H.DIS method [19]) or on different MCDA approaches (e.g., the ELECTRE TRI method [3, 11, 13]). Therefore, it would be interesting to perform a thorough investigation of the parameters involved in the model development process in these techniques and their impact on the classification performance and the stability of the sorting models that are developed. The extension of such an analysis considering a wide range of real-world data sets would also contribute to the enrichment and real-world validation of the findings of the present experimental study.

References

1. Bajgier, S.M. and Hill, A.V. (1982), "A comparison of statistical and linear programming approaches to the discriminant problem", *Decision Sciences*, 13, 604–618.
2. Belacel, N. (2000), "Multicriteria assignment method PROAFTN: Methodology and medical applications", *European Journal of Operational Research*, 125, 175–183.
3. Dias, L., Mousseau, V., Figueira, J. and Climaco, J. (2000), "An aggregation/disaggregation approach to obtain robust conclusions with ELECTRE TRI", *Cahier du LAMSADE*, No 174, Université de Paris–Dauphine.
4. Doumpos, M. and Zopounidis, C. (1998), "The use of the preference disaggregation analysis in the assessment of financial risks", *Fuzzy Economic Review*, 3/1, 39–57.
5. Horsky, D. and Rao, M.R. (1984), "Estimation of attribute weights from preference comparisons", *Management Science*, 30/7, 801–822.
6. Jacquet–Lagrèze, E. and Siskos, Y. (1982), "Assessing a set of additive utility functions for multicriteria decision making: The UTA method", *European Journal of Operational Research*, 10, 151–164.
7. Jacquet–Lagrèze, E. and Siskos, J. (1983), *Méthodes de Décision Multicritère*, Editions Hommes et Techniques, Paris.
8. Jacquet–Lagrèze, E. and Siskos, J. (2001), "Preference disaggregation: Twenty years of MCDA experience", *European Journal of Operational Research*, 130/2, 233–245.
9. Joachimsthaler, E.A. and Stam, A. (1988), "Four approaches to the classification problem in discriminant analysis: An experimental study", *Decision Sciences*, 19, 322–333.
10. Massaglia, M. and Ostanello, A. (1991), "N–TOMIC: A decision support for multicriteria segmentation problems", in: P. Korhonen (ed.), *International Workshop on Multicriteria Decision Support*, Lecture Notes in Economics and Mathematics Systems 356, Springer–Verlag, Berlin, 167–174.
11. Mousseau, V. and Slowinski, R. (1998), "Inferring an ELECTRE–TRI model from assignment examples", *Journal of Global Optimization*, 12/2, 157–174.

12. Mousseau, V., Slowinski, R. and Zielniewicz, P. (2000), "A user–oriented implementation of the ELECTRE–TRI method integrating preference elicitation support", *Computers and Operations Research*, 27, 757–777.
13. Mousseau, V., Figueira, J. and Naux, J.–Ph. (2001), "Using assignment examples to infer weights for ELECTRE TRI method: Some experimental results", *European Journal of Operational Research*, 130/2, 263–275.
14. Perny, P. (1998), "Multicriteria filtering methods based on concordance and non–discordance principles", *Annals of Operations Research*, 80, 137–165.
15. Roy, B. and Moscarola, J. (1977), "Procédure automatique d'examem de dossiers fondée sur une segmentation trichotomique en présence de critères multiples", *RAIRO Recherche Opérationnele*, 11/2, 145–173.
16. Yu, W. (1992), "ELECTRE TRI: Aspects methodologiques et manuel d'utilisation", *Document du Lamsade No 74,* Universite de Paris–Dauphine, 1992.
17. Vale, D.C. and Maurelli, V.A. (1983), "Simulating multivariate nonnormal distributions", *Psychometrika*, 48/3, 465–471.
18. Zopounidis, C. and Doumpos, M. (1999), "A multicriteria decision aid methodology for sorting decision problems: The case of financial distress", *Computational Economics*, 14/3, 197–218.
19. Zopounidis, C. and Doumpos, M. (2000), "Building additive utilities for multi–group hierarchical discrimination: The M.H.DIS method", *Optimization Methods and Software*, 14/3, 219–240.

This page is intentionally left blank

HIERARCHICAL CLASSIFICATION IN INSURANCE DATA ANALYSIS

NIKOS KOUTSOUPIAS

Department of Balkan Studies, Aristotle University of Thessaloniki,
P.O.Box 21, 53100 Florina, Greece

E-mail: nickk@auth.gr

IOANNIS PAPADIMITRIOU

Department. of Applied Informatics, Univ. of Macedonia, Thessaloniki, Greece.
Egnatia 156 P.O.Box 1591, 54006 Thessaloniki

E-mail: iannis@uom.gr

The private life insurance industry in Greece, as well as in the European region, is rapidly changing since it is challenged with new immigration, demographic and occupational circumstances. This paper is an attempt to demonstrate the use of Hierarchical Classification, a Statistical Multidimensional Data Analysis clustering method, implemented in a graphical environment using software built by the authors. The analysis of 1234 contracts based on 145 variables lead us into a collection of distinct, homogenous contract groups. We believe the method allows to efficiently obtain information about the portfolio picture of the life insurance company with no a piori assumptions and limitations.

1 Introduction

During the last few decades, the interest on private life insurance research is increasing and is focused mainly on financial, risk, management and actuarial aspects. Information published by national Census Organizations in most developed countries concentrate on quantitative analyses of their respective life insurance markets. In Greece, despite the fact that the life insurance sector is an important part of national industry, there is notable absence of research literature on qualitative study about life insurance.

In their work, DeVaney and Keaton [10], examine how the behavior of the insured changes with respect to family composition, using regression and classification trees. Mitchel [17] explores market trends of different generations based on a small number of insurance parameters and according to their wants and needs. Sabot [22] utilizes models based on data analysis to classify US insurance companies and Huttin [11] analyzes the relation between income and health services selection through single or group insurance programs. In general the insurance sector researcher will not find any unified approach to the life insurance, since all attempts focus partially on specific aspects of insurance markets. In the application section of this paper we describe the data set along with the parameters used for our research.

The utilization of computer power in all of the insurance administrative functions is indisputable. Mahaffie [16] writes about further changes on insurance services as a result of the intense technological progress and Zultowski [23] questions the need of insurance agent presence, since technology tends to substitute most of their responsibilities. According to Chookaszian [9] the changes taking place in the insurance industry due to the accelerated technology growth, during the next decade will be ten times more important than other market sectors.

2 Research Objective

Based on the above findings, along with Adler's [1] statement that the axiom "information is power" tends to become "information overload equals paralysis", we wish to propose that the concept of private life insurance should be treated as a whole, with no a priori assumptions or limitations.

We suggest the processing and management of life insurance contract portfolio as a unified phenomenon, in order to reveal the trends, formations and groupings that emerge directly from the available data, taking into account all life contract parameters. For this reason, we used data from the life portfolio for the year 1995 of an insurance company based in Northern Greece and applied the statistical data analysis method of Ascending Hierarchical Classification.

3 The Ascending Hierarchical Classification (AHC) method

A historical perspective of the conception and progress of data analysis methods was written by J.-P. Benzécri [6], a leading statistician and originator of AHC. Professor J.-P. Benzécri's [2, 4] work states the mathematical foundations of AHC - along with other methods- and accomplishes to show off their importance and advantages. Comprehensive descriptions of AHC are included in the works of J.-P. Benzécri [5,7], J.-P. Benzécri and Maiti [8] and F. Benzécri [3].

Data analysis with AHC as part of Descriptive Statistics [21] works directly on the available data set and explores the existence of motives, relations or trends that they may bear, emphasizes their graphical representation and reveals all possible simple, but functional descriptions of data characteristics. Moreover, it is possible to investigate indications of exceptional conditions and discover the existence of hidden subgroups and formations.

The basic characteristic of AHC is the absence of any a priori technical assumption and according to Lagarde [15], it is the researcher's responsibility to interpret the method's outcome and uncover the consequences of that outcome.

The final outcome of AHC is a dendrogram that shows how data groups are related to each other. Depending on the phenomenon analyzed, one can extract new, disjoint data groupings by cutting off the classification tree at appropriate levels. The methods works best when homogenized (0-1 valued) data is used, since this produces better group distance calculations. The dendrogram construction algorithm for the classification of n objects can be described as a procedure of four steps:

1. At first, the algorithm considers each object to be a distinct group, that is, there exist N groups with 1 object each. Then, for each object pair, the algorithm calculates the loss of extraclass inertia (I_E, see table 1) using the distance χ^2 (below) between them after their junction.

2. The objects with the smallest loss of extraclass inertia between them are agglomerated into one group, reducing the total number of groups to N-1.

3. Again, for each two of the remaining groups, the algorithm recalculates the loss of extraclass inertia between them and

4. Steps 2 and 3 are repeated and the algorithm terminates when only one group with N objects is formed.

The distance metric used is call χ^2 (chi-squared) [4] and is calculated as follows :

$$d^2(i,i') = \sum_{j=1}^{p} \frac{1}{P_{.j}} \left(P_j^{i} - P_j^{i'}\right)^2$$

where: $P_{.j} = \frac{P_j}{k}$ with $P_j = \Sigma_i T(i,j)$

and $P_j^{i} = \frac{T(i,j)}{P_i}$ with

$P_i = \Sigma_j T(i,j)$

$k = \sum_{i=1}^{n} \sum_{j=1}^{p} T(i,j)$ and

T(i,j) the value in line (or contract) i and column (or parameter) j,

Among others, examples of AHC application on data related to Greece can be found in the works of Papadimitriou [18] and Papadimitriou et. al. [19, 20].

4 Applying AHC on insurance data

We used a recent (graphical environment) software implementation of AHC [12,13,14] in order to apply the method on insurance data.

We classified a set of logical (0-1) data tables with 145 parameters and 1.234 objects that correspond to the life insurance contracts signed during the year 1995. These parameters fall into five major sets.

The first, relates to general insurance aspects. Specifically, this set contains the contract date, the insurance type, the current contract status and the company administrative region where it was signed.

The second set of parameters has to do with insured or covenanter (owner) "demographics". That is, insured occupation, place of residence, relationship with covenanter, covenanter age, gender and age of insured.

Another set of data parameters includes insurance premiums and insured amounts. There are ten parameters here: insured period, amount of death benefit, survivorship amount, life coverage amount, additional premium due to heath reasons, number of premium installments, additional life premiums due to occupation and due health condition and finally the amount of down payment.

The next set incorporates two subsets of four parameters each. The first subset consists of types of family relationships about the first four beneficiaries of the contract. The other subset has benefit percentage rates for each one of the four beneficiaries.

Last, but not least, comes the set of raider parameters. This set is comprised of three subsets of parameters: five accident type raiders, seven illness type raiders and three raiders additional to life coverage.

Despite such parameter groupings, the analyzed phenomenon is examined as a whole, with no assumptions or limitations on this grouping.

The total number of parameters sums up to 43. After categorization (each parameter producing 2 – 6 classes) the total number of parameters in the final data set turns out to be 145, all containing values 0 or 1 as follows:

ContractDate			InsuranceType					ContractStatus					AdminRegion			InsuredOccupation				...
1/1-15/6	16/6-15/9	16/9-31/12	A	B	C	D	E	1	2	3	4	5	1	2	3	1	2	3	4	...
1	0	0	1	0	0	0	0	1	0	0	0	0	1	0	0	1	0	0	0	...
1	0	0	0	1	0	0	0	0	0	1	0	0	0	0	1	1	0	0	0	...
1	0	0	0	1	0	0	0	0	0	0	1	0	0	1	0	0	1	0	0	...
1	0	0	0	0	1	0	0	0	0	1	0	0	1	0	0	0	0	0	1	...
1	0	0	1	0	0	0	0	1	0	0	0	0	1	0	0	1	0	0	0	...
1	0	0	0	1	0	0	0	0	0	1	0	0	0	1	0	0	0	1	0	...
1	0	0	0	0	0	1	0	0	0	0	0	1	1	0	0	1	0	0	0	...
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	

Figure 1 : Sample of the Input Data File

The resulting tree had 2467 nodes (2*1.234-1), but for the sake of practicality and because of space limitations we only worked on the upper 30 of them. In fact, when applying AHC on large data sets the most interesting information output usually emerges from the top 10 to 20 nodes of the resulting classification dendrogram.

The first table of AHC results, as produced by the software package mentioned above, had the following contents:

Table 1: Classification Results (I).

Node	A	B	Mass	Dist(D)	Inertia(I)	Inertia(E)	λ(r)
2438	2389	2401	0,026742	0,016732	1,234875	0,951172	0,435111
2439	2384	2434	0,042139	0,01774	1,252615	0,933432	0,426995
2440	2416	2360	0,030794	0,017984	1,270599	0,915448	0,418769
2441	2439	2348	0,045381	0,019227	1,289825	0,896221	0,409974
2442	2431	2405	0,226094	0,019365	1,309191	0,876856	0,401115
2443	2295	2337	0,007293	0,020169	1,32936	0,856687	0,391889
2444	2382	2440	0,04295	0,020555	1,349915	0,836132	0,382486
2445	2441	2438	0,072123	0,023506	1,37342	0,812626	0,371733
2446	2	2437	0,272285	0,023923	1,397344	0,788703	0,36079
2447	2444	1155	0,04376	0,02424	1,421584	0,764463	0,349701
2448	2436	2338	0,143436	0,024643	1,446227	0,73982	0,338428
2449	2373	2445	0,080227	0,024688	1,470914	0,715132	0,327135
2450	2442	2258	0,230146	0,025094	1,496008	0,690038	0,315656
2451	2449	2415	0,085089	0,025324	1,521333	0,664714	0,304071
2452	2426	411	0,051053	0,025659	1,546992	0,639055	0,292334
2453	2403	2343	0,012966	0,025747	1,572738	0,613308	0,280556
2454	2447	2356	0,047812	0,026164	1,598902	0,587144	0,268587
2455	2454	2080	0,049433	0,026599	1,625502	0,560545	0,256419
2456	2451	2421	0,103728	0,026913	1,652415	0,533631	0,244108
2457	2448	2443	0,150729	0,026943	1,679358	0,506688	0,231783
2458	2435	2456	0,166937	0,028927	1,708285	0,477761	0,21855
2459	2457	2452	0,201783	0,029563	1,737849	0,448198	0,205027
2460	2453	2409	0,017018	0,029575	1,767424	0,418623	0,191498
2461	2455	2433	0,1094	0,032923	1,800347	0,385699	0,176437
2462	2458	2446	0,439222	0,036037	1,836384	0,349663	0,159952
2463	2462	2461	0,548622	0,043872	1,880256	0,30579	0,139883
2464	2463	2418	0,551053	0,050399	1,930655	0,255392	0,116828
2465	2464	2460	0,568071	0,057624	1,988279	0,197768	0,090468
2466	2465	2459	0,769854	0,060063	2,048342	0,137704	0,062992
2467	2466	2450	1	0,137704	2,186047	0	0

For the purposes of this paper, we will only work on the first three columns of table 1 above. The first column refers to the node name (index or dipole) and the second and third columns contain the names of the resulting successor nodes of the corresponding dipole named in column 1. The rest of the information in the above table is if informative nature. That is, in column 4 the table contains the mass of each node (or object) given by:

$f_{i.}=k_{i.}/k$ (with k_i the sum of the i^{th} object),

in column 5 the above table contains the distance between the two sibling (A,B) nodes calculated as follows:

$$d^2(f_j^i, f_j^{i'}) = \sum_{j=1}^{p} \frac{1}{f_{.j}} \left(f_j^{\,i} - f_j^{\,i'}\right)^2$$

incorporating relative frequency values of the initial table.

The internal (intraclass) and external (extraclass) inertias of each node are shown in the two subsequent columns, 6 and 7. The internal inertia I_I is given by:

$$I_I = \sum_r \sum_i m_i d^2(i, g_i)$$

where m_i is the mass of each object i in group Kr.
The external inertia I_E is calculated as:

$$I_E = \Sigma_r m_i d^2(g_i, G)$$

with m_i the total mass of K_i class and g_i the mass center of class i.
Finally, column 8 shows λ(r), which gives the percentage of inertia interpretation of each node and is calculated as:

$$\lambda(r) = I_E/I_{tot}$$

where I_{tot} the total inertia produced during the partitioning of a class into two subsequent classes (going upwards the classification dendrogram). The total inertia is given by:

$$I_{tot}=I_I+I_E$$

In the upper 30 nodes we see that there are two dipoles connecting objects to nodes. These nodes are dipole 2446, that agglomerates object 2 to node 2437 and dipole 2452 that groups object 411 and node 2426. This means that at these points the closest (with respect to loss of extraclass inertia) groups contain a group of contracts on the one side (i.e. nodes 2437 and 2426) and a single contract on the other (2nd and 411th contract respectively).

The following table contains all object groups that correspond to each of the top 5 dendrogram nodes:

Table 2 : Classification Results (II).

Node	A	B	Mass	Objects
2467	2466	2450	1	A1 A2 A3 A4 A5 A6 A7 A8 A9 A10 A11 A12 A13 A14 A15 A16 A17 A18 A19 A20 A21 A22 A23 A24 A25 A26 A27 A28 A29 A30 A31 A32 A33 A34 A35 A36 A37 A38 A39 A40 A41 A42 A43 A44 A45 A46 A47 A48 A49 A50 A51 A52 A53 A54 A55 A56 A57 A58 A59 A60 A61 A62 A63 A64 A65 A66 A67 A68 A69 A70 A71 A72 A73 A74 A75 A76 A77 A78 A79 A80 A81 A82 A83 A84 A85 A86 A87 A88 A89 A90 A91 A92 A93 A94 A95 A96 A97 A98 A99 A100 A101 .
2466	2465	2459	0,769854	A1 A2 A3 A5 A6 A7 A8 A9 A10 A14 A16 A17 A18 A19 A20 A21 A22 A23 A24 A25 A26 A27 A28 A30 A31 A32 A33 A34 A36 A37 A38 A39 A40 A41 A42 A46 A47 A48 A49 A50 A51 A52 A53 A54 A55 A57 A58 A59 A61 A62 A63 A65 A66 A68 A69 A70 A71 A73 A74 A75 A76 A79 A80 A81 A82 A83 A84 A85 A86 A87 A88 A89 A90 A91 A93 A94 A96 A97 A99 A102 A103 A104 A105 A106 A107 A109 ...
2465	2464	2460	0,568071	A1 A2 A3 A5 A6 A7 A9 A10 A16 A21 A22 A23 A24 A25 A26 A27 A28 A30 A31 A33 A34 A36 A37 A38 A39 A40 A41 A42 A46 A47 A51 A52 A53 A54 A55 A57 A58 A59 A61 A62 A63 A68 A69 A75 A76 A79 A81 A83 A84 A85 A86 A90 A93 A96 A99 A102 A104 A105 A106 A107 A109 A111 A113 A114 A115 A116 A119 A120 A121 A124 A125 A129 A132 A133 A134 A135 A139 A142 A143 A144 ...
2464	2463	2418	0,551053	A1 A2 A3 A5 A6 A7 A9 A10 A16 A21 A22 A23 A24 A25 A26 A27 A28 A30 A31 A33 A34 A36 A37 A38 A39 A40 A41 A42 A46 A47 A51 A52 A53 A54 A55 A57 A58 A59 A61 A62 A63 A68 A69 A75 A76 A79 A81 A83 A84 A86 A90 A93 A96 A99 A102 A104 A105 A106 A107 A109 A111 A113 A114 A115 A116 A119 A120 A121 A124 A125 A129 A132 A133 A134 A135 A139 A142 A143 A144 A146 A147 A148 A154 A155 A157 A158 A160 A162 A163 A165 A166 A167 A168 A169 A170 A171 A172 A173 A175 A176 A178 A179 A180 A183 A186 A188 A192 A196 A197 A200 A201 A202 ...
2463	2462	2461	0,548622	A1 A2 A3 A5 A6 A7 A9 A10 A16 A21 A22 A23 A24 A25 A26 A27 A28 A30 A31 A33 A34 A36 A37 A38 A39 A40 A41 A42 A46 A47 A51 A52 A53 A54 A55 A57 A58 A59 A61 A62 A63 A68 A69 A75 A76 A79 A81 A83 A84 A86 A90 A93 A96 A99 A102 A104 A105 A106 A107 A109 A111 A113 A114 A115 A116 A119 A120 A121 A124 A125 A129 A132 A133 A134 A135 A139 A142 A143 A144 A146 A147 A148 A154 A155 A157 A158 A160 A162 A163 A165 A166 A167 A168 A169 A170 A171 A172 A173 A175 A176 A178 ...

As one can see, due to the extensive data size it is not possible to display the total group composition, although it is in fact stored if full length in the actual results data file.

The resulting dendrogram has the shape shown in figure 2. A closer examination of the tree reveals that the initial node (2467) breaks down into nodes 2466 and 2450. Likewise, node (2466) is composed of nodes 2465 and 2459 and subsequently, node 2450 into node 2258 and node 2442 and so on.

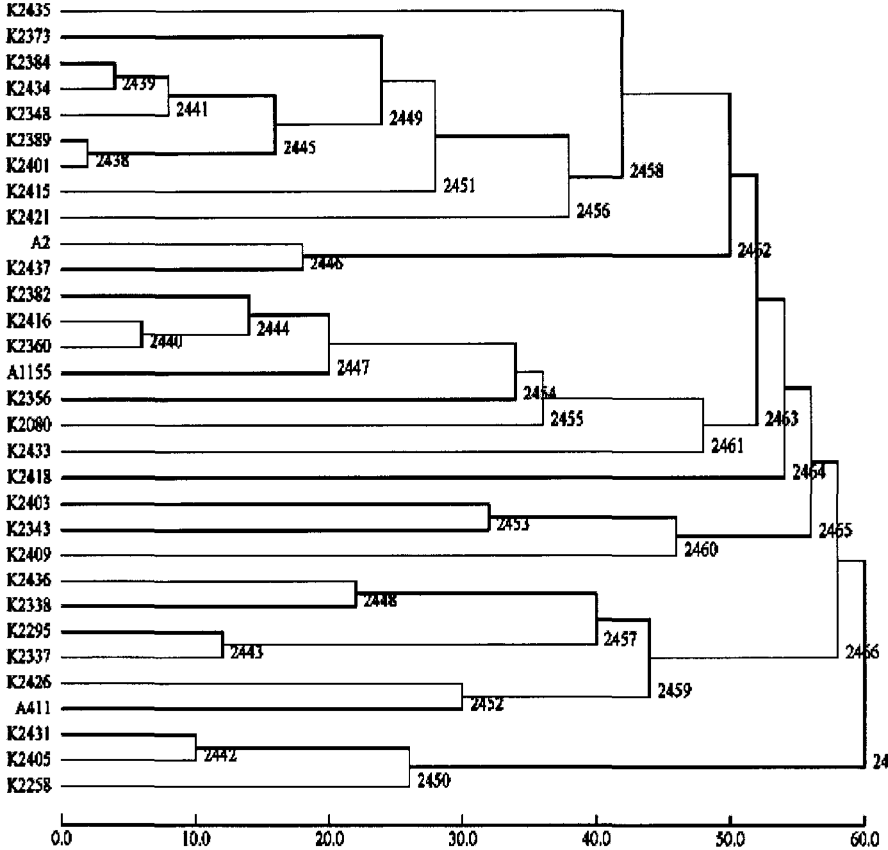


Figure 2 : The classification dendrogram.

At this point, we would like to note that the terminal objects (on the left side of the dendrogram) are denoted either with a "K" prefix, which corresponds to a node , or an "A" prefix, which refers to an object.

In order to determine which of the parameters contribute most in the distinction, or characterization of each node, it is important to examine each node's interpretation tables of characterization, as shown in table 3 below for node 2467 (after setting a selection criterion, to select the most important characterization parameters). The COR column in Table 3 refers to the correlation of the average contract vector of each node with respect to each one of the parameter classes examined.

This table asserts that node 2467 can be seen as a vector in the R^{145} vector space and that it is equivalent to the average life insurance contract.

Table 3 : The interpretation of node 2.467 based on a selection criterion (COR >= 30).

Node to be partitioned				1st Successor Node				2 nd Successor Node			
	CI	Cor	Ctr		CI	Cor	Ctr		CI	Cor	Ctr
2467				2466				2450			
M1	1	0	0	M28	20	33	133	M141	22	84	531
M2	4	0	0	M32	21	32	159	M30	23	79	508
M3	6	0	0	M140	21	32	159	M4	18	72	429
M4	5	0	0	M117	18	31	109	M34	15	58	333
M5	4	0	0	*M141*	*1*	*84*	*159*	M39	23	58	411
M6	2	0	0	*M30*	*2*	*79*	*152*	M116	23	50	365
M7	0	0	0	*M4*	*1*	*72*	*128*				
M8	9	0	0	*M34*	*1*	*58*	*99*				
M9	1	0	0	*M39*	*3*	*58*	*122*				
M10	0	0	0	*M116*	*4*	*50*	*109*				
M11	1	0	0								
M12	3	0	0								
M13	1	0	0								
M14	0	0	0								
M15	6	0	0								
M16	5	0	0								
M17	1	0	0								
M18	9	0	0								
M19	1	0	0								
M20	4	0	0								
:	:	:	:								
:	:	:	:								

The presence or absence of a parameter characteristic can be more obvious from the underlined (denoting presence) or italicized (indicating absence) letters in the parameter columns (columns 5 and 9).

The interpretation of the first successor node (2466) leads us to the conclusion that it is characterized (by 3.3%) from the presence of contracts where policyholder and insurer are the same person (parameter M28), include whole life programs (M32, by 3.2%) and the Premium Payment Exemption raider (M140, M117 by 3.2). On the other hand, in the group of node 2466 there is absence of contracts with Premium Payment Exemption raider (M141 and M116 by 8.4% and 5%) covering children (M4, M30 by 7.2% and 7.9% accordingly).

It is evident that node 2450 is characterized by the presence of parameters that where absent in node 2466 and by the absence of parameters that where present in node 2466.

For the successor (to 2466) node 2465 the resulting interpretation tables (not shown here) reveal that this node is characterized by the presence of Pension contracts (4.5%), and long insured periods (3.1%). On the other hand, in node 2465 there is absence of Term insurance contracts (M30, M4 by 4.7% and 3.9%).

In the same manner, we see that node 2459 (the second successor to 2466) is characterized by the presence of Whole Life insurance contracts (M5, M44 by 13%), Insured Amounts of less than 1M drachmas (M52, by 7.6%) and low Premiums (M64, 7.3%) that include the Major Medical raider (M139 by 4.2%).

As for node 2450, the successors are nodes 2442 and 2258. Node 2442 is characterized by the presence of the Premium Payment Exemption raider (M141, 8.4%), by contracts where Parents are insuring Children (M30 by 8%) with Term insurance programs (M4, 7.2%). On the other hand, successor node 2258 is only characterized by contracts with low Term Insurance Premiums (M56 and a frequency of 85.3%).

5 Conclusion

As a general remark on the results of the application of AHC on insurance data we can see that the method revealed two primary groups of contracts.

The first group refers to Term insurance contracts held by young parents covering their children.

The second group, on the contrary, includes life insurance contacts covering persons with dangerous occupations that with to combine their basic social security coverage with Pension or Whole Life private insurance programs, accompanied with a number of illness or injury raiders.

For an application of the AHC method (along with an application of Correspondence Analysis) on Northern Greece insurance data for the years 1992-1994 see N. Koutsoupias' thesis[13].

References

1. Adler, R. M., Personal Observations on Changes in Market Research, *Market Research Review*, **7** (1993) pp. 67-70.
2. Benzécri, J.-P., Analyse des Données, Tome 1: La Taxonomie, Tome 2: Correspondances (Dunod, Paris, 1973).
3. Benzécri, F., Introduction á la Classification Automatique d' arprés un exemple de données médicales, *Les Cahiers de l' Analyse des Données*, **5/3** (1980) pp. 311-340.
4. Benzécri, J.-P., Pratique ed l' Analyse des Données, Tome 1: Analyse des Correspondances, exposé élémentaire, Tome 2: Abrégé théorique (Dunod, Paris, 1980).
5. Benzécri, J.-P., Pratique ed l' Analyse des Données, Tome 3: Linquistique et Lexicologie (Dunod, Paris, 1981).
6. Benzécri, J.-P., Historie et Prehistoire ed l' Analyse des Données, (Dunod, Paris, 1982).
7. Benzécri, J.-P., et coll., Pratique ed l' Analyse des Données en économie (Dunod, Paris, 1986).
8. Benzécri, J.-P. and Maiti, G. D., Pratique ed l' Analyse des Données en médecine (Statmatic, Paris 1992).
9. Chookaszian, D. H., Keynote Address, The Insurance Data Network Forum, Chicago, Illinois, 9/10, (1997).
10. DeVaney, S. A. and Keaton, E. J., Determining Purchasers of Whole Life Insurance Using a Classification Tree, *Research Review of the Society of Insurance Research*, **3** (1994) pp. 33-45.
11. Huttin, C., Income distribution and consumer demand for health services. The case of prescribed medicines in the USA, *Applied Economics*, **29** (1997) pp. 497-503.
12. Koutsoupias, N. and Papadimitriou I., Assisting the Interpretation of the Results of Data Analysis Methods, *Proceedings of the 12th National Statistical Conference, E.S.I.*, Spetses (1999) pp. 318-326.
13. Koutsoupias, N., Statistical Analysis of the Northern Greek Life Insurance Market using Relational Databases (RDBMs) and Structure Query Language (SQL), Phd. thesis (dept. of Applied Informatics, University of Macedonia, Thessaloniki, 1999).
14. Koutsoupias, N., Data Analysis Software: Three Dimensional Contingency Tables, Correspondence Analysis, Ascending Hierachical Classification. User's Guide (Paratiritis Publications, Thessaloniki, 1999).
15. Lagarde, J., Initiation a l' Analyse des Donees (Dunod, Paris 1983).
16. Mahaffie, J. B., Forces of Change: Doing Business in the Next Millennium, *Market Research Review of the Society of Insurance Research*, **3** (1996) pp. 1-13.

17. Mitchel, J. O., Wants and Needs: Are They the Same from One Generation to the Next?, *LIMRA Market Facts*, **15** (1996) pp.1.
18. Papadimitriou, I., Applications of Data Analysis Methods in Economy, (A.V.S.TH, Thessaloniki, 1988) .
19. Papadimitriou, I., Karakos, A. and Koutroumanidis, G., The application of Data Analysis on land-planning of Greek perfectures according occupation (Scientific Yearbook, University of Macedonia, Thessaloniki, 1993).
20. Papadimitriou, I., Karakos, A. and Koutsoumanidis, G., Study of the Regional Structure of Occupation in the Industry and Services during the period 1969-1988, 6th International Symposium on Applied Stochastic Models and Data Analysis, Chania, Crete (1993).
21. Papadimitriou, I., Data Analysis Methods, University Lectures, (University of Macedonia Publishing, Thessaloniki, 1994).
22. Sabot, G., Modeling of 1994 A.M. Best Life/Health Insurer Ratings, *ARIA Research*, **4** (1996) p. 1.
23. Zultowski, W. H., Technology Will Not Replace the Agent (But Is the Agent of the Future Human?), *LIMRA Market Facts*, Life Insurance Marketing Research Association, **4** (1996) p. 15.

This page is intentionally left blank

CONSUMER PREFERENCES FOR EXTRINSIC VERSUS INTRINSIC QUALITY CUES FOR IMAGE PRODUCTS: THE CASE OF GREEK QUALITY WINE

EFTHALIA DIMARA

Department of Economics, University of Patras
University Campus – Rio, P.O Box 1391,
Patras, 26500, Greece

E-mail: dimara@econ.upatras.gr

GEORGE BAOURAKIS AND NIKOS KALOGERAS

Mediterranean Agronomic Institute of Chania, Department of Economic, Marketing and Finance, PO. Box: 85, Chania, 73100, Greece

E-mail: baouraki@maich.gr, kalogeras@maich.gr

The extrinsic quality cues of certification and packaging (bottling) and the intrinsic quality cues of aroma and taste are assumed to influence consumer preferences and the decision to purchase the product. Prior research suggests that, for most utilitarian products, intrinsic cues are more important than extrinsic cues. However, evidence suggests that this hypothesis might not be valid for image products. Quality wine, which is an image product, to a great extent, is used for examining consumer preferences towards these two sets of cues. A sample of 744 Greek wine consumers is used to assess the factors influencing consumer attitudes towards these quality cues. An ordered probit model with sample selectivity reveals that these quality cues are valued by consumers and are perceived to be highly important. Furthermore, they possess different socioeconomic and demographic characteristics. Results support the hypothesis that the market is highly fragmented where the importance of extrinsic and intrinsic quality cues on purchasing behaviour is concerned. The type and source of information received by consumers, their place of origin, disposable income, education and marital status all exert an independent effect on attitude formation. The use of quality cues such as certification, bottling, aroma and taste may be potentially useful in creating niche markets and advancing rural localities.

1 Introduction

Quality is an issue that ultimately refers to and affects the consumer, as it influences the consumer on which product to buy. Quality judgments depend on the perceptions, needs and goals of the consumer and formulate what is called perceived quality. Perceived quality has been variously defined as 'fitness for use, given the needs of the consumer' or 'the degree to which a product fulfills its functions, given the needs of the consumer' [1]. Monroe and Krishnan [2] have proposed that 'perceived product quality is the perceived ability of a product to provide satisfaction relative to available alternatives' and thus they recognize that perceived quality is a comparative concept in the sense that it is influenced by the perceived quality of other available products. Perceived quality is an overall evaluative judgment. As such, it is based on consumer's perception of the quality attributes, which the product possesses.

Literature on product quality makes a distinction between a product's quality attributes and cues. Quality cues are defined as 'informational stimuli that are, according to the consumer, related to the quality of the product, and can be ascertained by the consumer through the senses prior to consumption'. On the contrary, quality attributes are the functional and psychological benefits or consequences provided by the product and thus, they represent what the product is perceived to provide for the consumer. In that sense, quality attributes are unobservable prior to consumption. During purchase, a consumer will use quality cues in choosing among alternative products. Quality cues are categorized as either extrinsic or intrinsic [3]. Intrinsic cues are part of the physical product and cannot be changed without also changing the physical product itself. Intrinsic cues usually refer to features such as the color, aroma, etc. Extrinsic cues are related to the product but are not physically part of it. Examples of well-known extrinsic cues are the price, brand name, country or place of origin, packaging, etc.

The major hypothesis, supported by a wide range of research, and related to intrinsic and extrinsic cues is that for most products intrinsic cues are more important in the quality perception process than extrinsic cues [4]; [5]; [6]. However, for image products, such as carbonated bottled water, electronic lighters and others, extrinsic cues may be more important. Cues such as price, brand name and packaging may be more important in formulating the consumer's perceived quality than intrinsic cues such as aroma and color. For most utilitarian goods, firms have made extensive use of intrinsic cues and provided differentiated products by targeting segments of the market that highly value the respective cues. The purpose of this paper is to examine the factors that influence consumers to highly value extrinsic, as opposed to intrinsic characteristics, for quality wine, a very important Greek product possessing different images. The end point, of course, is to conclude whether firms can use extrinsic cue differentiation to target market segments and create highly specialized markets or 'niche markets'.

2 Factors Influencing Perceived Product Quality

In general, quality implies the satisfaction of consumer needs [7] and a consistent level of performance, taste, freshness, aroma and other properties of the product [8, 9]. Most post-harvest researchers, producers and handlers are product oriented in the sense that quality is described by specific attributes of the product [10]. From the consumer point of view, however, quality is a complex and contested notion, which is socially constructed and thus, highly variable among different socio-cultural contexts. From an economic perspective, quality, being a characteristic which is above minimum standards, assigns a product a cutting edge over its rivals and usually enables it to command higher prices in the market place [11]. In sociology, quality is a social construction dependent on the socio-cultural, political and economic contexts within which production-consumption relations exist. In marketing, quality itself implies a potential for powerful actors within the food-production-consumption chain to appropriate the term for their own products. In this way, quality may effectively confer property rights on primary (farmers) and intermediary (manufacturers) producers, wholesalers or retailers.

In the course of this study we have selected and examined three distinct quality cues, two of which are extrinsic and one which is intrinsic. Packaging is an extrinsic cue that has received attention in recent years, as far as imagery products (especially wine) are concerned. Packaging of precious wine has become very sophisticated and artistic in order to did recall. Wine packaging either depicts images of a lost and foregone countryside where values such as authenticity, wholesomeness, and tradition once prevailed or on the contrary, it illustrates the post-modern benefits of consuming wine. Certification is a multidimensional extrinsic cue, which does not refer explicitly to the product's feature but to a range of attributes that the product should posses before being granted certification. In the course of this study, certification refers to the quality marks awarded to agricultural and food products by the Commission of the EU according to the regulations introduced in 1992. These regulations define a broad set of four types of quality marks that aim to distinguish particular products that maintain special quality characteristics from similar products belonging to the same category. Currently, the EU recognizes four quality marks for Protected Designation of Origin (PDO), Protected Designation of Indication (PGI), Organic Products and Products of Specific Character. For wine, PDOs and PGIs are the usually awarded certifications. The awarding of such symbols embraces the notion of specification and association, i.e., the use of specific production methods and raw materials from a particular geographical area. Finally, the intrinsic cue of aroma is included as a typical cue as regards quality wine.

Economic, cultural, psychological and lifestyle factors as well as food trends are among the factors likely to influence consumer attitudes towards food attributes. The individual consumer's disposal income is a major determinant of the demand for food, and an important factor influencing price-quality effects among consumers and their attitudes towards certain food product properties [12, 13, 14]. Furthermore, income has been used as a major determinant of New Zealand consumers' perceptions of the taste of food products imported from different countries [15].

Information about the product has potentially important influences on perception of quality and preference [16]. Information refers to both the source and type of acquired knowledge. A wide range of sources of information on food are used by consumers including TV/Radio, magazines, newspapers, etc. [17]. The accuracy and source of information important because individuals process information according to its perceived cause and consider information provided by the factual performance of the entity in question more reliable than information provided by other factors

[18]. In general, personal or neutral information are more reliable and more influential than non-personal and market oriented information [19]. Information of a neutral nature, such as that provided by the news media or other objective sources, is more effective because it is considered more credible than advertising or other market oriented sources [20].

Finally, a number of social, demographic and cultural consumer characteristics may influence the formation of attitudes towards food choice and the demand for specialty products. Taking into account that certification, geographic association and traceability may act as risk reducing food choice strategies, socio-demographic characteristics associated with risky or riskless decisions may influence consumer's attitudes towards their importance. In many economic, psychological and social studies, the consumer's formal education, age, sex, marital status, occupation and place of origin, have been found to exercise an independent effect on attitude formation towards risk and food choice.

3 Data and Model

In order to examine consumer attitudes towards certification, geographic association and traceability, a survey of wine consumers was designed and executed, among others, in the framework of a European research project financed under the EU's FAIR program. Wine producers in Greece have taken advantage of recent EU regulations and shifted their production towards products with a PDO or a PGI mark. Furthermore, there is an increasing trend of producing brands with well designed and constructed packaging including hand-made artistic features such as wooden boxes, dried flowers, hand-made wrapping cloths, and even hand-made paintings.

Residents in the urban centers of Athens, Patras and Tripoli were selected as representative of consumers in very large (metropolitan), large and smaller towns. Data were collected by distributing questionnaires and conducting a face to face where trained personnel questioned the respondents. The questionnaire included both structured and semi-structured parts, in order to allow quantitative and qualitative analysis. A total of 750 questionnaires were collected in the three study regions. In order to achieve the highest possible coverage of the variance in purchasing behavior we decided to diversify questionnaire collection according to the place of purchase. Questionnaires were collected in regular places of wine purchase such as small and big supermarkets, specialist outlets such as cavas, and restaurants and tavernas.

Each surveyed consumer was offered to select a single category from a set of ordered responses (non-important, important, highly important) for each of the three product attributes, namely certification, packaging and aroma. Thus, the importance measurement better serves the aims of our survey than the 'like-dislike', 'good-bad', and 'favorable-unfavorable' measurements usually employed in marketing surveys. Each consumer is assumed to select a utility-maximizing position reflected by the ordered set of the j alternatives (J=0,1,2). Assume that the consumer's utility is represented by a well-behaved preference function, U^* as in the original model for polychotomous responses presented by [21]. With case subscripts suppressed, the maximum utility attained by selecting a choice j, U_j^*, is postulated to be linearly associated with exogenous variables:

$$U_j^* = \beta' \mathbf{x} + e_j \tag{1}$$

where $\mathbf{x}$ is a Kx1 vector of exogenous variables, β is a vector of unknown parameters to be estimated, and e_j is a random error assumed to be identically normally distributed with zero mean and unit variance. U_j^* is unobserved. What we do observe is:

$$\begin{aligned} &U_0 = 0 \text{ if } U_j^* \leq 0 \text{ (non - important)} \\ &U_1 = 1 \text{ if } 0 < U_j^* \leq \mu \text{ (important)} \\ &U_2 = 2 \text{ if } \mu \leq U_j^* \text{ (very important)} \end{aligned} \tag{2}$$

which is actually a form of censoring and the μ is an unknown parameters to be estimated with β.

In theory, a consumer of quality wine compares the three alternatives for each attribute and chooses the utility maximizing position. The probability that an alternative k is chosen is $P_k > P_j \ \forall \text{ j, j} \neq \text{k}$. Where:

$$P_k = \Pr\left[U_k > \max\left(U_0, U_1, U_2\right)\right] \tag{3}$$

For example, if we consider certification and the three alternative ranking choices, a consumer's choice that certification is non-important would imply that the utility derived by this choice is higher than the utility derived from the other two choices (important, highly important), respectively. The present survey was conducted among a randomly selected sample of consumers and thus, a model estimation based on data limited to consumers of quality wine may lead to a selectivity bias if the subsample is not random, and estimated model parameters would be inconsistent. The selection bias should be eliminated in order to assure randomness of responses. Selectivity bias may be corrected by adding a selection mechanism [22]:

$$
\begin{aligned}
&w^{*} = \alpha' \mathbf{z} + u \\
&w = 1 \text{ if } w^{*} > 0 \\
&w = 0 \text{ if } w^{*} \leq 0 \\
&\text{and} \\
&\text{Prob}(w = 1) = \Phi(a' \mathbf{z}) \\
&\text{Prob}(w = 0) = 1 - \Phi(a' \mathbf{z})
\end{aligned}
\tag{4}
$$

where w^{*} is the unobservable utility by the ith consumer from consuming a quality product instead of consuming the standard product, $\mathbf{z}$ is a set of exogenous variables, $\Phi(.)$ is the cumulative distribution function for a standard normal variable and $u \approx N(0,1)$. Let $w = 1$ if the consumer consumes quality wine and 0 otherwise. The model in equation (4) is a univariate probit model and U_k is observed if and only if $w = 1$. Equation (1) may be respecified as a sample selection problem [23]:

$$E\left[U \middle| \mathbf{x}, w^{*} > 0\right] = \beta' \mathbf{x} + E\left[e \middle| \mathbf{x}, w^{*} > 0\right] \tag{5}$$

If e and u are bivariate normally distributed with correlation coefficient ρ, then equation (5) becomes:

$$E\left[e \middle| \mathbf{x}, w^{*} > 0\right] = E\left[e \middle| \mathbf{x}, u > -\alpha' \mathbf{z}\right] = \rho\lambda \tag{6}$$

where λ is defined as the ratio of the density and the cumulative distribution function for a standard normal variable:

$$\lambda = \frac{\phi(-a' \mathbf{z})}{\Phi(a' \mathbf{z})} \tag{7}$$

Equation (1) for quality wine consumers becomes:

$$U_j^{*} = \beta' \mathbf{x} + \rho\lambda + e_j \tag{8}$$

This may be estimated if a consistent estimate of λ is obtained. Consistent estimates of λ are obtained by estimating the ordinary probit equation in (4) and all sample observations to compute consistent estimates of α that are used in equation (7) to estimate λ. Equation (8) is then estimated using the subsample of quality wine consumers as an ordered probit model by replacing λ with the consistent estimates derived by equation (7). The selectivity bias test is then equivalent to a t-test of the null hypothesis that ρ equals zero. The probability that a consumer chooses one of the three classes available for each quality attribute is given by:

$$\begin{aligned}
\text{Prob}(j=0) &= 1-\Phi(\beta'\mathbf{x}+\rho\lambda)\\
\text{Prob}(j=1) &= \Phi(\mu-\beta'\mathbf{x}-\rho\lambda)-\Phi(-\beta'\mathbf{x}-\rho\lambda)\\
\text{Prob}(j=2) &= 1-\Phi(\mu-\beta'\mathbf{x}-\rho\lambda)
\end{aligned} \tag{9}$$

where $\Phi(.)$ is the standard normal cdf and $0<\mu$ in order for the three probabilities to be positive. For estimation we let a binary variable S_{ij} take the value of one if the ith respondent chooses category j, zero otherwise. The likelihood function for equation (9) is:

$$L(\beta,\mu,\rho)=\prod_{i=1}^{n}\prod_{j=1}^{3}(P_{ij})^{S_{ij}} \tag{10}$$

and the log likelihood function is given by:

$$L^{*}=\log L(\beta,\mu,\rho)=\sum_{i=1}^{n}\sum_{j=1}^{3}S_{ij}\log(P_{ij}) \tag{11}$$

Consistent and asymptotically efficient estimates of the model parameters (β,ρ,μ) are obtained by maximizing the likelihood function [24, 25].

4 Results and Discussion

Table 1 presents definitions and descriptive statistics of the explanatory variables ($\mathbf{z}$ and $\mathbf{x}$) used in the probit model of equation (4) and the ordered probit model of equation (9), and of the sub-samples of quality wine consumers and non-consumers. The sub-sample of quality wine consumers contains, on average, younger people (AGE), and proportionately more males (SEX), single (MARST), and of urban origin consumers (ORIGIN). Furthermore, for quality wine consumers, information concerning the product is mainly received by reading the product's label (LABEL) and less by reading relevant articles (MEDIA) or talking with friends and sellers (PERSON). Table 2 presents estimates of the probit coefficients in equation (4). The chi-square test rejects the null hypothesis that all coefficients on the explanatory variables in $\mathbf{z}$, are simultaneously equal to zero. The model correctly predicted 88.8% of all cases (661 out of 744). The decision to consume quality wine is influenced by a wide range of respondent's socio-economic characteristics. The older the respondent the lower the probability that she is a quality wine consumer. Men are more probable to consume quality wine than women. Holding a higher education degree increases the probability to consume quality wine, while if the respondent's origin is from a rural area the corresponding probability decreases. Preference to white wine increases the probability that quality wine is consumed.

The ordered probit model of equation (9) was estimated on the subsample of the 640 quality wine consumers only, after correcting for selectivity bias. The results of estimating equation (9) for each of the examined quality cues, namely certification, packaging and aroma are presented in table 3. The t-values for ρ indicate the presence of selectivity bias for certification, and aroma but not for packaging. Therefore, the subsample limited to quality wine consumers is not a random sample, at least for two of the cues examined in this work. The model's parameter estimates indicate the direction of the effect of each continuous explanatory variable on the response probabilities, but do not directly represent the actual probability changes. By differentiating equation (9) and taking into account selectivity bias, we find the marginal effects of the regressors on the probability as:

$$\begin{aligned}
\frac{\partial \mathrm{Prob}(j=0)}{\partial \mathbf{x}} &= -\phi(\beta'\mathbf{x}+\rho\lambda)\beta \\
\frac{\partial \mathrm{Prob}(j=1)}{\partial \mathbf{x}} &= [\phi(-\beta'\mathbf{x}-\rho\lambda)-\phi(\mu-\beta'\mathbf{x}-\rho\lambda)]\beta \\
\frac{\partial \mathrm{Prob}(j=2)}{\partial \mathbf{x}} &= \phi(\mu-\beta'\mathbf{x}-\rho\lambda)\beta
\end{aligned} \tag{12}$$

where $\phi(.)$ is the standard normal pdf evaluated at the means, $\overline{x}$ and $\overline{\lambda}$. The change in probability is a function of the probability itself which, when multiplied by 100 is the percentage change in the probability that the specific response is selected given a change in the variable. For dummy independent variables the marginal effects are analyzed as discrete or relative changes when the respective dummy takes on its two different values, 0 and 1 respectively [24]. The marginal effects of the regressors on the probabilities are presented in table 4 for j=2 (the highly important category).

The probability that certification is a very important extrinsic quality cue that highly influences the decision to purchase the product increases if the consumer has higher education, is not married or is of a rural origin. On the contrary, for the extrinsic quality cue of packaging the exact opposite effects are derived. Packaging is valued as a highly important extrinsic quality cue by consumers of a lower level of education that are married. It is important to note that the intrinsic quality cue referring to aroma resembles the same behavior as certification and is highly valued by consumers of a higher education and not married. Income inversely affects the probability of stating that certification or aroma is a highly important cue and does not significantly influence attitudes towards packaging.

If the consumer receives information concerning wine from the media the probability that certification is a very important quality cue decreases while information received by personal contacts and the product's label, increases the respective probability. Again, packaging is inversely influenced by the factors influencing certification. Information reception from media does not significantly influences the consumer's attitudes towards packaging. However, information received from personal contacts decreases the probability that packaging is highly valued and influence the purchasing decision process. For aroma, information received from media increases the probability that the consumer values this quality cue as highly important, while information received from personal contacts decreases the respective probability. Respondents receiving information from specialist columns in magazines and newspapers may have a better knowledge of the wine's qualitative cues and attributes and thus, do not rely on certification as quality assurance schemes or packaging as a quality sign. The more information which is received from the product's label, the higher the probability that certification and packaging are highly valued factors influencing the decision to purchase the product. On the contrary the opposite holds true for aroma i.e., the probability decreases.

To summarize the discussion so far, one may argue that certification and packaging target two distinct segments of the market while the intrinsic quality cue of aroma partly targets the behavior of certification and partly that of packaging. We may assume that certification targets the 'specialist', highly educated, not married and possibly upwardly mobile consumer relying on personal and probably 'confidential' information about 'the good wine'. Consumers that highly value packaging may be termed the 'typical consumer' with a medium education, and who is married and who does not rely on information received from personal contacts.

5 Conclusions

Certification, packaging and aroma are valued as important quality cues in the decision to purchase a quality wine by segments of the market possessing different characteristics. Thus, the use of extrinsic and intrinsic cues may target different segments of the market and differentiate a firm's marketing strategy. It is important to note that packaging, a clearly extrinsic cue and aroma a clearly intrinsic cue target consumers with exactly opposite characteristics. Certification, an extrinsic quality cue which, may be used as a surrogate for many extrinsic and intrinsic quality cues partly targets the same segment of the market as aroma (educational and marital status and income classes) and partly the opposite (information received by media, personal contacts and the product's label).

The behavior of consumers towards the three quality cues examined in this paper provides very useful conclusions and points out possible marketing strategies for firms producing imagery products. The fact that extrinsic quality cues, i.e., cues that do not affect the product's physical state, affect different segments of the market implies that firms can produce the same, basic product but should change packaging or acquire certification in order to achieve larger market penetration. Furthermore, certification and differentiated packaging may act as a barrier to entry for new firms thereby enhancing the benefits attributed to an already existing firm. However, a certified product or a product with distinct packaging may be an opportunity for a new entrant to challenge and win a place in the wine market.

On the other hand, the same results are useful for the formulation of a rural development strategy and the creation of 'niche markets'. Locally produced quality foods and drinks with designations of authenticity and geographical origin are transferred to the regional and national markets [26, 27, 28]. In this endeavor, localities retain more economic benefits and control over the types of economic activity, which occur. Within this context, quality products and services have considerable potential as a rural development tool in lagging rural regions. However, the exact effects of linking product and place on production and demand have not been fully explored in the academic literature. Further research is required especially in the fields of psychology and sociology to identify the factors influencing quality perceptions, and in the fields of economics and rural development to identify the impacts of using quality assurance marks on the demand and production of such products.

Throughout the present analysis, a wide range of qualitative variables have been used as explanatory (independent) variables predicting consumer attitudes. In doing so, an arbitrary quantification was applied on the qualitative scales and numerical variables were assigned. However, there is a number of alternative approaches that do not require arbitrary transformation, such as fuzzy set theory and multicriteria analysis techniques. The comparative performance of such techniques should be examined in future research with the aim to enrich quantitative techniques in economics.

6 Acknowledgements

This work arises out of a programme of collaborative research by the following: the Department of Geography at the Universities of Coventry, Leicester, Lancaster, Caen, Valencia, Galway and Trinity College Dublin; the Scottish Agricultural College (Aberdeen); Institute of Rural Studies (Aberystwyth); CEMAGREF (Clermont-Ferrand); Teagasc (Dublin); Department of Economics (University of Patras); and Seinajoki Institute for Rural Research and Training (University of Helsinki). The research is being funded under the EU's FAIR programme (FAIR3-CT96-1827).

References

1. Steenkamp, J-B. (1990) Conceptual Model of the Quality Perception Process. *Journal of Business Research*, 21, 309-333.
2. Monroe, K. B. and Krishnan, R. (1985) The Effects of Price and Subjective Product Evaluations, in *Perceived Quality*, Jacob Jacoby and Jerry Olson, eds. Lexington Books, Lexington, MA, 209-232.
3. Ophuis, P., A. and van Trijp, H. (1995) Perceived Quality: A Market Driven and Consumer Oriented Approach. *Food Quality and Preference* 6, 177-183.

4. Olson, J. (1972) *Cue Utilization of the Quality Perception Process: A Cognitive Model and an Empirical Test*, unpublished Ph.D thesis, Purdue University.
5. Jun, W-J. and Jolibert, A. (1983) Revealed versus Hidden Attributes as Determinants of Perceived Product Quality, *Journal of Economic Psychology* 4, 263-272.
6. Davis, L. (1985) Effects of Physical Quality and Brand labeling on Perceptions of Clothing Quality, *Perceptual and Motor Skills* 61, 671-677.
7. Vastoia, A. (1997) Perceived quality and certification: the case of organic fruit. In G. Sciefer and R. Helbing (eds), *Quality management and process improvement for competitive advantage in agriculture and food. Proceedings of the 49th seminar of the European Association of Agricultural Economists.*
8. Rosen, D. (1984) Consumer Perceptions of Quality for Generic Grocery Products: a Comparison Across Product Categories. *Journal of Retailing*, 60, 64-80.
9. O' Neill, M. and Black, M. (1996) Current Quality Issues in the Northern Ireland Tourism Sector. *The Total Quality Magazine* 8, 15-19.
10. Shewfelt, R. L. (1999) What is quality, *Postharvest Biology and Technology,* 15, 197-200.
11. Quinn, M. (1994) Winning service quality – the PROMPT approach. *Journal of Retailing*, 60, 64-80.
12. Deaton, A. (1987) Estimation of Own-and Cross-price Elasticities from Household Survey Data. *Journal of Economics* 36, 7-30.
13. Deaton, A. (1988) Quality, Quantity, and Spatial Variation of Price. *American Economic Review* 78, 418-430.
14. Dong, D., Shonkwiler, J. and Capps, O. (1998) Estimation of Demand Functions Using Cross-Sectional Household Data: The Problem Revisited. *American Journal of Agricultural Economics* 80, 466-473.
15. Juric, B. and Worsley, A. (1998) Consumers' attitudes towards imported food products, *Food Quality and Preference*, 9(6), 431-441.
16. Dransfield, E., Zamora, F. and Bayle, M-C. (1998) Consumer Selection of Steaks as Influenced by Information and Price Index. *Food Quality and Preference* 9, 321-326.
17. Lappalainen, R., Kearney, J. and Gibney, M. (1998) A Pan EU Survey of Consumer Attitudes to Food, Nutrition and Health: An Overview, *Food Quality and Preference* 9, 467-478.
18. Mizerski, R. (1982) An Attribution Explanation of the Disproportionate Influence of Unfavorable Information. *Journal of Consumer Research*, 9, 123-40.
19. Engel, J.F. and Blackwell, R.D. (1982) *Consumer Behavior,* Dryden Press, Chicago.

20. Chang, H. and Kinnucan, H. (1991) Advertising, Information and Product Quality: The Case of Butter. *American Journal of Agricultural Economics* 73, 1195-1203.
21. Trost, R.P and Lee, L. F. (1984) Technical Training and Earning: a Polychotomous Choice Model with Selectivity. *Review of Economics and Statistics* 56: 151-156.
22. Greene, W. (1997) *Econometric Analysis.* Prentice Hall, New-Jersey.
23. Heckman, J. (1979) Sample Selection Bias as a Specification Error. *Econometrica* 46, 153-161.
24. Maddala, G. S. (1983) *Limited-dependent and Qualitative Variables in Econometrics*. Cambridge University Press, New York.
25. Battershill, M. and Gilg, A. (1996) Traditional Farming and Agri-environment Policy in Southwest England: Back to the Future? *Geoforum* 27, 133-47.
26. Young, C. and Morris, C. (1997) *Intra-industry Linkages in the Agro-food Chain: New Forces for Regulating Aspects of Food Production.* Paper presented to the *RGS-IBG conference*, Exeter, January.
27. Marsden, T. (1998) New Rural Territories: Regulating the Differentiated Rural Space. *Journal of Rural Studies*, 14, 107-117.

Table 1. Definitions and Descriptive Statistics of Explanatory Variables

Variable name	Variable definition	Quality Wine Consumers		Non-Consumers		*All*	
		Mean	St.Dev.	Mean	St.Dev.	Mean	St.Dev.
AGE	Respondent's age in years	36.02	11.30	42.11	16.67	36.87	12.37
SEX	*1 if respondent is female*	0.55	0.50	0.72	0.45	0.57	0.49
EDUC	1 if respondent has third level education	0.68	0.47	0.87	0.33	0.70	0.46
MARST	1 if respondent is not married	0.52	0.50	0.46	0.50	0.51	0.50
ORIGIN	1 if respondent comes from a rural area	0.38	0.48	0.60	0.49	0.41	0.49
INCOME	Respondent's family after tax income in million Gr.Dr.	5.09	3.33	5.13	2.99	5.10	3.28
WINTYPE	1 if respondent is red wine 'lover'	0.36	0.48	0.60	0.49	0.39	0.49
MEDIA	1 if respondent derives information on wine from media	0.39	0.49	0.48	0.50	0.40	0.49
PERSON	1 if respondent derives information from friends	0.34	0.47	0.43	0.50	0.35	0.48
LABEL	Number of information items read from wine labels	4.92	2.28	0.1	0.27	4.24	2.70
N	Sample size		640		104		744

Table 2. Parameter Estimates of the Probit Model

Variable name	*Parameter Estimates of the Probit Model*	
	Estimated Coefficient	Asymptotic t-value
Constant	2.886	11.080
AGE	-0.015	-3.102
SEX	-0.390	-3.093
EDUC	-0.801	-5.083
ORIGIN	-0.416	-3.185
WINTYPE	-0.407	-3.239
Log-likelihood	-261.377	
Restricted Log-likelihood	-301.002	
Chi-squared	79.251	
% Correct predictions	88.84	

Table 3. Parameter Estimates of the Ordered Probit Model

Variable names	Certification	Packaging	Aroma
Constant	1.418	0.127	1.738
	(4.672)**	(0.764)	(9.232)**
EDUC	-0.873	0.369	-0.210
	(-7.395)**	(3.585)**	(-1.866)*
MARST	-0.564	-0.263	-0.283
	(-4.955)**	(-2.607)**	(-1.801)*
ORIGIN	-0.203	0.032	-0.028
	(-1.926)**	(0.345)	(-0.280)
INCOME	-0.112	-0.020	-0.047
	(-6.323)**	(-1.255)	(-2.880)**
WINTYPE	-0.064	0.018	-0.091
	(-0.631)	(0.187)	(-0.889)
MEDIA	0.505	-0.097	-0.565
	(4.387)**	(-0.862)	(-4.420)**
PERSON	0.307	-0.175	0.475
	(2.765)**	(-1.653)*	(3.534)**
LABEL	0.237	0.145	-0.057
	(9.463)**	(6.399)**	(-2.417)**
μ	1.717	1.228	0.744
	(21.177)**	(18.709)**	(13.797)**
ρ	0.793	0.096	0.677
	(5.678)**	(0.173)	(2.827)**
Log-likelihood	-521.678	-651.517	-606.111
Restricted Log-likelihood	-607.353	-737.19	-691.786
% Correct Predictions	64.06	56.62	54.53

Note: Figures in parentheses are t-values, single and double asterisks indicate statistical significance at the a=0.10 and 0.05 levels, respectively.

Table 4. Marginal Effects of the Explanatory Variables in the Probability of Choosing the 'Highly Important' Category

Variable name	Certification	Packaging	Aroma
EDUC = 0	-0.459	0.126	-0.164
EDUC = 1	-0.113	0.107	-0.138
EDUC change	0.346	-0.233	0.026
MARST = 0	-0.107	-0.081	-0.053
MARST = 1	-0.057	-0.093	-0.049
MARST change	0.050	-0.012	0.004
ORIGIN = 0	-0.095		
ORIGIN = 1	-0.050		
ORIGIN change	0.045		
INCOME	-0.019		-0.016
MEDIA = 0	0.075		-0.223
MEDIA = 1	0.040		-0.162
MEDIA change	-0.035		0.061
PERSON = 0	0.026	-0.056	0.176
PERSON = 1	0.068	-0.065	0.169
PERSON change	0.042	-0.009	-0.007
LABEL	0.040	0.050	-0.021

Note: The table includes values only where the corresponding t-values reported in Table 3 are statistically significant.

IDENTIFICATION OF THE MARKET TRENDS FOR GREEK FRUIT JUICES

GEORGE BAOURAKIS AND IOANNIS APOSTOLAKIS

Mediterranean Agronomic Institute of Chania, Department of Economic, Marketing and Finance, PO. Box: 85, Chania, 73100, Greece.

E-mail: baouraki@maich.gr, yiannis@maich.gr

PERIKLIS DRAKOS

University of Crete, Department of Economics, School of Social Sciences, University Campus, Rethymno, 74100, Greece

e-mail: drakosp@mail.soc.uoc.gr

Consumers' preferences, attitudes and perceptions are examined with respect to fruit juices in Greece. An overview of the European and Greek Juice markets are presented. In an attempt to identify the main criteria and determining consumer purchasing behavior. An extensive number of consumers (800) was interviewed. Data analysis and multicriteria methodology is employed in order to retrieve and determine new market trends.

1 Introduction

Fruit juice sector in Greece has exhibited a significant growth rate in the area of utmost importance to the Greek economy. This sector occupies an essential place among the other agricultural products and there is great potential for further improvement of its market at the domestic as well as at the international level.

An increasing demand for traditional, and in recent years, for tropical fruit juices has been observed globally, with a corresponding growth in the juice industry, resulting in bigger plants, use of high-technology in process and packaging, and an increase in advertising expenses.

An increase in the consumption of fruit juice has taken place in Greece in recent years, a fact which is put down to changes in consumer preferences. Fruit juice consumption per capita in Greece more than doubled in the last decade.

Since the increase in diversification and expansion of dairy companies, their activities in juice production, and in the juice sector has witnessed an overall growth.

The present consumer-based market survey was carried out in three cities in Greece: Athens, Thessaloniki and Heraklion. The sample consisted of approximately 800 consumers who were randomly selected in a number of supermarkets in the aforementioned cities.

The aim of the present study was, firstly, to identify the market segments of juices and the factors that influence the buying behaviour of juice consumers. The study deals with consumer preferences, attitudes and perceptions with regard to the special characteristics pertaining to juices such as packaging, colour, price, taste, and advertisement. It further focuses on the exploration of consumer opinion in Athens, Thessaloniki and Heraklion. The second aim of the study was to identify the intended use of purchase differences in buying behaviour to describe distinguishing characteristics of the fruit juice market and finally to group these characteristics into "clusters".

2 European juice market

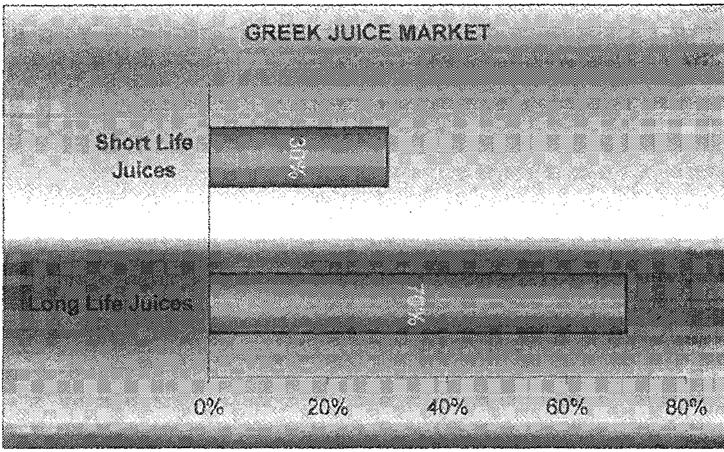
Germany, is the biggest consumer of fruit juices with the average German consuming 50 litres per head, followed by the Netherlands with 22 litres per head and the United Kingdom with 19 litres per head. France is the EU country with the strongest growth rate. This is due to the aggressive marketing strategies, which have been employed over recent years, even when there has been no product innovation. The EU's favourite flavour is orange. Despite the differences between the various countries in terms of the occasions when fruit juices are consumed, the demand for juices is rising. This is particularly true with regard to 100% products made from squeezed fruits rich in liquids and which are often also used to make long drinks and cocktails. The most important points as far as strengthening demand is concerned, are purity and the health aspect. In the case of all juices, nutritional value is also very important, and consequently, so is the possibility of these products having a real utility function for the consumer. Generally speaking, there has been a strong demand in the baby foods sector [7].

3 Greek juice market

The juice market is divided into two main categories, long-life juices, which dominate 70% of the market, and pasteurised or short-life juices, which comprise 30% of the market. Trends show that in the recent years there has been an increase in the demand for long life juices [10].

Juices can also be divided into three main subcategories, based on juice content. Thus, we have 100% juices, nectars, (i.e. those with a content of more than 50%) and fruit drinks with a content of more than 20% [10].

Pasteurised juices are mainly a product of Attica, where 75% of the volume of sales is made. Long-life juices have a more normal distribution since Attica consumes about 40% of the total of juices [10].



A survey identified when juices are usually consumed.

Between breakfast and lunch	7%
During lunch	6%
Between lunch and dinner	13%
During dinner	3%
After dinner	7%

The survey also showed that 82% of the respondents consume juices at home. According to the research, this great increase in consumption of most juices in Greece is mainly due to:

- Changes in consumption patterns and lifestyle (e.g. healthy diet)
- The introduction into the market of new, improved products
- Appealing packaging, which is mainly in carton form which facilitates their use
- Their increased availability- juices are available from more shops
- Intense advertising campaigns of the companies that work in this sector
- New non-traditional flavours that hold the market dynamic.

Recently, competition among big companies has increased with the introduction of products emphasising new tastes and more convenient packaging mainly from the four big companies.

4 The leading companies

The high competition in the juice and soft-drinks industry is easily identified. It is based on the introduction of a series of new products, marketing channels and intensive advertising. Within the next few years, it is expected that this competition will become more intensive [10].

Domestic production of fruit juices entered a new phase in 1984 with the introduction of the Amita long life brand by the Hellenic Bottling Company. The market grew strongly after 1989 when DELTA produced its "fresh" chilled juices and 3E responded with its Amita Cool and Evga with Refresh [4].

In 1992, Hellenic Bottling with its Amita and Amita Cool product line was in the lead with a total market share of 49 %. DELTA, with only its Life chilled range, trailed second with 21% while Florina, which had only long life products, was third with 14% because of its strong presence in the north of the country. Evga which only entered the market with its chilled Refresh brand in 1991 pushed into fourth place with 5% and IVI which had been struggling to restore its market position had only a 4,5% market share [4]. In 1996, Hellenic Bottling was still dominant, possessing 53% of the market followed by DELTA (25%). Florina and Evga followed with 14% and 6% of the market, respectively [9].

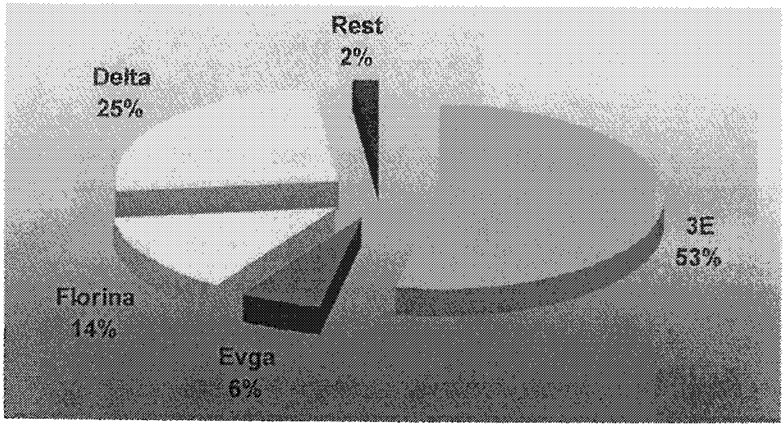


Figure 1: The Greek juice market shares in 1996
Source: IOBE, as shown in Industrial Review (1997)

5 Market survey

5.1 Data analysis methods

The frequency distribution and proportions can be used in simple tabulation. The other way to derive indicate the pattern of responses from a single question is to compute a statistic. The most common statistics are mode, median, and mean. To assess the accuracy of the mean in representing responses, a measure of central tendency is needed. Thus, measures of dispersion (e.g. standard deviation and variance) are required.

Bivariate analysis deals with two variables at a time. The choice of the proper procedure, which examines the relationship between the two variables, depends on the scale level of the variables and on whether the researcher requires a descriptive statistic or an inferential test [6]. At this stage, cross-tabulation between questions and difference in means, which is tested with a t-test statistic, are calculated. The objective of cross-tabulation is to repeat the analysis of a single question for various subgroups [11].

A measure of association indicates the extent to which characteristics of different sorts occur together [2]. Measures of association differ depending on the kind of data used. The data that are collected in the present research are nominal, ordinal and ratio scale data. Thus the following measures of association were used: Phi, Cramer's V Contingency Coefficient Lambda, Tau b, Tau c Gamma.

Correspondence analysis identifies interrelationships among possible alternative answers. It analyses correspondence tables (such as two-way frequency tables) and produces a graphical display that characterizes the relation between the categorical variables and computes a score for each row and each column category, displaying them in a scatterplot. Categories that are similar to each other appear close to each other in the plot. Mathematically, this procedure decomposes a chi-square measure of association for the table into components in a manner similar to that of principal component analysis for continuous data [5].

The Non-linear Principal Components Analysis procedure (known by the acronym PRINCALS - Principal Components Analysis by means of Alternating Least Squares) allows for the reduction of an original set of variables into a smaller set of uncorrected variables that represents most of the information in the original variable set, while accounting for as much of the variation as possible in the original set of variables. Non-linear principal component analysis allows for the examination of any mix of nominal, ordinal, and numerical variables. Non-linear principal component analysis is similar to homogeneity analysis, but it allows inclusions of ordinal and numerical variables [1].

5.2 *Sample structure*

From the whole sample, 88.8% of consumers drink juices and just 11.2% do not, showing that juices is a widely consumed good. The respondents were asked what their favorite category among very fresh juices was: 2-3 days, the ordinary fresh juices and long life juices. Most of the consumers, 60.3%, prefer short-life juices, A second preference, 20.9%, is for long life juices and finally very short life (2-3 days) with 18,8%.

For the frequency of consumption a representative quantity was chosen, so the respondents gave their preference in glasses of juice for a certain period of time (day, week). Illustrating the results, 24.5% of the whole population claimed that they drink more than one glass per day, 30.8% said that they drink approximately one glass per day, 26.6% said that they drink two to three glasses per week, 9.2% claimed to drink approximately one glass per week and 8.9% less than one glass per week. It was assured that juices are widely consumed goods since 80% of the whole population drink one to two glasses everyday.

Where packaging is concerned, the juices were divided into three groups: • 250ml or 330ml, the small packages, • 500ml, the medium package and • 1000ml the big package.

In terms of volume, most respondents (54.5 %) preferred the big package of 1000ml. Seventeen point four percent (17,4%) preferred the smallest package of 250ml or 330ml and 12.8% preferred the medium package of 500ml. The remaining 15.2% didn't care about the volume.

Concerning the time of consumption within the day there was the opportunity for more than one answer to be given, so the results are:

Time of Day	%
At breakfast	53.9
Before lunch	20.6
During lunch	22.3
Evening	**58.6**
Night	33.4

6 Consumers' preference in the brands

In total, eight brands were chosen to be surveyed six of which have the largest market shares in the Greek market. These are Amita (3E), Ivi, Life (DELTA), Florina, Refresh (EVGA) and Frulite. The seventh juice, BIOXYM/CRETA FRESH, is a Cretan product and it was chosen as a regional interest product. Finally the last juice, Fresh Juice, is an imaginary product which was included among the target brands in order to evaluate the knowledge of and the attention paid by the respondents.

6.1 Cross-tabulation analysis

Upon aggregating and re-coding the questions «what kind of juices do you drink regarding duration» and «the sex», a new variable was constructed which deals with the buying behavior of the Greek consumers concerning the sex and the duration of the branded juices. Thus, respondents were divided into 3 groups according to their preferences regarding duration (very short life, short life and long life) and into two groups according to their sex, as presented in Table 1. Cross-tabulation shows the combination of these two variables and the significant results are presented in the chi square test.

Table 1: Cross-tabulation of sex and duration of branded juices

			SEX		
			male	female	Total
Duration	Very short life	Count	66	77	143
		% of Total	8.7%	10.2%	18.9%
	short life	Count	234	225	459
		% of Total	30.9%	29.7%	60.6%
	Long life	Count	68	88	156
		% of Total	9.0%	11.6%	20.6%
Total		Count	368	390	758
		% of Total	48.5%	51.5%	100.0%

Chi-Square Tests	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	2.951[a]	2	.049

Juice brand special characteristics

Consumers' preferences, attitudes and perceptions concerning eight different brands of fruit juice (Amita, Ivi, Life, Florina, Refresh, Frulite, Creta Fresh and Fresh Juish) were examined. Consumers were given five basic criteria such as packaging, colour, price, taste and advertisement to assess.

Table 2, shows consumers' preferences for the aforementioned brands in a ranking order according to the score (ratio index). It is computed by assigning certain weight to significance levels that were attributed to these criteria. A final approach for indicating the relative score was to recalibrate the scale, which was done by first finding the typical average importance score for an attribute by taking the average of the average importance scores. Averages were then recomputed as an index of the ratio of the average for the particular attribute to the overall average (ratio index). The advantage of this index is that it highlights extreme cases and quickly indicates relatively important and unimportant attributes [12].

The end of this table presents the consumer overall purchasing order according to the ratio index method and which of the eight Juice brands the consumer would purchase in a ranking order. "Amita" held first place in consumers' purchasing preference, with "Life" in second, "IVI" in third, "Frulite" and "Refresh" in fourth, "Florina" in sixth, "Creta Fresh" in seventh and finally "Fresh Juice" in last place.

Table 2: Consumers' preference in ranking order according to the following criteria

		Product							
Criteria		Amita	Ivi	Creta Fresh	Fresh Juice	Life	Florina	Refresh	Frulite
Price	*Ratio Index*	1.26	1.18	0.57	0.53	1.08	1.20	1.09	1.06
	Rank	*1 st*	*3 rd*	*7 th*	*8 th*	*5 th*	*2 nd*	*4 th*	*6 th*
Taste	*Ratio Index*	1.51	1.15	0.26	0.23	1.41	1.07	1.19	1.15
	Rank	*1 st*	*4 th*	*7 th*	*8 th*	*2 nd*	*6 th*	*3 rd*	*5 th*
Adverti-sement	*Ratio Index*	1.69	0.97	0.23	0.24	1.41	0.64	1.15	1.63
	Rank	*1 st*	*5 th*	*8 th*	*7 th*	*3 rd*	*6 th*	*4 th*	*2 nd*
Packa-ging	*Ratio Index*	1.23	1.14	0.58	0.55	1.17	1.04	1.12	1.12
	Rank	*1 st*	*3 rd*	*7 th*	*8 th*	*2 nd*	*6 th*	*5 th*	*4 th*
Color	*Ratio Index*	1.46	1.18	0.34	0.29	1.32	1.03	1.17	1.17
	Rank	*1 st*	*3 rd*	*7 th*	*8 th*	*2 nd*	*6 th*	*5 th*	*4 th*
Overall	Ratio index	0.22	0.80	1.60	1.73	0.62	1.15	0.93	0.93
	Order	1st	3 rd	7 th	8 th	2 nd	6 th	4 th	4 th

6.2 *Multivariate analysis*

6.2.1 Correspondence analysis

The Correspondence Analysis procedure allows for the examination of the relationship between two nominal variables, which are graphically displayed in a multidimensional scatterplot. It computes row and column scores and produces plots based on the scores. Both variables must be numeric, and the values should be integers. Fractional data values are truncated in the analysis.
The underlying relation between the variables «what is your favorite package (volume)» and the demographics variables of "age", "monthly family income", "education level" and "occupation" is displayed in Figure 2a, 2b, 2c, 2d below.

It was observed that the age group below twenty is indifferent towards any kind of juice packaging, while the age group 20-30 shows a strong preference for small packages. Finally, the rest of the age groups involved in the research preferred the one litter package (figure 2a).

With regard to the analysis of the relation between income and packaging, the low-income category seems to present a strong relation with the small package volumes. Regarding the remaining income categories the results are not so clear (figure 2b).

As far as package and occupation are concerned only respondents who have multi-employment seem to display an opposite behavior to those having other occupations (figure 2c).

With respect to the analysis of package and education the only useful outcome is that of respondents with a low level of education. They demonstrated an opposite behavior towards packaging as opposed to the behavior of the higher educated respondents (figure 2d).

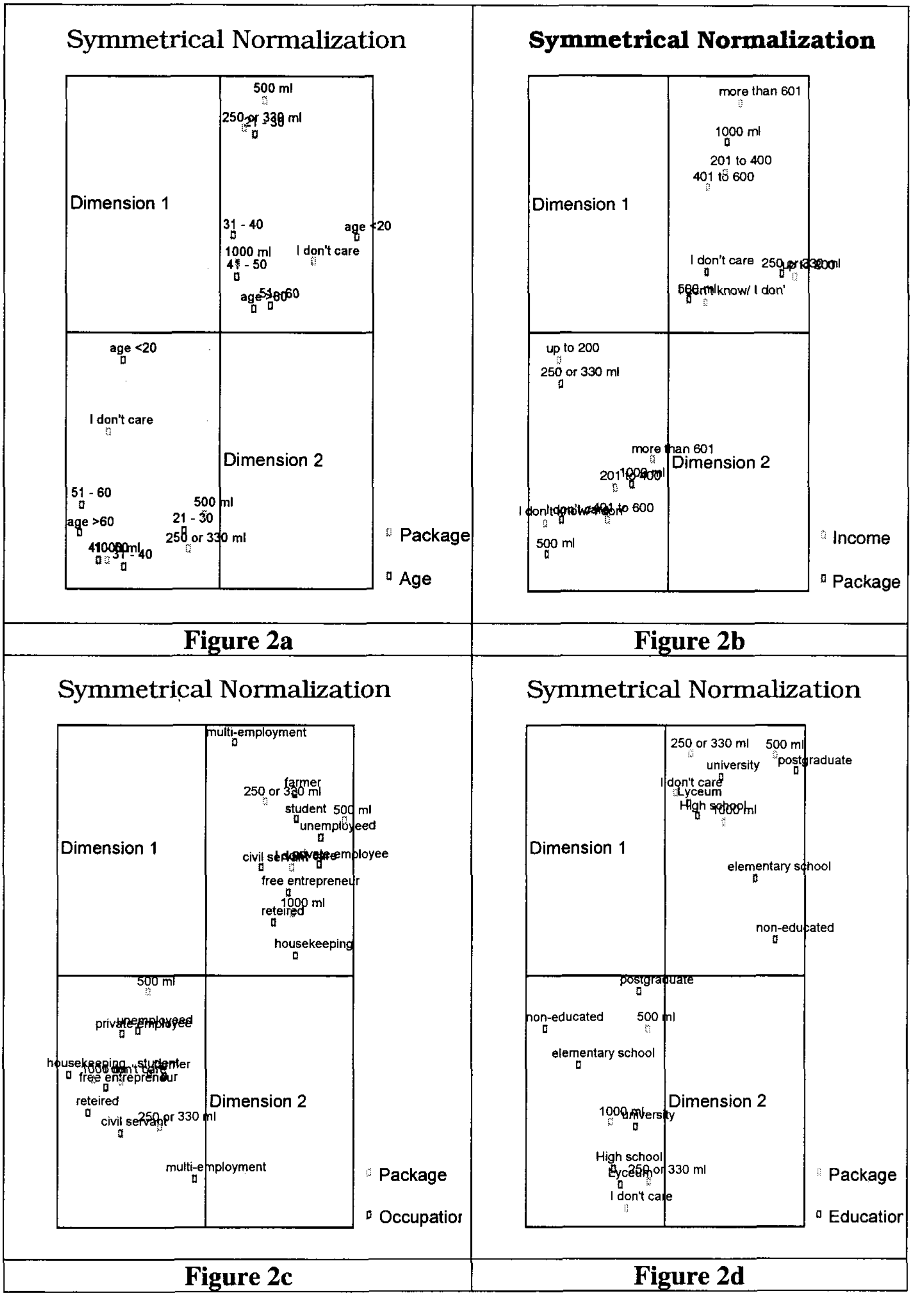


Figure 2a

Figure 2b

Figure 2c

Figure 2d

Figure 2: Package (volume) preference with respect to demographics

The interdependence between price perceptions with respect to the demographics variables of "age", "monthly family income", "occupation" and "education level" are shown in figures 3a and 3b.

In the first diagram, the category "very cheap as a price" perception does not apply to any age categories, as it appears alone regarding both dimensions examined in the correspondence analysis. The category of "cheap" seems to be close to the age category of young people less than 20 years of age, according to dimension one and two, while the categories "expensive" and "very expensive" related strongly with the remaining age categories (figure 3a).

The second diagram depicts the relation between income and the price perception. All the income categories above 200,000 drachmas per month seem to relate with the categories of "cheap" and "indifferent" while the low-income people present a different behavior regarding the price perception (figure 3b).

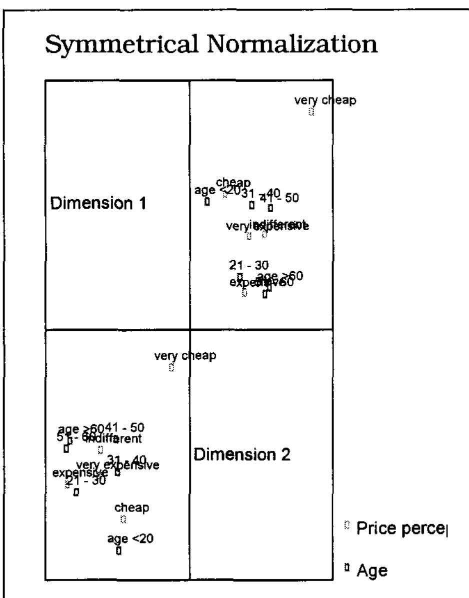


Figure 3a

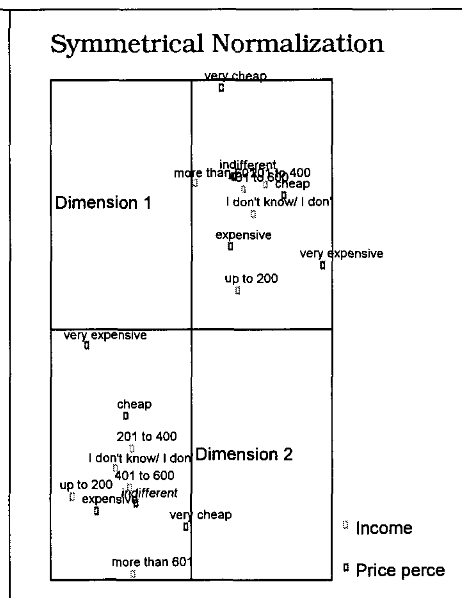


Figure 3b

Figure 3: Price perception with respect to demographics

6.3 *Non-linear Principal Components Analysis*

Non-linear principal components analysis (known by the acronym PRINCALS – PRINcipal Components analysis by means of Alternating Least Squares) reduces a set of variables (any mix of nominal, ordinal, and numerical variables) into a smaller set of uncorrelated variables that represents most of the information in the original variables. The scale (nominal, ordinal, and numeric) and range of each variable are defined (when all variables are quantitative, this analysis is analogous to classical principal components analysis). Plots of object scores and component loadings are available.

Non-linear principal components analysis was applied to the multi-criteria table of the questionnaire. The analysis took consumer preference for 8 different juice brands of the Greek market with regard to 5 different criteria such as price, taste, advertisement, packaging, and color into consideration. Forty variables were used (the variables start at pric_ami and finish at colr_frul). - The study covers:

- Analysis of the relations among the 40 variables. This includes a two-dimensional ordinal PRINCALS solution. The two dimensional solution will be compared with the numerical solution.
- Analysis of relations between the brands in each characteristic. Five different two-dimensional ordinal PRINCALS were applied to the eight brands.

Analysis of responses on brands and characteristics

Number of dimensions
We first performed a two-dimensional PRINCALS solution with all variables (m=40 ordinal). This resulted in eingenvalues of 0.2415 and 0.1263. Since the last eingenvalue is still larger than 1/m=0.025, a three dimensional ordinal solution was also considered. Its last eingenvalue appeared to be 0.0866.

7 Analysis of relations between the brands in each characteristic

In the present analysis eight different brands were examined (two-dimensional PRINCALS solution) with a view to the special characteristics attached to branded juices such as price, taste, advertisement, packaging, and color.

The graph plot for the Amita brand suggests a difference between the advertisement (ADV_AMI) on the one hand and the price, taste, and package (PRIC_AMI, TAST_AMI, PACK_AMI) characteristics on the other, according to dimension 2. However, the advertisement and color of the brand AMITA appeared on the positive side while the others were placed on the negative side and seem to affect consumer preference for this brand.

In the case of the IVI brand the variables react in exactly the same way for the consumers as in the case of the AMITA brand.

The graph plot of the brand name VIOCHYM indicates an opposite impact in preference of the consumer between the set of variables concerning price, package and taste (PRIC_VIO, TAST_VIO, PACK_VIO) with the variable advertisement (ADV_VIO). The color (COLR_VIO) of this brand does not affect consumer opinion since it is placed near the zero value according to dimension 2.

The other examined brand Fresh juice seems to exhibit the greatest percentage of total fit among the brands examined (showing that these two dimensions account for 82.82% of the total variance). The advertisement of fresh juice brand on the one hand and the group of characteristics including packaging, price, taste and color on the other suggest a difference according to the second dimension. These conclusions show us how the consumers group the characteristics for this brand of juice.

Where the brand name LIFE is concerned, the plot shows that advertisement (ADV_LIF) is one of the main characteristics, which influence consumer opinion while packaging (PACK_LIF), taste (TAST_LIF) and price (PRIC_LIF) suggest a difference according to the second dimension. This brand has a particular significance since it is the brand that does not exist in the Greek market but it is an "imaginary" brand invented in order to see how the respondents will react to it.

For the next product examined, the FLORINA brand, the two-dimensional ordinal Princals solution finds a distinction between the variable advertisement (ADV_FLO) on the one hand and the remaining variables on the other. According to dimension 1, none of the variables seem to have a special meaning in consumer preference. In dimension 2 a difference exists as described above between advertisement and the other examined characteristics of the brand.

In the case of REFRESH, the characteristics examined are grouped in the same way as the previous brand, with advertisement placed in a different place from the other examined criteria, as far as the second dimension is concerned.

For the last brand of juice, FRULITE, the two dimensional ordinal PRINCALS solution shows that in the first dimension of the solution all the characteristics behave in a similar way and a strong interdependence exists among them. The second dimension suggests a difference among the set of price, package, and taste (PRIC_FR, PACK_FR, TAST_FR), with the advertisement (ADV_FR) presenting the same trend as the previous brands.

8 Conclusions and recommendations

This study attempts to identify and determine consumer behaviour with respect to the attributes and special features displayed by the product examined. Market analysis was feasible through the use of multicriteria preference and data analysis methods and techniques. Consumer profile and the factors contributing to the selection of a product are determined by data analysis methods. Multicriteria preference analysis identifies and further evaluates the significant criteria that, to a large extent, determine the purchasing behaviour of the consumer.

The Greek fruit juice market is a relatively restricted market in comparison with the northern countries. Market figures have revealed an increasing tendency to consume fruit juices during the last 10 years, mainly due to the changes in lifestyles and the availability of widespread information on the nutritional and health aspects of fruit juices and aggressive marketing strategies of the leading companies.

In the present study eight different brands were examined in the light of the special characteristics of fruit juice such as packaging, color, price, taste and advertisement.

According to their purchasing preferences, consumers fall into two groups, one strongly faithful to the taste and package of the brand's name and the other to the price of the brands.

A new pricing policy for a product or a product with distinct packaging may be an opportunity for a new entrant to challenge and win a place in the fruit juice market.

The same policy would be useful in the formulation of a marketing strategy and the creation of "niche markets".

Despite the fact that the Fruit Juice market in Greece is stagnant and is dominated by four large companies, the two chosen segments are the ones that have flourished in the last few years and are, thus the most promising ones for the future (especially for new firms).

Finally, these results could prove useful as a basis for more detailed studies aimed at the successful marketing of fruit juice products and the efficient expansion of the industry.

9 References

1. Gifi, A. (1985), PRINCALS, Department of data theory, Leiden
2. Dempy, E. (1997) Success in marketing research? Here are 15 key factors that will help you. (Marketing News, Jan. 28, 1997, pp. 74-88)
3. FAO, (1992-1997) [URL http ://apps.fao.org] IOOC . International Olive Oil Council (1996). International trade standard applying to olive oil and olive pomace oil, Madrid.
4. Industrial Review (1997). Profitable Potables: a survey of the Greek drinks industry. Special Survey Series, No 10, pp. 7-32. [Supplement to Industrial Review, Vol. 60, No 680]
5. Lebart, L., A. Morineau and K. Warwick, (1984)Multivariate Descriptive Statistical Analysis: Correspondence Analysis (Wiley Inc., 1984, New York, pp. 143-148)
6. Kinnear, T. and J. Taylor, (1987) Marketing Research: an applied approach (3rd edition Mc Graw-Hill International, Singapore, 1987, pp. 235-241)
7. PANORAMA 1995. Fruit and vegetable processing and conserving. Panorama of EU Industry 1995/96. EUROSTAT: Brussels
8. Parras Rosa, M. (1996). World demand for olive oil. Olivae/No 63, pp. 24 – 33.
9. Pliagou N. (1997). Back to Roots: Mediterranean diet: the most successful diet model. Industrial Review, Vol. 64, No 721, pp. 838-844 [In Greek]

10. Psallas, Z. (1996). An x-ray of a "cool" market. Industrial Review Vol. 62, No 708, pp. 96-99 (596-599) [In Greek]
11. Siegel, S., (1991) Nonparametric statistics (New York: Mc Graw-Hill, 1991, p.254)
12. Wiers, R., Marketing Research (1988) (Prentice Hall International, 2nd edition, Englewood Cliffs, N.J., 1988).

This page is intentionally left blank

THE TOURISM INDUSTRY IN CRETE: THE IDENTIFICATION OF NEW MARKET SEGEMENTS

GEORGE BAOURAKIS

Mediterranean Agronomic Institute of Chania, Department of Economic, Marketing and Finance, PO. Box: 85, Chania, 73100, Greece.

E-mail: baouraki@maich.gr

The principle objective of this research was to provide an analysis and an assessment of both the market for Cretan agricultural products and of the latter's distribution in the local tourist market. A secondary objective was to provide information about Cretan products to foreign clients and to also guarantee the variety and quality of these commodities thereof. The identification of potential new markets and the promotion of Cretan products, adapted to consumer attitudes and beliefs, was to be a further objective of this research. Correspondence Analysis and Non-linear Principal Component Analysis were implemented. The former was conducted to identify the existing relationships among variables, and the latter to provide a thorough examination of the following factors: an analysis of the distribution channels, consumer behaviour, product characteristics and consumer willingness to buy the aforementioned products.

1 Introduction

The primary aim of obtaining information obtained from the current survey was to provide a thorough insight into the behaviour, attitudes and knowledge of Tourism enterprises concerning Cretan products, in particular, Cretan Extra Virgin Olive Oil, Fresh Vegetables, Fresh Juice, and Wine and Olive Oil Soap. More specifically, the goal of the present study was to assist in clarifying marketing issues concerning the products in question which were of interest to both the production and marketing sides of the market. The major issues focused upon in the current study were an analysis of the current market position of Cretan agricultural products and an assessment of the potential of this specific sector for further expansion.

The findings of the market survey, which was conducted in the four prefectures of Crete, afforded the information required to draw relevant conclusions and to allow recommendations to be made concerning the topic of this research.

2 Research methodology

The drawn up questionnaire was designed to cover several issues, which were important for the achievement of the objective of the research. The first part of the questionnaire sought and obtained information about the general attitudes and perceptions of the enterprises regarding the products that they used. Issues relating to the purchasing habits of the enterprises and the weight that they assigned to the attributes of the aforementioned products were included in this part.

The second part of the questionnaire focused on the nutritional habits and preferences of the respondents as a means of attaining a more precise view of consumer beliefs and attitudes. The last part was devoted to the behaviour of enterprises towards Cretan products. Due to the widespread distribution of these products, the awareness of the consumers who participated in the survey and the level of consumption was initially identified. Other issues focused upon in the questionnaire were the intention of enterprises to acquire Cretan products, their opinions concerning the use of Cretan products, as well as the degree of importance attached to the use of Cretan products by their consumers.

Finally, a further objective of this study was to attempt to identify possible new market segments and target marketing (Kotler, 2000).

The target population comprised enterprises (Hotels and Restaurants) in the four prefectures of Crete, in which the survey was carried out. The sample consisted of 100 randomly selected enterprises. Personal interviews were carried out with the directors of the hotels and the owners or those in charge of the restaurants, after completion of the structured questionnaire. The number of the sample respondents in each prefecture was chosen according to its proportion of the Cretan population in Crete. The hotels had to include a restaurant which did not serve frozen or fast food meals, in other words, the hotels were 5, 4, and 3 star category hotels.

The data analysis techniques employed in this research study were those of Cross-tabulation, Correspondence and Non-Linear Principal Component Analysis.

In view of the fact that agriculture and tourism are closely related in Crete, a first-time study of this nature was considered to be essential.

3 Tourist sector of Crete

The contribution of tourism to the economic growth of many areas, especially regional ones, is very important to the country. Crete is one of the main target locations of tourists in Greece. Therefore, the interest in the performance of the sector is significant for the economy not only at the local level, but also at the national level. The mild climate of the island and the beautiful landscape, along with the remarkable tourist resorts, attract as many as 2,400,000 visitors every year.

General governmental policy decisions about the development of tourism have been based on the existence of comparative advantages (in terms of natural resources) of many regions in Greece, and on the fact that this kind of developmental activity has needed just the basic infrastructure networks. The subsequent radical increase in the overall demand for tourism services and the income level of the regions explains the importance of tourism in the regional analysis. On the other hand, the anarchic pattern of development and the lack of regional planning and interference has led to such phenomena as the shrinkage of the agricultural sector and small industry and the enlargement of the service sector with all its socio-economic and cultural consequences at the regional level, (Vafeiadis et al., 1992).

Regarding the charter arrivals in Crete, it can be observed in the figure below there has been a high increase in charter flights. In 1986, there were 881.992 charter flights to Crete which last year, in 1999 almost tripled with 2.449.986 arrivals.

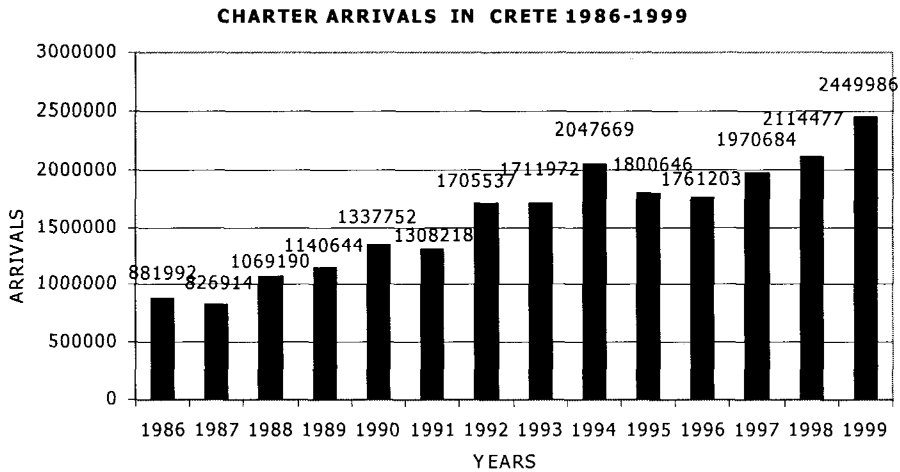


Figure 1: Number of Charter flight arrivals

In a depth analysis, we examined the arrivals by country. In the figure below we can see the percentages of these arrivals by country; the highest percentage was achieved by Germany with 31%, followed by Scandinavia, 18%, then England with a substantial percentage of 15%.

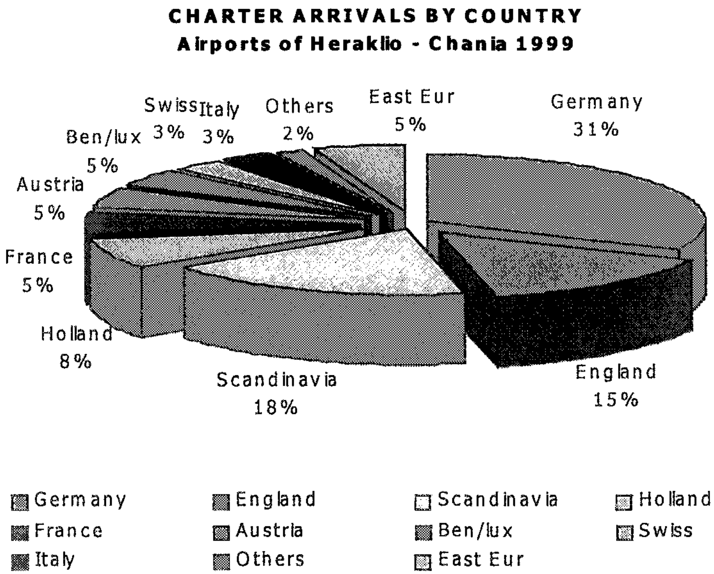


Figure 2: Charter arrivals by country

Regarding the infrastructure in the four prefectures of Crete, it is obvious from the next figure that the prefecture of Heraklio is in first position regarding bed capacity. It accounts for almost half the bed capacity, followed by the prefecture of Lassithi. Third position is held by the prefecture of Rethymno, followed by the prefecture of Chania in last position. However, percentages appear to converge over time between the four regions. The decrease in the percentage for the region of Heraklio since 1989 can be attributed to the development of tourism operations by some hotel units. Hence, the development of the infrastructure continues at rates which are higher for the other three prefectures than for the region of Heraklio, a fact that may indicate the diminishing probability of further development in the aforementioned prefecture.

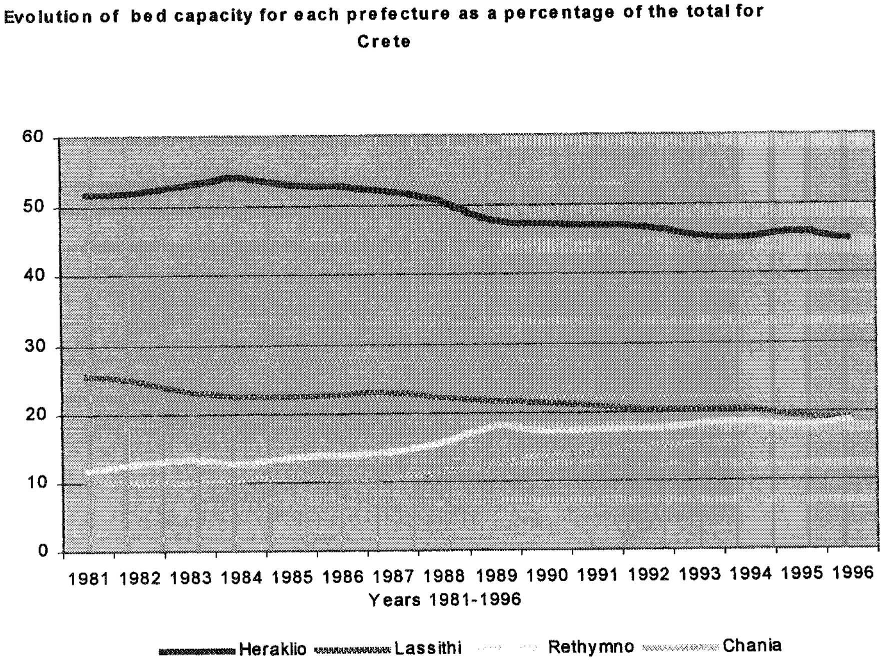


Figure 3: The evolution of the bed capacity of each prefecture as a percentage of the total for Crete
Source: Choice 1996,1997

The figure that follows is an index of productivity; up to 1993, the most productive seems to have been the prefecture of Heraklio, the least productive the prefecture of Chania (but with an improvement in the last years), with the other two prefectures fluctuating in between. After 1993, however, the relation among the prefectures obviously changed. The prefecture of Chania now almost coincided with the prefecture of Heraklio and the other two prefectures lost their leading positions. There is evidence to suggest that the prefecture of Chania developed in a more "productive" way than the other three prefectures during this time period.

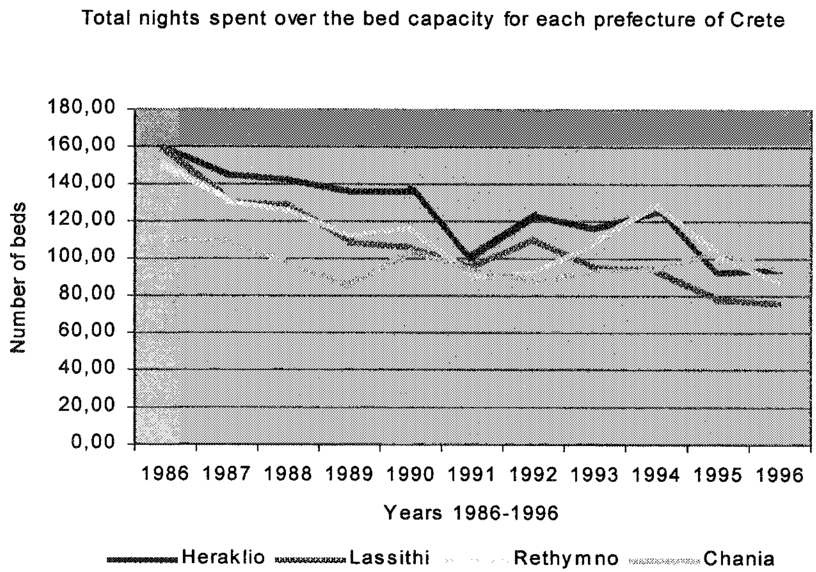


Figure 4: Total nights spent over bed capacity for each prefecture of Crete
Source: Choice 1996,1997

4 Market survey

The analysis provides descriptive statistics frequency distributions and cross-tabulation of the respondents' answers which reflect general information about consumer (Hotels and Restaurants) preferences with regard to Cretan agricultural products.

4.1 Analysis of general information

The questionnaire helped to map consumer preferences and attitudes on topics that are related to the consumption of Cretan products. The table below indicates the consumers' frequency distribution of the sample by Prefecture, Enterprise Type, Working Period and Capacity of the enterprises.

The "Working Period" information comes from an open-ended question and has accordingly been categorized into three periods.

Table 1. Frequency distribution of "Prefecture", "Enterprise Type", "Working Period" and "Capacity".

	Count/Percent (%)
"Prefecture" Frequencies	
Chania	26
Rethymno	13
Heraklio	48
Lassithi	13
"Enterprise Type" Frequencies	
Hotel	47
Restaurant	53
"Working Period" Frequencies	
7 months	67
8-10 months	21
12 months	12
"Capacity" Frequencies	
70-150	25
151-200	22
201-350	16
351-600	17
601-1100	20

Eighty-eight percent (88%) of the enterprises closed for the winter period (December-February), working only from March until late November. The same procedure was followed for the "Capacity of the Enterprise" question which was classified into five categories (See Table 1).

The comparison that followed was between "Enterprise Type" and "Working Period". There were no big differences in the "Working Period" among the enterprises that were interviewed. A total of 67% of respondents worked "7 months". Twenty one percent (21%) worked for "8-10 months" and 12% worked for "12 months" (See Table 2).

Table 2. Cross-tabulation of "Enterprise Type" * "Working Period"

		"Working Period"			
"Enterprise Type"		**7 months**	**8-10 months**	**12 months**	**Total**
Hotel	Count/Percent	33	11	3	47
Restaurant	Count/Percent	34	10	9	53
Total	Count/Percent	67	21	12	100

The same procedure was carried out for the cross-tabulation of questions "Enterprise Type" and "Capacity". As can be seen in Table 3, "Hotels" accounted for 47% of the total sample. The results indicated that "Hotels" which belonged to the "Capacity" category "351-600" beds comprised 16% of the sample and the "Capacity" category "601-1100" accounted for 20%. The "Restaurants" sample which was close to the "Capacity" categories "70-150", "151-200" represented 23% and 19%, respectively, of the total sample.

Table 3. Cross-tabulation "Enterprise Type" * "Capacity"

		"Capacity"					
"Enterprise Type"		**70-150**	**151-200**	**201-350**	**351-600**	**601-1100**	**Total**
Hotel	Count/Percent	2	3	6	16	20	47
Restaurant	Count/Percent	23	19	10	1	-	53
Total	Count/Percent	25	22	16	17	20	100

5 Multivariate analysis

5.1 Correspondence analysis

The Correspondence Analysis procedure allows examination of the relationship between two nominal variables plotted graphically in a multidimensional scatter plot. It computes row and column scores and produces plots based on the scores. Both variables must be numeric, and the values should be integers. Fractional data values are truncated in the analysis. The underlying relation between the variables "What is your opinion of the cost of Cretan Agricultural Products?" and "Prefecture" has been analyzed.

Specifically, for Extra Virgin Olive oil, upon individual examination of both dimensions in the correspondence analysis, the categories "Very Expensive", "Very Cheap" and "Cheap" (as a cost perception) were not related to any of the examined prefecture categories. Combining the variables of "What is your opinion of the cost of Cretan Fresh Juice?" and the variable "Prefecture", it was observed that the categories "Cheap", "Very Cheap", "Very Expensive" and "I don't know" did not indicate different behaviour according to the correspondence analysis diagram.

The interdependence between the cost of Cretan Wine and the "Prefectures" variable showed that the category of "cheap" seemed to represent prefecture "Chania".

The underlying relations among the variables "Cretan Olive Oil Soap" and "Prefecture" indicated that the categories relating the cost of "Cretan Olive Oil Soap" appeared alone, regarding both dimensions examined in the correspondence analysis.

5.2 Non-linear principal components analysis

Non-linear principal components analysis (known by the acronym PRINCALS – PRINcipal Components analysis by means of Alternating Least Squares) reduces a set of variables (any mix of nominal, ordinal and numerical variables) into a smaller set of uncorrelated variables that represents most of the information in the original variables. The scale (nominal, ordinal and numeric) and range of each variable are defined (when all variables are quantitative, this analysis is analogous to classical principal components analysis). Plot of component loadings are available.

Non-linear principal components analysis was applied to the products of Extra Virgin olive oil, Vegetable, Fresh Juices and Olive Oil Soap. The analysis took into consideration the preference for each product regarding the ranking of quality criteria in terms of how important these criteria were in the consumer choice patterns (question 3- Q3) and according to their opinions (question 5 – Q5).

5.2.1 Extra virgin olive oil

We first performed a two dimensional PRINCALS analysis with all variables (m=13 ordinal). This resulted in eingenvalues of 0.2754 and 0.2123. It was found that the last eingenvalue was still larger than 1/m=0.08 and total fit was 0.4877.

It is observed that the squared lengths of the arrows in the plot corresponded to row sum single fit. Projections on the horizontal axis corresponded to correlations between object scores on the first dimension and each of the qualified variables. Projections on the vertical axis corresponded to correlations between quantified variables and object scores on the second dimension.

According to the first and second dimensions and the projected variables along the horizontal axis, it was observed that the variables concerning the "Packaging" and "Country of Origin" criteria of Extra Virgin Olive Oil were located near zero and, therefore, were not such strong indicators of consumer opinion and choice patterns.

It was clear that, in the first dimension, the respondents' opinion criteria regarding Extra Virgin Olive Oil showed a difference between "Health Attribute" and "Nutritional Values" on the one hand, and "Smell", "Colour" and "Taste" characteristics on the other.

As far as the second dimension was concerned, the only important conclusion that could be drawn was that the consumer choice criteria of "Quality", "Label" had the opposite impact on "Price", "Variety" consumer preferences.

5.2.2 Cretan vegetables

We first performed a two dimensional PRINCALS analysis including all variables (m=13 ordinal). This resulted in eingenvalues of 0.2955 and 0.1847. The last eingenvalue still remained larger than 1/m=0.08 and total fit was 0.4802.

The two-dimensional ordinal principal components solution showed that there was a distinction between the variables examined in terms of consumer choice and respondents' opinion. Thus, the consumer choice criteria was placed along the vertical axis (dimension 1), while the respondents' opinion was placed along the horizontal axis (dimension 2), thereby depicting the difference between them.

5.2.3 Cretan Fresh Juice

We performed a two dimensional PRINCALS analysis including all variables (m=13 ordinal). This resulted in eingenvalues of 0.2785 and 0.2314. The last eingenvalue still remained larger than 1/m=0.08 and total fit was 0.5099.

The two dimensional ordinal principal components solution showed that the first dimension indicated a strong relationship among the set of "Health Attribute", "Nutritional Value" and "Country of Origin" criteria as characteristics influencing consumer choice preference. Also, this set indicates the opposite impact to that of the quality criteria. According to the second dimension, it was observed that the variables concerning the "Packaging", "Price" and "Label" criteria of fresh juice were placed together and a strong indicator of correlation was found among them.

5.2.4 Cretan Wine

The graph plot for Cretan Wine indicated that the "preference" criteria had the opposite effect than that of "quality" criteria. The criteria of "variety", "country of origin" and "label" had the opposite effect on both consumer choice preferences. The eingenvalues were 0.4119 and 0.1918 and exhibited a greater percentage of total fit among the products examined (indicating that these two dimensions accounted for 61.17% of the total variance).

5.2.5 Cretan Olive Oil Soap

Finally in the last comparison, the plot derived from the non-linear principal components analysis was not so clear. This could be attributed to the fact that many of the respondents were not knowledgeable about Olive Oil Soap. Thus, as far as the first dimension was concerned, "Quality" and "Price" were placed together but indicated the opposite effect than did the variables "Country of Origin", "Package" and "Label". Regarding the second dimension, the variable "Package" had the opposite impact than the "Variety" variable.

6 Conclusions

The primary aim of this current market research was to present the results of a survey on market attitudes with respect to Cretan agricultural products and their promotion through the tourism market. Within the framework of the present study, some important conclusions can be drawn.

Our analysis of the current market position of Cretan agricultural products has indicated that, despite the major importance of the sector for the total Greek agricultural market, is now relatively high. This can be attributed to several factors, such as the lower transportation costs incurred as a result of enterprises obtaining their products from within Crete, and to the high levels of per capita consumption of Cretan and Greek agricultural products, compared to those of other EU countries.

The majority of the respondents interviewed in this research study were found to use Cretan agricultural products. Regarding the places of purchase, the majority of the sample most commonly chose to buy Cretan products from the producer, the wholesaler and the cooperative.

Taking into account the frequency of purchase, most of the interviewees representing the sample purchased Cretan agricultural products on a daily basis, with the most dominant purchased quantities ranging from "20-40 kilos" for Extra Virgin Olive Oil and "100+" for the remaining products.

According to the findings, the origin of products exercised an unconditional impact on consumer decision-making concerning their purchases. Crete was the most dominant preference among those stated by the respondents.

The findings also indicated health attributes and nutritional value as being very important factors in determining consumers' purchasing decisions.

7 Recommendations

Agriculture is the main occupation of Cretans and the Tourism market is considered to be a secondary source of income for the majority of inhabitants in Crete. It is a foregone conclusion, therefore, that future studies need to focus on a combination of these two main markets.

An effort has to be made to support Cretan agricultural products through a combination of advertising, quality control and promotion through tourism enterprises.

The main issues that require promotion are:

The application of quality control of production, taking into consideration the advances in new agricultural practices (organic farming; biological control; insertion trademark for Cretan products, etc.)

Improvements in marketing strategies at local and international levels which will promote the benefits of the Cretan diet in international exhibitions with themes on diet, health, nutrition, etc.

The promotion of the benefits of the aforementioned products at a local level through the distribution channels.

Quality control has to be a first priority in the production stage in order to ascertain whether producers are using pesticides, insecticides etc. Secondly, to verify that the entrepreneurs are using Cretan products, it is necessary to introduce a trademark.

Advertising campaigns should concentrate on the targets and require the inclusion of local people, farmers and entrepreneurs to ensure the use and promotion of Cretan products. The focus point should be international tourism and food exhibitions as a means of familiarising as many people as possible with the benefits of the products consumed in the Cretan diet.

The tourism sector in Crete needs to shift to other types of tourism, such as nature-orientated holidays and agro tourism. That is in general, tourism that promotes a natural lifestyle and a good nutritious diet. These forms of tourism will not only promote the Cretan diet through tourism here in Greece but also abroad. In conclusion, it is evident that greater thought needs to be given to the economic advantages to be gained from combining the two main markets of agriculture and tourism in Crete.

References

1. Baker M.J. (1992). Marketing Strategy and Management. 2nd Edition. London: Macmillan.
2. Choice (1996), "Greek Economy", Special annual editions, Economic Review, Athens, (GR).
3. Choice (1997), Special annual editions, Economic Review, Data Athens, (GR).
4. Gifi A., PRINCALS. Department of Data Theory, Leiden, 1985.
5. Jobber D. (1995). "Principles and Practice of Marketing". London: Mc Graw-Hill.
6. Kotler P., (2000). "Marketing Management". The Millennium Edition.
7. Manly B., Multivariate Statistical Methods: A Primer. Chapman and Hall Editions, London, 1986.
8. National Statistics Service of Greece, (1981-1996), Tourist Statistics, Greece (GR).
9. Tagg S., (1996). Breaching the Qualitative/Quantitative Divide: Repertory Grid Analysis with SPSS Princals. MEG 96 Conference Research Methods Track Session.
10. Vafeiadis G., Kontogeorgos C., Papakonstantinidis L., (1992), Agro-tourism and equal development, Agricultural Bank of Greece, Bureau of Studies and Planning, Athens, (GR).

STOCK MARKETS AND PORTFOLIO MANAGEMENT

This page is intentionally left blank

THE USE OF FUZZY PROGRAMMING FOR THE MANAGEMENT OF IMMUNISED FIXED INCOME PORTFOLIOS

JORGE DE ANDRÉS, M. GLÒRIA BARBERÀ, ANTONIO TERCEÑO

Faculty of Management and Economic Sciences. Rovira i Virgili University. Spain
E-mail: jdas@fcee.urv.es; mgbm@fcee.urv.es; atg@fcee.urv.es

The programming models for selecting fixed income portfolios that are usually proposed in the literature have a number of limitations. For example, they consider profit and immunising risk to be mutually exclusive or suppose a clearly defined planning horizon. In this paper we propose the use of fuzzy programming to solve these limitations and to formalise the problem in a more realistic and flexible way.

1 Introduction

Mathematical programming is one of the most useful techniques for selecting portfolios, whether these are of stocks or fixed income securities. Its use in the management of stock portfolios; Markowitz's model [11] is the paradigm however, it is also very common in the management of fixed income portfolios. In this paper we will propose the use of fuzzy programming for managing immunised fixed income portfolios.

All the programming models for managing immunised fixed income portfolios propose maximising the return of the portfolio or, alternatively, minimising its immunising risk. The key constraint is a consequence of the equality between the portfolio duration and the planning horizon. Therefore, these programming models imply an acceptance that the planning horizon is quantified clearly and that the investor is not disposed to change it.

These models have some limitations that can be solved by fuzzy mathematical techniques. The first of these is the way in which the investor quantifies the investment horizon. It is supposed that the planning horizon is crisp and that the portfolio duration must exactly match the planning horizon. We believe that these postulates are not realistic because an investor may be disposed to accept some deviation of the duration about the planning horizon in order to obtain more profit on his investment. Likewise, the planning horizon is usually quantified in a vague way. The period may be, for example, "approximately three years" and not "exactly three years".

Another limitation of the usual mathematical programming models is the acceptance that the profit and the immunising risk of a portfolio are mutually exclusive management objectives. Actually, an investor tries to satisfy both objectives. This problem is therefore better conceptualised if we consider it as a multiple objective problem in which we are optimising profit and immunising risk simultaneously.

In this paper we propose to solve some of these problems by using fuzzy programming (FP). Firstly, this instrument allows us to homogenise the magnitudes related to the objective functions in multiple objective programming, because the value of the objectives is expressed as satisfaction levels. Secondly, we will obtain greater flexibility and realism when formalising an investor's decision making process because the assessment of the objectives and constraints with planning horizon is both subjective and vague.

In the next sections we present the basic concepts in the management of immunised fixed income portfolios and analyse the models on which we will base our proposals for using FP.

2 Price, return and risk of a fixed income security

A fixed income security provides its owner with several cash flows (coupons and principal) before its maturity. These can be represented as: $\{(C_s, t_s)\}_{s=1,2,\ldots,m}$, where C_s is the amount of the s–th cash-flow and t_s is its maturity measured in years. The price of one security, P is:

$$P = \sum_{s=1}^{m} C_s \left(1 + i_{t_s}\right)^{-t_s}$$

where i_{t_s} is the spot rate for the maturity t_s expressed as effective annual rate.

We can obtain the internal rate of return (IRR) of a bond from its price. The IRR is an average of the spot rates that apply in the market until the bond matures. If we denote the IRR as r, it can be obtained from the following equation:

$$P = \sum_{s=1}^{m} C_s (1 + r)^{-t_s}$$

As well as the issuer default risk, which will not be discussed in this paper, bonds are associated with the interest rate risk produced by movements in the temporal structure of interest rates (TEIR) throughout the planning horizon. This risk is made up of two compensating effects. The first one is the price risk, i.e. the fall in the price of the bonds when the spot interest rates increase. The second one is the risk of reinvesting the cash flows, because when the interest rates decrease, these cash flows will be reinvested at a lower interest rate than the one that existed when the investment started. These two risks have opposite effects. Investing the intermediate cash flows at a higher interest rate can compensate the drop in the price of a bond due to an increase in interest rates. If the TEIR fluctuations are parallel, for one bond the Fisher and Weill's duration [7], denoted as D, compensate both risks. This is expressed as:

$$D = \frac{\sum_{s=1}^{m} t_s C_s \left(1 + i_{t_s}\right)^{-t_s}}{P}$$

Duration is an additive measure. If D_i is the duration of the i-th bond, and x_i is the proportion of the i-th bond in a portfolio, the duration of a portfolio made up of n bonds, D_C, is:

$$D_C = \sum_{i=1}^{n} x_i D_i$$

We know that the return of a portfolio will be immunised if the planning horizon is equal to the duration of the portfolio. So if we denote this temporal horizon as H, an immunised portfolio must satisfy $D_C=H$.

When a portfolio is immunised by making its duration equal to H, we cannot assume that it is absolutely immunised. The profit from the portfolio is only immunised at the beginning of the investment because as time goes by the continuous movements of the TEIR eliminate the perfect matching of the duration with the portfolio duration. This concept of duration immunises a fixed income portfolio if the TEIR shifts are parallel, but not in other cases. These problems are known in the literature as immunising risk.

There are a number of ways in which this risk can be partially avoided. Firstly, we can use more sophisticated measures of duration, (e.g. multiplicative shocks measures) if the investor expects this kind of shift. Another way is to use a duration vector that reflects the different movements of the TEIR –see, for example, [2], [5] or [15]. However, Bierwag shows in [4] that in practice these concepts are not better than Fisher and Weill's duration.

A second solution is to restructure the portfolio continuously so that it is immunised throughout the planning horizon. However, we do not think this is suitable because of the transaction costs involved in the continuous buying and selling which would obviously reduce the final return of the portfolio considerably.

A third solution is to reduce the immunising risk by minimising it. The investor must minimise the dispersion of the cash flows around the planning horizon. If it were possible to select a portfolio made up of zero coupon bonds with maturity H, the portfolio return would be absolutely guaranteed and independent from the movements of the TEIR during the entire investment period. The most commonly used dispersion measure, which we have also used in this study, is Fong y Vasicek's M^2 (see [8] for a more detailed analysis). Its expression for a bond is:

$$M^2 = \frac{\sum_{s=1}^{m} \left(t_s - H\right)^2 C_s \left(1 + i_{t_s}\right)^{-t_s}}{P}$$

Dispersion is also an additive measure. The dispersion of a portfolio, M_C^2, can therefore be calculated from:

$$M_C^2 = \sum_{i=1}^{n} x_i M_i^2$$

Another application of duration we have used in this paper is the numerical approximation of the portfolio return. If r_i is the IRR of the i-th bond, and r_C is the portfolio IRR, according to [6] the estimated r_C is:

$$r_C \approx \frac{\sum_{i=1}^{n} x_i D_i r_i}{\sum_{i=1}^{n} x_i D_i}$$

Table 1 shows the TEIR for the Spanish public debt market on January 22nd 1999 from McCulloch's method [12]. Table 2 shows the characteristics of 5 hypothetical bonds from the TEIR in Table 1. These bonds will be used in this paper for numerical applications.

Table 1. TEIR January 22nd 1999

t	i_t	t	i_t
0.5	0.0280	8	0.0492
1	0.0304	9	0.0505
2	0.0347	10	0.0516
3	0.0383	11	0.0526
4	0.0413	12	0.0535
5	0.0438	13	0.0542
6	0.0459	14	0.0548
7	0.0477	15	0.0554

Table 2. Characteristics of the sample bonds

	B1	**B2**	**B3**	**B4**	**B5**
Annual coupon	0%	3%	3.5%	4%	5%
Maturity (years)	1	3	5	10	15
R_i	0.0304	0.0382	0.0434	0.0505	0.0534
D_i	1.00	2.91	4.66	8.31	10.67
M^2_i (H=3)	4.00	0.15	3.68	36.45	84.52
P	97.048	97.730	96.283	91.926	96.532

3 Mathematical programming models for selecting immunised fixed income portfolios

3.1 A static model for selecting immunised fixed income portfolio with a return criterion

As suggested by [6], it seems logical if the manager wants to immunise the portfolio to choose the one with the highest return. Specifically, the mathematical programme proposed by these authors, which we shall call [P1], is:

$$\text{Max } r_C \approx \frac{\sum_{i=1}^{n} x_i D_i r_i}{\sum_{i=1}^{n} x_i D_i}$$

subject to:

$$\sum_{i=1}^{n} x_i D_i = H \text{ , } \sum_{i=1}^{n} x_i = 1,\ x_i \geq 0 \qquad i = 1, 2, \ldots, n.$$

This programme is transformed into a linear programme simply by taking $z = \sum_{i=1}^{n} x_i D_i r_i$ as the objective function, since $\sum_{i=1}^{n} x_i D_i = H$.

With the sample of Table 2 and H= 3, [P1] produces Table 3. We can see that the selected portfolio has a dumbbell structure. This is because the objective is to maximise the return and so, the manager must buy the greatest possible proportion of the bonds with the highest profit (ones that matures after 15 years), and to immunise the portfolio, this long duration should be compensated with bonds of a shorter duration.

Table 3. Selected portfolio using [P1]

Bonds	**B1**	**B2**	**B3**	**B4**	**B5**
Percentage	79.33%	0.00%	0.00%	0.00%	20.67%
Portfolio characteristics	R_C	D_C	M^2_C		
	4.73%	3	20.65		

3.2 Selecting an immunised portfolio with a minimum risk of immunisation

[P1] obtains a portfolio with an immunised return if the planning horizon is H. However, the investment is only immunised at the beginning. Once it has started, the TEIR continuously shifts and changes shape, so that the price of the selected bonds and the conditions in which the intermediate cash-flows must be reinvested, etc. are continuously being modified.

However, if the investor sacrifices a return objective, the immunisation risk can be reducing minimised by the dispersion of the cash flows around the planning horizon. In [2], therefore the following mathematical programme, denoted as [P2], is proposed:

$$\text{Min } M_C^2 = \sum_{i=1}^{n} x_i M_i^2$$

subject to:

$$\sum_{i=1}^{n} x_i D_i = H \text{ , } \sum_{i=1}^{n} x_i = 1 \text{ , } x_i \geq 0 \qquad i=1,2,\ldots,n.$$

We can see that [P2] is similar to [P1], but in [P2] the immunising requirements are stronger. In [P1] the portfolio is only immunised when the investment begins and the selected portfolio has the highest potential return. On the other hand, by solving [P2] we will select the immunised portfolios with the lowest immunising risk.

From the sample Table 2, and with H=3, the characteristics of the selected portfolio after solving [P2] are given in Table 4. We can see that, in this case, the portfolio has a bullet structure and its immunised risk is minimum, but the IRR is noticeably lower than the return on the selected portfolio with [P1].

Table 4. Characteristics of the selected portfolio using [P2]

Bonds	**B1**	**B2**	**B3**	**B4**	**B5**
Percentage	0.00%	94.96%	5.04%	0.00%	0.00%
Portfolio characteristics	r_C	D_C	M^2_C		
	3.86%	3	0.33		

It is difficult to find an investor who only considers the objective of [P1] or [P2]. It is more common to try to meet the objectives of profit and risk immunisation simultaneously. It is therefore more realistic to construct a multiple objective programme, as [P3], to take into account both [P1] and [P2]:

$$\text{Max } r_C = \frac{\sum_{i=1}^{n} x_i D_i r_i}{\sum_{i=1}^{n} x_i D_i} \text{ , Min } M_C^2 = \sum_{i=1}^{n} x_i M_i^2$$

subject to:

$$\sum_{i=1}^{n} x_i D_i = H \text{ , } \sum_{i=1}^{n} x_i = 1 \text{ , } x_i \geq 0 \qquad i=1,2,\ldots,n.$$

Following on from this idea, the objective proposed in [14] is to maximise the terminal value of the investment, but to introduce M^2_C into the objective function with a penalty coefficient. In this way the portfolios with the most dispersed cash flows are penalised despite having more profit. In the next section we will show that fuzzy programming is a good instrument for solving this problem.

4 Using fuzzy programming to select fixed income portfolios

This section analyses the advantages of using fuzzy programming in the passive management of fixed income portfolios. We will show that it makes models [P1] and [P2] much more flexible. More specifically, the normal criteria for managing portfolios and handling information about objectives and constraints will be more realistic and more in tune with how investors make decisions.

We will use Zimmerman's fuzzy programming model (see [16] and [17]), which is based on Bellman and Zadeh's concept of fuzzy decision proposed in [3]. Objectives and constraints will receive the same treatment. In this way, the selected portfolio should maximise the satisfaction that is provided by meeting both objectives and constraints. This level of satisfaction is expressed via fuzzy subsets. We believe that this is a more accurate way of quantifying subjective concepts such as "satisfaction".

4.1 Using fuzzy programming for selecting fixed income portfolios with a fuzzy planning horizon and a profit objective

Investors are usually willing to partially sacrifice an immunisation objective, in order to obtain a greater yield. It is normal, therefore, to introduce some level of active criteria into the portfolio management. On the other hand once the investment has begun the need to match the duration with the planning horizon perfectly losses some of its sense. Likewise in many circumstances the planning horizon is ill-defined. For example, many fixed income mutual funds define their planning horizon through linguistic variables such as "middle term", "long term", etc. In such cases is better to quantify the temporal horizon with fuzzy subsets.

Thus, in [18] the portfolio return is the main objective but the deviations of the planning horizon from the planning horizon that are permitted will be penalised. More specifically, the authors suggest maximising the "weighted profit of the portfolio", which is obtained as the potential return of the portfolio, less a function of its duration deviations from the planning horizon.

FP is a good alternative for solving this problem. If an investor is willing to sacrifice some of the immunisation objective or if the planning horizon is defined by a linguistic variable, it is natural to express H as a fuzzy triangular number: $\tilde{H} = (H_1, H_2, H_3)$ in which is H_2 the most reliable planning horizon and H_2-H_1 y H_3-H_2 are the permitted portfolio duration deviations from the aforementioned temporal horizon.

From $\tilde{H}$, we can construct two membership functions to indicate the manager's satisfaction with the duration of the portfolio. If the duration is below H_2, this membership will be written as $\mu_{D_{C_1}}(x)$. Otherwise it will be written as $\mu_{D_{C_2}}(x)$. For these functions we propose a linear expression:

$$\mu_{D_{C_1}}(x)=\begin{cases} 1 & \sum_{i=1}^{n} x_i D_i \geq H_2 \\ \dfrac{D_C - H_1}{H_2 - H_1} & H_2 \geq \sum_{i=1}^{n} x_i D_i \geq H_1 \\ 0 & H_1 \geq \sum_{i=1}^{n} x_i D_i \end{cases} \tag{1}$$

$$\mu_{D_{C_2}}(x)=\begin{cases} 1 & \sum_{i=1}^{n} x_i D_i \leq H_2 \\ \dfrac{H_3 - D_C}{H_3 - H_2} & H_2 \geq \sum_{i=1}^{n} x_i D_i \geq H_3 \\ 0 & H_3 \leq \sum_{i=1}^{n} x_i D_i \end{cases} \tag{2}$$

The functions are illustrated in Figure 1.

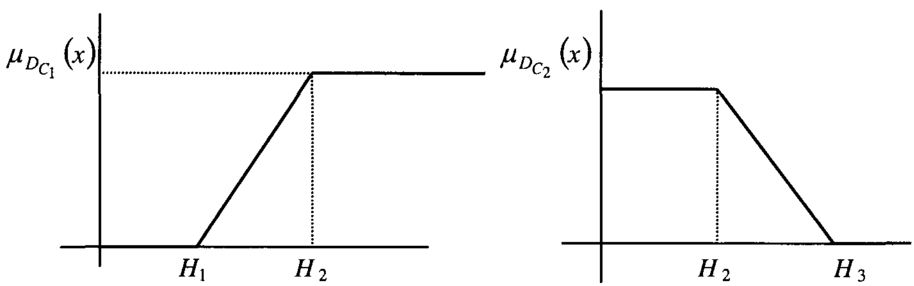


Figure 1. $\mu_{D_{C_1}}(x)$ and $\mu_{D_{C_2}}(x)$

With a fuzzy subset the investor can also quantify satisfaction level with any IRR. Its membership function is $\mu_{r_C}(x)$. We must remember that the IRR of the portfolio cannot be greater that the yield to maturity of the bond with the greatest return. The largest profit (r_{max}) one can demand of a portfolio, therefore, must satisfy $r_{max} \leq \text{Max}\{r_1, r_2, \ldots, r_n\}$.On the other hand, the lowest profit to demand, the r_{min}, must not be smaller than the one from an immunised portfolio, the IRR obtained by solving

[P1]. In this way, the membership function for the IRR objective may be expressed as:

$$\mu_{r_C}(x)=\begin{cases}1 & r_C \geq r_{max} \\ \dfrac{r_C - r_{min}}{r_{max} - r_{min}} & r_{max} \geq r_C \geq r_{min} \\ 0 & r_{min} \geq r_C\end{cases} \tag{3}$$

Using Zimmerman's FP model, the selected portfolio must maximise return and immunisation simultaneously. The linear programme [P3] must therefore be reformulated as [P4a] below:

$$\text{Max } \alpha = \text{Min } \{\mu_{D_{C_1}}(x), \mu_{D_{C_2}}(x), \mu_{r_C}(x)\}$$

subject to:

$$\mu_{D_{C1}}(x) \geq \alpha\,,\ \mu_{D_{C2}}(x) \geq \alpha\,,\ \mu_{r_C}(x) \geq \alpha\,,\ \sum_{i=1}^{n} x_i = 1\,, x_i \geq 0 \quad i=1,2,\dots,n,\ \alpha \in [0,1]$$

If we substitute the linear membership functions indicated, [P4a] is equivalent to the following auxiliary crisp programme, which we shall call [P4b]:

$$\text{Max } \alpha$$

subject to:

$$\sum_{i=1}^{n} x_i D_i r_i - (r_{max} - r_{min})\alpha \sum_{i=1}^{n} x_i D_i - r_{min} \sum_{i=1}^{n} x_i D_i \geq 0$$

$$\sum_{i=1}^{n} x_i D_i - (H_2 - H_1)\alpha \geq H_1\,,\ \sum_{i=1}^{n} x_i D_i + (H_3 - H_2)\alpha \leq H_3$$

$$\sum_{i=1}^{n} x_i = 1\,, x_i \geq 0, \qquad i=1,2,\dots,n,\ \alpha \in [0,1]$$

Table 6 shows the results obtained using the bonds in Table 2 and the data in Table 5. As expected, the portfolio maintained the dumbbell structure of [P1], but the duration of the portfolio is longer than the preferred (or most feasible) horizon planning. Consequently in the acquired portfolio the percentage of the bonds with maturity at 15 years is higher.

Table 5. Additional data to solve [P1] through [P4b]

Planning horizon	r_{min}	r_{max}
$\tilde{H} = (2.25, 3, 3.75)$	4.73% [a)]	5 % [b)]

[a)] Result after solving [P1],

[b)] The bond with the largest yield to maturity is B5 (5.34%).

Table 6. Characteristics of the optimum portfolio from [P4b]

Bonds	B1	B2	B3	B4	B5
Percentage	75.37%	0.00%	0.00%	0.00%	24.63%
Portfolio characteristics	α	r_C	D_C	M^2_C	
	0.37	4.83%	3.38	23.83	

4.2 *Using fuzzy programming to select immunised portfolios with a return objective and a minimised immunising risk*

In this subsection we assume that the planning horizon is done by a crisp number H and that the portfolio duration must match it exactly H. These assumptions are typical of mutual funds with guaranteed return if the manager wants to offer as high return as possible but takes into account the immunising risk.

FP is also a good alternative in this case. The decision-maker should specify a membership function, like (3), which quantifies the satisfaction level of satisfaction for a particular level of profit. One possible interest rate of reference for r_{max} is the positive ideal solution of r_C, i.e. its value in [P1]. Similarly, he/she should demand for r_C at least, its value from [P2] (the negative ideal solution), and this can be taken as a reference for r_{min}.

We must know the membership function that indicates the level of satisfaction for a certain M^2_C. To construct this the investor should not accept a greater dispersion than its negative ideal solution –the dispersion from the selected portfolio with the programme [P1]. It should also be borne in mind that a smaller dispersion than the one obtained by using the linear programme [P2] is not possible.

Thus we can construct the membership function for the dispersion of a portfolio as:

$$\mu_{M_C^2}(x)=\begin{cases} 1 & \sum_{i=1}^{n} x_i M_i^2 \le M_{min}^2 \\ \dfrac{M_{max}^2 - M_C^2}{M_{max}^2 - M_{min}^2} & M_{max}^2 \ge \sum_{i=1}^{n} x_i M_i^2 \ge M_{min}^2 \\ 0 & M_{max}^2 \le \sum_{i=1}^{n} x_i M_i^2 \end{cases} \tag{4}$$

where M_{min}^2 is the ideal dispersion and dispersions greater than M_{max}^2 are not possible.

Using Zimmerman's FP model, the multiple objective programme [P3] can be formulated as the following programme, which we shall call [P5a]:

$$\text{Max } \alpha = \text{Min } \left\{\mu_{r_C}(x), \mu_{M_C^2}(x)\right\}$$

subject to:

$$\mu_{r_C}(x) \geq \alpha,\ \mu_{M_C^2}(x) \geq \alpha,\ \sum_{i=1}^{n} x_i D_i = H,\ \sum_{i=1}^{n} x_i = 1,\ x_i \geq 0,\ i=1,2,\dots,n,$$

$\alpha \in [0,1]$

If the characteristics functions of $\mu_{r_C}(x)$ and $\mu_{M_C^2}(x)$ are as above, [P5a] is equivalent to the next programme, which we shall call [P5b]:

Max α

subject to:

$$\sum_{i=1}^{n} x_i D_i r_i - (r_{max} - r_{min})\alpha \sum_{i=1}^{n} x_i D_i - r_{min} \sum_{i=1}^{n} x_i D_i \geq 0$$

$$\sum_{i=1}^{n} x_i M_i^2 + (M_{max}^2 - M_{min}^2)\alpha \leq M_{max}^2$$

$$\sum_{i=1}^{n} x_i = 1,\ \sum_{i=1}^{n} x_i D_i = H,\ x_i \geq 0 \quad i=1,2,\dots,n,\ \alpha \in [0,1]$$

It is easy to prove that [P5b] is a linear programme since $\sum_{i=1}^{n} x_i D_i = H$

Table 8 shows the results of a numerical application that uses the additional data of Table 7. We can see that the consequence of considering simultaneously the profit and the risk immunisation taking a fixed duration is that the selected portfolio has an intermediate structure (ladder).

Table 7. Additional data for solving [P4] via [P5b]

H	M_{min}^2	M_{max}^2	r_{min}	r_{max}
3	0.33 [a)]	20.65 [b)]	3.86% [c)]	4.73% [d)]

[a)] The value of M_C^2 in [P2] [b)] The value of M_C^2 in [P1]
[c)] The value of r_C in [P2] [d)] The value of r_C in [P1]

Table 8. Characteristics of the selected portfolio with [P5b]

Bonds	**B1**	**B2**	**B3**	**B4**	**B5**
Percentage	60.13%	0.00%	25.07%	14.80%	0.00%
Portfolio characteristics	α	R_C	D_C	M^2_C	
	0.59	4.37%	3	8.72	

4.3 *Using fuzzy programming to select fixed income portfolios with a fuzzy planning horizon and with return and immunising risk objective*

As we pointed out in the subsection 4.1, if the investor is willing to renounce a perfect matching of the duration with H to obtain a higher IRR or if the planning horizon is ill-defined, the exact matching of the duration with the planning horizon is not realistic. In this section we will generalise programmes [P5a] and [P5b] by introducing the possibility of quantifying the planning horizon with a triangular fuzzy number $\tilde{H} = (H_1, H_2, H_3)$. Membership functions (1) and (2) will therefore define to what extend the immunising portfolio is met when the portfolio duration is respectively, above or below the preferred planning horizon, H_2.

The decision-maker should express his/her satisfaction with the IRR of a portfolio via a membership function that is analogous with (3). The reference IRR for r_{min}, may correspond to that of an immunised portfolio with the minimum dispersion i.e. the one obtained with the optimum in [P2]-, or if the investor is more ambitious, the one obtained with [P5b]. For r_{max}, we can take the value obtained by solving [P1], or [P4b] if a higher profit is sought.

Likewise, we believe that the decision-maker also considers the immunising risk as an objective. He/she should therefore to construct a membership function that is analogous with (4). It seems logical to take as the representative value for the planning horizon the value with the highest level of membership H_2. The reference value for M^2_{max} is obtained by solving [P1] or [P4b] and for M^2_{min} is obtained by solving [P2] or [P5b] for M^2_{min}. Of course, whether one or another reference is taken depends on the decision-maker's risk aversion.

This programme, which we shall call [P6a] is:

$$\text{Max } \alpha = \text{Min } \{\mu_{D_{C_1}}(x), \mu_{D_{C_2}}(x), \mu_{r_C}(x), \mu_{M_C^2}(x)\}$$

subject to:

$$\mu_{M_C^2}(x) \geq \alpha,\ \mu_{D_{C_1}}(x) \geq \alpha,\ \mu_{D_{C_2}}(x) \geq \alpha,\ \mu_{r_C}(x) \geq \alpha,$$

$$\sum_{i=1}^{n} x_i = 1,\ x_i \geq 0 \qquad i=1,2,\dots,n,\ \alpha \in [0,1]$$

This is equivalent to the following programme, which we shall call [P6b]:

$$\text{Max } \alpha$$

subject to:

$$\sum_{i=1}^{n} x_i D_i r_i - (r_{max} - r_{min})\alpha \sum_{i=1}^{n} x_i D_i - r_{min} \sum_{i=1}^{n} x_i D_i \geq 0$$

$$\sum_{i=1}^{n} x_i M_i^2 + (M^2_{max} - M^2_{min})\alpha \leq M^2_{max}$$

$$\sum_{i=1}^{n} x_i D_i - (H_2 - H_1)\alpha \geq H_1,\ \sum_{i=1}^{n} x_i D_i + (H_3 - H_2)\alpha \leq H_3$$

$$\sum_{i=1}^{n} x_i = 1,\ x_i \geq 0\ \ i=1,2,\dots,n,\ \alpha \in [0,\ 1]$$

Table 10 shows the results of the numerical application the additional data of Table 9. Under this assumption the results are similar to [P5b], but as a margin to mismatch the duration with H is allowed, the investor takes a greater proportion of bonds with maturity at 10 years at the expense of securities which have a shorter duration and lower IRR.

Table 9. Additional data for solving [P6b]

Planning horizon	M^2_{min}	M^2_{max}	r_{min}	r_{max}
$\tilde{H} = (2'25, 3, 3'75)$	0.33 [a]	20.65 [b]	4.37% [c]	4.83 % [d]

[a] The value of M_C^2 in [P2]   [b] The value of M_C^2 in [P1]

[c] The value of r_C in [P5b]   [d] The value of r_C in [P3b]

Table 10. Characteristics of the selected portfolio with [P6b]

Bonds	**B1**	**B2**	**B3**	**B4**	**B5**
Percentage	59.01 %	0.00%	15.18%	25.81%	0.00%
Portfolio characteristics	α	R_C	D_C	M^2_C	
	0.41	4.56%	3.44	12.33	

We finally want to point out that is not our aim to analyse detailed the sensitivity of the mathematical programmes proposed in the prior subsections when changing the tolerances in profit, immunising risk or planning horizon. Nevertheless, we feel that this is an interesting question that deserves, at least, some comments. Firstly, let us remark that the sensibility of the objective function (that quantifies the investor's satisfaction) over changes in the parameters (r_{min}, r_{max}, H_1, etc.) is easily reachable by the Kuhn and Tucker's multipliers of the constraints (or the corresponding ones to their algebraic ad-hoc modification if it is necessary). In any case, it is easy to check that when increasing (decreasing) investor's exigency about the accomplishment of the objectives, the value of the investor's satisfaction decreases (increases). Secondly, the sensibility analysis about the final composition of the portfolio when changing the tolerances becomes a harder task because of the well-known technical difficulty of this analysis and moreover we must take into account that it is conditioned to the shape of the TSIR when the investment begins. So, the result of this analysis will depend on the context and then it is not possible to extract any general conclusion .

5 Conclusions

We have shown that we can formalise the passive management of fixed income portfolios more realistically if, in conjunction with traditional concepts, we use fuzzy programming. This avoids some of the limitations of using traditional concepts alone.

Fuzzy subsets are suitable for the decision-making process involved in selecting portfolios because they quantify the level of satisfaction associated with meeting objectives or constraints in a natural way. Obviously, the final level of satisfaction will depend on the permitted deviations of each objective from its preferred value. The more flexible the investor is, the higher his/her level of satisfaction will be.

Fuzzy programming has a number of other advantages. Firstly, like multiple objective programming, it simultaneously integrates more than one criterion into a unique decision-making process. This is clearly more realistic for selecting a fixed income portfolio, and indeed for most other economic phenomena, because several objectives are taken into account when taking the decisions. Secondly, we can express an investor's level of satisfaction with every objective via membership functions. Therefore, although the objectives are heterogeneous variables, they are quantified in terms of homogeneous ones (levels of satisfaction), which range from zero (if the objective is not met satisfactorily at all) to one (if the objective is met optimally).

Moreover, an important limitation of using mathematical programming models to select fixed income portfolios is that the planning horizon must be clearly defined, and this is usually not the case. This may be because an investor is prepared, to a certain extent, to abandon an immunising objective in favour of greater profit, or because the planning horizon is defined too vaguely. Fuzzy programming deals more suitably with this variable. A further advantage of the exposed methodology is that it considers the possible instability of the investor's preferences, which is consistent with evidence from several economic studies.

6 Acknowledgements

The authors wish to thank two anonymous reviewers for their valuable comments.

References

1. Balbás, A. and Ibáñez, A., El riesgo de inmunización y su sensibilidad frente al horizonte planificador del inversor, *Análisis Financiero* **70** (1996) pp. 50-59.
2. Barber, J.R. and Cooper, M.L., Immunization using principal component analysis. *Journal of Portfolio Management* **fall 1996** (1996) pp. 99-105.

3. Bellman, R.E. and Zadeh, L.A., Decisión Making in a fuzzy environment. *Management Sciencies* **17** (1970) pp. B141-B164.
4. Bierwag, G.O., *Duration Analysis. Managing Interest Rate Risk.* (Bellinger, Cambridge (Massachusetts), 1987).
5. Chambers, D.R., Carleton, W.T. and McEnally, R.W., Immunizing default-free bond portfolios with a duration vector. *Journal of Financial and Quantitative Analysis* **23** (1988) pp. 89-103
6. Dahl, H., Meeraus, A. and Zenios, S.A., Some financial optimization models: I risk management. In *Financial Optimization,* ed. by S.A. Zenios (Cambrigde University Press, Cambrigde, 1992) pp. 3-36.
7. Fisher, L. and Weill, R.L., Coping with the risk of interest rate fluctuations: returns to bondholders from naive and optimal strategies. *Journal of Business* **October** (1971) pp. 408-431.
8. Fong, H.G. and Vasicek, O.A., A risk minimizing strategy for portfolio immunization. *The Journal of Finance* **39** (1984), pp. 1541-1546.
9. Kaufmann, A., Gil Aluja, J. and Terceño, A., *Matemática para la economía y la gestión de empresas* (Foro Científico, Barcelona, 1994).
10. Lai, Y.L. and Hwang, C.L., *Fuzzy multiple objective decisión making* (Springer-Verlag, Berlin, 1996).
11. Markowitz, H.M., Portfolio selection. *The Journal of Finance* **7** (1952) pp. 77-81.
12. McCulloch, J.H., The tax-ajusted yield curve. *The Journal of Finance* **30** (1975) pp. 811-829.
13. Meneu, V. and Navarro, E., Una aproximación a la duración generalizada y su aplicación a la inmunización financiera. *Cuadernos Actuariales* **5** (1991) pp. 149-166.
14. Navarro, E. and Nave, J.M., Inmunización dinámica, riesgo de inmunización y costes de transacción. *Análisis Financiero* **64** (1994) pp. 54-65.
15. Reitano, R.R., Non-Parallel yield curve shifts and durational leverage. *Journal of Portfolio Management* **Summer** (1991) pp. 62-67.
16. Zimmermann, H.-J., Description and optimization of fuzzy system. *International journal of general system* **2** (1976), pp. 209-216.
17. Zimmermann, H.-J., *Fuzzy Set Theory and Its Applications* (Kluwer Academic, Dordrecht 1991).
18. Zoido, F.J. and Terriente, J., Modelo de optimización de carteras de renta fija. *Papeles de Economía. Suplementos del Sistema Financiero* **10** (1985) pp. 213-229.

This page is intentionally left blank

THE CAUSALITY BETWEEN INTEREST RATE, EXCHANGE RATE AND STOCK PRICE IN EMERGING MARKETS: THE CASE OF THE JAKARTA STOCK EXCHANGE

JYOTI GUPTA, ALAIN CHEVALIER
ESCP-EAP, 79 avenue de la République, 75011 Paris, France
E-mail: Gupta@escp-eap.net, Chevalie@escp-eap.net

FRAN SAYEKT
University of Jog Jakarta, Indonesia

This paper examines the relationship between the interest rate, exchange rate and stock price in the Jakarta stock exchange. This was felt timely, as the Indonesian economy is under-going difficult times and there are numerous and conflicting reports on the effect of interest rate and exchange rate on the stock market price. The study was conducted for a five year period from 1993 to 1997 which was divided into three sub periods. Depending on the sub periods being considered, sporadic unidirectional causality from closing stock prices to interest rates and vice versa and weak unidirectional causality from exchange rate to stock price were found. The overall evidence, however, failed to establish any consistent causality relationships between any of the economic variables under study. Hence it seems that Jakarta market efficiently incorporated much of the interest rate and exchange rate information in its price changes at closing stock market index. These results can be used as a measure of stock market efficiency, however with caution, as there are many other dimensions that have to be studied before arriving at any definite conclusion about the efficiency.

1 Introduction

The stock markets are becoming an integral part of the economies of many countries. With the introduction of free and open economic policies and advanced technologies, investors are finding easy access to stock markets around the world. The fact that stock market indices have become an indication of the health of the economy of a country indicates the importance of stock markets. This increasing importance of the stock market has motivated the formulation of many theories to describe the working of the stock markets.

One piece of information which arrives quite often to the stock markets are interest rates and exchange rate fluctuations. In theory the interest rates and the stock price have a negative correlation. This is because a rise in the interest rate reduces the present value of future dividends income which should depress stock prices. Conversely, low interest rates result in a lower opportunity cost of borrowing. Lower interest rates stimulate investments and economic activities which would cause prices to rise. On the other hand according to the parity conditions the interest rates and the exchange prices should be related with a negative coefficient. Hence we would expect a relationship between exchange price and stock price with a positive coefficient.

However the empirical studies carried out in the various markets had revealed conflicting results on causality between stock prices and the above economic variables. Mok [5], verified the causality of daily interest rate, exchange rate and stock prices in Hong Kong for the period 1986 to 1991. The results indicate, that the HIBOR (Hong Kong Inter Bank Offered Rate) and the price indices are independent series. As a further extension to the study the relationship between exchange rate and stock price was examined, the research concluded that those series are also independent.

Hashemzadeh and Taylor [3] have found bi-directional causality present in regression models between money supply and stock returns using stock indexes to estimate market returns. As regards, the interest rate the results are not as conclusive. The direction of causality seems to be mostly running from interests rate to stock price but not the other way. Solnik [8] found a weak positive relation between real stock return differential and the changes in the real exchange rate and he also found that a real growth in the stock market also has a positive influence on the exchange rate.

Very few studies have been carried out so far to test the causality between the interest rate, exchange rate and the stock prices in the emerging markets. Hence to fulfil this apparent lack of in-depth study in emerging markets the present study was carried out. The objective of this study to is test the existence of causal relationship between stock price movements and interest rates and to verify, whether changes in exchange rates affect the stock prices or vice versa. The results were further used as a surrogate to test the efficiency of the Jakarta Stock Exchange.

The paper is divided into five parts. The second part discusses the methodology, including the model used to test the causality. In the third part we deal with the data used for the analysis, in the fourth section we discuss the results of the tests and ,in the last part we analyse the results of the findings.

2 Methodology

The generally accepted definition of causality is given by Granger and is based on the time series notion of predictability, that is given a set of variables, variable x causes variable y if the present value of y can be predicted more accurately by using only the past value of x than by using all of many combinations of other variables in the set that includes x and y. This paper uses the Granger approach[1].

The methodology adopted here was developed as a combination of the ARIMA approach described by Markridakis [6] and the definition for the direction of

[1] For more information see Granger [9] (pp. 237-352).

causality proposed by Harvey [2]. The ARIMA[2] procedure is designed to fit time series into appropriately selected models which are known as Auto-regressive Integrated Moving Averaged process of order p,d,q or ARIMA (p,d,q).

ARIMA (p,d,q) model represents integration of three parameters indicated by (p,d,q). Here p represents order of autoregressive, d represents differencing used to make the series stationary and q is the order of moving average operator. With these definitions the form of ARIMA (p,d,q) can be written as:

$$(1\text{-}B)(1\text{-}\phi_p B)\ x_t = \mu + (1\text{-}\theta_q B)e_t$$

The use of ARIMA for forecasting involves initial transformation of input series to a stationary series by use of the differencing method. The differencing process can be described by the following equation.

d^{th} order difference = $(1\text{-}B)^d X_t$, where B is the back shift operator[3]. The stationary input series is used to build the ARIMA model. The model building involves determining the order of autoregressive and moving average operators, using auto correlation and partial auto correlation diagrams. The model is used to prewhiten the input and output series before subjecting it to sample cross correlation. The cross correlation between the prewhitened input and output series conveys the information about the existence and direction of the causality, according to the definition proposed by Harvey(1981). According to this definition the theoretical cross-correlation is given by:

$$\Upsilon_{xy}(k) = \rho_{xy}(k) = \frac{(X_t - \mu_x)(Y_t - \mu_y)}{\sqrt{S_X^2 S_Y^2}}$$

the pattern of causality can be classified as follows [2]:

a. $\Upsilon_{\alpha\beta} \neq 0$ for some $\tau > 0$ indicate $x \Rightarrow y$.
b. $\Upsilon_{\alpha\beta} \neq 0$ for some $\tau < 0$ indicate $y \Rightarrow x$
c. $\Upsilon_{\alpha\beta} \neq 0$, for $\tau = 0$ indicates instantaneous causality

where α represents the prewhitened input series and β represents the prewhitened output series, Υ is the cross correlation between them and τ shows the lag.

For the case of $\Upsilon_{\alpha\beta} \neq 0$ at some $\tau > 0$ means that the causality is running from the input series to the output series. Whenever none of these conditions are satisfied we will conclude that there is no causality between the input and output series. The

---

2 For more detail please see Makridakis, S., " Forecasting : Methods and Applications,2nd Edition, New York: Wiley, 1983.

3 The back shift operator (B) is a convenient way to describe the process of differencing. $Bx_t = X_{t-1}$ This was revealed by the pattern observed from the auto correlation and partial auto-correlation diagrams after differencing. Please refer to Markridakis (1982) for more information.

definition of $\Upsilon_{\alpha\beta} \neq 0$ means that the sample cross correlation is significant (i.e. lies out side the two standard error limits).

We decided to use stock prices instead of log of stock prices which corresponds to using of returns on stocks as the basis of testing the causality between the stock market returns and the interest rates and exchange rates. The returns on stock would have been more suitable if we had used daily stock indexes.

3 Data

The major difficulty in any study related to the emerging markets is the availability of relevant data and its reliability. For our study, the following data was gathered from Jakarta stock exchange (JSX) : Daily and weekly closing market index, Overnight inter-bank offered rate and Daily and weekly closing exchange rate between Rupiah and US $. The closing market index is used as the indication of the stock price fluctuation at the JSX, while overnight inter-bank rate and daily closing exchange rate between Rupiah and US $ is taken as the representative of the interest rate and exchange rate fluctuations (Selection of the variables was based on the investors' confidence and usage and the practical considerations such as availability and easy access).

Information about stock prices is provided in the form of Jakarta weakly (Monday to Friday) weighted average stock exchange composite indices for each trading day in the week. Exchange rate information is obtained from the median rate of Rupiah to US Dollar. The Central bank publishes this information daily. The inter bank interest rate for trading and clearing operations between banks in the money market (which is also known as call money) has been used to get the information on interest rates. The data was collected to cover a five year period, from January 1993 till August 1997.

The period under consideration was from January 1993 to December 1997, which represents nearly five years. The period is divided into three sub periods. This was necessary to accommodate and filter the external influences on the market and possibly on the economic variables under study. The first subperiod is from January 1993 to March 22 1994, which was the date when the monetary policy was abandoned. This process of monetary policy abolishment started with the September 1993 Government announcement of a regulation proposing the development of a capital market in Indonesia. The second sub group is the period from March 22 to Dec 1995, in which the interest rate increased sharply. Finally the third sub group is from August 1997 till before the monetary crisis in Indonesia. However the data close to the beginning and end of the sub periods were removed from the analysis to avoid influence from these external incidents.

4 Analysis and results

For each sub period the input series are made stationary. The pre-whitened series are cross correlated and determined for causality and direction.

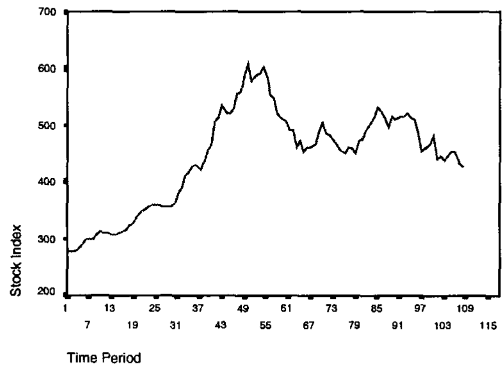


Figure 1. Sub period 1 (January 1993 to March 1995) the stock price.

In this period the stock price index showed an upward trend as illustrated in the figure 1. The trend was a result of increase in activity at the Jakarta Stock Exchange as reflected by the rising number of shares traded each day and the surging index. Analysts believed that this was primarily caused by the Government's effort to encourage the capital market activity and to maintain lower domestic interest rates. But the market had slowed down in the first quarter of 1994. However this decline in trend was dampened by several large listings such as Indosat which reduced the sharp downward trend of the Composite Stock Price Index in the Jakarta Stock Exchange.

In 1993 there were three significant events which might have had an effect on the behaviour of the stock price movement in this period. Firstly, the loosening of the monetary policy at the beginning of the year, secondly in March brokers had uncovered counterfeit shares of five companies in circulation. The third event was the issuing of a decree by Bank of Indonesia authorising the use of shares as supplementary collateral for credit.

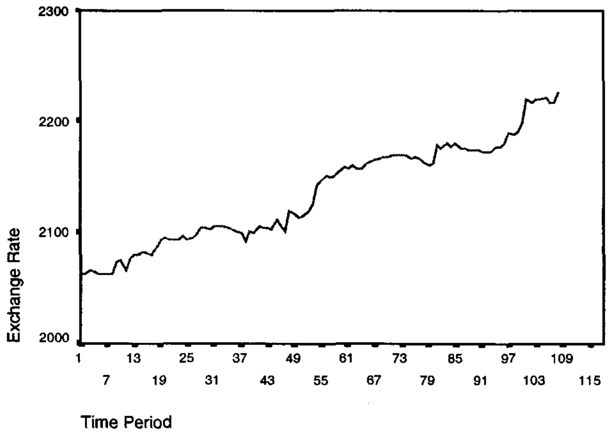


Figure 2. Sub period 1 (January 1993 to March 1995) the stock price and exchange rate.

The graph of raw data of exchange rate series showed an upward trend (fig. 2). The auto-correlation coefficient diagram showed a non-stationary series. The fact that all auto correlation coefficients till lag 16 were significantly different from zero revealed the series was non-stationary. The series was transformed to a stationary series by taking the first order difference. After the first differencing the series was observed to be stationary and basically autoregressive in nature[4]. The same diagrams were used to determine the order of the auto regressive operator and it was found that the series was first order autoregressive. Hence the series was pre-filtered by a model with a constant, a non-seasonal autoregressive operator of p=1 and moving average non seasonal operator of q= 0 which was ARIMA(1,1,0). The pre filtering equation used in prewhitening the input and output series was as follows:

$$\alpha_t = x_t - 0.10199998\ x_{t-1}$$

where α_t is the white noise series, x_t is the transformed input series and the numeric is found from the ARIMA model building and represents first order autoregressive coefficient.

[4] This was revealed by the pattern observed from the auto-correlation and partial auto-correlation diagrams after differencing. Please refer to Markridakis (1982) for more information.

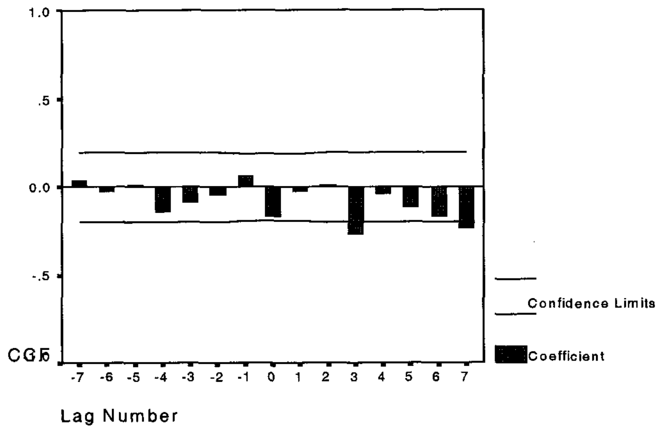


Figure 3. Sample cross correlation diagram between prewhitened exchange rate and stock price.

The Sample Cross Correlation (SCC) is calculated using the prewhitened input and output series. The result of SCC calculation is shown in figure 3. The figure shows that most of the significant coefficients have negative values. This indicates that there was an inverse relationship between exchange rate and stock price series, which means the exchange rate depreciation (appreciation) was associated with a fall (rise) in stock prices.

The Sample Cross Correlation function also shows two significant coefficients at lag = 3 (0.271) and lag = 7 (0.242). The diagram also includes relatively large values at lag =5 (0.149) and lag = 6 (0.165) although they still lie within the two standard error limits. The past lag values include relatively large coefficients while future lag values include insignificant coefficients. According to Harvey's definition [2], this pattern suggests that there was a unidirectional causality running from exchange rate to stock price.

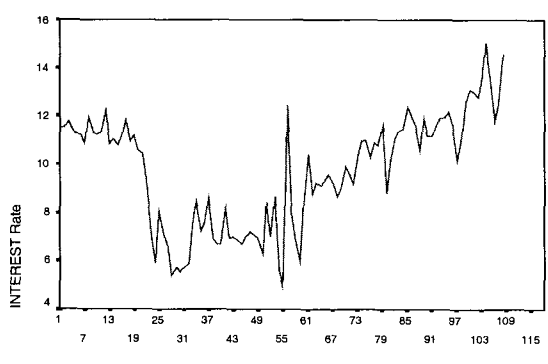


Figure 4. Sub period 1 (January 1993 to March 1995) the stock price and interest rate.

The line graph of interest rate series shows an upward trend after sharp initial decline (fig. 4). The same diagram indicates that the series contains a seasonal component. The auto-correlation diagram and partial auto-correlation diagram confirmed the series was non-stationary. The series was made stationary by taking the first difference.

The examination of the auto correlation and partial auto correlation diagrams of the stationary series resulted in the identification of the input series as basically auto regressive in nature with an order of three. The same diagrams showed a seasonality of five weeks. The initial model selected was subjected to iterative improvements using SPSS software until the parameters, loglikelihood, standard error and correlation between variables showed best fit values. Based on these observations the model selected was ARIMA$(3,1,0)(1,0,0)^5$ Hence the series was pre-filtered by a model with a constant, a seasonal auto-regressive operator of p=3, seasonal auto-regressive operator = 1 and moving average non seasonal and seasonal operator of q= close to zero. The final prefilering equation took the form as follows:

$$\alpha_t=X_t+0.22684096x_{t-1}+0.12977215x_{t-2}+0.09367329x_{t-4}-0.02124894x_{t-5}-0.01215618x_{t-6}$$

The Sample Cross Correlation was calculated and the resulting graph is shown in figure 5. By examining figure 5 we can see that the SCC coefficients are mostly negative. This supports the theoretical belief of inverse relationship between the stock price index and interest rates. Hence the increase (decrease) of interest rate will be associated with fall (rise) in the stock price index. However the fact that all the coefficients except one lag lie within the error limits indicates that the relationship was not strong.

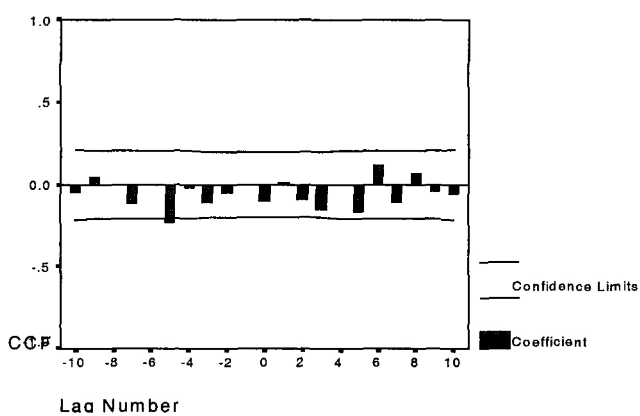


Figure 5. Cross correlation diagram between prewhitened interest rate and prewhitened stock index.

The cross correlation coefficients between prewhitened interest rate and the stock index (figure 5) was only significant at $\tau = -5$. This suggests that even though there was one significant cross correlation coefficient, as a group there was no causality relationship between the interest rate and stock price. Hence the hypothesis of inverse relationship between the interest rate and the stock price has to be dropped for this period. The sporadic causality is the sign of a puzzling relationship running from stock price to interest rate.

In the first quarter of 1994 the Indonesian capital market started slowing down after a bullish trend till the end of 1993. The main reason attributed for this declining trend was the selling spree by foreign investors, because of the Mexican crisis. The analysts believed that this declining of the stock price index caused an increase in inter bank interest rate to 11 % in September 1994 and to 13 % in March 1995. This scenario might explain the puzzling causality relationship indicated by the significant SCC coefficient at lag $\tau = -5$.

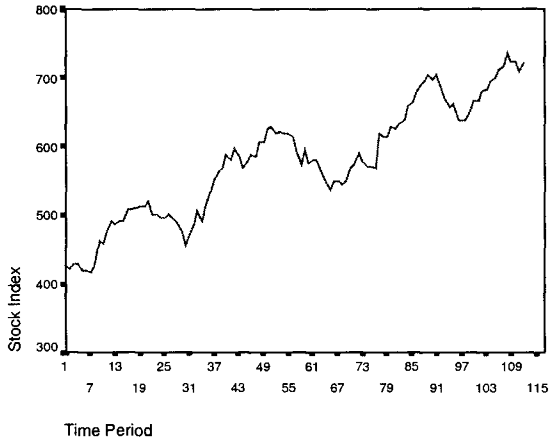


Figure 6. Sub period 2 (April 1995 to July 1997) the stock price.

In this period the capital market showed a bullish trend (figure 6). Market capitalisation and the share trading volume was also on the upward trend. This favourable development was due to the strengthening of price of some blue-chip stocks in the stock market, including PT Telkom, PT Timah and several cigarette and finance companies.

Despite the favourable conditions in the stock market in Dec 1995, The Composite Stock Index plunged below the psychological limit of 500 points. This was believed to be mainly due to the selling of shares by foreign investors as a result of weak trading activity. However at the beginning of 1996 the stock market quickly recovered. The bullish capital market saw The Composite Stock Price Index increase to 623.9 points in April 1996.

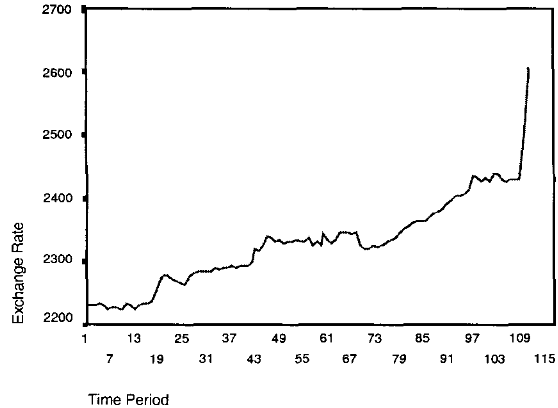


Figure 7. Sub period 2 (April 1995 to July 1997) the stock price and exchange rate.

Rising capital inflows encouraged appreciation of the Rupiah against some major currencies during 1996, reflecting the market's confidence in the fundamentals of the Indonesian economy. Another factor which stimulated capital inflows was the attractive interest rate differentials, despite some declines.

Rupiah exchange rate movements, however were not only affected by developments in fundamentals, but also by rumours. Despite temporary upheavals, Bank Indonesia refrained from market intervention since the Rupiah fluctuation was short lived and did not exceed the limits of the intervention band.

The graph of exchange rate series shows an upward trend with a sudden jump at the latter stages of the period (fig. 7). Examination of the auto correlation and partial auto correlation diagrams confirmed the non stationary characteristics observed in the line diagram. First order difference was taken to remove the non stationarity of the series. The auto- correlation and partial autocorrelation diagrams after first differencing revealed that the series had become stationary. Hence the differencing process was terminated after the first differencing and the resulting series was used to build the ARIMA model.

These diagrams revealed that the series was basically non seasonal and auto regressive in nature with a first order autoregressive operator. Hence the series was pre filtered by an ARIMA model with a constant, a non seasonal auto regressive operator of p=1 and moving average non seasonal operator of q= 0 which was ARIMA(*1,1,0*). The final version of the prefiltering equation was found to be :

$$\alpha_t = x_t + 0.8070418\ x_{t-1}$$

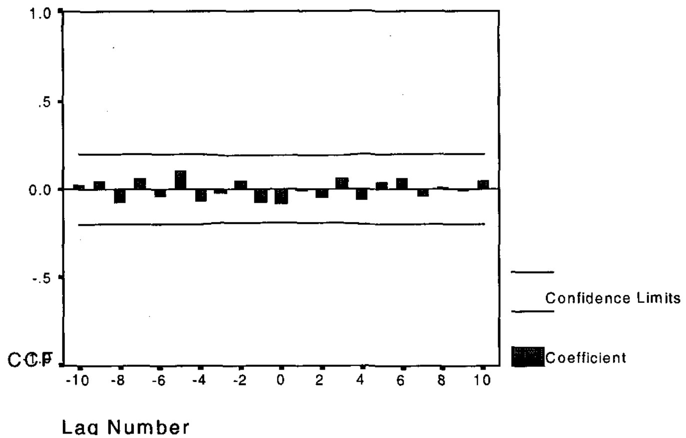


Figure 8. Cross correlation diagram between prewhitened exchange rate and prewhitened stock index.

The sample cross correlation of the prewhitened exchange rate series and the stock price series (fig 8) are insignificant at all lags. This means that stock price series and exchange rate series in this period were independent and historical information of exchange rate was not significantly predictive for changes in stock prices.

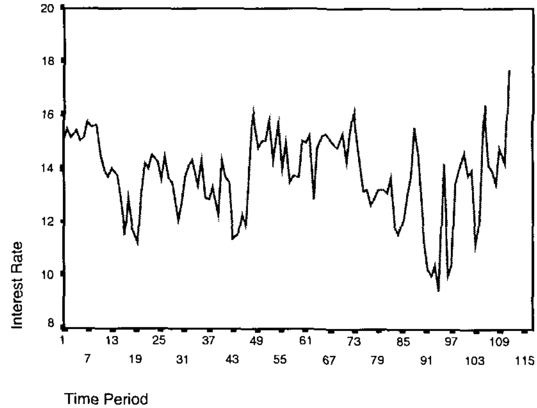


Figure 8. Sub period 2 (April 1995 to July 1997) the stock price and interest rate (in this period interest rates were generally on an upward trend).

The Rupiah liquidity of commercial banks declined in the reporting year. Total bank reserves, as a percentage of current liabilities, and liquid assets as a percentage of total deposits generally declined, encouraging the bank to adopt a tighter monetary policy. The tighter monetary policy in turn brought about higher transaction volumes and pushed up Inter bank rates from 13.1 % in the previous year to 13 .9 % (figure 8)

The series showed non-stationary characteristics and this was confirmed by the auto correlation and partial auto correlation diagrams. The series was transformed using first order differencing. An examination of the autocorrelation and partial autocorrelation diagrams of the stationary series revealed that the series had a seasonality component with a cycle time of 4 weeks. Based on the autocorrelation, and partial autocorrelation graphs of the stationary series, the appropriate model for the interest rate is estimated by a constant, non seasonal auto regressive operator p = 1 and seasonal moving average operator of Q = 1. Hence ARIMA(1,1,0)(0,0,1)^4 model was used to prewhiten the interest rate series and the stock price series. The final form of the pre-filtering equation can be rewritten as follows:

$$\alpha_t = x_t + 0.27185015\, x_{t-1} + 0.18058151\, \alpha_{t-4}$$

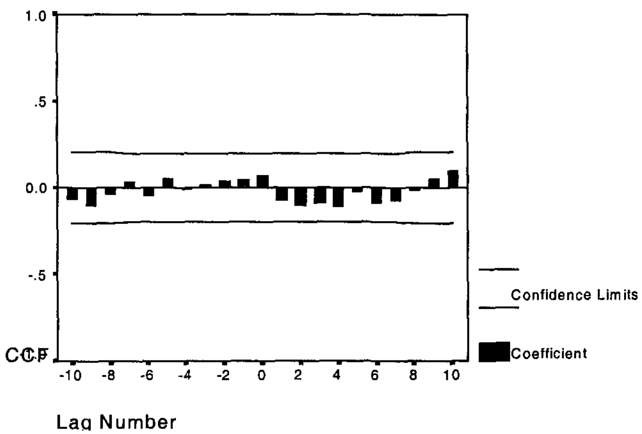


Figure 9. Cross correlation diagram between prewhitened interest rate and prewhitened stock index.

The sample cross correlation between prewhitened interest rates series and the stock indices (figure 9) resulted in insignificant coefficients on all lags, in the past and future values. The values are small and lie within the boundaries of the two standard error terms. However figure 9 contains more negative coefficients than positive ones. This feature supports the theoretical negative relation between stock price and interest rate. But these coefficients lie within the two standard error limits indicating that they are insignificant. Hence we can conclude that there were no significant causal relationships between interest rate and stock price within the sub period under study.

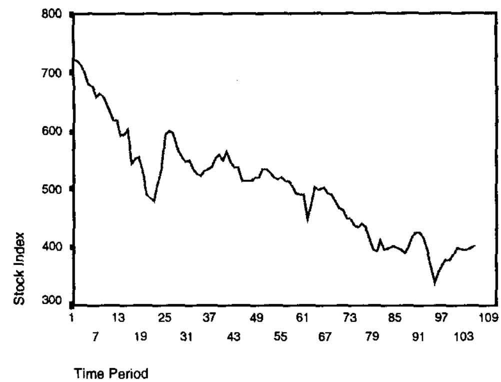


Figure 10. Sub period 3 (August 97 to December 97) the stock price.

Starting from July 1997, the stock price index has shown a downward trend. This period was the beginning of the economic crisis in Indonesia. Parallel to the exchange rate crisis the stock market showed a sharp down-ward trend, mainly due to the loss of investors' confidence. This was basically due to the weakening foreign reserves and the rising balance of payment figures. However the severity of the stock market crisis was fuelled by the rumours about the lack of stability of the government and the intended intervention measures.

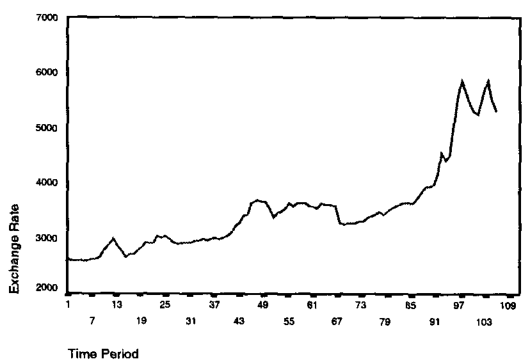


Figure 11. Sub period 3 (August 97 to December 97) the stock price and exchange rate.

The crisis was initiated by the sharp depreciation of the Rupiah against the US Dollar exchange rate, which forced the government to relax the flotation band of the exchange rate. However the currency crisis continued to worsen and on August 14, 1997 Bank of Indonesia was left with no other option than to abolish interventions in the money market in order to manage the exchange rate. In other words from this day Indonesia started adopting the flotation of Rupiah. The severity of the problem persisted and in the midst of the crisis the Rupiah had depreciated almost 100% against the US dollar.

The autocorrelation and partial autocorrelation diagrams of the exchange rate series proved to be nonstationary, as there were significant values in most of the lags and these decreased gradually to zero as the number of lags increased. The nonstationary data series was transformed into stationary by taking the first difference. The examination of autocorrelation and partial autocorrelation diagrams of the stationary series revealed seasonality of a 4 days' cycle. It also revealed that the non seasonal component is basically autoregressive in nature with an order of two and the seasonality part is a moving average with an order one. Hence the series was pre filtered by a model with a constant, a non seasonal autoregressive operator of p=2 and seasonal moving average operator of q=1 which was ARIMA *(2,1,0)(0,0,1)*^{4}. The final version of the prefiltering equation was found to be:

$$\alpha_t = x_t - 0.468721\ x_{t-1} + 0.294241\ x_{t-2} + 0.033249\ \alpha_{t-4}$$

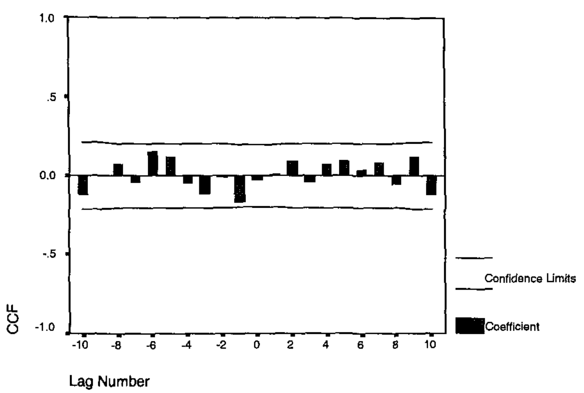


Figure 12. Cross correlation diagram between prewhitened exchange rate and prewhitened stock index.

The sample cross correlation diagram (figure 12) between prewhitened exchange rate and stock price index resulted in coefficients of both positive and negative values. This suggests that there is no significant negative or positive correlation between exchange rate and stock price index in this period. Figure 12 shows insignificant correlation coefficients at all lags. The values are small and lie within the boundaries of two standard error. The pattern suggests that stock price and exchange rate series in this period were independent series and the historical information of exchange rate was not significantly predictive for changes in stock prices. This finding does not support the findings of Solnik [8] which stated a significant correlation between exchange rate and stock price index.

Many recent newspaper and journal articles have suggested that the Jakarta Stock Index dropped due to the currency crisis in the money market. However a closer look at the situation suggests that the stock price decline was a result of the events triggered by the currency crisis rather than by the currency crisis itself.

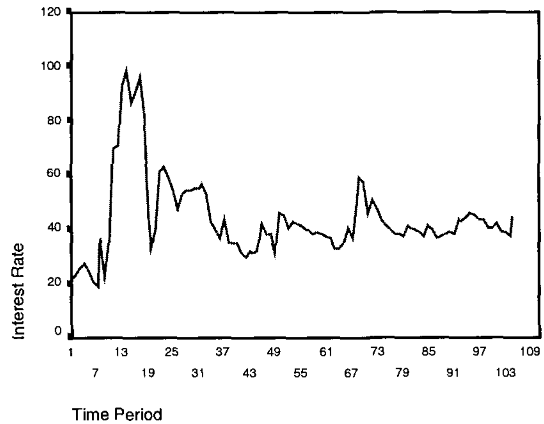


Figure 13. Sub period 3 (August 97 to December 97) the stock price and exchange rate.

In order to overcome the crisis condition, the Government initiated a tight money policy by increasing the interest rate (in the beginning) for the government bonds (SBI) up to 30 %. Which caused the inter bank interest rate to rise as high as 94% within the study period. This measure was taken to protect the money market in Indonesia. This resulted in the steep rise of the interest rate as well as frequent fluctuations. Figure 13 clearly shows these events.

In order to remove the nonstationarity the data series was transformed using first differencing process. The autocorrelation and partial autocorrelation diagrams after first differencing showed that there was seasonality with a cycle time of 7 days. The seasonality seems to be roughly autoregressive in nature. However this fact was confirmed by repetitive processing of different ARIMA models with a changing seasonality component. The process reveals that the seasonality component was in fact autoregressive in nature with an operator of first order. It was found that the nonseasonal component is also autoregressive in nature with a second order. Hence the final ARIMA model selected for prewhitening process was ARIMA$(2,1,0)(1,0,0)^7$. Based on these the final prewhithening equation took the form:

$$\alpha_t = x_t - 0.13273408\, x_{t-1} + 0.12903496\, x_{t-2} + 0.112461 x_{t-7} - 0.01476217 x_{t-8} + 0.01435076 x_{t-9}$$

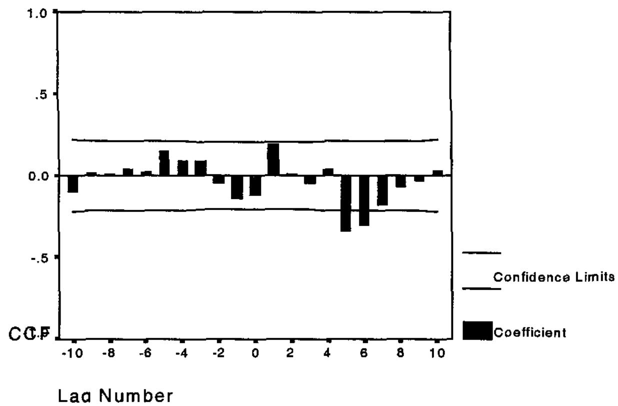


Figure 14. The cross correlation diagram between prewhitened interests rate and stock index.

The sample cross correlation between prewhitened interest rates series and the stock indices (fig 14) contains more negative coefficients than positive ones. This feature supports the theoretical negative relation between stock price and interest rate. The figure also shows that more than half of the coefficients have negative values and two of them are significantly larger than zero. This suggests the existence of a negative relationship between the interest rate and stock price, as one would expect.

The two significant spikes at lags $\tau = 5$ and 6 indicate the direction of this relationship. Since these spikes lie in the past lag values, according to our definition the direction of causality is running from interest rate to stock price.

At the beginning of August 1997, The Government introduced a tight money policy in order to overcome the crisis caused by the depreciation of the Rupiah. The government increased the government bond rate (SBI) to 30 %. This increase in the SBI may have caused the subsequent increase in other interest rates in the Indonesian capital market, including inter bank interest rates with peaks to a maximum of 125 % for the overnight rate.

5 Conclusions

The research findings show that in most of the instances there were no strong causal relationship between stock price and the interest rate or exchange rate. These findings are summarised in the table below

Table 1. The summary of research findings

The sub period	Variables	Existence of causality	Direction of causality
Sub period I	Exchange rate Vs Stock price	Yes	Unidirectional from exchange rate to stock price
	Interest rate Vs Stock price	Yes (weak)	Unidirectional from stock price to interest rate
Sub period II	Exchange rate Vs Stock price	No	-
	Interest rate Vs Stock price	No	-
Sub period III	Exchange rate Vs Stock price	No	-
	Interest rate Vs Stock price	Yes	Unidirectional from Interest rate to stock price

The above table illustrates that only in sub period one and three we have observed causality relationships between the variables under study. The other interesting finding is that the interest rates trend to have more causality relationship with the stock price rather than exchange rate. However findings fail to indicate strong causality relationship consistently running between the variables during the whole period of study. Hence we can safely conclude that this causality evidence fails to establish that historical information generally offers sufficient significant a significant short-term predictive content for stock price.

In general it seems that the exchange rate and the stock price series are generally quite independent as a group. There was, however a possible unidirectional causality running from exchange rate to the stock price in the sub period one. The large number of negative cross correlation coefficients suggests the existence of a positive correlative relationship between exchange rate and stock price, meaning that when the Rupiah devalues against the Dollar, the stock price also declines, and vice versa. This may be due to the liquidity pressure bought in by Rupiah devaluation which encourages firms to release some stocks in the market to encounter the short term liquidity problems. It is a generally accepted fact that Indonesian firms held very high amount of Dollar debts and had to service them in Dollar, making their liquidity position quite sensitive to the Dollar to Rupiah exchange rate fluctuations.

These results are in line with the findings from a similar study conducted in Hong Kong for the period of 1986 to 1991 (five years) using daily market data on exchange rate, interest rate and stock price using ARIMA and Granger and Sim test. In conclusion the study reveals that interest rate and stock price are independent series for most of the time even though there is indication of some causality relationship running from stock price to the interest rates. The results also confirm that exchange rate and stock price are strongly independent series.

There were however other studies which found contradicting results. In the United States Hashemzadeh and Taylor [3] using a Granger-Sims test on weekly data for the period 1980 to 1986, causality found to run mostly from the interest rates to stock price changes but not vice versa. Cheung et al [1], using monthly data for the period 1984 - 1989, identified a "puzzling" unidirectional causality running from stock price to interest rate with no feedback in Hong Kong.

However the stock price is influenced by numerous factors so that predictions using selected variables may give incorrect forecasting. Many researchers and mathematicians have striven hard to build models which incorporate a diverse array of variables to predict the stock market price but have failed. Stock markets are very complex and organic phenomena. Hence finding a pattern in stock price is quite difficult when using a limited amount of data. These studies need a huge amount of data on numerous variables. In this context, it should be of no surprise that the research study has not been able to reveal any significant pattern using five years of data. The other major criticism is the selection of the time period. Stock markets are also influenced by economic cycles as some analysts believe, hence in our study, the selection of time period may have affected the results.

The selection of the variables could also be a concern. There are many interest rates in use and hence nobody can be sure which interest rate the investors are looking at. It may be that the government bond rate is of more interest to the investors. However the government bond rate in Indonesia is not frequently changed and hence its usefulness as a time series variable is questionable. For the exchange rate also there is no composite index representing all the fluctuations in the currency market. Hence as the next best alternative we have selected the Rupiah to US $ exchange rate.

The findings of this research may be used as a surrogate measure to test the efficiency of Jakarta Stock Market. The research findings failed to establish any significant causality relationship between the economic variables. This indicates that past information does not have a significant information content to be used as a predictor of stock price. Hence the findings reveal that Jakarta Stock Market is relatively efficient in this regard.. It should be noted here that the efficiency of the stock market cannot be judged based only on causality (which indicates the use of past information to determine the stock price) as there are many other factors which should be considered before concluding about the level of efficiency of the stock market.

Some of the possible further research in line with the present study is listed below.

- The present research can be extended to cover large period under study and can be verified using other multivariate statistical forecasting models.
- Some sectors of the stock market can be more sensitive to the exchange rate and interest rate fluctuations than composite stock index. Hence further research to find out causality between select the sectoral Indexes, interest rate and exchange rates may reveal some pattern.
- In this study we only looked at the linear relationships, the study could be extended to include non-linear relationships between the variables.

- An alternative methodological approach could also be used: to estimate a VAR with all three variables and then use the Granger causality test.

Acknowledgements

The authors would like to thank the anonymous referees for their helpful comments.

References

1. Cheung Y. L and Ho Y. K., The intemporaly stability of the relationship between the Asian Emerging markets and developed equity markets, *Journal Of Business Finance and Accounting* **18** (1991), pp. 235-254.
2. Harvey A., *The Econometric Analysis of Time Series* (Philip Allan, 1981).
3. Hashemdah N. and Taylor P., Stock prices, money supply and interest rates: The question of causality, *Applied economics* **20** (1988), pp. 1603-1611.
4. Kanas A., Volatility spillovers between stock returns and exchange rate changes: International evidence, *Journal of Business Finance and Accounting* (2000), April/ May.
5. Mok H. M. K., Causality of interest rate, exchange rate and stock prices at stock market open and close in Hong Kong, *Asia Pacific Journal Of Management* **10** (1993), pp. 123-143.
6. Makridakis S., *Forecasting: Methods and Applications*,2nd edition (Wiley, New York, 1983).
7. Rogalski R., New findings regarding day of the week on stock returns, *Journal of Finance* **39** (1984), pp. 1603-1614.
8. Solnik B., Using financial prices to test exchange rate models: A note, *Journal of Finance* **42** (1987), pp. 141-149.
9. Granger C. W. J., Generating and Mechanism models and Causality. In *Advanced of Econometrics*, ed. by W. Hildenbrand (Cambridge University, Oxford, 1982) pp. 237-352.
10. Sorensen E. H., Rational expectation and the impact of money upon stock prices, *Journal of Financial and Quantitative Analysis* **XVII**.

This page is intentionally left blank

FUZZY COGNITIVE MAPS IN STOCK MARKET

KOULOURIOTIS D.E., DIAKOULAKIS I.E., EMIRIS D.M.
Department of Production Engineering and Management
Technical University of Crete, Chania, 73100, Greece
E-mails:{ jimk, dgiannis, emiris}@dpem.tuc.gr

Stock price forecasting constitutes a challenging research area. Diverse schemes (such as regression models, neural networks, neurofuzzy systems etc.) have been developed and applied; yet, the overall prediction behavior of such systems is questionable in real world conditions. The major reason limiting the accurate stock price predictions is the existence of a plethora of interrelated agents (quantitative and qualitative) affecting stock price movements and fluctuations. Fuzzy Cognitive Maps (FCMs) seem to constitute a useful modeling tool for the development of a forecasting model, which takes into account the characteristics of the stock market. The main purpose of this work is thus to present analytically the FCM operation mode and the potential extensions of the underlying inference mechanism, and to describe possible applications of FCMs in the domain of stock market.

1 Introduction

Stock market has been a research area during the recent decades, with particular effort being observed only during the last decade, due to the enhancement of information technology and the continuous improvement of modeling tools. The main topics that have been extensively analyzed are: (1) the portfolio selection principles, (2) the analysis and evaluation of stock market and economic factors effect on stock price movements and fluctuations, (3) the research for methodologies that discover and exploit statistically significant stock price patterns, and (4) the development of systems for forecasting purposes. The main restrictions in the aforementioned research directions are the numerous factors affecting investors' behavior, the interaction and interrelations among these agents, the weakness of an accurate estimation of their effect on stock prices and the deficient integration and implementation of highly sophisticated techniques. The factors that affect and determine investors' behavior belong to domains that are not only related directly to stock market concepts but also to macroeconomic and business matters, as well as national and international political affairs. In addition, the way in which these factors act usually changes; therefore, the development of deterministic systems for stock price predictions remains extremely difficult.

The complexity of the problem at hand renders the *technical analysis* the most widely accepted technique, as its inference mechanism depends on users' (analysts, investors, researchers) subjective interpretation, knowledge and opinion. Users' knowledge and instinct, somehow, substitutes the necessity for estimation, in qualitative terms, of the impact of various factors; nevertheless, the lack of systematic knowledge, substantiated reasoning and robust inference constitute critical drawbacks of the theoretical and practical framework.

The necessity for accurate stock price predictions triggered the conduction of several studies during the last decade. The main contribution of these studies is that they set the bases for a systematic approach of the existing problem, combining and applying advanced methodologies and exploited past information in order to achieve a satisfactory performance of the developed prediction models.

This paper attempts to further advance the research in the domain of stock market prediction models. For this purpose, initially, the most significant works on the subject were identified and analyzed. Literature review helps for a better understanding of stock market behavior-characteristics and the difficulties that arise trying to predict stock price movements. The results and conclusions of the literature review, led to the search of an appropriate method for a forecasting model development. Fuzzy Cognitive Maps (FCMs) were selected, as they constitute a new, yet attractive, approach that encompasses extremely useful modeling features and appears to be a powerful platform for the development of a stock-price forecasting model. At the same time, in order to demonstrate the modeling capabilities of FCMs, potential improvements were investigated, the most important of which is the description of an algorithm managing multi-stimulus situations.

The development of an FCM-based stock market model is accompanied by the enumeration of a series of stock market factors along with an approximation of the market structure and its operation mode. Finally, an FCM based stock price forecasting system is proposed and existing problems are discussed.

The remaining of this article is organized as follows: Section 2 presents the literature review that comprises a brief reference of the conclusions drawn by previously published works on the development of stock market prediction models. The methodology of Fuzzy Cognitive Maps and the improvements of FCM inference mechanism are presented in Section 3, while in Section 4, the basic parameters affecting stock prices are listed and a possible application of the FCM methodology using these parameters is described. Finally, in Section 5, future research topics related to the implementation of FCMs in stock market are discussed.

2 Literature review

The development of stock price prediction models has drawn the interest of numerous researchers who produced a large number of worth mentioning works on the subject. In general, four major categories of such forecasting systems exist:

1. technical analysis based systems;
2. forecasting systems derived from typical linear, non-linear methodologies and intelligent techniques;
3. hybrid systems, mainly genetic, neural and fuzzy combinatory approaches; and

4. a distinct and very promising cluster of forecasting systems that enable utilization of qualitative information.

Technical analysis based systems [LL97, LCC95, SDK92] utilize analysts' experience in order to extract patterns that have been recognized in the past and formulate indicators, thus enabling the design and implementation of trading strategies on them. These systems are usually rule-based, incorporating simultaneously fuzzy reasoning mechanisms. Technical analysis indicators are mostly moving averages and oscillators, price and volume indices, support/resistance and trend lines.

The second category encompasses linear models (regression and ARMA), exponential smoothing techniques, and diverse types of neural networks (feed-forward, recurrent, probabilistic, etc) [Bas95, BM91, HH95, JL93, KAY90, KC98, KED01, KGW93, KW97, LDC00, MW00, Phi96, Pod98, RZF94, SD96, SPW98]. These techniques have been used to construct stock trading strategies, to forecast absolute stock prices or/and trends, or even to approximate price volatility. The input data-set usually includes stock market indices, macroeconomic variables and financial measures. Despite the improvement of the attained profits obtained by models of this category, the trading risk remains at significantly high levels.

The third category of stock price forecasting models concerns hybrid systems, like neuro-fuzzy and rule-based networks [CYC96, WWG92, ZL93, ZXM97]. The achievements of hybrid techniques compared to linear and non-linear models are indisputably far more satisfying, even though significant weaknesses remain.

Finally, forecasting systems exploiting qualitative information rather than using as input strictly quantitative variables constitute the fourth category [KIF97, KLL96, YS91]. The plethora of practical difficulties managing qualitative information and especially the weakness of modern techniques (linear models, neural nets, hybrids etc.) to provide the appropriate capabilities for handling these set restrictions to such an approach.

The survey of Diakoulakis et al. [DKE01] adequately analyzed and extracted the basic points of the most important studies in the domain of stock market forecasting. A first conclusion that was drawn is that the forecasting ability and accuracy is related to the forecasting horizon; long-term predictions are much more effective than short-term ones. The main reason for this distinction is that the factors affecting stock prices in a long-term basis are only a few and easily recognized macroeconomic and business parameters, while in a short-term basis, investors' behavior is influenced by numerous temporary agents. A second conclusion was that the produced results in the case of forecasting of combined indices were significantly better than in cases where single stocks were examined.

As far as the data used are concerned, most of these are quantitative variables related to macroeconomic concepts and financial ratios. Despite the fact that qualitative data were used only in a few papers, the importance of this kind of information proved to be valuable. Overall, in order to accomplish better (or the best) results, the use of qualitative data constitutes a promising direction.

Additionally, the intelligent methods (neural networks, hybrid systems etc.) proved to include better predictive capabilities than their counterparts (multiple regression, ARMA etc.); however, the application of advanced techniques didn't prove completely satisfactory. This conclusion enforces the argument for searching new forecasting techniques and modeling approaches.

To summarize, according to the findings of the published works/systems, future research should focus on the short-term prediction of single stock prices; moreover, the need for a systematic search, representation, storage and handling of business and economy data is imperative. Through this information, all the factors affecting stock prices can be determined and assessed. Finally, an innovative technique able to manage diverse (quantitative and qualitative) interrelated variables is necessary.

The above main conclusions set the prerequisites for the development of an effective stock-price forecasting model.

3 Method selection

One of the conclusions of the literature review was the necessity for a method that would provide the means for handling qualitative data and representing relationships among various variables. In the work of Lee and Kim [LK97], a new approach, based on the theory of Fuzzy Cognitive Maps, was presented. Although their work focused on the further development and establishment of the FCM theory, an innovative idea concerning stock price forecasting was presented. The idea of Lee and Kim constitutes the motive for the conduction of this study.

3.1 Cognitive Maps

Cognitive Maps (CMs) were first presented by Axelrod [Axe76]. Their early use mainly concerns cognitive representation in political and social sciences. Recently, many studies have been conducted both for the enhancement of the CM (and the FCM) principles and the presentation of innovative applications using them [Bru96, BWH89, KL98, Kos86, LA87, ML00, SG98, SSK98, SM88, Tab91, Tab94, TMM95, TS87, Yea88, ZC88, ZC89]. Admittedly, there is a lack of a stable theoretical basis that becomes easily recognized by the various differences among the published papers discussing CMs and FCMs.

The scope of CMs is the synthesis and analysis of systems described by diverse interrelated concepts. The basic function of a CM is the graphical and mathematical representation of cause-effect relationships (measured by causality) that are perceived to exist among specific concepts of a given system. Cognitive maps provide information about the effect of the state (or the change of the state) of some elements on the state of the total elements of given systems.

As far as its structure is concerned, a cognitive map is a network where the nodes represent the concepts and the links represent the cause-effect relationships between the concepts of the given system. An example of a CM is depicted in Figure 1.

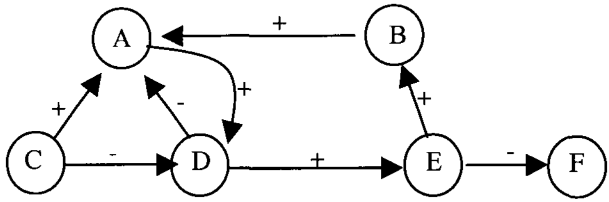


Figure 1. Cognitive Map

In CMs, links between nodes may obtain only two values, +1 (or simply +) and -1 (or simply -). In addition, nodes may obtain the values -1, 0 and +1. A value of +1 in a node indicates the increase or improvement of the concept state that the node represents while a value of -1 indicates the decrease of the corresponding concept state. The value +1 in the link between nodes A and D (Figure 1), for example, means that: (a) increase of concept A state causes increase of concept D state, and (b) decrease of concept A state causes decrease of concept D state. On the other hand, the value -1 in the link between nodes C and D (Figure 1) means that: (a) increase of concept C state causes decrease of concept D state and (b) decrease of concept C state causes increase of concept D state. It is important to emphasize that cause-effect relationships are only these depicted with the links on the CM, with a specific direction at all times.

3.2 Fuzzy Cognitive Maps

Fuzzy Cognitive Maps (FCMs) constitute a functional extension of cognitive maps. This approach was developed in order to avoid the binary logic that CMs enclose. FCMs emanate from the combination of CM theory and Fuzzy Logic principles and provide a more reliable and effective representation of the real world systems than CMs do. Nodes and links in an FCM obtain values in the interval [-1,1]. Positive and negative values have the same meaning as that described in Section 3.1; therefore, the advantage of FCMs is their ability to apply a wide range of values on their nodes and links overcoming the constraints of binary logic.

As far as the nodes are concerned, if the value of a node is positive and very high (e.g. +0.9) then the ensuing conclusion is that the state of the corresponding concept increased or improved significantly, while, if the node value is positive and relatively low (e.g. +0.2) then the increase or improvement of the concept state is relatively small. Respectively, a high (absolute) negative node value (e.g. -0.9) means a significant decrease in a concept state, while a relatively low (absolute) negative node value (e.g. -0.2) means that the decrease in the corresponding concept state is relatively small.

As far as the links are concerned, high (absolute) values signify strong cause-effect relationships among the concepts. For instance, in Figure 1, relationship from

C to A may be less strong (in absolute values) than the relationship from D to A. In the case of a medium positive change of concept C state, the state of concept A will change positively only a little, while in the case of a medium positive change of concept D state, the state of concept A will change negatively at a high degree. The necessity for the use of different (absolute) values for the links among the concepts is thus obvious. This modification is a prerequisite for an accurate approach of any given system. In case of such a modification, the CM depicted in Figure 1 may turn to the FCM depicted in Figure 2.

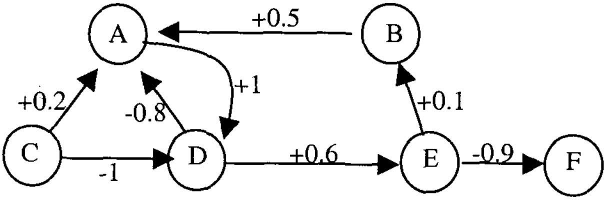


Figure 2. Fuzzy Cognitive Map

3.3 Inference Engine

FCM inference engine in its simplest form constitutes a mathematical matrix manipulation and is based on the *concept vector* and the *adjacency matrix*. The concept vector includes as many elements as the number of concepts in the given system. The values of the vector elements represent the concept states. The adjacency matrix is used for the mathematical representation of cause-effect relationships among concepts. Its structure for the FCM in Figure 2 is:

		To					
		A	**B**	**C**	**D**	**E**	**F**
From	**A**	0	0	0	+1	0	0
	B	+0.5	0	0	0	0	0
	C	+0.2	0	0	-1	0	0
	D	-0.8	0	0	0	+0.6	0
	E	0	+0.1	0	0	0	-0.9
	F	0	0	0	0	0	0

Assume a change (stimulation) of concept C state with a value +0.8. Firstly, the initial values of concept vector elements are determined and the concept vector becomes [0 0 +0.8 0 0 0]; it is then multiplied with the adjacency matrix to yield:

$$\begin{bmatrix} 0 \\ 0 \\ +0.8 \\ 0 \\ 0 \\ 0 \end{bmatrix}^T \times \begin{bmatrix} 0 & 0 & 0 & +1 & 0 & 0 \\ +0.5 & 0 & 0 & 0 & 0 & 0 \\ +0.2 & 0 & 0 & -1 & 0 & 0 \\ -0.8 & 0 & 0 & 0 & +0.6 & 0 \\ 0 & +0.1 & 0 & 0 & 0 & -0.9 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} = \begin{bmatrix} 0.16 \\ 0 \\ 0 \\ -0.8 \\ 0 \\ 0 \end{bmatrix}^T$$

On the ensuing concept vector, two procedures apply: (a) the element that corresponds to the concept that was initially stimulated takes the initial value, so the concept vector becomes: [+0.16 0 +0.8 -0.8 0 0], and (b) if the value of any element lies outside the interval [-1,1], the value is adjusted according to the corresponding limit (if value>1, it becomes +1, while if value<-1, it becomes -1). The adjusted concept vector is multiplied again with the adjacency matrix:

$$\begin{bmatrix} 0.16 \\ 0 \\ +0.8 \\ -0.8 \\ 0 \\ 0 \end{bmatrix}^T \times \begin{bmatrix} 0 & 0 & 0 & +1 & 0 & 0 \\ +0.5 & 0 & 0 & 0 & 0 & 0 \\ +0.2 & 0 & 0 & -1 & 0 & 0 \\ -0.8 & 0 & 0 & 0 & +0.6 & 0 \\ 0 & +0.1 & 0 & 0 & 0 & -0.9 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} = \begin{bmatrix} 0.8 \\ 0 \\ 0 \\ -0.64 \\ -0.48 \\ 0 \end{bmatrix}^T$$

On the ensuing concept vector, the aforementioned two procedures apply again.

This algorithm follows the previous steps until the values of the concept vector remain constant after each multiplication with the adjacency matrix. The final concept vector expresses the condition of the whole given system as it gives the states of all the concepts of the system after the initial stimulation.

The inference mechanism may undergo several adjustments. At first, the upper and lower limits (thresholds) of the nodes may be adapted according to problem-specific prerequisites and take values other than +1 and -1. For instance, a possible node interval may be the following: [-0.1,+0.8]. In addition, diverse functions may correspond to each cause-effect relationship. These functions are applied to each element-by-element multiplication. Possible forms are the linear, the stepwise, the sigmoid, the Gaussian or the hybridization of these (their output always lies in the interval [-1,+1]).

Apart from the above adjustments, there is another issue in the FCM theory that requires further investigation, and concerns the case in which more than one concepts are stimulated. According to previous work, all stimulations act together. That is, at each step of the inference algorithm, the elements of the concept vector that correspond to the stimulated concepts take as value the initial values of the concept states. This approach includes a substantial weakness if the initially stimulated concepts have cause-effect relationships among them. This weakness becomes obvious with the following example, based on the 3-concept system of Figure 3.

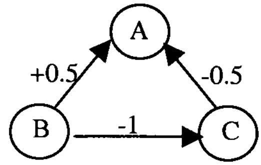


Figure 3. Example of a 3-concept system

Assume first that the upper and lower limits of the node interval obtain the typical values +1 and -1, respectively, and that the applied functions in the inference mechanism are linear. If concept C state takes the value +1, then concept A state will take the value -0.5 and concept B state will remain 0. If concept B state takes the value +1, then concept A state will be influenced not only by concept B but also by concept C, through the relationship B-C-A. So, the node A value is expected to be greater than +0.5. This weakness arises when concepts B and C are stimulated simultaneously. If all stimulations act together, the final value of node A will be 0 (!). This happens because the algorithm ignores the cause-effect relationship between B and C.

The above drawback may be resolved considering the stimulation of each concept apart from the remaining ones and assessing the final state of each concept of the given system taking into account the diverse distinct results. The proposed algorithm steps are the following:

1. Stimulate each of the initially stimulated concepts separately;
2. For each initial stimulation, store the final states for all the concepts;
3. For each concept separately, calculate the maximum positive (*max(positive)*) and the maximum (absolute) negative (*max(negative)*) final state from the set of the distinct final states;
4. For each concept separately, calculate the sum of the positive final states except from *max(positive)*, multiply with a parameter β ($0<\beta<1$) and calculate the sum of the ensuing product and the *max(positive)*. The resulting sum is *total(positive)*.
5. For each concept separately, calculate the sum of the negative final states apart from *max(negative)*, multiply with the predetermined parameter β and calculate

the sum of the ensuing product and the *max(negative)*. The resulting sum is *total(negative)*.

6. For each concept separately, its final state equals: *total(positive)+ total(negative)*.

This algorithm constitutes an innovation in that it ensures the effectiveness of the FCM inference mechanism in cases of multi-stimulus initial conditions.

4 Stock market structure and determinant factors

In this section, an approximation of the way through which economy, companies profile and international political and social situation affect stock prices with an FCM-based model is attempted. The proposed scheme constitutes a theoretical framework which could possibly be brought into action as a real world application. Such an effort necessitates two core tasks:

- The determination of the FCM structure (how many nodes/concepts compose the system and what are the cause-effect relationships between them) and the estimation of the causality for each relationship through the intervention of analysts/experts or alternatively the design of a process for automating these tasks, in any degree.
- The systematic observation, acquisition, analysis and finally exploitation of relative information for the determination of the initial values of the concept nodes.

Getting into details, the major factors affecting stock prices could be classified into 3 clusters: (1) *National Economic and Political Condition*, (2) *International Economic and Political Condition*, (3) *Company Condition*. In detail, these clusters include the following concepts:

- *National Economic and Political Condition*. National economy as a whole exerts a significant effect on stock market operation. To be specific, the health and growth of the national economy is directly reflected on the stock market. Indeed, the changes of the core economic agents, that are *inflation*, *interest rates* and *money flow*, instantly affect the economic status and subsequently the perspectives for the future stock market progress. In addition, many cause-effect relationships take place among these agents. For instance, an increase of inflation causes through the intervention of the central bank a proportionate increase of the level of interest rates, which in turn leads to a decrease of the money flow. Finally the downward movement of the money flow reduces the product prices and subsequently the inflation. In parallel, a concept named *economic events* is needed in order to encompass diverse economic activities conducted by the state or the private sector, such as privatization programs, changes of the tax rates, strategic alliances etc. Last, the *political stability*, expressing constitutional firmness, public opinion etc., is chosen to accompany the economic structure as it is tightly related to it.

- *International Economic and Political Condition.* The international political and economic environment unavoidably affects the operation of stock market especially in the contemporary era where the economy globalization, the technological advancement and the liberalization culture have been established. Therefore, the *international economic condition* and the *international political situation* constitute core concepts of the specific subsystem while in parallel special attention must be given to *inter-country relations* that play a significant role in areas where political instability or economic dependence are strong.
- *Company Condition.* This subsystem is the most crucial when analyzing the stock market operation as a whole. The major determinants for a firm condition are its *financial performance* (profitability, liabilities, capital structure etc.), the *management* (business collaborations, stockholders opinion, HRM profile etc.), the existing *core competencies* (technology, know-how, human resources etc.) and the applied *business strategies* (objectives, short/mid/long term plans, competition tactics, joint ventures etc.). Getting into details, the core competencies undoubtedly affect the firm's financial performance, while on the other hand direct to some extent the formulation of the business strategies. The financial performance affects the management style and the business strategies create/adapt the core competencies and influence the management profile.

The mechanism through which the diverse factors affect the stock prices can be constructed using the methodology of FCMs. Such a model could be used as a stock price forecasting system. Taking into account the major findings of the literature review (…"future research should focus on the short-term prediction of single stock prices and qualitative data should be exploited"…), FCM forecasting systems can adequately fulfill these specifications through further extensions and practical implementation and experimentation.

The designed FCM may have as nodes the entire set of concepts related to the above three clusters of stock market factors. In general, the FCM may be separated in three distinct subsystems (Figure 4), each of which may be an autonomous FCM (Figures 5 to 7), but the inference mechanism will act to the entire set of concepts and not separately to each subsystem.

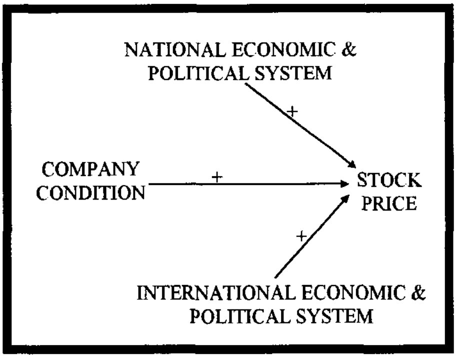


Figure 4. Basic Structure

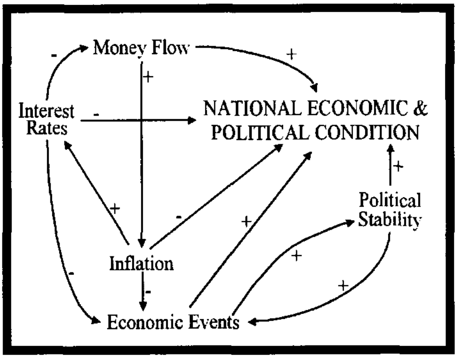


Figure 5. National Economic & Political Condition

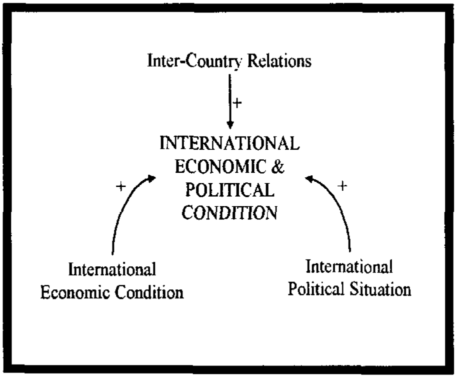


Figure 6. International Economic & Political Condition

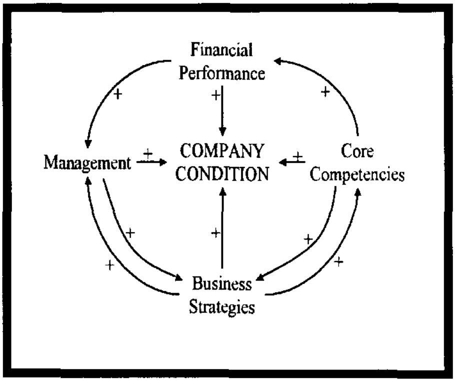


Figure 7. Company Condition

5 Conclusions - future research

This paper sheds light on an innovative approach concerning the development of a stock price forecasting system using the theoretical framework of Fuzzy Cognitive Maps. At first, the most important published papers on stock price forecasting applying advanced techniques are outlined; crucial conclusions about the existing weaknesses and possible solutions are extracted. Considering the various specifications for achieving more satisfactory results (especially the need for qualitative data), FCM theory arises as a powerful approach for the development of a stock price forecasting system. For this purpose, three clusters of economic, political, financial and company factors were selected in order to formulate an integrated FCM structure for a forecasting system.

Despite the significance of the proposed methodological framework, research must continue not only towards the implementation of the proposed scheme but also for the further enhancement of the underlying theory that, admittedly, has still many deficiencies. First and foremost, a methodology based on fuzzy logic principles must be developed, in order to obtain an effective and more reliable methodology for assessing the initial FCM concept states. This can be accomplished mainly by applying proper selected membership functions for each concept and also by developing a rule-based mechanism; furthermore, the FCM inference mechanism may undergo a number of modifications that may have a general form or may just fit to the needs of the specific application. A potential extension and fine tuning of the stock market structure presented in Section 4, calls for the participation and especially the cooperation of agents-experts-analysts that belong to different research fields, the experimentation with real data and the combinatory application of quantitative and qualitative factors e.g. technical analysis tools combined with fuzzy rule bases handling qualitative variables.

6 REFERENCES

1. [Axe76] Axelrod, R. (1976), *Structure of Decision: The Cognitive Maps of Political Elites*, Princeton University Press, New Jersey.
2. [Bas95] Bassi, D.F. (1995), "Stock Price Predictions by Recurrent Neural Network Architectures", *3rd International Conference on Neural Networks in the Capital Markets*, p.331-341.
3. [BM91] Boehm, E.A. and Moore, G.H. (1991), "Financial Market Forecasts and Rates of Return Based on Leading Index Signals", *International Journal of Forecasting*, vol. 7, p.357-374.
4. [Bru96] Brubaker, D. (1996), "Fuzzy Cognitive Maps", *EDN*, April, pp. 209-212.
5. [BWH89] Burns, J.R., Winstead, W.H. and Haworth, D.A. (1989), "Semantic Nets paradigms for both causal and judgmental knowledge representation", *IEEE Transactions on Systems, Man and Cybernetics*, vol. 19, pp. 58-67.
6. [CYC96] Chou, S.T., Yang, C., Chen, C. and Lai, F. (1996), "A Rule-based Neural Stock Trading Decision Support System", *IEEE/IAFE Conference on Computational Intelligence for Financial Engineering*, pp.148-154.
7. [DKE01] Diakoulakis, I.E., Koulouriotis, D.E. and Emiris, D.M (2001), "A Review of Stock Market Prediction using Computational Methods", in *Computational Methods in Decision-Making, Economics and Finance*, Kontoghiorghes E., Rustem B. and Siokos S. (editors), *Kluwer Applied Optimization Series*, in press.

8. [HH95] Hiemstra, Y. and Haefke, C. (1995), "Predicting Quarterly Excess Returns: Two Multilayer Perceptron Training Strategies", *3rd International Conference on Artificial Intelligence Applications on Wall Street*, p. 212-217.
9. [JL93] Jang, G.S. and Lai, F. (1993), "An Intelligent Stock Market Prediction System Using Dual Adaptive-Structure Neural Networks", *2nd International Conference on Artificial Intelligence Applications on Wall Street.*
10. [KAY90] Kimoto, T., Asakawa, K., Yoda, M. and Takeoka, M.; (1990), "Stock Market Prediction System with Modular Neural Networks", *International Joint Conference on Neural Networks*, p.1-6.
11. [KC98] Kim, S.H. and Chun, S.H. (1998), "Graded Forecasting Using an Array of Bipolar Predictions: Application of Probabilistic Neural Networks to a Stock Market Index", *International Journal of Forecasting*, vol.14, p.323-337.
12. [KED01] Koulouriotis, D.E., Emiris, D.M., Diakoulakis, I.E. and Zopounidis, C. (2001), "Comparative Analysis and Evaluation of Intelligent Methodologies for Short-Term Stock Price Forecasting", *Fuzzy Economic Review*, to be appeared.
13. [KGW93] Kryzanowski, L., Galler, M. and Wright, D.W. (1993), "Using Artificial Neural Networks to Pick Stocks", in *Neural Networks in Finance and Investing*, Trippi, R.R. and Turban, E. (editors), IRWIN Professional Company, Chicago, p. 525-541.
14. [KIF97] Kohara, K., Ishikawa, T., Fukuhara, Y. and Nakamura, Y. (1997), "Stock Price Prediction Using Prior Knowledge and Neural Networks", *Intelligent Systems in Accounting, Finance & Management*, vol.6, pp.11-22.
15. [KL98] Kim, H.S. and Lee, K.C. (1998), "Fuzzy implications of fuzzy cognitive map with emphasis on fuzzy causal relationship and fuzzy partially causal relationship", *Fuzzy Sets and Systems*, vol.97, pp.303-313.
16. [KLL96] Kuo, R.J., Lee, L.C. and Lee, C.F. (1996), "Intelligent stock market forecasting system through ANN and fuzzy Delphi", *World Congress on Neural Networks*, pp.886-889.
17. [Kos86] Kosko, B. (1986), "Fuzzy Cognitive Maps", *International Journal of Man-Machine Studies*, vol.24, pp.65-75.
18. [KW97] Kai, F. and Wenhua, X. (1997), "Training Neural Network with Genetic Algorithms for Forecasting the Stock Price Index", *IEEE International Conference on Intelligent Processing Systems*, p.401-403.
19. [LA87] Looney, C.G. and Alfize, A.A. (1987), "Logical Controls via Boolean Rule Matrix Transformations", *IEEE Transactions on Systems, Man and Cybernetics*, vol. 17, pp. 1077-1082.
20. [LCC95] Lam, K.P., Chiu, K.C. and Chan, W.K. (1995), "An Embedded Fuzzy Knowledge Base for Technical Analysis of Stocks", *3rd International Conference on Neural Networks in Capital Markets*, p.317-330.

21. [LDC00] Leung, M.T., Daouk, H. and Chen, A.S. (2000), "Forecasting stock indices: a comparison of classification and level estimation models", *International Journal of Forecasting*, vol.16, pp.173-190.
22. [LK97] Lee, K.C. and Kim, H.S. (1997), "A Fuzzy Cognitive Map-Based Bi-Directional Inference Mechanism: An Application to Stock Investment Analysis", *Intelligent Systems in Accounting, Finance & Management*, vol. 6, pp. 41-57.
23. [LL97] Liu, N.K. and Lee, K.K. (1997), "An Intelligent Business Advisor System for Stock Investment", *Expert Systems*, vol. 14, p.129-139.
24. [ML00] Miao Y. and Liu Z.Q. (2000), "On causal inference in Fuzzy Cognitive Maps", *IEEE Transactions on Fuzzy Systems*, vol. 8, no. 1, pp. 107-119.
25. [MW00] Motiwalla, L. and Wahab, M. (2000), "Predictable variation and profitable trading of US equities: a trading simulation using neural networks", *Computers & Operations Research*, vol.27, pp.1111-1129.
26. [Phi96] Philips, H. (1996), "Adaptive Forecasting", *World Congress on Neural Networks*, pp.504-7.
27. [Pod98] Podding, T. (1998), "Developing forecasting models for integrated financial markets using artificial neural networks", *Neural Network World*, January, pp.65-80.
28. [RZF94] Refenes, N.A., Zapranis, A. and Francis, G. (1994), "Stock performance modelling using neural networks: A Comparative Study with Regression Models", *Neural Networks*, vol.7, pp.375-388.
29. [SD96] Schierholt, K. and Dagli, C.H. (1996), "Stock Market Prediction Using Different Neural Network Classification Architectures", *IEEE/IAFE Conference on Computational Intelligence for Financial Engineering*, pp.72-78.
30. [SDK92] Siriopoulos, C., Doukidis, G., Karakoulas, G., Liakopoulou, T. and Skeuofilax, E. (1992), "An Intelligent Advisor System for Stock Market Investments". *EET- Bulletin Of Greek Banks Association*, 3rd Quarter 1992, p.78-83, in Greek Language.
31. [SG98] Stylios, C.D. and Groumpos, P.P. (1998), "The challenge of modeling supervisory systems using fuzzy cognitive maps", *Journal of Intelligent Manufacturing*, vol.9, pp.339-345.
32. [SM88] Styblinski, M.A. and Meyer, B.D. (1988), "Fuzzy cognitive maps, signal flow graphs and qualitative circuit analysis", *Proceedings of the 2nd IEEE International Conference on Neural Networks*, vol. 2, pp. 549-556.
33. [SPW98] Saad, E.W., Prokhorov, D.V. and Wunsch, D.C. (1998), "Comparative Study of Stock Trend Prediction Using Time Delay, Recurrent and Probabilistic Neural Networks", *IEEE Transactions on Neural Networks*, vol.9, pp.1456-1469.

34. [SSK98] Schneider, M., Shnaider, E., Kandel, A. and Chew, G. (1998), "Automatic Construction of FCMs", *Fuzzy Sets and Systems*, vol.93, pp.161-172.
35. [Tab91] Taber, R. (1991), "Knowledge Processing with Fuzzy Cognitive Maps", *Expert Systems with Applications*, vol.2, pp.83-87.
36. [Tab94] Taber, R. (1994), "Fuzzy cognitive maps model social systems", *AI Expert*, vol.9, pp. 18-23.
37. [TMM95] Tsadiras, A., Margaritis, K. and Mertzios, B. (1995), "Strategic planning using extended fuzzy cognitive maps", *Studies in Informatics and Control*, vol.4, no.3, pp.237-245.
38. [TS87] Taber, W. and Siegel, M. (1987) "Estimation of experts' weights using fuzzy cognitive maps", *IEEE International Conference on Neural Networks*, vol.2, pp.319-326.
39. [WWG92] Wong, F.S., Wang, P.Z., Goh, T.H. and Quek, B.K. (1992), "Fuzzy Neural Systems for Stock Selection", *Financial Analysts Journal*, Jan-Feb, pp.47-52.
40. [Yea88] Yeap, W.K. (1988), "Towards a computational theory of cognitive maps" *Artificial Intelligence*, vol. 34, pp. 297-360
41. [YS91] Yoon, Y. and Swales, G. (1991), "Predicting Stock Price Performance: A Neural Network Approach", *IEEE 24th AIC of Systems Sciences*, pp.156-162.
42. [ZC88] Zhang, W. and Chen, S. (1988), "A logical architecture for cognitive maps", *Proceedings of the 2nd IEEE International Conference on Neural Networks*, vol. 1, pp. 231-238.
43. [ZC89] Zhang, W.R. and Chen, S.S. (1989), "A Generic System for Cognitive Map Development and Decision analysis", *IEEE Transactions on Systems, Man and Cybernetics*, January/ February.
44. [ZL93] Ye Zhongxing, Y. and Liting, G. (1993), "A hybrid cognition system: application to stock market analysis", *IEEE International Joint Conference on Neural Networks*, pp.3000-3003.
45. [ZXM97] Zuohong, P., Xiaodi, L. and Mejabi, O. (1997), "A Neural Fuzzy System for Forecasting", *30th Hawaiian International Conference on System Sciences*, pp.549-558.

This page is intentionally left blank

NEURAL NETWORK VS LINEAR MODELS OF STOCK RETURNS: AN APPLICATION TO THE UK AND GERMAN STOCK MARKET INDICES

ANGELOS KANAS
Department of Economics, University of Crete, 74100 Rethymnon Crete, Greece
E-mail: a-kanas@econ.soc.uoc.gr

We compare the out-of-sample performance of monthly returns forecasts for two indices, namely the FAZ and the FT. A linear and a nonlinear artificial neural network (ANN) model are used to generate out-of-sample competing forecasts for monthly returns. We consider two fundamental variables as the explanatory variables in the linear model and the input variables in the ANN model, namely the trading volume and the dividend. The comparison of out-of-sample forecasts is done on the basis of forecast encompassing. The results suggest that the out-of-sample ANN forecasts encompass linear forecasts of both indices. This finding indicates that the inclusion of nonlinear terms in the relation between stock returns and fundamentals is important in out-of-sample forecasting.

1 Introduction

It is now well documented that aggregate stock returns cannot be satisfactorily explained by the Present Value (PV) model. To account for these failures, several new theoretical models were introduced. Froot and Obstfeld [14] introduced the intrinsic bubbles specification in which the bubbles are driven by the dividends. An important property of an intrinsic bubble is that for a given level of dividends the bubble will remain constant over time. Stable and highly persistent dividends lead to stable and highly persistent departures from the linear PV model, thereby entailing nonlinearity in the stock price - dividend relation. Summers [23] proposed the fads model which also leads to nonlinearity between stock prices and dividends. If there are fads in the stock market, one may observe long temporary price swings which can be modelled as a slowly decaying stationary component in prices. The decay over time in the transitory component will entail mean reversion in stock prices. van Norden and Schaller [24] show how this fads model entails regime switching and thus, nonlinearity in the stock price - dividend relation. Krugman [21] showed that trigger strategies in stock markets also entail nonlinearities between stock prices and fundamentals. Trigger strategists are market participants who buy or sell when the price reaches a predetermined level (e.g. participants in portfolio insurance schemes). Thus, stock market equilibrium is established at two regimes, one with insurers being in the market and another with insurers having dropped out of the market. Under rational expectations, the stock price - dividends relation is tied down (and thus nonlinear) by the fact that when the price falls below the trigger level there is a transition to the other regime. Moreover, a strand of empirical research has introduced models in which trading volume is an important determinant of stock prices [19]. These models depart from a linear relation between stock prices and fundamental variables and establish a nonlinear relation between stock returns and trading volume.

The purpose of this article is to compare the forecasting performance of linear and nonlinear models of monthly aggregate stock returns. Our aim is to examine whether forecasts from a nonlinear stock returns model are preferable to forecasts from a linear stock returns model. We employ the forecast encompassing principle, recently advanced by Harvey et al. [18], Clements and Hendry [8], and Granger and Neubold [17], to compare out of sample stock return index forecasts from a linear and a nonlinear model. We examine whether a nonlinear stock returns model encompasses a competitor linear model, in the sense of being able to explain the forecast errors made by the linear model. A neural network (NN) methodology is employed to estimate a nonlinear model for stock returns, and out-of-sample (nonlinear) stock return forecasts are obtained from this model. The NN methodology is preferred to other nonlinear models because it is nonparametric, and thus appropriate here since we do not want to examine a specific nonlinear functional form between stock prices and fundamentals. The input layer of the NN contains two input variables, namely trading volume and dividends. The specification of these variables is postulated by recent work establishing a nonlinear relationship between stock returns and these variables [4, 14, 19, 21]. Its forecasting competitor is a linear model with the same explanatory variables. We seek to examine whether out-of-sample short-run (1-step) forecasts generated by the NN model encompass out-of-sample forecasts generated by the linear model, following the testing procedure in Clements and Hendry [8]. It should be noted that although the ANN model is expected to encompass the linear model *in-sample*, there is no guarantee that it will do so *out-of-sample* [10].

The remainder of this article is organised as follows. In the second section, we outline the empirical models employed in this study, namely the neural network and the linear model. In the third section, we discuss the forecast encompassing tests used. In the fourth section, we outline the data. The fifth section reports the results from the forecast encompassing tests. The final section provides a summary and concludes.

2 Empirical models for forecasting stock returns

2.1 A nonlinear neural network model for stock returns

We employ the technique of Artificial Neural Network (ANN) estimation to obtain out-of-sample forecasts from a nonlinear model. The specific type of ANN employed in this study is the Multilayer Perceptron (MLP)[1].

---

[1] For a discussion of MLPs, see Campbell, Lo and MacKinlay [2].

The architecture of the MLPs trained includes 1 hidden layer and 6 hidden units. Such an MLP is denoted as MLP(1, 6). The output variable is the monthly aggregate (i.e. index) stock returns. The input variables in the input layer include the lagged percentage change of trading volume (denoted by X1) and the contemporaneous percentage change of dividends (denoted by X2). We also set a third input variable X3 which takes the value of 1, and plays the role of a constant in a regression setting. Moreover, a link was introduced between the input variables and the output variables[2]. As there is no reliable method of specifying the optimal number of hidden layers, we specified one hidden layer on the basis of White's [25] conclusion that single hidden layer MLPs do possess the universal approximation property, namely they can approximate any nonlinear function to an arbitrary degree of accuracy with a suitable number of hidden units. The pictorial representation of the MLP(1, 6)'s architecture employed is given in Figure 1. The algebraic expression for the MLP(1,6) pictured in Figure 1 is given by equation (1), where the subscripts t from the output and input variables are suppressed to ease the exposition. Thus:

$$y = \sum_j a_j x_j + \sum_i b_i f\left(\sum_j c_{i,j} x_j\right) \quad j = 1, 2, \text{ and } i = 1,2,\ldots 6 \tag{1}$$

where f(.) is the activation logistic cumulative distribution function[3], a_j are the weights for the direct signals from each of the two input variables to the output variable, b_i is the weight for the signal from each of the six hidden units to the output variable, and $c_{i,j}$ are the weights for the signals from each of the two input variables to the hidden units. The MLP model given by Equation (1) nests the simple linear regression model and collapses to a linear model with the same explanatory variables if one sets b_i = 0, or if the activation function f(.) becomes the identity function [i.e. f(u) = u]. The nesting of the linear model within the MLP in Equation (1) ensures that the MLP will perform in-sample at least as well as the linear model. Although the ANN model is expected to perform better in-sample, since it nests the linear model, there is no guarantee that it will dominate the linear model out-of-sample [10]. The method of estimation adopted is the on-line error backpropagation.[4] To avoid overfitting, we adopt the cross-validation strategy in training suggested by Kavalieris [20].

---

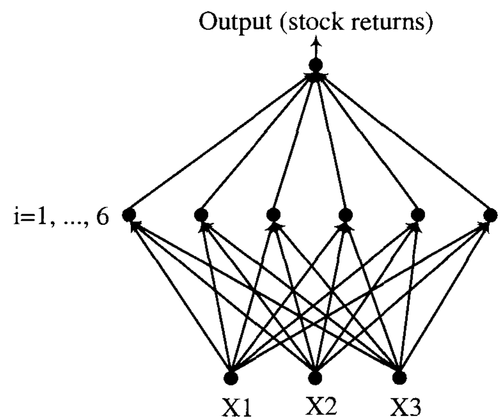
[2] Another variable which could potentially be included in the input variable layer is the lagged stock return. The inclusion of this variable would be justified if there were evidence of nonlinearities in the stock return series. As indicated by the BDS test in Table 1, however, there is no evidence of nonlinearities in the monthly returns series of the FAZ and the FT indices. Thus, the inclusion of a lagged stock return cannot be justified.

[3] The algebraic expression of this function is: f(u) = 1/[1+exp(-u)].

[4] The weights are updated by the following method:

ANN models have recently been used extensively to forecast financial time series. ANN models have been applied to forecast both the mean and the variance of stock market returns [11, 13, 15]. Empirical evidence has indicated that ANN methodology is very promising in forecasting stock returns, especially for high frequency data. Importantly, the ANN methodology has been also extensively applied for forecasting exchange rates, interest rates and futures and options prices.

Figure 1. An MLP(1,6) for stock returns.



2.2 *A linear model for stock returns and fundamentals*

A competitor to the nonlinear ANN model is the linear model for stock returns and fundamentals given by Equation (2).

$$y_t = \delta_0 + \delta_1 X_{1,t-1} + \delta_2 X_{2,t} + w_t \tag{2}$$

where y_t is the stock returns series, $X_{1,t-1}$ is the lagged percentage change in trading volume, $X_{2,t}$ is the percentage change in dividends, and w_t is the error term. The explanatory variables in the linear model are the same as those in the

---

$w_{i,t} = w_{i,t-1} + n[\partial E / \partial w] + a(w_{i,t-1} - w_{i,t-2})$ where n is the learning parameter and a is the momentum coefficient. A high learning rate may speed up convergence, but it may also lead to over-correction and failure to converge. By contrast, low learning rate may prolong convergence. The momentum value, a, determines how much of the previous update should be carried on the current stage. The larger the momentum value, the greater the influence of the last update error. The actual value of the weights, once convergence is achieved, may be sensitive to the choice of the learning and momentum parameters. In this study, the value of the learning rate was set equal to 0.5, and the value of momentum equal to 0.8.

nonlinear ANN model in order for the nonlinear model to nest the linear model[5]. Different explanatory variables in the linear and nonlinear models would entail non-nesting of the models. Moreover, the use of dividends in the linear model is theoretically supported by the PV model, and the inclusion of the lagged trading volume is supported by the findings of Campbell et al. [4].

3 Forecast encompassing

We employ the forecast encompassing principle, recently advanced by Harvey et al. [18], Clements and Hendry [7], and Granger and Neubold [17], to compare out of sample stock return index forecasts from a linear and a nonlinear model. Granger and Neubold [17] suggested that the comparison of forecasts from two competing models should be based on whether competing forecasts embody no useful information absent in the preferred forecasts. Clements and Hendry [7] refer to this situation as forecast encompassing. To illustrate the essence of forecast encompassing, consider two competing models, namely A and B. A is said to encompass, in terms of forecast, B if A can explain the forecast error of B while B cannot explain the forecast error of A. Formal tests of forecast encompassing have been recently developed by Clements and Hendry [8]. These tests are based on a set of OLS regressions. To illustrate, let 1 = ANN model, 2 = linear model, E_i denote the forecast error for model i (i =1,2), and D denote the difference between the forecasts from the two models. Given forecasts from the two models, we can test the null hypothesis that neither model encompasses the other by running two regressions: the first involves regressing the forecast error from the ANN model on the difference of forecasts, i.e.

$$E_{1,t} = \beta_0 + \beta_1 D_t + w_t \tag{3}$$

and obtain the estimated coefficient $\hat{\beta}_1$. The second involves regressing the forecast error from the linear model (i.e. model 2) on the difference of forecasts, i.e.:

$$E_{2,t} = d_0 + d_1 D_t + u_t \tag{4}$$

and obtain the estimated coefficient $\hat{d}_1$. If $\hat{\beta}_1$ is not statistically significant and $\hat{d}_1$ is statistically significant, then we reject the null hypothesis that neither model encompasses the other in favour of the alternative hypothesis that the ANN model encompasses the linear model. If $\hat{\beta}_1$ is significant and $\hat{d}_1$ is not significant, then the linear model encompasses the ANN. If both $\hat{\beta}_1$ and $\hat{d}_1$ are significant or if both

[5] The inclusion in the linear model of the same explanatory variables does not by itself entail nesting of the linear model by the nonlinear model. The nonlinear ANN model will nest the linear one if all three conditions are met. First, the explanatory variables are the same. Second, in equation (1) the function f(.) is not the identity function. Third, in equation (1), $b_i = 0$.

are not significant, then we fail to reject the null hypothesis that neither model encompasses the other. Chong and Hendry [5] have shown that the t-statistics for $\hat{\beta}_1$ and $\hat{d}_1$ are valid forecast encompassing tests when forecasts are made out of sample.

4 Data and preliminary statistics

This study uses monthly data for aggregate stock returns, trading volume and dividends for two European countries, namely the UK and Germany. The stock indices used are the FT All Share index (UK), and the FAZ General (Germany). As in Campbell et al. [4], trading volume is the number of shares traded on the London and Frankfurt Stock Exchanges. The dividend series for each country is the dividend index constructed by Datastream. All series are expressed in natural logarithms. The monthly stock returns are continuous rates of return, computed as 100 times the first difference of the natural logarithm of the monthly stock price in successive months. Similarly, the percentage change in trading volume and the percentage change in dividends are computed as 100 times the first difference of the natural logarithms of the trading volume and dividend index series in successive months. The period under examination extends from January 1980 to June 2000, with a total of 246 observations for each series. The period from January 1980 to December 1995 is treated as the 'training' in-sample period for the ANN and the estimation of the linear model. The subsequent period from January 1996 to June 2000 is the 'testing' out-of-sample period.

Descriptive statistics for stock returns, percentage changes in trading volume and dividends for the training period are reported in Table 1. As shown in this Table, the sample mean of monthly stock returns is positive and statistically different from zero. The variances are 0.003 (FT) and 0.001 (FAZ). The measures for skewness and kurtosis indicate that the distributions of all the series are different from the standard normal. This Table also reports the Augmented Dickey Fuller (ADF) statistics for nonstationarity for all the series, as well as the BDS statistic for nonlinearities in the univariate stock return series[6]. The ADF statistics indicate that all series are stationary. Finally, the BDS test indicates that there is no nonlinear

---

[6] The purpose of conducting the BDS statistics for the univariate stock return series is to test for nonlinear structure in the monthly stock returns series. If there are no nonlinearities in the monthly return series, then the non-inclusion of the lagged stock returns in the input layer of the ANN is justified.

structure in the univariate series of stock return series[7]. The importance of the latter result is that the inclusion of the lagged stock return in the input layer of the ANN is not justified.

Table 1. Preliminary Statistics (Period: January 1980 - December 1995).

Statistics	FT Returns	% Change in LSE Trading Volume	% Change in Dividend in UK	FAZ Returns	% Change in Frankfurt Trading Volume	% Change in Dividend in Frankfurt
Sample Mean	0.01 [a]	0.014	-0.002	0.01 [a]	0.003	-0.003
(t-statistics)	(2.52)	(-0.32)	(0.49)	(2.50)	(0.8)	(-0.54)
Variance	0.003	0.11	0.009	0.001	0.001	0.008
Skewness	-1.60. [a]	2.11. [a]	0.14	-1.13. [a]	0.006	-0.12
Kurtosis	7.23. [a]	26.1. [a]	3.35. [a]	14.43. [a]	0.5	4.99. [a]
ADF statistic (number of lags)	-15.4 (0)	-18.94 (0)	-18.5 (0)	-14.99 (0)	-17.55 ((1)	-18.6 (1)
BDS statistic	-0.0034	------	-------	-0.0026	-------	------

Notes

1. ADF is the augmented Dickey-Fuller test for the stationarity of the returns series. The critical value (5%) is -2.86. This test indicates that all returns series are stationary I(0).
2. [a]: Deviations from normality.
3 * : Statistically significant at 5 percent significance level.

5 Empirical results

The empirical results from training the ANN model given in equation (1) for the training period (i.e. from January 1980 to December 1995) are given in Table 2. This Table reports the estimated coefficients a_j, b_i, and $c_{i,j}$ for both the FT and the FAZ. This Table also reports the Root Mean Squared Error (RMSE) of the generated output for the validation in-sample period, namely January 1995 to December 1995. Further training of the ANN model would result in a higher RMSE, thereby indicating overfitting. As shown in this Table, the estimated weights for both ANN models are of similar value. This indicates that the nonlinear relation between stock returns and fundamentals is similar for both indices.

---

[7] The BDS test has been widely used to test the null hypothesis that a series is iid. See, for example Brock et al. (1991). The empirical findings suggest that the null is rejected when the data are daily. In our study, the relatively low values of the BDS test can be explained by the fact that we deal with monthly data.

Table 2. ANN model estimation (Period : January 1980 - December 1995).

ANN parameters	FAZ index	FT index
a_1	-0.09	0.11
a_2	-1.50	-1.425
b_1	-1.80	-2.16
b_2	-0.251	-0.237
b_3	0.082	0.051
b_4	0.171	0.256
b_5	-0.014	-0.051
b_5	-0.331	-0.343
$c_{1,1}$	-0.57	-0.56
$c_{1,2}$	-0.038	-0.037
$c_{1,3}$	-0.121	-0.118
$c_{1,4}$	30.91	30.72
$c_{1,5}$	8.712	8.529
$c_{1,1}$	-57.50	-57.6
$c_{1,1}$	-0.73	-0.722
$c_{2,2}$	-0.068	-0.065
$c_{2,3}$	-0.081	-0.076
$c_{2,4}$	-0.40	-0.44
$c_{2,5}$	-0.248	-0.241
$c_{2,6}$	-0.929	-0.925
RMSE	0.0371	0.0377

Notes:

1. a_i are the weights for the direct signals from each of the two input variables to the output variable, b_j is the weight for the signal from each of the six hidden units to the output variable, and $c_{i,j}$ are the weights for the signals from each of the two input variables, j, to the hidden units, i. i = 1,2..., 6, and j = 1, 2.

The linear model in equation (2) is estimated using OLS for the same period as the 'training' period of the ANN. Table 3 reports the results. The Table reports the Ramsey's RESET test for correct functional form. The null hypothesis that the examined (linear) functional form is correct is rejected for both stock return indices. A comparison of the results between the linear model and the linear component of the ANN (parameters a_1 and a_2) reveals that the value of the estimated coefficients differs between the two models. The coefficients for the dividend variable are -0.192 (FAZ) and -0.31 (FT) for the linear model whereas they are -1.50 (FAZ) and -1.42 (FT) for the ANN. Similar comments apply to the trading volume variable. All coefficients, however, have the same sign across both models. On the basis of these estimated coefficients, out-of-sample forecasts (i.e. linear forecasts) were generated and compared to the ANN-based forecasts, using forecast encompassing tests.

Table 3. Linear model estimation (Period: January 1980 - December 1995)

	FAZ index	Financial Times index
Constant	0.01 *	0.009 *
	(2.51)	(2.54)
δ_1	-0.009	0.001
	(-0.51)	(0.09)
δ_2	-0.192 *	-0.31 *
	(-4.91)	(8.40)
DW	2.01	2.26
Q	24.65	38.95
[p-value]	[0.32]	[0.23]
LM	0.072	2.38
[p-value]	[0.99]	[0.99]
RESET	5.77 *	5.53 *
[p-value]	[0.001]	[0.001]
R^2	0.26	0.35

Notes:
1. Q is the test for higher-order serial correlation in the residuals.
2. t-statistics in the parentheses.
3. * denotes statistically significant at the 5% level of significance.
4. LM is the Lagrange Multiplier test for heteroscedasticity.
5. RESET is the F-test version of the Ramsey RESET test for correct functional form. The null hypothesis is that the functional form (linearity) is correct. As shown in the Table, the null hypothesis is rejected for both indices, as the corresponding p-values are lower than 0.05.

Table 4 reports the results from the forecast encompassing tests for the out-of-sample forecasts. The out-of-sample forecasts refer to the period January 1996 - June 2000, which is the 'testing' period for the ANN. On the basis of the estimated coefficients of the ANN and the linear model for the 'training' period (January 1980 - December 1995), one-step ahead forecasts were generated from both models. This table reports the heteroscedasticity-robust t-statistics of the estimated coefficients $\hat{\beta}_1$ and $\hat{d}_1$ from regressions (3) and (4) and the corresponding p-values. If the p-values of both estimated coefficients are lower than 0.05 then the null hypothesis should be accepted (namely, neither model encompasses the other). If the p-value of

$\hat{\beta}_1$ is lower than 0.05 and the p-value of $\hat{d}_1$ is higher than 0.05, then the null should be rejected in favour of the alternative hypothesis that the ANN model encompasses the linear. In the opposite case where the p-value of $\hat{\beta}_1$ is higher than 0.05 and the p-value of $\hat{d}_1$ is lower than 0.05, the null is rejected in favour of the alternative that the linear model encompasses the ANN.

Table 4. Forecast encompassing tests (Period: January 1996 - June 2000)

$\hat{\beta}_1$	p-value of t-statistic of $\hat{\beta}_1$	$\hat{\delta}_1$	p-value of t-statistic of $\hat{\delta}_1$	Outcome
		Financial Times index		
-0.43 (-0.94)	0.34	-1.43 * (-3.17)	0.0015	ANN *model forecast* encompasses the linear model
		FAZ index		
-0.31 (-1.05)	0.29	-1.49 * (-3.18)	0.002	ANN model forecast encompasses the linear model

Notes

1. This Table reports the estimated coefficients from equations (3) and (4) in the text, namely

$E_{1,t} = \beta_0 + \beta_1 D_t + w_t$ (3 in text)

$E_{2,t} = d_0 + d_1 D_t + u_t$ (4 in text)

where 1 = ANN model, 2 = linear model, E_i is the forecast error for model i (i =1,2), and D is the difference of the forecasts from the two models.

2. If $\hat{\beta}_1$ is not statistically significant and $\hat{d}_1$ is statistically significant then we reject the null hypothesis that neither model encompasses the other in favour of the alternative hypothesis that the ANN model encompasses the linear model. If $\hat{\beta}_1$ is significant and $\hat{d}_1$ is not significant then the linear model encompasses the ANN. If both $\hat{\beta}_1$ and $\hat{d}_1$ are significant or if both are not significant then we fail to reject the null hypothesis that neither model encompasses the other.
3. The values in parentheses are heteroscedasticity-robust t-statistics.
4. * denotes statistically significant coefficient at the 5% level.

As shown in Table 4, the null hypothesis is rejected for both indices, in favour of the alternative that the ANN model encompasses the linear model. This implies that in those cases where the linear model fails to forecast both indices correctly, this failure can be accounted for by the ANN model. Moreover, the fact that the ANN model is not encompassed by the linear model suggests that in the cases where the ANN model fails to correctly forecast the indices, this failure cannot be accounted for by the linear model. The conclusion from the forecast encompassing tests is that the ANN model explains the forecast error of the linear model in both cases, whereas the linear model cannot explain the forecast error of the ANN in either case. This indicates the superiority of the nonlinear ANN-based stock index forecasts over the linear-based forecasts. These findings suggest that the inclusion of nonlinear terms in the linear model, as is done by the ANN model, is important in terms of out-of-sample monthly stock return forecasting, and are consistent with the view that the relation between stock returns and fundamental variables is nonlinear and not linear[8].

6 Conclusions

The paper compared out-of-sample forecasts of monthly returns for the FT and FAZ stock index returns, generated by two competing models, namely a linear model and a nonlinear ANN model. We consider two fundamental variables as the explanatory variables in the linear model and the input variables in the ANN model, namely the trading volume and the dividend. The comparison of out-of-sample forecasts is carried out on the basis of forecast encompassing. The results suggest that the out-of-sample ANN forecasts can explain the forecast errors of the linear model in for both indices, while the linear model cannot explain the forecast errors of the ANN in either of the two cases. Overall, the results indicate that the inclusion of nonlinear terms in the relation between stock returns and fundamentals is important in out-of-sample forecasting. This conclusion is consistent with the view that the underlying relation between stock returns and fundamentals is nonlinear.

Acknowledgments

I wish to thank two anonymous referees for several constructive comments and suggestions, and Yue Ma and Laurence Copeland for helpful discussions on the topic. Partial financial assistance from the Research Committee, University of Crete is gratefully acknowledged. The usual disclaimer applies.

References

1. Brock, W A, D A Hsieh, and B LeBaron (1991), '*Nonlinear Dynamics, Chaos and Instability*', MIT Press, Cambridge, Massachusetts.
2. Campbell, J Y, A W Lo, and A C MacKinlay (1998), '*The Econometrics of Financial Markets*', Princeton University Press, Princeton, NJ.
3. Campbell, J Y and R J Shiller (1988), ' The dividend-price ratio and the expectations of future dividends and discount factors', *Review of Financial Studies*, 1, pp. 195-228.

[8] An alternative criterion upon which comparison of out of sample forecasts can be carried out is the RMSE. For an interesting exposition of this, see Siriopoulos et al. (1995).

4. Campbell, J Y, S J Grossman, and J Wiang (1993), 'Trading volume and serial correlation in stock returns', *Quarterly Journal of Economics*, November, pp. 905-939.
5. Chong, Y. Y., and D F Hendry (1986), 'Econometric evaluation of linear macroeconomic models', *Review of Economic Studies*, 53, pp. 671-690.
6. Clemen R T (1989), 'Combining forecasts: A review and annotated bibliography', *International Journal of Forecasting*, 5, pp. 671-690.
7. Clements, M. P. and D F Hendry (1993), 'On the limitations of comparing Mean Square Forecast Errors', *Journal of Forecasting*, 5, pp. 559-583.
8. Clements M P and D F Hendry (1998), *'Forecasting economic time series'*, Cambridge University Press, Cambridge, UK.
9. Diebold, F X and R S Mariano (1995), 'Comparing predictive accuracy', *Journal of Business and Economic Statistics*, 13, pp. 253-263.
10. Donaldson, R G., and M Kamstra (1996), 'Forecast combining with Neural Networks', *Journal of Forecasting*, 15, pp. 49-61.
11. Donaldson, R G and M Kamstra (1997), 'An artificial neural network-GARCH model for international stock return volatility', *Journal of Empirical Finance*, 4, 17-46.
12. Flood, R P and P M Garber (1980), 'Market fundamentals versus price-level bubbles: The first tests', *Journal of Political Economy*, 88, pp. 745-770.
13. Flower, B T Cripps, M Jabri, and A White (1995), 'An artificial neural network based trade forecasting system for capital markets', in A P Refenes, Y A Mostafa, J Moody, and A Weigend (eds), 'Neural Networks in Financial Engineering', World Scientific.
14. Froot, K A and M Obstfeld (1991), 'Intrinsic bubbles: the case of stock prices', *American Economic Review*, 81, pp. 831-842.
15. Gencay, R (1995), 'Forecast performance of moving average rules with stock returns', in A P Refenes, Y A Mostafa, J Moody, and A Weigend (eds), 'Neural Networks in Financial Engineering', World Scientific.
16. Granger, C. W. J. (1989), 'Invited review: combining forecasts - twenty years later', *Journal of Forecasting*, 8, pp. 167-173.
17. Granger, C. W. J. and P Newbold (1986), 'Forecasting Economic Time Series', 2nd Ed. Orlando, FL, Academic Press.
18. Harvey, D.I, S. Leybourne, and P. Neubold (1998), Tests for forecast encompassing, *Journal of Business and Economic Statistics*, April, 16, pp. 254-259.
19. Hiemstra C and J D Jones (1994), 'Testing for linear and nonlinear Granger causality in the stock price-volume relation', *Journal of Finance*, XLIX, 5, pp.1639-1664.
20. Kavalieris, L. (1989), 'The estimation of the order of an autoregression using recursive residuals and crossvalidation', *Journal of Time Series Analysis*, 10, pp. 271-178.

21. Krugman, P. (1987), Nonlinear dynamics in stock and currency markets, Working Paper, MIT
22. Siriopoulos C., R. N. Markellos, and K. Sirlantzis (1995), 'Applications of Artificial Neural Networks in Emerging Financial Markets', in A P Refenes, Y A Mostafa, J Moody, and A Weigend (eds), 'Neural Networks in Financial Engineering', World Scientific.
23. Summers, L H (1986), 'Does the stock market rationally reflect fundamental values', *Journal of Finance*, XLI, 3, pp. 591-603.
24. van Norden, S and H Schaller (1994), 'Fads or Bubbles', Wor. Paper, Bank of Canada, pp. 53.
25. White, H. (1992), *'Artificial Neural Networks: Approximation and Learning Theory'*, Blackwell Publishers, Cambridge.

This page is intentionally left blank

HOW WAVELETS AND GENETIC ALGORITHMS CAN ASSIST INTELLIGENT HYBRID METHODOLOGIES IN HANDLING DATA DRIVEN STOCK EXCHANGE DAILY TRADING

A.TSAKONAS, G. DOUNIAS, A.MERIKAS

University of the Aegean, 8 Michalon Street, 82100 Chios, Greece

Email: tsakonas@stt.aegean.gr, g.dounias@ba.aegean.gr, a.merikas@ba.aegean.gr

In this paper is explored the suitability of genetic algorithms for constructing fuzzy rule bases, as part of an hybrid decision support architecture, involving neural networks for wavelet-filtered daily stock rates of change. Specifically, the main structure of the suggested methodology combines a wavelet-based noise removal system, a "multilayer perceptron feedforward neural network" and finally a fuzzy system, which provides the trader with both, linguistic and numerical output, representing a buy/hold/sell strategy. The use of wavelet filtering in data pre-processing, improves the predictability of neural networks, however, it involves the selection of proper wavelet bases. Therefore, by applying genetic algorithms in fuzzy rule bases for optimizing the decision policy, the paper aims at offering a decision support, independent of the selection of the wavelet basis. It is also demonstrated how, based on the test results, the overall system is able to make successful trend prediction, which is then used to create an output similar to the policy that traders would apply if forward price movement was considered to be known.

1 Introduction

During the last decade, scientists more and more often acknowledge innovative methodological schemes that combine traditional computational intelligence tools and techniques, in order to provide successful decision support in complex and vague domains of real world application. Most of these methods are trying to face uncertainty, complexity, imprecision and lack of mathematical modeling, in other words they consist in fact a supplement, or an improvement of corresponding traditional mathematical approaches for approximation, model based reasoning and optimization. Under the conceptual umbrella of computational intelligence, literature basically means research areas such as fyzzy sets, soft computing, neural networks, machine learning and genetic algorithms. On the other side, numerous applications exist, that still have not been faced adequately by decision making and data analysis methodologies, such as real-time monitoring of complex dynamic systems, medical diagnosis, managerial decision making, forecasting on time series data.

In this work the well-known challenge of discovering information hidden inside financial time series data is studied, in order to draw short-term policies for trading decision support on the complex domain of stock exchange markets. The present work does not consist the prime attempt of the authors in the field of stock exchange decision support using computational intelligence [26] [27], nevertheless, it consists of their first detailed and extensive description of the particular role that genetic algorithms and wavelets – a mathematical technique for handling noise in signal processing– are playing in the effective analysis of stock exchange decision making. Through the paper, the main parts of the proposed hybrid computational intelligence based methodology are reviewed and explained, as well as their importance and impact in the overall success of the approach in terms of gained income.

Neural networks are often used as a standard non-linear approximation procedure for classification and prediction. However, the strong presence of noise in daily stock market data, affects dramatically the efficiency of intelligent systems. According to [4], this noise in Stock Markets is usually characterized by a large number of small transactions, which often determine a short-term equity price evolution rather than a smaller number of large transactions. Therefore, training with such raw data, often produces inefficient generalization [3] for computational intelligence systems such as neural networks. Based on recent studies, it is demonstrated that in non-stationary data, when proper assumptions have been made, the application of wavelet thresholding seems to dramatically remove the noise content of a time-series signal [12]. The use of such a filtering system as data pre-processor, makes possible the training of a neural network efficiently with this signal, while also offers the ability of a noise-free future prediction. Nevertheless, the trend prediction requires that the data used are locally stationary, therefore efficiency in this prediction may be expected for a small number of forward values. Thus, the short-term trend prediction by a neural network cannot offer on its own a decision support for a trader. The need for short-term decision support, led to the introduction of a fuzzy system in order to supply the trader with a buy-hold-sell policy. Though noise elimination is near optimal, in order to enhance the decision ability of this fuzzy system, training is required after the initial selection of antecedent sets. To train such a fuzzy system, primarily means to form an objective function. Fuzzy set theory has achieved a high level of efficiency when dealing with noisy values. Using this conclusion, the authors suggest that it is possible to use the defuzzified output for a given data set, as an objective function when a fuzzy system is applied to forward noisy time-series data.

The general implementation suggested in this paper, involves genetic algorithms for configuring a rule base and neuro-fuzzy techniques for adjusting the shapes of the antecedent sets[1]. The input data represent daily returns of stock equities. The whole procedure may be summarized in the following steps:

- Remove the noise of the daily rates of change of closing prices (the initial signal) using wavelet thresholding, obtaining a de-noised signal.
- Train a feedforward neural network using the de-noised signal.

[1] The software was coded in Microsoft VC/C++ and Vbasic as an end-user application. No other domain library or known software was used.

- Feed the neural network output into a fuzzy system (hereafter called as fuzzy system II) in order to obtain a linguistic (i.e. buy/hold/sell) and a defuzzified output to the trader/investor.

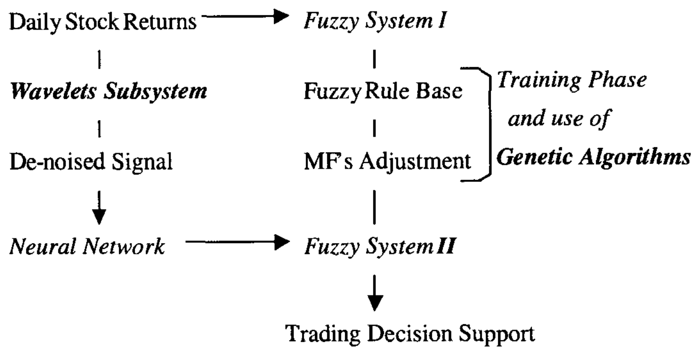


Figure 1. Overall System Architecture highlighting the involvement of wavelets and genetic algorithms

The shapes and ranges of the input membership functions for fuzzy system II may affect dramatically the output of such a system, therefore a special six-step procedure is followed during the training phase, in order to improve the decision making effectiveness of fuzzy system II. The task of determining the above specific settings of the fuzzy system II, becomes particularly advantageous when assisted by genetic algorithms. The suggested six-step procedure as well as the involvement of genetic algorithms in it, is described below and also represented in Figure 1:

1. Select the inputs, membership functions and crisp outputs for two fuzzy systems (called as fuzzy system I, fuzzy system II).
2. Generate a rule base for fuzzy system I using common sense rules (consisting of all possible rules).
3. Create an objective function applying the fuzzy system I to daily rates of change of closing price (the initial signal).
4. Generate a rule base for fuzzy system II (consisting of all possible rules).
5. Feed fuzzy system II with neural network's output and configure its rule base, using genetic algorithms, evaluating the fuzzy system II output with the objective function.
6. Feed fuzzy system II with neural network's output and adjust the shapes of fuzzy system II, using neuro-fuzzy techniques, evaluating the fuzzy system II output with the objective function.

The description of the system's structure on wavelet filtering as well as a brief reference to the specific neural network approach, follows in the next section. Then the model of the proposed objective function is presented in detail, which is used by a genetic algorithm and neuro-fuzzy techniques, in order to train the fuzzy rule based system. At last, the results of the tests performed are presented and discussed, and the paper concludes by proposing further research attempts on the examined domain.

2 System's Structure and Training Methods

2.1 Wavelet Filtering

In this paragraph we discuss extensively the role of wavelet filtering applied on the acquired daily stock exchange data, in order to try and get "clean and transparent" data, free -as much as possible- of noise contained in the initial signal. Due to the nature of the specific application domain, this kind of analysis applied in the initial signal, will offer the ability of identifying real patterns and information hidden inside these data, which then can be modeled and processed properly in order to assist daily trading decision support.

Time-series data are typically considered to be stationary, i.e. their characteristics are homogeneous over time [19]. However, when examining financial time-series data, due to the noise[2] embedded in these time-series, they are assumed to be non-stationary, a case in which might be applied wavelet analyses of long-memory processes that follow scale behavior, including the 1/f (power law) processes. When time-series data are non-stationary, may be assumed that the characteristics (e.g. mean and variance) of the examined signal change slowly over time, hence these time-series data are considered locally stationary. In other words, when analyzing a short range of these data, they seem to be stationary, therefore might be suggested that their characteristic properties (variance, etc.) do not change over a small period of time. In this case wavelet analysis might be applied, in order to obtain a noise free signal. The term wavelet is used to describe families of basis functions having special features. Due to their structure, wavelets have various fundamental properties rendering them highly useful in signal analysis [1]. For a given function space, different wavelet orthonormal bases exist, which enable the representation of functions uniquely and sparsely. A common approach to wavelets is the Ingrid Daubechies [7], which initiates with the use of two orthonormal and related functions that are usually referred as the scaling function φ (or father wavelet), and the mother wavelet ψ. The other wavelets are generated by translating first the scaling function and then by translating and dilating the mother wavelet. The existence of such wavelets depends upon the following conditions on the recursion coefficients:

$$\phi(x) = \sum_{k=0}^{N} h_k \phi(2x - k) \qquad \text{(recursion)}$$

[2] For the "noise" determination in stock exchange, see [4]

$$\psi(x)=\sum_{k=1-N}^{1}(-1)^k h_{1-k}\phi(2x-k) \qquad \text{(definition)}$$

$$\sum_{k=0}^{\lfloor N/2 \rfloor} h_{2k} = 1 = \sum_{k=0}^{\lfloor (N-1)/2 \rfloor} h_{2k+1} \qquad \text{(existence)}$$

$$\int_{\Re} \phi(2x-k)\phi(2x-\ell)dx = 0, k \neq \ell \quad \text{(orthogonality)}$$

$$\sum_{k=\max\{0,2m\}}^{\min\{N,N+2m\}} h_k h_{k-2m} = \begin{cases} 2, m = 0 \\ 0, m \neq 0 \end{cases} \qquad \text{(orthogonality)}$$

In order to perform the wavelet decomposition the fast wavelet transform [30], was used. Therefore, the selected sample must contain a number of values equal to an integral power of two, $N=2^n$ [23] :

$$\vec{s} = \left(s_0, s_1, \ldots, s_{2^n-2}, s_{2^n-1}\right)$$

As a next step the $2N=2^{(n+1)}$ building blocks are replaced:

$$\tilde{f}(r)=\sum_{k=0}^{2^{(n+1)}-1} a_k \phi(r-k))$$

by an equivalent combination of $N=2^n$ slower building blocks $\varphi([r/2]-k)$ and $N=2^n$ slower wavelets $\psi([r/2-1]-k)$:

$$\tilde{f}(r)=\sum_{k=0}^{2^n-1} a_k^{(n-1)}\phi([r/2]-k))+\sum_{k=0}^{2^n-1} c_k^{(n-1)}\psi([r/2-1]-k))$$

where the superscripts $^{(n-1)}$ indicate a frequency lower than for the initial coefficients. In electrical engineering this algorithm is referred as *subband filtering*, and the filters are known as *quadrature mirror filters* [1]. In order to perform noise-reduction using wavelets is being considered the principle that noise contributes to many coefficients, while the trend contributes to only a few coefficients. Thus, if the smaller coefficients are set to zero, the noise can nearly optimally be considered eliminated, preserving at the same time the underlying trend. The denoising method used in this paper is described by Donoho and Johnstone [8] and it is called *wavelet shrinkage,* which is a general method for curve de-noising. This technique offers a non-linear estimation of the initial signal using the *universal threshold* in order to obtain "thresholded" wavelet coefficients. The threshold used, derives from the following equation [13]:

$$\lambda_j = s_j\sqrt{2\log n}$$

Where, j is the level of decomposition, s_j is the standard deviation of the coefficients at that level and n is the overall sample size. In this paper two different thresholding policies were applied in order to test the denoising effect: the soft thresholding and the hard thresholding. Hard thresholding is derived by the equation:

$$\delta^h(d,\lambda) = d\mathbf{1}(|d| > \lambda), \lambda \geq 0, d \in \Re$$

where λ is the threshold, d are the values of the coefficients considered to be processed and $\mathbf{1}(A)$ is the indicator of a set A, i.e. $\mathbf{1}(A)=1$, if $x \in A$, and $\mathbf{1}(A)=0$, if $x \in A^c$. Soft thresholding (or *clipping*) is derived by the following equation:

$$\delta^s(d,\lambda) = (d - \operatorname{sgn}(d))\mathbf{1}(|d| > \lambda), \lambda \geq 0, d \in \Re$$

Due to the different characteristics of the various wavelet families, experiments were performed with only those wavelets that were proper for the type of data used through this paper. While a finite set of data is used as input, the selected wavelet system requires wavelets having compact support. Three wavelet families were tested:

- Daulets, an extremal phase Daubechies family,
- Symmlet wavelets, a minimal phase Daubechies family or least asymmetric family, and
- Coiflets, a rather more symmetric family than the two previous ones, but with a larger support [29].

The denoising effect was examined, using these wavelets with various *vanishing points* and levels of decomposition, facing the trade-off that as the smoothness of the wavelet family increases, the support correspondingly increases [1]. Though incrementing the levels of decomposition, offers better scaling and smoother results, in the present case the tradeoff of over-smoothing was encountered. Over-smoothing is a direct side effect of applying the universal threshold with relatively small size of data (n=256), while its value is considered ideal [29], when $n = \infty$. Moreover, due to the basic assumption made, that noise variance equals the signal variance, daily rates of change are considered very noisy, interpreted as SNR =1.

2.2 *Neural Network*

The main idea for involving a neural network is to provide to the system a robust forecasting ability. A simpler predicting system consisting only of wavelet pre-filtering and trainable fuzzy rule-based system has its limitations regarding the anticipation power due to the size (and the degrees of freedom) of the fuzzy rule base adopted. Therefore, in this work was suggested the assignment of an "easier" task to the fuzzy system, which is the interpretation of a neural network's numerical output to a trader's strategy. Thus, this may result to the advantage of the relatively higher forecasting ability of a feedforward neural network when applied to numerical denoised data. Although in general the output of neural networks often has a smoothing effect, in the present case, where the learning curve is already smooth due to wavelet processing, the training curve fitting is accurate and there is not any added curve smoothing by the involvement of neural network.

Neural networks perform the learning task from the examined numerical stock exchange data. The network generally functions as a black-box mechanism which learns from a set of classified past training data modeling their behavior, in order to be able to properly handle (i.e. make predictions) new input data in the future. In literature there are numerous types of neural networks and general principles for constructing them. The neural network model used in the current approach, is a feed-forward multi-layer perceptron with full connectivity between nodes. Considering the ability of this type of nets to work as global (universal) approximators [25] is acknowledged the fact that other types of neural networks might work as well. For simplicity in this paper, is being considered the case where the network consists of 4 inputs and 4 outputs with 2 hidden layers, each of them containing 8 neurons. The "training set" inputs are built upon a moving time-window, where each next record input set represents a "one-day delayed" data set, which is an architecture popular to bibliography [3], [5], [9] often called as "time-delay networks".

As transfer functions the hyperbolic tangent and the symmetrical sigmoid were preferred due to their ability to handle negative numbers while they had better performance (they work better in [-1,1] range). For the selection of the number of outputs, the corresponding number of inputs was followed. As training algorithm, a fast-convergence algorithm was applied, such as Quickprop, or Rprop [10] using batch processing in the data set used.

2.3 Rule Base Generation and Objective Function Definition

Although the fuzzy system II may be configured empirically from domain knowledge, best results are taken when both the rule base and the membership functions are trained using an objective function. In order to proceed when training a fuzzy system, the objective function is derived by the output of a second fuzzy system, called hereafter as *fuzzy system I.* The training procedure may be summarized for a fuzzy system consisting of n inputs and an objective fuzzy system consisting of k inputs, in the following steps:

1. *Select a training data set.*
2. *Select the vector size for system inputs and outputs.*
3. *Apply evolutionary training using genetic algorithm for rule base determination*
4. *Apply neuro-fuzzy training for membership function refinement*

Each training vector consists of a time window containing subsequent values and its size is equal to the sum of vector sizes of both fuzzy systems' inputs. This objective function is calculated by applying the objective fuzzy system into the last values of the training vector. This function is created based on the cumulative prediction returns over a future step and may be easily derived by a heuristic model which extends as a chain rule the following set of rules in the form of a decision output of the objective fuzzy system, when one day prediction is needed (considering here three antecedent sets):

IF tommorow daily return IS high THEN buy;
IF tommorow daily return IS zero THEN hold;
IF tommorow daily return IS low THEN sell;

For two days prediction this set of rules becomes:

IF tommorow daily return IS high AND next day after tommorow daily return IS high THEN strong buy;
IF tommorow daily return IS high AND next day after tommorow daily return IS zero THEN buy;
IF tommorow daily return IS zero AND next day after tommorow daily return IS high THEN buy;
IF tommorow daily return IS zero AND next day after tommorow daily return IS zero THEN hold;
IF tommorow daily return IS low AND next day after tommorow daily return IS zero THEN sell;
IF tommorow daily return IS zero AND next day after tommorow daily return IS low THEN sell;
IF tommorow daily return IS low AND next day after tommorow daily return IS low THEN strong sell;

It may be easily shown that the possible outcomes using this concept, are given by the following formula:

$$U = a * (g - 1) + 1 \tag{2.2}$$

where, a is the desirable forward number of days (and so the objective fuzzy system inputs), g is the number of membership functions (number of premise sets[3]), and U the possible outcomes. By assigning to these outcomes a crisp value may be obtained a defuzzified output when the objective value is applied to the last values of a training vector. In general, the above scheme, may be considered as a common sense set of rules, if forward price returns were known *a-priori*. The concept behind the neuro-genetic training is to obtain a rule base as well as membership functions' shapes for a fuzzy system, which will have the same or almost the same[4] defuzzified output with the objective fuzzy system applied to forward values using this common sense set of rules.

The use of the initial signal values as inputs (prior to wavelet processing) offers the chance to introduce back the noise content of the processed signal in an aggregate manner, which may make use of any knowledge of the noise distribution. For example, if the signal noise follows normal distribution having zero median and finite variance, there might be encountered better results using Gaussian type membership functions [11], for the inputs of the fuzzy system I.

[3] In the example above, three
[4] For a given training data set, the success of a low RMSE during the training phase, depends on the model selection itself.

2.4 *Genetic Algorithm*

The training of the fuzzy system II consists of two phases: determination of the output values of the rule base using a genetic algorithm and calculation of each input's membership function parameters, using neural techniques such as error backpropagation. The need for determining the fuzzy rules derives from the fact that they may not necessarily be derived from a common sense rule set. The predicted values of the neural network used, do not correspond to actual future price movements but they represent a trend, which is constructed after a noise reduction. Considering the given training set, alternative rule bases may perform better than a common sense rule base for the fuzzy system II. According to the implementation adopted, this rule base will be selected by evaluating the output of fuzzy system II, as compared to the output of fuzzy system I (which is applied to the initial signal). In fact, the redefinition of the rule base reflects the possibility of unwanted removal of trend content during the wavelet procedure. It also mirrors possible remaining noisy content in this signal in spite of the applied wavelet process and yet, it accepts the fact that the noise reduction may not be completely accurate but due to the unknown noise variance, trend content might be removed by this procedure. Thus, should be adopted the rule base that offers the optimum evaluation values depending on the fuzzy system I output for the given signal. The genetic algorithm was selected as the training method of choice, due to the fact that derivative-based methods for minimizing the evaluation error cannot be applied. To accomplish this task, the structure of a chromosome was introduced into the proposed system. By coding the input values in a fuzzy associative memory (FAM) the outcome is the number of all possible rules r as a function of the used inputs n and their membership functions m (assuming there is only one such output in the present case):

$$r = m^{n}$$

The number of the output crisp values depends on the degrees of freedom [11] allowed to the system by the user. For a simple decision system such as the Buy/Hold/Sell strategy, three crisp output values are enough, though more values may be more useful for a trader/investor, depending on the policy that might want to apply. After selecting the number of the desired crisp output values, each rule may be encoded using binary representation. The number of bits b needed to represent a rule, when τ possible crisp output values are selected, is calculated by the following formula:

$$b = x : \; 2^{x} > \tau > 2^{x-1}$$

Hence, the chromosome consisting of the full set of rules is then constructed, each of them containing one value of the desired crisp output set (Figure 2).

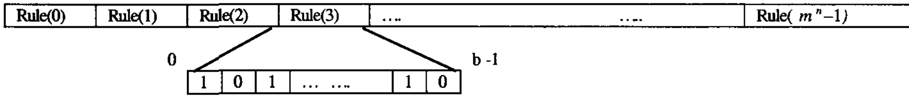


Figure 2. The fuzzy rule chromosome

During the genetic process, special care has to be taken to the chromosome genes produced, in case they are not contained in the crisp output value set *H*. As an example, when three crisp outputs are selected as the crisp value set, two bits for representation are needed, which produce four possible values, one of them not contained in the crisp value set, *H*. The genetic algorithm operators were recombination, crossover and mutation. A population of 30 chromosomes, crossover rate 0.9 and mutation rate 0.001 was adapted while it offered fast convergence. The selection technique applied was the stochastic sampling with replacement (SSR). To avoid premature convergence and speed up the search [17], the fitness after each epoch was re-scaled. While the position of the genes considered plays important role as it affects dramatically the output of fuzzy system II, special care has to be taken in the crossover technique selection to avoid poor performance when the examined population has largely being converged. Hence, there was experimentation with various crossover techniques, and finally was adapted a heuristic variable crossover scheme, starting with two-point split and doubling the points each time that the average error spread was reduced by half [26]. In all of the tests performed, this rule base was different from the common sense rule base and it offered better prediction results, demonstrating this way the need for the training of the fuzzy system II.

2.5 *Neuro-Fuzzy Training*

When the rule base configuration is completed, the whole process continues with the training phase, by modifying the input membership function parameters of the fuzzy system II, in order to minimize the prediction error. This procedure consists of the tuning of the fuzzy system II over the outputs of the fuzzy system I by reducing the overall cumulative error. A neuro-fuzzy training method was implemented, based on the model described by Nauck and Kruse in [22], in order to adapt better shapes and locations for the considered membership functions. The rule base generation method introduced by Nauck and Kruse was not implemented finally, as the fuzzy systems used, are small and easily manageable by current technology. For the main reason rule pruning was not considered. In addition to the triangular delta parameters presented by Nauck and Kruse in [21], heuristic delta calculations were applied, in order to train Trapezoid, Gaussian, Close Sigmoidal and Cauchy functions. For more details on the procedure and the algorithm used, the reader should refer to

[20]. In the above-mentioned training process a constraint set Φ is assumed, in order to keep the MF modifications in low levels. These constraints mirror the need for preserving the explanatory scheme for fuzzy system II and for keeping the meanings of the antecedent sets. In these constraints, the limits of each MF's position and the size of each MF were included.

3 Results and Discussion

Experimental data (Table I) were chosen according to the principle that the selected equities should belong to different sectors and capitalization classes, because these factors usually affect price behavior through time. The equities chosen were:

1. Atticat S.A., a constructions company, since such companies have shown in the past their own price behavior, based on macroeconomic/financial factors.
2. Alco S.A., a metallurgical company of the parallel market of Athens Stock Exchange (ASE), belonging to a sector of enterprises with large price variations.
3. Hellenic Bottling S.A., a large capitalization enterprise belonging to the food sector, that usually shows "defensive" behavior against exogenous factors of the stock exchange, even in serious situations (e.g. ASE: April 1998).

The testing period (see Table I) should reflect both a 'bullish'[5] market and a 'bearish' market. Therefore two testing periods were selected, (a) from 26/4/99 to 25/10/99, a typically bullish market period, and (b) from 1/11/99 to 3/5/00, a typically bearish market period. Note (Table I) that the 'buy and hold' strategy -also reflecting the price evolution over time- for the equities of (a) period, resulted in 234.49%, 224.24% and -10.07% respectively, while for the (b) period resulted in -36.5%, -64.95% and -37.64% respectively. Concluding, the selection of equities and periods described above, is very strict in its definition, compared to any other random selection process of equities.

A 4X8X8X4 neural network was trained, using symmetrical sigmoid function and the quickprop algorithm for 30 epochs. The wavelet noise reduction methodology was applied using Daubechies extremal phase wavelets with 3 vanishing moments and 2 levels of decomposition and soft thresholding was used. The fuzzy system II consisted of 3 inputs and 3 triangular membership functions (High, Zero, Low) and 7 possible outputs were used. Also, for the training was used a fuzzy system I consisting of three inputs, 3 gaussian membership functions and 7 possible outputs. The training phase involved genetic-based learning of the fuzzy system's II rule base, using population of 30 chromosomes, the suggested variable point split crossover, crossover rate = 0.9, mutation rate = 0.001, generation gap=0.5 for 100 epochs. It also involved error backpropagation with a neuro-fuzzy technique for the inputs' membership function parameters using on-line learning rate

---

[5] bearish = descending market behavior, bullish = ascending market behavior

0,00001 for 10 epochs. The overall system achieved to predict all major and some minor trend changes. The results present the percentage of investment earning or loss for the testing set data. They are also compared with the Buy/Hold Strategy (assuming that an investor buys a stock equity at the first day of the test set and sells it at the last day of the test set).

Table I. System evaluation on test set for three equities and two market periods

Stock Equity	Training Set:	Test Set:	System Strategy Income % in Test Set :	Buy & Hold Income % in Test Set:
ATTIKAT S.A.	From 2/4/1998 To 23/4/1999 (256 values)	From 26/4/1999 To 25/10/1999 (128 values)	476,27 %	234,49 %
ATTIKAT S.A.	From 16/10/1998 To 29/10/1999 (256 values)	From 1/11/1999 To 3/5/2000 (128 values)	173,37 %	-36,50 %
ALCO S.A.	From 2/4/1998 To 23/4/1999 (256 values)	From 26/4/1999 To 25/10/1999 (128 values)	315,02 %	224,24 %
ALCO S.A.	From 16/10/1998 To 29/10/1999 (256 values)	From 1/11/1999 To 3/5/2000 (128 values)	426,38 %	-64,95 %
Hellenic Bottling S.A	From 2/4/1998 To 23/4/1999 (256 values)	From 26/4/1999 To 25/10/1999 (128 values)	19,39 %	- 10,07 %
Hellenic Bottling S.A	From 16/10/1998 To 29/10/1999 (256 values)	From 1/11/1999 To 3/5/2000 (128 values)	101,56 %	- 37,64 %

In order to apply the system's decision support, for these examples, the following strategy was selected: If the system suggests an action (i.e. buy or sell) the investor opens (or closes accordingly) one position[6]. If the system suggests a "strong" action (i.e. strong buy or strong sell) the investor opens (or closes accordingly) two positions. In other signals (hold, possible buy, possible sell) the investor holds. In all testing cases the system's output proved to be superior to the Buy & Hold Strategy. Its rate of success was depending on the noise reduction method adapted, the neural network used, the Fuzzy System I & II setups and training, and on the general stock behavior during the test period[7].

---

[6] Opening one position is equivalent with buying a quantity of the equity. Closing one position is equivalent with selling the same quantity of this equity. Opening (closing) two positions, refers to buying (selling) the double quantity of this equity.

[7] The data used are freely available by http://www.naftemporiki.gr and http://www.9in.gr

4 Conclusions and Further Research

Computational intelligence tools and techniques seem to rapidly become lately the methods of choice in complex time-series data domains with near-stochastic behavior, as they can discover patterns and find best policies in handling decision making, in a manner that only expert humans could approximately perform with success up to now. In this paper was presented the use of genetic algorithms and wavelet filtering as parts of a wider hybrid computational intelligence methodology for assisting the daily trading decision support in stock exchange transactions. The main framework of the overall decision-making methodology was designed prior to this paper, but the role and appropriateness of wavelets and genetic algorithms was described and discussed in detail for the first time through this work. Specifically, in order to improve the learning capability of the main neural network, a pre-processing of the data using wavelet filtering was considered. The application of the wavelet subsystem enhanced the anticipation power of the neural network, however it offered a short-term predicted trend, which on its own could not be used for short-term decision support. The reasons were the strong noise content of the raw data used, which should lead to a selection of the best wavelet basis for the best noise removal. Consequently, the outputs of the neural network were driven to an adaptive fuzzy-rule based system, which in its turn was trained using genetic algorithms and neuro-fuzzy techniques. The goal of this training was to achieve a linguistic and a defuzzified output, which would be similar to the output of a fuzzy system, if it had been applied 9to real-world forward rates of change, using a common sense rule base. Results demonstrated that the proposed system is capable in providing efficient decision support for short term trading. Moreover, it is clear that either wavelets or genetic algorithms have both a strong impact on the improvement of the final success of the overall architecture in terms of potential income. Concluding, wavelets are strongly suggested as method of choice when dealing with large amounts of "noisy" time-series data. On the other hand, genetic algorithms should also be considered as method of choice when adjusting fuzzy rule bases contained within hybrid computational intelligence methodologies, for two main reasons. First, due to their ability to find near-optimal solutions in such complex situations where derivative based methods for minimizing the evaluation error are not applicable. And then because they are also capable of overcoming limitations of common sense reasoning when trying to determine specific fuzzy rules for setting alternative decision policies.

Further research may involve the selection of the more suitable input membership functions for both fuzzy systems of the proposed architecture, possibly with the aid of a computational intelligence based algorithm. Fuzzy clustering, evolutionary algorithms and evolutionary programming could be used in this topic as well as for the rule base configuration of fuzzy system II. Moreover, the application of recurrent -or other type- neural networks [15], instead of the feed forward structure used in this paper, might offer better generalization, hence improving the network prediction. The choice of the more suitable financial data to be used as inputs in this system instead of daily returns, is another open topic, which deserves further exploitation. Finally, the adoption of different thresholding methods, the introduction of the second-generation wavelets, and at last, the selection of other threshold values, may affect the effectiveness of the noise reduction procedure.

References

1. Abramovich Felix, Baily C. Trevor, Sapatinas Theofanis, (1999), "Wavelet analysis and its statistical applications", The Statistician, Vol. 49, 1-29.
2. Baker, J.E., (1987), "Reducing bias and inefficiency in the selection algorithms", in Proc. 2nd International Conf. on Genetic Algorithms, Lawrence Erlbaum Ass., Hillsdale, NJ, pp.14-21.
3. Bishop, C.M., (1995), Neural Networks for Pattern Recognition, Oxford University Press.
4. Black Fischer, (1986), "Noise", (Goldman Sachs & Co.), The Journal of Finance, Vol. 41 (XLI), No. 3, July '86, pp.529-543.
5. Chester, D.L., (1990),"Why Two Hidden Layers are Better than One", IJCNN-90-WASH-DC, Lawrence Erlbaum, 1990, Vol.1, pp 265-268.
6. Cody, M.A., (1992), "The Fast Wavelet Transform: Beyond Fourier Transforms", Dr.Dobb's Journal, April '92, pp.16-31.
7. Daubechies, I., (1990), Ten Lectures on Wavelets, SIAM Press, Philadelphia.
8. Donoho, D.L. and Johnstone, I.M., (1995), "Adapting to Unknown Smoothness via Wavelet Shrinkage", Jour. of Amer. Stat. Association 90, pp.1200-1224.
9. Georg Dorffner, (1995), "Neural Networks for Time Series Processing", Dept. of Medical Cybernetics and Artificial Intelligence, University of Vienna, 1995.
10. Fahlman Scott, (1988), "An empirical study of learning speed in backpropagation networks", Technical Report CMU-CS-88-162, Carnegie Mellon University, Pittsburgh, PA 15213, September '88.
11. Jang, J.-S.R., Sun C.-T., Mizutani E., (1997), Neuro-Fuzzy and Soft Computing, Matlab Curriculum Series.
12. Jansen Maarten, (2000), "Wavelet Thresholding and Noise Reduction", Ph.D. Dissertation, Katholieke Universiteit Leuven, Celestijnenlaan 200-A-B-3001, Heverlee, Belgium.
13. Johnstone, I.M. and Silverman, B.W., (1997), "Wavelet threshold estimators for data with correlated noise", Jour. Royal Stat. Soc., Series B, 59, pp. 319-351.
14. Kruse Rudolf and Nauck Detlef, (1995), "Learning Methods for Fuzzy Systems", in Proceedings of the 3rd German GI-Workshop "Neuro-Fuzzy Systeme", Nov. '95, pp. 7-22.
15. S. Lawrence, A. Chung Tsoi, C. Lee Giles, (1996), "Noisy Time Series Prediction using Symbolic Representation and Recurrent Neural Network

Grammatical Inference", Technical Report, UM IACS-TR-96-27 and CS-TR-36-25, Institute for Advanced Computer Studies, University of Maryland, College Park, MD 20742, USA.

16. S. Lawrence, C. Lee Giles, A. Chung Tsoi, (1996), "What size neural network gives optimal generalization? Convergence properties of backpropagation", Technical Report UMIACS-TR-96-22 and CS-TR-36179, Institute for Advanced Computer Studies, Univ. of Maryland, College Park, MD20742, USA
17. Man, K.F, Tang, K.S and Kwong, S., (1999), Genetic algorithms: concepts and designs, Springer-Verlag, London.
18. P. Mars, J.R. Chen, R. Nambiar, (1996), Learning algorithms: theory and applications in signal processing, control and communications, CRC Press.
19. Nason P. Guy and Sachs Von Rainer, (1999), "Wavelets in time series analysis", University of Bristol, Phil. Trans. Royal Soc. Lond. A, 357, pp. 2511-2526.
20. Nauck Detlef and Kruse Rudolf, (1994), "Choosing Appropriate Neuro-Fuzzy Models", in Proceedings of EUFIT'94, Aachen, Germany, pp. 552-557.
21. Nauck Detlef and Kruse Rudolf, (1995), "NEFCLASS – a Neuro-Fuzzy approach for the classification of data", in K. M. George, Janice H. Carrol, Ed. Deaton, Dave Oppenheim and Jim Hightower, Editors, Applied Computing, 1995, ACM Symposium on Applied Computing, Nashville, USA, Febr. 26-28, pp. 461-465, published by ACM Press, New York, USA.
22. Nauck Detlef and Kruse Rudolf, (1996), "Designing Neuro-Fuzzy Systems Through Backpropagations", in Witold Pedrydz editor, Fuzzy Modeling – Paradigms and Practice, pp. 203-231, Kluwer Academic Publishers.
23. Nievergelt Yves, (1999), Wavelets made easy, Birkhauser, Boston.
24. Procyk, T.J and Mamdani, E.H., (1979), "A linguistic self-organizing process controller", Automatica, 15, pp.15-30.
25. Sarle, W.S. (1994), "Neural Networks and Statistical Models", in SAS Institute Inc., Proceedings of the 19th Annual SAS Users Group International Conference, Cary, NC: SAS Institute Inc., pp. 1538-1550.
26. A. Tsakonas, G. Dounias and G. Tselentis, (2000), "Using Fuzzy Rules in Multilayer Perceptron Neural Networks for Multiresolution Processed Signals: A Real World Application in Stock Exchange Market", in Proceedings of the Int. Symposium on Computational Intelligence and Learning, CoIL 2000, Chios, Greece, pp. 154-170.
27. A. Tsakonas and G. Dounias, (2000), "Decision making on noisy time-series data under a neuro-genetic fuzzy rule-based system approach", in Proc. of 7th UK Workshop on Fuzzy Systems, pp. 80-89.
28. G. Tselentis and G. Dounias, (1998),"A Neuro-Fuzzy Approach for Forecasting Exchange Rates", in "Managing in Uncertainty: Theory and Practice", Eds.: Zopounidis K. and Pardalos J., Kluwer Academic Publishers.

29. Vidakovic Brani, (1999), Statistical Modeling by Wavelets, Duke University, Wiley Series in Probability and Statistics.
30. Uytterhoeven Geert, (1999), "Wavelets: Software and Applications", Ph.D. Dissertation, Katholieke Universiteit Leuven, Celestijnenlaan 200-A-B-3001, Heverlee, Belgium, April '99.

CORPORATE FINANCE AND BANKING MANAGEMENT

This page is intentionally left blank

EXPERTONS AND BEHAVIOUR OF COMPANIES WITH REGARD TO THE ADEQUACY BETWEEN BUSINESS DECISIONS AND OBJECTIVES

ALAIN COUTURIER

Conservatoire National des Arts et Métiers (CNAM) Nante 25 boulevard Guy Mollet 44071 NANTES Cedex 03 France
Email : a.couturier@cnam-paysdelaloire.fr

BERNARD FIOLEAU

Faculté des Sciences Économiques et de Gestion de Nantes 110 Boulevard Michelet BP 52231 44322 NANTES Cedex 03 France
Email : fioleau@sc-eco.univ-nantes.fr

In this article we propose a model for measuring the adequacy of business decisions with a given standard or set of standards they are designed to achieve. These standards or set of standards can either be external contraints or self-imposed objectives (internal constraints). The model, which is presented in the first part, is based on the theory of expertons. The model is then tested on a sample of 161 companies. Finally, the results are presented and discussed.

1 Introduction

Observation of the behaviour of operating businesses shows that their activities are strongly influenced by the many constraints to which they are subjected. These constraints can be considered as norms, either explicit or implicit: for example, accounting and financial norms, technical and environmental norms, and so on. The successful implementation of any operation or strategy, possibly even the survival of the business, invariably depends on the pertinence to these norms of the decisions taken.

In this context, it must be emphasised that while these norms are mainly expressed in terms of performance targets, they do not define the means of achieving those targets. For a given objective, companies therefore appear to be free to choose both the objective itself and the most suitable technical means of achieving it. Let us take as an illustration the launching of a new car model (the objective). The budget which would normally be needed for the promotional advertising campaign would be in the region of $1.5 million (the norm). The company has to decide on the date, duration, choice of media, launch budget, target, geographic spread, etc. (the choices). The company cannot be absolutely certain of achieving a successful launch in any case, but one or more bad decisions (an insufficiently long campaign, underfunding, badly chosen media, etc.) will probably result in a failure.

As these norms are inherent in all business activities, we are not here concerned with discussing their merits but with investigating the possibility of assessing the pertinence of the behaviour and the policies carried out in order to achieve the stated objective. The fact that all businesses are subject to these norms generates the risk that decisions may be echoed from one company to the next. However, varying factors such as size, maturity, management style, means of finance and the like lead to some variation in companies' behaviour.

In this article we propose a model for measuring the adequacy of the decisions taken by companies, and hence the consequences of such decisions, with regard the objectives they have set themselves. In the first part, after a brief presentation of the experton theory, we propose two models which enable to position the business decisions of companies compared with these objectives. The criteria which enable the measurement of each company's behaviour evolve over time. Observation over time of a given criteria is made by a set of values which can be expressed as an interval. The second part therefore discusses the necessary adaptation of the model so as to process these intervals as fuzzy data. We then illustrate the model by applying it to a sample of companies in the wholesale industrial equipment sector. The last part is an analysis and commentary of the results obtained.

2 The model

Experton:

If A is a predicate of fuzzy valuation $v(A) \in [0,1]$, the associated experton Exp(A) is a pair (X_m, X_M) of fuzzy random subsets representing the lower and upper bounds of v(A).

Estimating an experton:

Let A be a predicate, v(A) its valuation and (x, y) a pair of n-uples:

$$x = (x_1, x_2, ..., x_i, ..., x_n), y = (y_1, y_2, ..., y_i, ..., y_n)$$

are samples of the upper and lower bounds respectively of v(A) where (x_i, y_i) may, for example, represent the minimum and the maximum values of a financial ratio.

If $C_m(x)$ and $C_M(x)$ are the empirical cumulative functions of the upper and lower bounds then, the estimation of the experton is formed from the pair $(1 - C_m, 1 - C_M)$.

Constructing an experton from empirical data:

Let:

X be a random support variable in $\Re$ of density f(x) and distribution function F(x), and

$x = (x_1, x_2, ..., x_i, ..., x_n)$ is an n-sample.

Note: variable X is assumed to characterise a certain predicate A. The problem of this characterisation will be considered in the applications presented in part three.

We are therefore seeking a function X $\rightarrow$ v(A) where v(A), valuation of A, is a random variable with a distribution function L.

We need to be able to configure the transformation in order to obtain the bounds of v(A). In practice this involves determining a transformation T : F(x) $\rightarrow$ L

In this article, two transformations T are envisaged:

. Transformation 1: sigmoid adjustment

$$f(x) = \frac{1}{1 + e^{-ax-b}}$$

$x \in \Re$ and $f(x) \in [0, 1]$

. Transformation 2: exponential adjustment

$$f(x) = e^{-e^{-ax-b}}$$

$x \in \Re$ and $f(x) \in [0, 1]$

Whichever model is chosen, recoding to [0, 1] has two main advantages:

- it enables distribution queues to be processed non-selectively;
- it opens up the possibility of using the concept of expertons (Kaufmann [5], [6], [7], Gil Aluja [4]).

We can obtain a minimum and maximum value for each individual and criteria. It is hence possible to build characteristic minimum and maximum distribution of the experton model.

Coefficients a and b in both models may be :

- estimated by adjustment, or
- estimated by expertise. In this case, a specialist (or group of specialists) is asked to set the limits beyond which a value is recoded to either near 0 or near 1 (Céruti, Gatino, [1]).

For example: if we wish to analyse the behaviour of clients in an economic sector regarding the settling of bills, the most widely used ratio is the average trade credit period: $\frac{\text{Accounts receivable}}{\text{Sales inc VAT}} * 360$

We may wish to assimilate a period of under 10 days to a cash payment, in which case the value of the ratio will be recoded as 0. Conversely, a period of over 180 days may be considered excessive and be recoded as 1. This recoding to [0, 1] enables us :

- to synthesise a set of criteria and compare them to a norm or set of norms,
- to evaluate some implications of criteria, and,
- to measure degrees of pertinence to the norm, and detect both pertinent and non-pertinent behaviour.

3 Application and adaptation of the model

3.1 Sampling

The model was tested on a random selection of French companies in the wholesale industrial equipment sector. The sample of 161 companies with at least 100 employees was selected from a list of 6679 eligible companies. The small sample size is explained by the need to have available a complete set of financial data for a four-year period (1994 - 1997)[1].

For each of these years and for each company we calculated 6 criteria representative of the working capital requirement typifying their activities and main constituent elements. These criteria are listed in Table 1.

Table 1: List of criteria characterising the activity

1 - $\frac{\text{Inventory}}{\text{Cost of sales}} *360$	4 - $\frac{\text{Added value}}{\text{Sales}}$
2 - $\frac{\text{Accounts receivable}}{\text{Sales inc VAT}} *360$	5 - $\frac{\text{Working capital}}{\text{Manpower}}$
3 - $\frac{\text{Suppliers}}{\text{Cost of goods inc VAT}} *360$	6 - $\frac{\text{Working capital}}{\text{Sales}}$

The fact that all the companies in the sample work come from the same sector should guarantee a broad homogeneity in economic activity. Table 2 summarises the main characteristics: mean values, standard deviation and range, for the six criteria. For most of them it shows in fact a very high deviation between the highest and lowest values observed. It should be noted however that these extremes are reported for only a very small number of companies, usually one or two, and that these companies are different from one ratio to another. Careful analysis shows that, whatever the year, around twenty companies show one ratio with a value different by

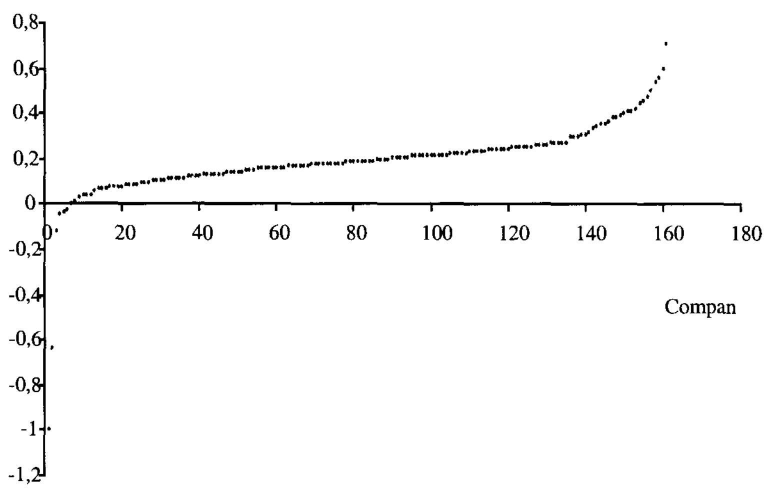
[1] This sample has been taken from the SCRL " Diane " database ; sector 516 in the NAF classification of INSEE.

more than three standard deviations from the average value. Graph 1 illustrates this phenomenon for criterion 6.

Table 2. Mean values and dispersion of ratios

	Min.	Max.	Mean	Std deviation	Std dev/Mean
C_1	1,33	329,00	89,21	51,95	0,58
C_2	11,00	235,33	83,75	27,20	0,32
C_3	19,67	234,00	75,12	28,89	0,38
C_4	0,05	0,68	0,25	0,11	0,44
C_5	-1015,07	2004,33	318,79	337,53	1,06
C_6	-1,01	0,71	0,19	0,17	0,87

Graph 1. Working capital/Sales



Each ratio for the 161 companies is expressed in the form of an interval comprising the minimum and maximum values observed during the period. Graphs 2 and 3 plot each of these values for ratio 1, in ascending order.

Thus in graph 2:

- the lower curve represents minimum values in descending order,
- the upper curve represents associated maximum values.

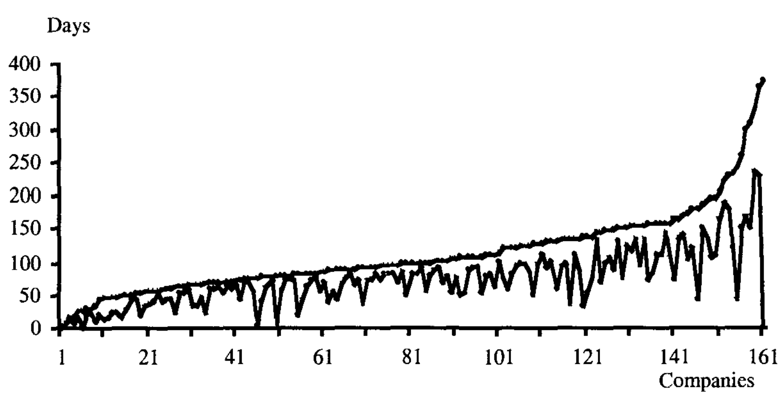
Beyond a certain value, each ratio is considered to be either excessive or negligible. Therefore, all criteria exhibit " bilateral limits " (Céruti, Gatino, [1]). These limits are given in Table 3. Depending on the ratio, when data are recoded in the interval [0, 1] between these limits, three cases may be observed:

- The higher the value of the ratio, the nearer the recoded value will be to 1 (cf. criteria 3 & 4);
- The lower the value of the ratio, the nearer the recoded value will be to 1 (cf. criteria 2, 5 & 6);
- the further the value of the ratio from the average value, the nearer the recoded value will be to 0 (cf. criterion 1)

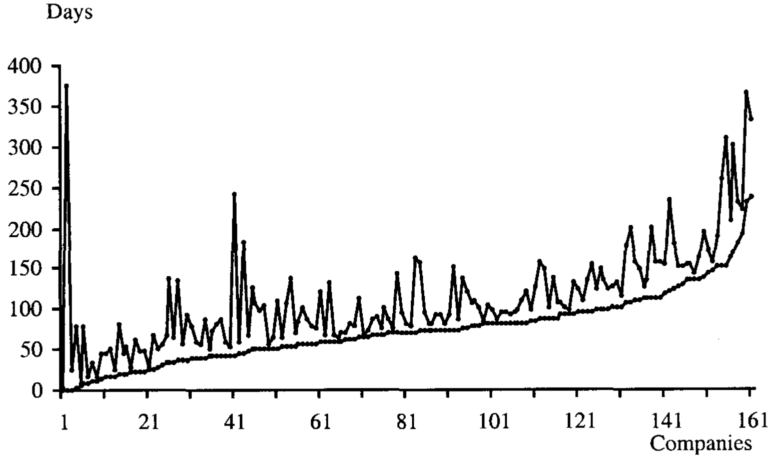
Table 3. Admissible limits for fluctuations of ratios

	Minimum	Maximum
C_1	30	150
C_2	30	140
C_3	45	120
C_4	0,1	0,5
C_5	-600	+1000
C_6	-0,15	0,5

Graph 2. Minimum storage times



Graph 3. Maximum storage times



The principle of adjustment of model 1 was already described and tested on two samples of companies in a preceding work (Couturier, Fioleau [2] [3]). For the model using the equation (2) parameters a and b are determined by the system:

$z = ax + b$ with $z = -\log(-\log(y))$

By setting two thresholds (minimum, maximum) we obtain, for example for the criterion 3 the following system:

$x = 45 y = 0.05$
$x = 120 y = 0.95$
or :
$y = 0.05 z = -1.09719$
$y = 0.95 z = 2.9702$
then :
$-1.09719 = 45a + b$
$2.9702 \quad = 120a + b$

$a = 0.054232$
$b = -3.53764$

So, for this criterion the norm is 0.45 on a scale [0, 1].

For each criterion the experton is built by using for each company the minimum and the maximum of the intervals observed during the period.

4 Results and comments

The pertinence of companies' behaviour to a norm profile for their economic sector was measured according to norms determined from the average activity profile of the 20 best-performing companies in the sector. These norm values are set out in Table 4.

Table 4. Norm values for the six ratios

	Norms
C_1	86
C_2	76
C_3	63
C_4	0,30
C_5	350
C_6	0,22

Table 5 synthetizes the results obtained[2].

Table 5. Position of criteria on a scale [0, 1]

Criteria	Type of ajusting	a	b	Position on a scale [0, 1]	Expertons	
					mini	max
C_1	symmetrical	150	50	***0.7***	0.65	0.8
C_2	decreasing	-0.0369	4.07949	***0.75***	0.7	0.8
C_3	increasing	0.0542	-3.53764	***0.4***	0.3	0.5
C_4	increasing	0.0025	0.428	***0.75***	0.7	0.8
C_5	decreasing	-10.1685	3.98705	***0.7***	0. 7	0.8
C_6	decreasing	-6.25754	2.03197	***0.6***	0.55	0.7

The preceding table indicates :
- the type of adjustment which was carried out,
- coefficients of adjustments a and b,
- the average of experton's highest and lowest limits,

[2] The method proposed by Céruti and Gatino [1] consists in adjusting the data to a symetrical parametric distribution of the following type: $y = e^{-\left|\frac{2x-a}{b}\right|^n}$ with $n = 2.14$

- the measure of the norm value for each criterion.

Thus for criterion 5, 0.7 is the measure of the norm 350.

0,7∈[0, 1] can be interpreted as shown in table 6 as " fairly strong "

The observation of various positionings of the norms for the six criteria highlights the fact that it seems to exist a relatively strong adequacy of the behavior of the companies to the norms of the sector. It can be seen that all the recoded values lie between 0.6 and 0.75 (with the exception of criterion 3) corresponding to a degree of adequation of the behaviors that one can describe, as " somewhat true " or " almost true " in comparison with the norms of the sector.

Table 6. Valuation of a predicate

v(A)	Interpretation
0	false
0.1	quasi false
0.2	almost false
0.3	fairly false
0.4	somewhat false
0.5	neither false nor true
0.6	somewhat true
0.7	fairly true
0.8	almost true
0.9	quasi true
1	true

The preceding application shows that with the chosen adjustment model, when it is applied to a group of companies in the same sector, enables to obtain a fuzzy evaluation of the level and structure of their working capital requirements in relation to sector based norms. This method may also be applied to other situations which companies frequently encounter. It can be used, for example, to measure the effectiveness of and advertising campaign, or recruitment in regard to predetermined profile.

The data treated in this experimentation are of a symmetrical character. A possible further application of this method is to apply it to assymetrical data : to do so, an adjustment by blocks must be carried out. In this case, data should be divided into intervals with curves of different parameters.

References

1. Cérutti O.and Gattino B., Indicateurs et tableaux de bord. (AFNOR Gestion. Paris. 1997)
2. Couturier A. and Fioleau B., Expertons and management decision making. Criteria interdependence and implications valuation *Fuzzy Economic Review.* N° 1 **1**. (1996) p. 31-46.
3. Couturier A.and Fioleau B., Stable set internally maximal: a classification method with overlapping in *Operational Tools in the Management of Financial Risks,* C. Zopounidis (ed.), (Kluwer. Academic Publishers, Dordrecht. 1998). p. 91-106.
4. Gil Aluja J. , Elementos par una teoría de la decisión en incertidumbre. (Ed. Milladoiro. Santiago de Compostella. 1999).
5. Kaufmann.A. , Les expertons. (Hermes. Paris. 1987).
6. Kaufmann.A., Nouvelles logiques pour l'intelligence artificielle. (Hermes. Paris. 1987)
7. Kaufman A. and Gil Aluja J., Téchnicas especiales para la gestión de expertos. (Ed. Milladoiro. Santiago de Compostella. 1993)

MULTIPLE FUZZY IRR IN THE FINANCIAL DECISION ENVIRONMENT

GONZÁLEZ SANTOYO F., FLORES ROMERO J., FLORES ROMERO B., MENDOZA RAMÍREZ J.

Universidad Michoacana de San Nicolás de Hidalgo
Morelia, Michoacán, México

email:{fsantoyo, juanf}@zeus.ccu.umich.mx

In this article we present a new fuzzy methodology to determine multiple IRRs, based on J.T.C. Mao's algorithm. Also, we present an alternative algorithm that exhibits a high level of efficiency and efficacy to solve the multiple IRR problem. The analysis and algorithms presented here have not been reported so far in the fuzzy literature.

1 Introduction

Terceño Gómez A. et al [Terceño 1999] define that the goal of any evaluation of an investment project is to determine a measure of the investment. That measure is an indicator that leads to a decision to reject or accept the investment.

The financial evaluation of any company requires [González 1998] the determination of the cash flow, the planning horizon (lifetime), the interest rate, and the behavior of the cash flow with time to efficiently guide the evaluation criteria to be applied.

According to Mendoza [Mendoza 2000], all companies search the efficient assignment of financial resources (the necessary assets to be productive), pursuing the goal at long term, from a financial perspective. Many investment projects can be justified, but not all of them can be accomplished. That is the main reason to establish a hierarchy and select to most profitable ones. To reach this goal, you need to evaluate each of the multiple investment possibilities present to the company at a given moment.

The traditional criteria are efficient when the information is well-behaved or it can be analyzed with probabilities. Nevertheless, this perception has taken place in several occasions [Gil 1998], through reasoning based in the concept of precision and have been formalized through the classical mathematical schemes. The result is a set of models that constitute a modified reality that adapts to our mathematical knowledge, instead of the other way around, an adaptation of model to the facts. That is the reason why the main mathematical tool to handle uncertainty is fuzzy theory, with all of its variants. On the other hand, we treat likelihood with probability theory. Fuzzy theory has not been well known, or even unknown, as much by the pure mathematicians as by the specialists in applied mathematics. Recently, there has been a change and we now know better how to separate and associate, when necessary, what is measurable, and what is not.

In this paper, we present an analysis of investment evaluation in the presence of multiple financial decisions in a fuzzy environment.

2 Fuzzy IRR with multiple cash flow

In this paper, we will use fuzzy cash flow and interest rates to determine the Internal Revenue Rate (IRR). The analysis uses the fuzzy number criteria in the analysis.

Carlsson and Fuller [Carlsson 2000] define: a fuzzy number A is a fuzzy set of the Real line with a normal convex membership function of bounded support. The set of fuzzy numbers is denoted by **F**. A fuzzy number with a single maximal element is called a quasi-triangular fuzzy number. A fuzzy set A is called a symmetric triangular fuzzy number with center a and width $\alpha > 0$ if its membership function has the following form:

$$A(t) = \begin{cases} 1 - \dfrac{|a-t|}{\alpha} & \text{if } |a-t| \leq \alpha \\ 0 & \text{otherwise} \end{cases} \tag{1}$$

Following Carlsson and Fuller, we will use the notation A=(a, α) to denote such symmetric triangular fuzzy number. If α=0 then A collapses to the characteristic function of $\{ a \} \subset IR$, and we write A = a. A triangular fuzzy number with center a may be seen as a fuzzy quantity "x is approximately equal to a".

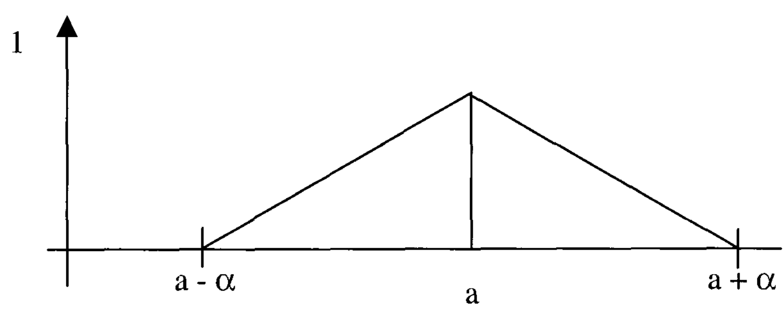


Figure 1. Representation of a quasi-triangular fuzzy number

This will be the basic information representation for the fuzzy cash flow. To compute the multiple IRR the projects need to be classified into simple investments (which implies a single IRR), pure not simple investments (a single IRR), and mixed not simple investments (several IRRs).

3 Multiple fuzzy IRR

This criteria is used in the evaluation of investment projects where there are two or more interest rates for the same proposal. This happens due to sign changes in the cash flows within the planning horizon, which in turn is due to the fact that cash expenditures are not restricted to the beginning of the investment, but can happen through their lifetimes. Graphically, this can be depicted as shown in Figure 2.

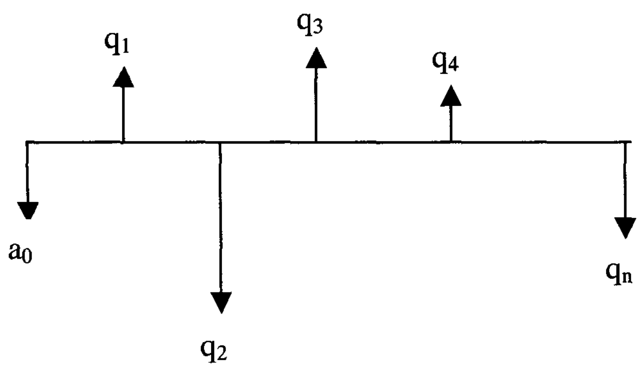


Figure 2. A project's cash flow

Investments can be classified in simple and non-simple ones. Simple investments can have only one sign change; this guarantees a single interest rate. Non-simple investments are classified in two sets: pure and mixed. Pure investments also have a single IRR despite the sign changes. Mixed investments can have several IRRs. The complexity of mixed investments lies in the existence of multiple IRRs.

To provide an answer to projects with multiple solutions, we present two alternatives. The first one is an extension of the fuzzy theory for the IRR and the second one is the fuzzification of James C. T. Mao's algorithm.

4 First criteria

The first thing that must be done is the application of the equations to find the fuzzy IRR r =[$\underline{r(\alpha)}, \overline{r(\alpha)}$] such that it yields a zero Neat Present Value (NPV). The equation that represents the fuzzy IRR is:

$$\overline{a_o} = \sum_{i=1}^{n} \overline{q_i}(1+\overline{r})^{-i} \tag{2}$$

where $\overline{q_i}$ represents the cash flow for period i. For cases where there are sign changes the equation is:

$$\overline{a_0} = \pm\sum_{i=1}^{n} \overline{q_i}(1+\overline{r})^{-i} \tag{3}$$

To determine the lower and upper bounds of the α-cut, we use equations 4 and 5, respectively

$$\underline{a_0(\alpha)} = \pm\sum_{i=1}^{n} \underline{q_i(\alpha)}[1+\underline{r(\alpha)}]^{-i} \tag{4}$$

b) To find the upper bound of the alpha cut:

$$\overline{a_0(\alpha)} = \pm\sum_{i=1}^{n} \overline{q_i(\alpha)}[1+\overline{r(\alpha)}]^{-i} \tag{5}$$

With the obtained result, that is, the triangular approximation $r = \left[\underline{r(\alpha)}, \overline{r(\alpha)}\right]$, you evaluate the pending balances of the investment project.

Under this criterion, a pure investment is defined as an investment where the pending balances evaluated with the triangular approximation $r = \left[\underline{r(\alpha)}, \overline{r(\alpha)}\right]$ are negative or zero, through the proposal's lifetime and positive at the end. That way, an investment is pure if and only if $r = \left[\underline{r(\alpha)}, \overline{r(\alpha)}\right] \leq 0$, t=0, 1, 2, …, n-1. On the other hand, a mixed investment is a project where $r = \left[\underline{r(\alpha)}, \overline{r(\alpha)}\right] \geq 0$ for some values of t, and $r = \left[\underline{r(\alpha)}, \overline{r(\alpha)}\right] \leq 0$ for the rest.

The procedure to determine the conditions that define pure and mixed investments are established in steps 3 and 4 of the algorithm shown in Figure 3.

1.- Apply equations for the multiple fuzzy IRR:

$$a_0 = \pm\sum_{i=1}^{n} q_i (1+\bar{r})^{-i}$$

For the lower bound of the α-cut:

$$\overline{a_0(\alpha)} = \pm\sum_{i=1}^{n} \underline{q_i(\alpha)}[1+\underline{r(\alpha)}]^{-i}$$

For the upper bound of the α-cut:

$$\underline{a_0(\alpha)} = \pm\sum_{i=1}^{n} \overline{q_i(\alpha)}[1+\overline{r(\alpha)}]^{-i}$$

2.- Determine the triangular approximation:
$r = \left[\underline{r(\alpha)}, \overline{r(\alpha)}\right]$

3.- If $r = \left[\underline{r(\alpha)}, \overline{r(\alpha)}\right] \leq 0$ t= 0,1,2,.....n-1; the investment is pure.

4.- If $r = \left[\underline{r(\alpha)}, \overline{r(\alpha)}\right] \leq 0$ for some values of t and $r = \left[\underline{r(\alpha)}, \overline{r(\alpha)}\right] \geq 0$ for the rest, then the investment is mixed.

Figure 3. Alternate Algorithm

5 Fuzzification of James C.T. Mao's algorithm

Since the initial investment is a negative cash flow, any investment can be forced to satisfy the condition that pending balances $F_t(i) < 0$ $(t = 0,1,2, \ldots, n-1)$ by increasing i to a critical value known as $\overline{r_{min}}$. To achieve the fuzzyfication of this second criteria, $\overline{r_{min}}$ must be determined for each of the components of a fuzzy triangular number. The obtained result can be used to evaluate the pending balances of the project, such that $\overline{r_{min}} = \left(\overline{r_{min1}}, \overline{r_{min2}}, \overline{r_{min3}}\right)$. The expression to obtain the pending balances f the investment is $F_t\left(\overline{r_{min}}\right) = q_{t-1}\left(1+r^*\right) + q_t$. So the pending balances must always satisfy the condition $F_t(\overline{r_{min}}) \leq 0$ $t = 0,1,2, \ldots, n-1$.

Once evaluated the pending balances and the condition given above is satisfied, you need to determine if the investment is pure or mixed. You use the obtained value for $\overline{r_{min}}$ to evaluate the final result of the project's planning horizon. If

$F_n(\overline{\mathrm{r}_{\min}}) > 0$, then the investment is pure. If we determine that $F_n(\overline{\mathrm{r}_{\min}}) < 0$, the investment is mixed. In the case of a pure investment no multiple IRRs exist. If the investment is mixed, we need to determine r^* (the revenue of the invested assets) such that $F_n(r^*, TREMA) = 0$[1]. If r^* > TREMA, the project is accepted.

The revenue of the invested assets r^* can be represented as a fuzzy triangular number, $\overline{r^*} = (\overline{r_1^*}, \overline{r_2^*}, \overline{r_3^*})$. To determine r^*, we start the computation with TREMA, so we can determine whether to accept or reject the project. Besides, for mixed investments, the revenue varies with TREMA.

1.- Iteratively determine $\overline{\mathrm{r}_{\min}}$.

$\overline{\mathrm{r}_{\min}} = (\mathrm{r}_{\min 1}, \mathrm{r}_{\min 2}, \mathrm{r}_{\min 3})$

2.- $\overline{\mathrm{r}_{\min}}$ must satisfy $F_t(\overline{\mathrm{r}_{\min}}) \leq 0$ t = 0,1,2,......, n-1

3.- If $F_n(\overline{\mathrm{r}_{\min}}) > 0$, then the investment is pure, so there exists a single IRR, and must be compared with the TREMA.

4.- If $F_n(\overline{\mathrm{r}_{\min}}) < 0$, then the investment is mixed

5.- Determine $\overline{r^*}$, such that:

$F_n(\overline{\mathrm{r}^*}) = 0$

$\overline{r^*} = (r_1^*, r_2^*, r_3^*)$

6.- If $\overline{r^*}$ > TREMA, the project is accepted.

Figure 4. J.T.C. Mao's Algorithm

6 Application cases of fuzzy theory in multiple investments

To illustrate both algorithms, in this section, we present two cases with multiple investments.

[1] The term TREMA comes from Spanish "Taza de Recuperación Mínima Atractiva", and can be translated as Capital Cost. Throughout the paper, it will be treated as TREMA.

6.1 Case 1

Company "W" is analyzing an investment proposal. The investment analysis is being handled by a consulting group of experts. The amounts of cash flow have been defined, for a planning horizon of 2 years, as fuzzy triangular numbers, as shown in Table 1.

Table 1. Cash Flow (case 1)

Year	Cash Flow
0	(-4 000, -4 000, -4 000)
1	(24 500, 25 000, 25 300)
2	(-24 000, -25 000, -26 000)

The company has a minimum reference rate (TREMA) of 20% ad wants to know whether it is convenient or not to accept the project, using the IRR criteria.

The first thing to be done in the analysis of an investment proposal, is to verify if there exist sign changes in cash flows. In this case sign changes do occur, since we have withdraws at the beginning and at the end of the project; we also have incomes in the first year. There is a cash flow behavior with a pattern (-, +, -), so it is possible to find multiple IRRs.

According to the first criteria, we apply equation 3 for multiple fuzzy IRRs, yielding

$$\overline{a_0} = \pm\sum_{i=1}^{n} \overline{q}_i (1+\overline{r})^{-i} \tag{6}$$

Next, we determine the lower and upper bounds of the α-cut, using equations 4 and 5. Substituting the values for each of the cases, we have:

$$\underline{a_0(\alpha)} = 4000 = \frac{[24500+500(\alpha)]}{[1+\underline{r(\alpha)}]} \quad \frac{[24000+1000(\alpha)]}{[1+\underline{r(\alpha)}]^2} \tag{7}$$

$$\overline{a_0(\alpha)} = 4000 = \frac{[2530 \quad 300(\alpha)]}{[1+\overline{r(\alpha)}]} \quad \frac{[26000 \quad 1000(\alpha)]}{[1+\overline{r(\alpha)}]^2} \tag{8}$$

Doing the corresponding computations by an iterative process, we obtain the following results, for each of the α-cuts

Table 2. Results for each alpha-cut

α	$\underline{r}(\alpha)$	$\overline{r(\alpha)}$
1	25.00 %	25.00 %
0.9	24.75 %	25.42 %
0.8	24.50 %	25.83 %
0.7	24.24 %	26.25 %
0.6	23.99 %	26.66 %
0.5	23.73 %	27.07 %
0.4	23.47 %	27.49 %
0.3	23.22 %	27.90 %
0.2	22.95 %	28.31 %
0.1	22.69 %	28.72 %
0	22.43 %	29.13 %

The triangular approximation we get is

$$r = \left[\underline{r}(\alpha), \overline{r(\alpha)}\right] = (22.43\% \,,\, 25\% \,,\, 29.13\%) \tag{9}$$

Due to the sign change, the system also exhibits the results shown in Table 3, having the triangular approximation $r = \left[\underline{r}(\alpha), \overline{r(\alpha)}\right]$ = (390.06% , 400%, 403.37%).

Table 3. Results

α	$\underline{r}(\alpha)$	$\overline{r(\alpha)}$
1	400.00 %	400.00 %
0.9	398.99 %	400.34 %
0.8	397.89 %	400.67 %
0.7	397.00 %	401.02 %
0.6	396.01 %	401.34 %
0.5	395.01 %	401.68 %
0.4	393.85 %	402.00 %
0.3	393.02 %	402.34 %
0.2	392.04 %	402.69 %
0.1	391.05 %	403.03 %
0	390.06 %	403.37 %

As you can see, this is a mixed investment, since there exist two results to solve these equations.

To complete our analysis, and to verify the first criterion, let us assume the first result to be the triangular approximation r = [22.43%, 25.00%, 29.13%]. With this result, we evaluate the pending balances of the investment project.

$$\begin{aligned} F_0(0.2243, 0.25, 0.2913) &= (-4\,000, -4\,000, 4\,000) \\ F_1(0.2243, 0.25, 0.2913) &= (19\,602.80, 20\,000, 20, 134.80) \\ F_2(0.2243, 0.25, 0.2913) &= (0, 0, 0) \end{aligned} \tag{10}$$

There exist some values for $\overline{F_m} \leq 0$ for years 0 and 2, and $\overline{F_m} \geq 0$ for year 1, so the investment is mixed.

To demonstrate the method's consistency, we verify the results with the other triangular approximation we got

$$r = \left[\underline{r(\alpha)}, \overline{r(\alpha)}\right] = (390.06\%, 400\%, 403.37\%) \tag{11}$$

The pending balances for each year can be computed as follows:

$$\begin{aligned} F_0(3.9006, 4, 4.0337) &= (-4\,000, -4\,000, -4\,000) \\ F_1(3.9006, 4, 4.0337) &= (4\,897.60,\ 5\,000,\ 5\,165.20) \\ F_2(3.9006, 4, 4.0337) &= (0,\ 0,\ 0) \end{aligned} \tag{12}$$

There exist some values that verify $\overline{F_m} \leq 0$ and others for $\overline{F_m} \geq 0$, so the investment is mixed.

No matter what result we used, we have come to determine we are dealing with a mixed investment. The analyst has now to determine whether to accept or reject the project. When there are multiple IRRs, that decision must agree with the value of the TREMA.

$$r = \left[\underline{r(\alpha)}, \overline{r(\alpha)}\right] > \text{TREMA} \tag{13}$$

For our case:

$$r = (22.43\%, 25\%, 29.13\%) > \text{TREMA} = (20\%) \tag{14}$$

So the project must be accepted.

According to the second criteria, the results we get must be the same as with the first criteria. First, we need to determine the value of rmin for year 0 and 1.

$$\overline{r_{min}} = \left(\overline{r_{min1}}, \overline{r_{min2}}, \overline{r_{min3}}\right) \tag{15}$$

For $\overline{r_{min1}}$:

$$-4\,000\ (1+\overline{r_{min1}}) + 24\,500 = 0$$
$$(1+\overline{r_{min1}}) = 6.125 \tag{16}$$
$$\overline{r_{min1}} = 5.125$$

For $\overline{r_{min2}}$:

$$-4\,000\ (1+\overline{r_{min2}}) + 25\,000 = 0$$
$$(1+\overline{r_{min2}}) = 6.25 \tag{17}$$
$$\overline{r_{min2}} = 5.25$$

For $\overline{r_{min3}}$:

$$-4\,000\ (1+\overline{r_{min3}}) + 25\,300 = 0$$
$$(1+\overline{r_{min3}}) = 6.325 \tag{18}$$
$$\overline{r_{min3}} = 5.325$$

So we have $\overline{r_{min}}$ = (5.125, 5.25, 5.325). We use this value to determine the pending balances to fulfill the condition $F_t\left(\overline{r_{min}}\right) \leq 0$ t = 0, 1, 2, …, n-1.

$$\begin{aligned} F_0(5.125,5.25,5.325) &= (-4\,000,-4\,000,-4\,000) \\ F_1(5.125,5.25,5.325) &= (0,0,0) \\ F_2(5.125,5.25,5.325) &= (-24\,000,-25\,000,-26\,000) \end{aligned} \tag{19}$$

The condition has been satisfied, and we can continue with our example.

The balance at the end of the project is negative (we determined the values -24 000, -25 000, -26 000). We can say the investment is mixed, since $F_n\left(\overline{r_{min}}\right) < 0$.

In the last step we determine the value of $\overline{r^*}$. We start with the TREMA, so $\overline{r^*}$ = TREMA = (20%, 20%, 20%)

$$\begin{aligned} F_0(0.20,0.20,0.20) &= (-4\,000,-4\,000,-4\,000) \\ F_1(0.20,0.20,0.20) &= (19\,700,20\,200,20\,500) \\ F_2(0.20,0.20,0.20) &= (-360,-760,-1\,400) \end{aligned} \tag{20}$$

Since the obtained values are negative, we need to increase the TREMA to get to know the triangular approximation. If $\overline{r^*}$ =30%

$$\begin{aligned} &F_0(0.30, 0.30, 0.30) = (-4\,000, -4\,000, -4\,000) \\ &F_1(0.30, 0.30, 0.30) = (19\,300, 19\,800, 20\,100) \\ &F_2(0.30, 0.30, 0.30) = (1\,090, 740, 130) \end{aligned} \tag{21}$$

Since we have gotten negative and positive results, we can now interpolate to obtain:

$$\overline{r^*} = (22.43\%, 25\%, 29.13\%) \tag{22}$$

The results are similar to those determined using the first criteria, so we can use either method. It is obvious that the company's decision must be to accept the project, since $\overline{r^*}$ > TREMA.

6.2 Case 2

Company "X" is analyzing the following investment proposal, provided by the expert group of the corporation.

Table 4. Cash Flow (case 2)

Year	Cash Flow
0	(-600, -600, -600)
1	(750, 800, 830)
2	(-590, -600, -650)
3	(680, 700, 710)
4	(100, 200, 285)

The company is interested to know if the investment is classified as pure or mixed, and to obtain the triangular approximation to decide whether it is convenient or not to accept the project. The company uses a TREMA of 25% to take decisions. As you can see in Table 6, the cash flow pattern is of the form (-, +, -, +, +), so there may exist multiple IRRs.

Applying equations 4 and 5 to determine the lower and upper bounds of the α-cut, and substituting values we get

$$600 = \frac{[750+50(\alpha)]}{[1+\underline{r(\alpha)}]} - \frac{[590+10(\alpha)]}{[1+\underline{r(\alpha)}]^2} + \frac{[680+20(\alpha)]}{[1+\underline{r(\alpha)}]^3} + \frac{[100+100(\alpha)]}{[1+\underline{r(\alpha)}]^4} \tag{23}$$

$$600 = \frac{[830-30(\alpha)]}{[1+\overline{r(\alpha)}]} - \frac{[650-50(\alpha)]}{[1+\overline{r(\alpha)}]^2} + \frac{[710-10(\alpha)]}{[1+\overline{r(\alpha)}]^3} + \frac{[285-85(\alpha)]}{[1+\overline{r(\alpha)}]^4} \tag{24}$$

The obtained results are shown in Table 5.

Table 5. Results of the alpha-cuts

α	$\underline{r(\alpha)}$	$\overline{r(\alpha)}$
1	36.08 %	36.08 %
0.9	35.14 %	36.40 %
0.8	34.21 %	36.72 %
0.7	33.26 %	37.03 %
0.6	32.32 %	37.35 %
0.5	31.35 %	37.66 %
0.4	30.38 %	37.97 %
0.3	29.40 %	38.28 %
0.2	28.41 %	38.58 %
0.1	27.41 %	38.89 %
0	26.39 %	39.19 %

Consequently, the triangular approximation is:

$$r = \left[\underline{r(\alpha)}, \overline{r(\alpha)}\right] = (26.39\%, 36.08\%, 39.19\%) \tag{25}$$

Comparing this result with the TREMA, we conclude that the project must be accepted, since $r = \left[\underline{r(\alpha)}, \overline{r(\alpha)}\right] >$ TREMA. That is, (26.39%, 36.08%, 39.19%) > 25%.

The second criterion to solve cases with multiple IRRs indicates we need to compute $\overline{r_{min}}$. So for years 0 and 1,

For $\overline{r_{min1}}$:

$$\begin{aligned} &-600\ (1+\overline{r_{min1}}) + 750 = 0 \\ &(1+\overline{r_{min1}}) = 1.25 \\ &\overline{r_{min1}} = 0.25 \end{aligned} \tag{26}$$

For $\overline{r_{min2}}$:

$$-600\ (1+\overline{r_{min2}}) + 800 = 0$$

$$(1+\overline{r_{min2}}) = 1.3333 \tag{27}$$
$$\overline{r_{min2}} = 0.3333$$

For $\overline{r_{min3}}$:

$$-600\ (1+\overline{r_{min3}}) + 830 = 0$$
$$(1+\overline{r_{min3}}) = 1.3833 \tag{28}$$
$$\overline{r_{min3}} = 0.3833$$

So we have:

$$\overline{r_{min}} = \left(\overline{r_{min1}}, \overline{r_{min2}}, \overline{r_{min3}}\right) = (0.25, 0.3333, 0.3833) \tag{29}$$

With the obtained value we need to determine the pending balances to satisfy the condition $F_t\left(\overline{r_{min}}\right) \leq 0$ t = 0,1,2,..........,n-1

$$\begin{aligned}
&F_0(0.25, 0.3333, 0.3833) = (-600, -600, -600)\\
&F_1(0.25, 0.3333, 0.3833) = (0, 0, 0)\\
&F_2(0.25, 0.3333, 0.3833) = (-590, -600, -650)\\
&F_3(0.25, 0.3333, 0.3833) = (-57.50, -100, -189.16)\\
&F_4(0.25, 0.3333, 0.383) = (28.13, 66.67, 23.32)
\end{aligned} \tag{30}$$

These results indicate that condition $F_t\left(\overline{r_{min}}\right) \leq 0$ for t = 0, 1, 2, ...,n-1, is satisfied. Besides, the final balance of the project is (28.13, 66.68, 23.33), so we can conclude we have a pure investment, in spite of the sign changes in the project's lifetime. There exists a single IRR, since $F_n\left(\overline{r_{min}}\right) > 0$.

As a final step, we determine $\overline{r^*}$, starting with the TREMA,

$$\overline{r^*} = \text{TREMA} = (25\%, 25\%, 25\%) \tag{31}$$

$$\begin{aligned}
&F_0(0.25, 0.25, 0.25) = (-600, -600, -600)\\
&F_1(0.25, 0.25, 0.25) = (0, 50, 80)\\
&F_2(0.25, 0.25, 0.25) = (-590, -537.50, -550)\\
&F_3(0.25, 0.25, 0.25) = (-57.50, 28.13, 22.50)\\
&F_4(0.25, 0.25, 0.25) = (28.13, 235.16, 313.13)
\end{aligned} \tag{32}$$

The computed value for $\overline{r^*}$ is $\overline{r^*} = (26.39\%, 36.08\%, 39.19\%)$
To verify, we obtain:

$$
\begin{aligned}
&F_0(0.2639, 0.3608, 0.3919) = (-600, -600, -600)\\
&F_1(0.2639, 0.3608, 0.3919) = (-8.34, -16.48, -5.14)\\
&F_2(0.2639, 0.3608, 0.3919) = (-600.54, -622.43, -657.15)\\
&F_3(0.2639, 0.3608, 0.3919) = (-79.02, -147.00, -204.69)\\
&F_4(0.2639, 0.3608, 0.3919) = (0, 0, 0).
\end{aligned} \tag{33}
$$

Since $\overline{r^*}$ = (26.39%, 36.08%, 39.19%) > TREMA = (25%), we conclude that the project must be accepted.

7 Conclusions

Fuzzification of J.T.C. Mao's Algorithm shows to obtain efficient and high-quality solutions for multiple IRRs.

The new methodology, proposed here, has been tested with a classical case and shows a higher computational efficiency than J.T.C. Mao's algorithm. The efficiency was measured by the number of iterations. The proposed algorithm provides high quality in the results for the fuzzy evaluation case.

The alternate algorithm shows high efficiency for he solution of problems with multiple IRRs. This makes it competitive with the state of the art for the deterministic case. It also allows the extension for the fuzzy case. To the best knowledge of the authors, this algorithm, applied to multiple IRRs, has not been reported in the literature.

References

1. [Terceño 1999]. Terceño Gómez A., Saez B.J. *Criterios de valoración de inversiones*. Notas de la Facultat de Ciences Economiques i Empresarials. Universitat Rovira i Virgili. Reus, España, 1999.
2. [González 1998]. González Santoyo F. *Fuzzy sets versus enfoques tradicionales en la evaluación de inversiones*. Tesis de grado Maestro en Administración. DEPFCA - UMSNH, Morelia, Michoacán, México, 1998.
3. [Mendoza 2000]. Méndoza Ramírez Jorge A. *Decisiones Financieras múltiples en ambiente difuso*. Tesis de grado Maestro en Administración. DEPFCA - UMSNH, Morelia, Michoacán, México, 2000.
4. [Carlsson 2000]. Carlsson Christer, Fuller Robert. *On Fuzzy Internal Rate of Return*. Turku Centre for Computer Science, TUCS technical report No. 211. Finlandia, 2000.
5. [Buckley 1992]. Buckley J.J. *Solving fuzzy equations*. Fuzzy sets and systems, 50, 1-14. 1992.

6. [Buckley 1990]. Buckley J.J., Qu, Y. *Solving linear and quadratic fuzzy equations*. Fuzzy sets and systems 38, 43-59, 1990.
7. [Mao 1986]. Mao, J. *Análisis financiero.* Buenos Aires. El Ateneo, 1986.
8. [Gil 1998]. Gil Aluja J. *Nuevas técnicas para la gestión de empresas.* Separata Management & Empresa, 289-313. Universitat Barcelona, 1998.

This page is intentionally left blank

AN AUTOMATED KNOWLEDGE GENERATION APPROACH FOR MANAGING CREDIT SCORING PROBLEMS

M. MICHALOPOULOS, D. HATAS AND C. ZOPOUNIDIS

Technical University of Crete, Dept of Production Engineering and Management, Financial Engineering Laboratory, Chania 73100, Greece, kostas@ergasya.tuc.gr

G. DOUNIAS

University of the Aegean, Dept of Business Administration, Chios 82100, Greece

The aim of this article is to propose an example-based intelligent approach for classifying enterprises into different categories of credit risk. The data used are of both numerical and linguistic nature. The methodology used for the rule based categorization task is the well-known inductive machine learning approach, based on entropy information. The drive for this paper was the application domain, which is very common in banking management worldwide, moreover very often turns to be a confusing and time-consuming situation. The goal is to obtain a model that correctly classifies a training sample of 130 enterprises with 76 decision variables to the predetermined classes, using a substantially less amount of attributes, trying at the same time to minimize the error rate. Data are transformed into a proper input database, and then training experiments take place in order to find the optimal settings for the training phase. After having obtained the right adjustments, the training phase initiates. The aim is to produce a comprehensible decision tree as an output, which can then be transformed to a set of simple IF/THEN rules. The specific decision tree produced from the training data, uses only 16 attributes to be formed, and is equivalent to a few comprehensible and short rules consisting of 2 to 10 premise parts. As a result, the classification task is made easier to perform and check, the amount of required data is minimized, and finally the whole process is easier to use in decision-making. Furthermore, the produced decision tree works as a knowledge generator and thus, reflects the banking organization's expertise on the application domain, represented by a handy and meaningful set of rules. Finally, the classifier could be continuously reformed, by adding to it every new credit-risk case, becoming a more and more accurate and robust classification model with time.

1 Introduction

In the very recent years, a rise in the use of different kinds of financial transactions is observed, either in the banking level, or in the private business sector, a fact that makes imperative the automation of some decision making processes. However, this automation should not incorporate high error rates during the data processing phase, while it should be also as less time consuming as possible. Furthermore, the decisions taken with use of an automated process, should approximate as much as possible the corresponding decision of an expert for the same problem. This is the reason why, methods that imitate human thought and model human expertise and abilities, have become very popular in the last two decades. Usually, all these "intelligent" methodologies for data analysis, decision making, and model based reasoning, are placed under the umbrella of Artificial Intelligence [7]. Machine learning, is that part of artificial intelligence tools and techniques that seems to perform in the most promising way, when dealing with real world applications [16]. Lately, various approaches have been appeared in literature, that work as knowledge generation approaches for real world application domains [1, 5]. Particularly famous as the reader should expect, are those applications dealing with financial domains and business applications [12].

This paper deals with such a financial application domain, that of a loan, which is one of the most commonly used financial transactions in banking organizations. There is a large number of criteria, according to each case, that in practice determine the restrictions and the final outcome of the decision process for granting a loan. For the fundamentals of credit scoring the reader should advice [10]. The subject of credit scoring is not so famous in literature as one should expect [11]. Most of the papers attempt statistical approaches for classifying candidates to be loaned in two classes: positive and negative, see [2, 11, 26]. They usually perform discriminant analysis in order to produce a discriminant rule for these two classes [25]. Other approaches use count data models in order to predict the number of times that an applicant for credit will not pay the accorded amount to return the credit [14]. Other classic methods used to face the credit scoring problem, include classification and regression trees, Markov chains, linear models and linear programming, graphical models, logistic regression, and many more [11, 26]. On the other hand, a limited number of intelligent methodologies has been appeared during the last years focusing on credit scoring, such as neural networks [4], genetic algorithms [4], machine learning [3], as well as data mining and bayesian networks, see also [11, 26].

In that sense, the present paper is not exactly the very first attempt of facing credit risk with the aid of computational intelligence techniques. Still, it is the ultimate one and more sophisticated, as it deals with a strong and robust methodology of inductive learning, equipped with new features and processes for learning from examples, such as boosting, pruning, etc. (see section 4 for details). On the other hand, special attention has been paid through this paper, to carefully model the application domain, a fact that contributes to a high degree, in the final success of the overall attempt. Another clear aim of this paper is to introduce to banking experts, a category of comprehensible new methodologies, for handling a very difficult financial decision making problem. Especially in banks, which are the main loan makers, there is a strong need for quick and valid evaluation of the applicant, in order for some key decisions to be made. Some of them are whether to grant the loan or not, what its amount would be if granted, for how long its maturity should be, what the payback terms would be, in order to ensure the minimum possible risk for the bank. Several bank executives devote much time to this kind of decision making, in an effort to give a scientific and fully understandable way of dealing with the problem that at the same time must fulfil all the requirements that the bank sets for the risk level and provides for speed and clarity of the decisions.

The paper is organized as follows: The basics of the credit scoring application (ACS) are presented in section 2. Then, in section 3 a clear view of the data that constitute the problem is given, as well as the way they are modeled and organized. In section 4 follows inductive learning theory and in section 5 the proposed overall methodology is presented and the acquired results are discussed. Finally, in section 6 are given concluding remarks drawn from the presented approach.

2 Credit Scoring

The basic concept behind the methodology of ACS is that the near future has a high degree of similarity to the near past. One can come to this conclusion even by using his everyday experience, since every new day the things expected to occur as much in regular business as in personal life, do not differ significantly from those occurred the previous day or days. If this conclusion is applied in the field of credit risk assessment, where a bank -or any other loan maker- always tries to grant from a loan given to a new customer, it is quite certain, that each new potential customer will behave similarly to a previous one having the same characteristics with him. For instance, if the customers' demographic and financial characteristics appear a substantial degree of similarity, then all the necessary requirements to reach the above conclusion are met. This projection of the near past in the near future is usually accomplished by using proper statistical models [11, 26], such as multiple regression and discriminant analysis, whose results are valid only if a set of demands and criteria are met (see also section 1).

ACS is a process during which the potential client is graded, with the use of the proper statistic methods [10]. This is accomplished by multiplying the various kinds of information with a different weight according to the field and summing up at the end. This score represents a specific possibility for the potential client to behave in a certain way, according to a clearly stated criterion given by the bank. The information about the client is gathered mainly from the loan application filled in at the time of the request, from previous data that the bank maintains about that client or similar types of clients and finally from an outside information database. The final outcome of the entire process is to automatically reject or approve the application, without the need of the human factor, after having taken into consideration the total of the individual score that the client has accumulated and the cutoff base predetermined by the bank according to its demands.

The history of credit scoring is relatively short and begun in the US. Many executives had tried to find a scientific approach to an evaluation process for credit risk applications based on quantitative criteria, but all efforts had been unsuccessful since there were not any computers that had the processing power of today's personal computers. Henry Wells, an executive of Spiegel Inc., may as well be considered as a pioneer in this field. During the Second World War he used a number of statistical techniques to form a credit scoring system for the purpose of replacing the credit analysts that the companies used at that time. Many experts tried now and then to develop similar methods, but the appearance and activation of the Fair Isaak firm was undoubtedly the key point.

The basic setback in the evolution of credit scoring was to conviction of the executives specializing in the field of credit not to be so reluctant and to stop dealing with the system as if it was something that its solely purpose was to leave them without a job. However, the circumstances that followed in the decades of 70's and 80's, forced all the major financial organizations to use quantitative techniques in order to assess the credit risk in hand. One problem that still remains unsolved during the process of creating the credit scoring system, is that of the use of the data representing the cases already rejected, for which their behavior in case they had not been rejected is not known.

The use of the aforementioned quantitative methods is being spread lately in banking. Its main application areas range from a basic evaluation of credit risk for a potential client using the data gathered from his application to a wide variety of products that covers all the forms of consumer credit and generally every financial activity. On the Internet one can find information[1] ranging from specialized software for a specific application of credit scoring, to instructions on how to achieve great score numbers in order for his application for a consumer credit product to be eligible for grant.

3 Presentation of the problem

In this section is given a complete description of the problem of granting a loan, while the application area and the variables that consist the data set, are presented.

3.1 Description of the problem

The problem analyzed in this paper, lies in the decision making process from the bank's viewpoint, on whether to grant a loan to an enterprise and with what amount of certainty, based only on the data gathered from the application of the potential client. The bank from which were gathered the data, required the potential client to bring forward the following items:

- To fill in the application that the bank branch provided, which required some general fields to be filled in. These fields were information not available through the normal financial statements such as the balance sheet. Examples are, if the company uses a certain kind of software, how many people does it employ, what are the main activities and other information of similar nature.
- A balance sheet.
- An income statement.

---

[1] See the site www.creditscoring.com

- Also with the return of the above documents the bank executive that handles the certain case adds some information gathered from information agencies. This information includes what the firm's position is among its peers, its potential, the skill level of its executives and other similar information.

As it is already obvious the experimental data were gathered from 4 different sources, which had both numerical data in the form of percentages, indexes, or simple numbers, and qualitative ones consisted of terms. In the classic system that banks use up to date, the qualitative data are transformed into numerical with the use of certain transformation scales. Here lies one of the major disadvantages of the conventional methodologies, which try to convert linguistic terms and concepts into numerical equivalents, losing at the meantime the essence of words.

3.2 Application area

As far as the application area is concerned, it is none other than the relevant department of any financial organization that grants loans to firms. In this area highly qualified experts are used, that have a firm knowledge of the problem to deal with, i.e., the granting of a loan to firms that issue accounting books of specific categories[2], and have a lot of field experience. Finally, there is a need for a computer system with increased processing power that would allow the safe and quick processing of the data.

3.3 The data

In order to achieve the goal of managing credit scoring with the use of computational intelligence techniques, a set of data that fulfill the criteria below, should be available:

- To represent the general population
- To be homogenous
- To have as less as possible noise
- There should be no empty fields (although inductive learning deals with them)
- There should be a substantial amount of cases
- All the different classification categories should be covered in the sample.

Having in mind all the above criteria, the phase of gathering data followed, concerning data that were involved in the bank's decision on whether to grant or not a loan to a firm that keeps accounting books of specific categories. The certain bank examined in this paper, has already been using a classic application credit scoring system. The system keeps as input, 76 fields in total. For each of the 130 cases that had been granted a loan during the past from the examined bank, there were 76 attributes that covered an area of up to three years before granting the loan. Below, an analysis follows, of the basic elements of the specimen that allows the bank to

---

[2] in Greek accounting system these categories correspond to A and B class books.

classify them in one of the following classes: "Risk-free", "average risk", "under surveillance", "high risk", "weak" (i.e. impossible to receive a loan).

4 Brief reference to fundamentals of inductive machine learning

The methodology used for producing automated domain-dependent expert knowledge by mining the data of various credit scoring profiles, belongs to the area of inductive machine learning [13]. Inductive learning tools and techniques have been widely applied during the last two decades in various domains [16], for their comprehensibility, as well as for their ability to generalize from processing with large databases and high complexity domains of application [5]. Usually, but not exclusively, the inductive learning approaches construct decision trees [1, 13], by applying an intelligent approach for reducing either the complexity of the search space, or the size of the tree produced, known as the "divide-and-conquer" approach. Regarding the handling of complexity by inductive decision trees, note for example that, according to combinatorial theory, there are more than 10^{13} ways to partition a set containing 20 items, while an inductive tree forms a near perfect classifier for these items in only a few nodes and in a very short time, by observing a number of pre-classified examples for these items [8].

The present work applies Quinlan's approach [19, 20, 21], the most widely used in machine learning for its comprehensibility and simplicity in data processing. The algorithm (former ID3, [19], later called C4.5, [21], and now called C5.0, or See5,[3] depending on the operating system and version), works as follows:
Given (1) a set of observational statements (i.e. attribute value vectors) each of which is assigned to a certain class, and (2) a universe of classes, find a set of discriminative descriptions between classes.
The algorithm leads to the generation of a decision tree, in which leaves are class names and nodes represent attribute-based tests with branches for each possible outcome. Since all available cases belong to different classes, the algorithm attempts to split them into subsets, by the "divide and conquer" principle [21]. The quantitative criterion for splitting the set of the initial statements to subsets and forming the tree, is based on information entropy measurements. Specifically, by selecting one case at random from a set S of cases and denoting that it belongs to some class C_j, the probability (*freq* means frequency) of this message is

$$freq\,(C_j,S)\,/\,|S|$$

and thus, the information that this message conveys, is calculated as

$$-\log_2 [\,freq\,(C_j,S)\,/\,|S|\,]\text{, in bits.}$$

Then the "expected information" is calculated, from such a message pertaining to class membership, also known as the entropy of the set S, which is

---

[3] Download demo version of see5 software and source code of C4.5 / C5.0, from http://www.rulequest.com/See5-examples.html, Application Examples, J.R.Quinlan.

$$\inf o(S) = -\sum_{j=1}^{k} \frac{freq(C_j, S)}{|S|} \times \log_2 \left(\frac{freq(C_j, S)}{|S|}\right) \text{ bits,}$$

also known as the entropy of the set *S*. For the construction of the subsets of the tree, a simple test is being used. Considering a test *X* with n outcomes which partitions a training set *T*, into *i*-subsets T_i, the expected information requirement can be found as the weighted sum over the subsets, as

$$\inf o_X(T) = \sum_{i=1}^{n} \frac{|T_i|}{|T|} \times \inf o(T_i)$$

The known as "gain criterion" then selects a test to maximize the mutual information between the test *X* and the class:

$$gain(X) = info(T) - info_X(T)$$

In order to avoid difficulties with data having a very large number of different values and thus the highest information gain, a sensitive measure -on how broadly and uniformly the attribute splits the data- is used, called *SplitInformation*:

$$SplitInformation(S, A) = -\sum_{i=1}^{c} \frac{|S_i|}{|S|} \log_2 \frac{|S_i|}{|S|}$$

where S_i through S_c are the *c* subsets of examples resulting from partitioning *S* by the *c*-valued attribute *A*. Then the *GainRatio* measure is defined in terms of the Gain measure as well as this *SplitInformation* as follows:

$$GainRatio(S, A) = \frac{Gain(S, A)}{SplitInformation(S, A)}$$

The above algorithmic description is taken from [21], where the reader can find detailed information on the software implementation of the overall approach, as well as detailed descriptions of all the additional functions embodied in the tool. Quinlan's monograph, [21], was published in 1993 containing also methods for pruning the produced tree [20], cross-validation techniques for evaluating results, special features for the handling of continuous valued attributes, unknown attribute values treatment, etc. Specifically, evaluation of the tree output usually takes place by a cross-validation process, where the initial training set is divided *n*-times randomly, to *m*-subsets[4], *m-1* of which are used for training and the *m-th* subset is used to test the classification success on unseen data. Then the mean and variance of the total number of repeated cross-validation tests is calculated, and the learning rate of the decision tree classifier for a given task arises.

Since 1993, although the main idea of this inductive learning scheme has been always remained very same, many additional features have been given in literature. In this sense, the reader should find:

- numerous techniques for effectively proposing ways of shortening (i.e. pruning) the produced decision tree [9],

[4] *n, m*, are user defined, having both a default value of 10.

- alternative criteria for splitting and stopping the algorithm when forming the tree [8], as well as decades of different methods for calculating the partial information gain [15],
- other ideas for handling continuous valued attributes [27],
- use of soft or fuzzy thresholds, in other words an attempt to prevent small change in the continuous attributes from creating a drastic change in the outcome by defining subsidiary cutoff points below and above a threshold value[5], [18],
- methodologies for testing the algorithmic efficiency of inductive learning, by use of cross-validation processes [23], and
- boosting techniques, in other words methodologies for improving the performance of the learning algorithm by combining different classifiers, successively concentrating on the training examples that are difficult to classify correctly [6, 17, 18, 24].

Another very important statement for inductive learning techniques that form decision trees from examples, is that special care should be given in the phase of data collection, encoding, preparation, and cleaning [22], preceding the algorithmic execution and experimentation. The help and support of domain experts is always necessary, in order to model and prepare correctly the problem and set the conditions for a successful and useful, practical outcome.

5 The application

For the application of the inductive learning methodology, the available data supplied from the collaborative bank organization were used, in order to form the classifier's knowledge base. As mentioned before, the sample consisted of 130 firms – cases, with a total of 76 attributes, numerical and linguistic. Data were properly encoded and a specific set of options was used, briefly explained below:

- Subsets of values: Attribute values were grouped into subsets and thus, each produced subtree is associated with a subset of values rather than with a single value (see Fig. 2-3).
- Ignoring hypothetical costs of different misclassification types: as experts stated, all misclassifications are considered of equal importance (though there is such an option within the system for adjusting different misclassification costs).
- Pruning level of 51%: A large tree is first grown to fit the data closely and is then pruned by removing parts that are predicted to have a relatively high error rate or low importance in the final outcome (51% level results in less pruning).
- Minimum of 2 cases: At each ending point of the decision tree, 2 training cases must be assigned (single case decision rules are discarded).

---

[5] For the soft thresholds, see also J.R. Quinlan, (1997), "Is C5.0 better than C4.5?", www.rulequest.com/see5-comparison.

After all the options had been adjusted with the proper values, the tree classifier was constructed. The outcome was a decision tree with a size of 19 branches (see Figure 1, continued in Figures 2 and 3). In these 19 branches only 16 out of the whole 76 attributes can be found. The depth of the produced decision tree varies from 2 to 10 levels. This means that for a certain case to be classified only 2 to 10 attributes are used, which corresponds to a very low complexity solution for the decision maker.

The processing time for the system to extract the classifiers from the database using a conventional PC is very small, only 0.3 sec., and it reaches to an error rate of 6,2%. Out of the 130 cases only 8 were misclassified in total and those 8 cases were given a class one level higher or lower. That fact could suggest that the initial decision might have been wrong, or that it was on the limits between two classes and a human intervention decided the final class.

The decision tree that has been extracted can be used in two different ways. The first idea is to visualize the actual tree using nodes and branches that end to a decision point. In order for a decision to be made, one has only to follow the flow of the tree based on the values of the attributes that are present on it. As stated above the minimum number of attributes would be only 2 and the maximum is again only 10 out of the initial 69. The only practical difficulty is to visualize the decision tree in a way friendly and readable to the employee particularly when it becomes large, but the overall decision process is quite obvious and one can fully understand the bank's policy. An alternative method for utilizing the constructed classifier is to incorporate it to an interactive interpreter, and use it to automatically or manually assign new cases to classes. Since the values of all attributes may not be needed, the attribute values requested will depend on the case itself (2 to 10). When all the relevant information has been entered, the most probable classes are shown, each with a certainty value (this is a standard feature of the tool used). Before the presentation of the actual results, a brief description of the attributes is given and their range of values so that the reader can understand the meaning of the outcome of this application of inductive learning in the specific field of credit scoring.

- Debt / Equity a year ago: Index in the form of percentage (Continuous).
- Annual change in Sales a year ago: Percentage (Continuous).
- Products – Services (Quality): Bad, Average, Good, Exceptional.
- Sector's Net Profit Margin: Index average (%) for the sector (Continuous).
- Net Profit Margin a year ago: Index (%) (Continuous).
- Geographical Coverage: Certain Areas, Local, Widely Local, National.
- Years in business: Years that the firm is in business (Integer).
- Net Income a year ago: The firm's net income (Continuous).
- Sector's Average Inventory (Continuous).
- Number of Products: Number of products-services the firm provides (Cont.).
- Business's Future: How the executives of the bank foresee the firm's future in the sector. Range: Insufficient to Adequate, Adequate, Good, Exceptional.
- Sector's Debt / Equity: Sector index average in percentage (Continuous).
- Security Margin 2 years ago: Index in the form of percentage (Continuous).

- Sector's Accounts Receivable: (Continuous).
- Accounts Receivable a year ago: (Continuous).
- Quick Ratio: Index in the form of percentage (Continuous).

Note here that all the numbers are based on calculations done with million of drachmas. The output of the inductive learning software tool that was used after all the options had been set, is represented in Figures[6] 1,2,3.

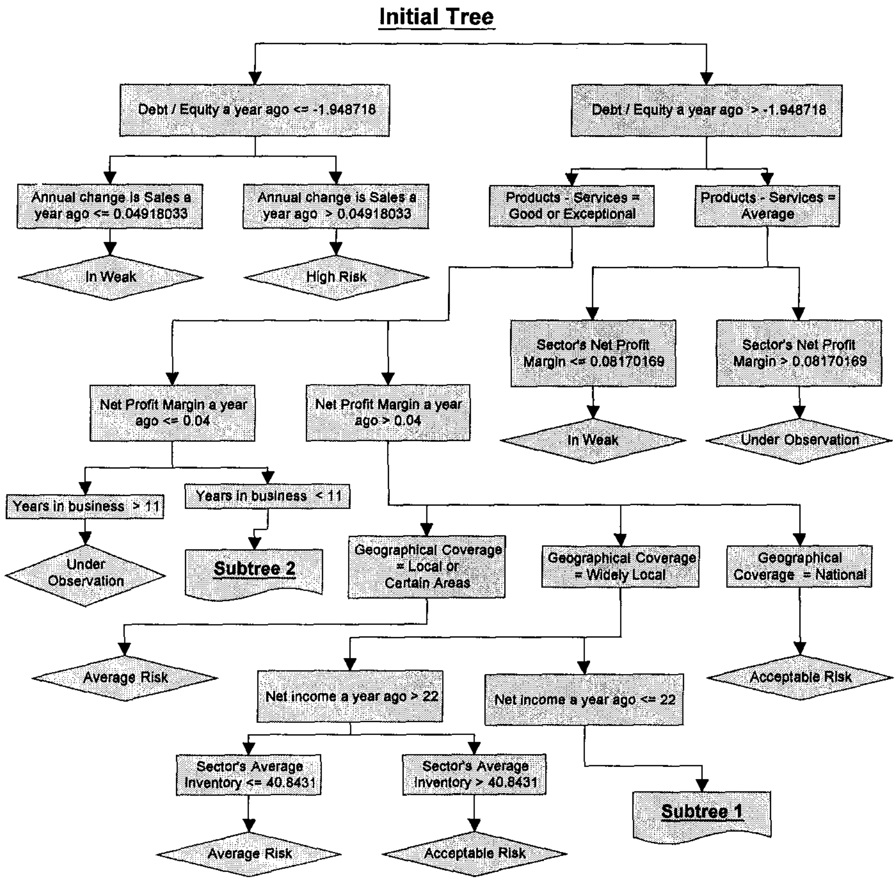


Figure 1: Representation of the Initial Decision tree

Experts acknowledge the tree as a really suitable and simple way of representing the criteria for credit risk problems. Higher importance was given by the experts to

[6] In Figures 1,2,3, rectangular shapes denote tests, rhombus shapes denote leaf nodes, i.e. classes assigned to the examples

short rules, i.e. those having a few only premise parts, and they are easier to remember. Experts also agree with the content of the most powerful decision rule produced by the tree:

IF Debt / Equity a year ago > -1.948718 AND Products/Services (Quality) in [Good-Exceptional] AND Net Profit Margin a year ago > 0.04 AND Geogra- phical Coverage in {Local, Certain Areas} THEN ➔ AVERAGE RISK (41/3)

The specific rule examines a very well aimed combination of attributes and criteria used in credit scoring according to bank managers, when dealing with average risk applicants. Observe within the brackets, that the rule covers 41 positive examples and is falsified by only 3 exceptions i.e., it represents almost one third of the sample used in only 4 conditions. The same rational and easy to use and interpret, were found to be the two prime rules of the tree, concerning low and high risk applicants:

IF Debt / Equity a year ago ≤ -1.948718 AND Annual change in Sales a year ago ≤ 0.04918033 THEN "IN WEAK" (7/0)

IF Debt / Equity a year ago ≤ -1.948718 AND Annual change in Sales a year ago > 0.04918033 THEN "HIGH RISK" (2/0)

Although not covering many cases, these two rules apart from simple in structure are also covering only positive examples of the training set. The experimental results presented in this paper are indicative of what knowledge can generate an inductive learning approach. To be consistent with the standard practice usually followed in machine learning when presenting classification results, a few other related issues should be also mentioned.

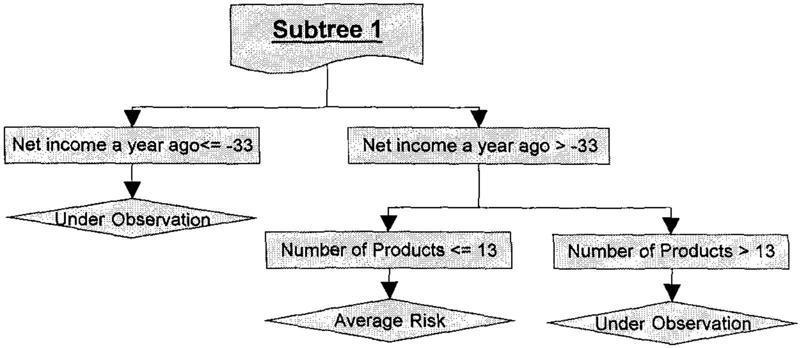


Figure 2: Representation of the Sub-tree 1 (cont. from Fig.1)

First of all, Table I, represents the training classification accuracy and the kind of misclassification occurred. Cross-validation (see details in section 4) of results was exceeding 90% (93.8%) of classification accuracy on new cases, showing the ability of the classifier to correctly classify new applicants in credit scoring. The default option of 10 repeated cross-validation tests was used in order to obtain the above percentage, while 10% of the data was kept out of the training phase at random each time to be used then, as test file. Moreover, the "boosting" technique for tree generation was also applied (see details in section 4), and reduced the errors in classification of the training set to almost 0% (~0.2%), thus forming an almost perfect classifier. Remind that boosting technique consists an improved methodology of producing decision trees, which is more accurate than the traditional technique but is useful only when constructing automated classification tools, due to the large size of the rule base it produces [17, 18].

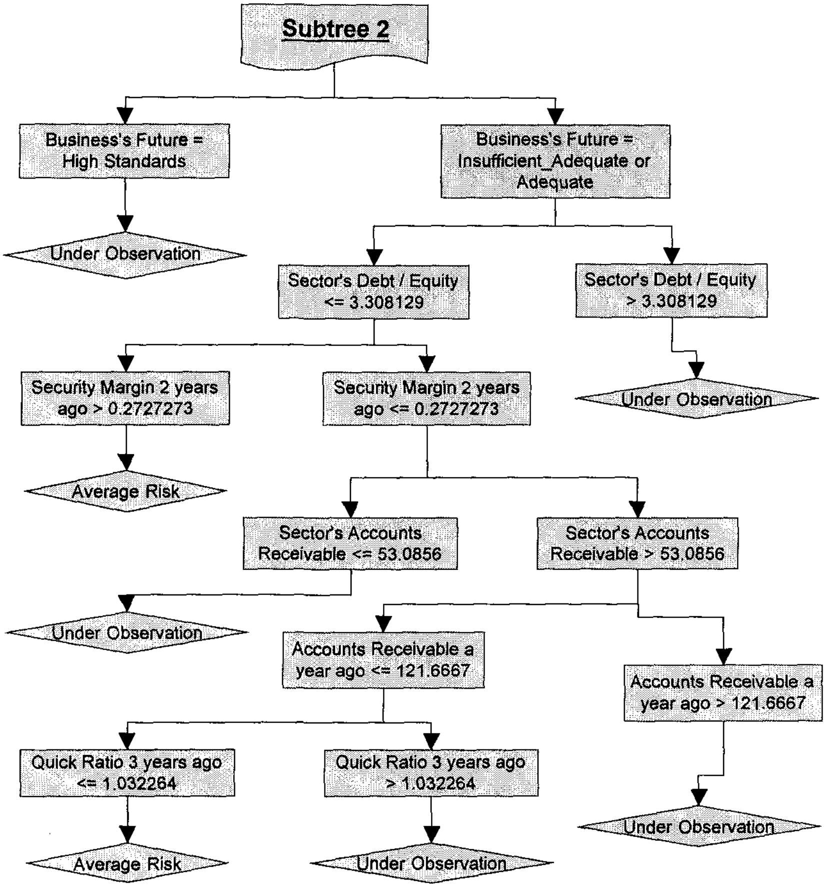


Figure 3: Representation of the Sub-tree 2 (cont. from Fig.1)

Note also that according to literature [11], classification accuracy for different conventional scoring approaches, ranges between 43.3-93.2%, in any case is considerably lower that the accuracy obtained by the suggested approach. Concluding, results are more than encouraging in facing the credit scoring problems of banking organizations safely, accurately and fast. Moreover, the proposed methodology seems to be accepted by the expert staff, as easy to understand and reliable to use.

Table I: Classification rate of training data (130 cases)

Cases	Attr.	Classes	Tree Size	Errors	Accuracy
130	76	7	19	8 (6.2%)	93.8 %

(a)	(b)	(c)	(d)	(e)	(f)	(g)	← classified as
----	----	----	----	----	----	----	
							(a): class WITHOUT RISK
	7	1					(b): class ACCEPTABLE RISK
	2	75	1				(c): class AVERAGE RISK
		3	29				(d): class UNDER OBSERVATION
			1	2			(e): class HIGH RISK
					9		(f): class IN WEAK
							(g): class INSUFFICIENT

6 Concluding remarks

After having completed the implementation of the above described tools we realized that the use of the overall inductive learning based knowledge generation approach to the specific banking organization, was an experience by which the collaborating expert staff changed its entire way of thinking on credit scoring problems. It is now obvious to bank decision makers that, when dealing with problems involving a great number of attributes to be considered in order to make the final decision, there is a much easier alternative for achieving the task. The inductive learning approach proved relatively easy to use. One of the typical problems usually met through the data modeling phase when applying inductive learning, is that one concerning the format of the linguistic values and the way they should be introduced in the application. The comparison of the outcome of the inductive learning software and of the classic statistical model that is currently used by most banks, resulted in the following advantages for the use of the inductive learning approach:

1. There is no need for a huge amount of cases for the extraction of the classifiers. Only a few representative cases for each class can be enough.
2. No requirement applies, of any mathematical knowledge from the user.
3. No transformation of the linguistic variables is required.
4. The classifiers can be extracted both in the form of a set of rules or a decision tree using the labels stated in the data file. A fact that makes it easier to understand them and of course use them.
5. The processing time of the database in order to extract the classifiers is limited to a number of seconds even for a large database.

6. Due to the above fact it is easy to incorporate new cases to the database in order to extract new classifiers.
7. The approach makes uses an interactive interpreter which can be used after the training as an on-line assistant for the decision making process of new cases.
8. A much smaller amount of attributes for the classification process, is used.
9. The decision process is clear and obvious.

Currently, the authors work in two parallel research directions: At first, the drive is given from the fact that the produced decision tree works as a knowledge generator and thus, reflects the banking organization's expertise on the application domain, represented by a handy and meaningful set of rules. In this sense, numerous other banking applications are examined, where the automated knowledge extraction would be of great assistance in managing decision making processes. Typical domains of application such as credit scoring, should be the ranking of bank branches according to their annual performance in order to grant productivity bonus, staff performance evaluation, bankruptcy prediction of potential future clients, etc. At last, the classifier could be continuously reformed, by adding to it every newly appearing case, resulting in more and more accurate and robust classification rules with time. The authors work in the direction of automating such a tool for decision making with the supplement of a user-friendly interface for direct and effective use of the methodology by non-expert financial staff.

References

1. Adriaans P. and Zantinge D., (1996), Data Mining, Addison-Wesley Public.
2. Boyes W.J., Hoffman D.L. and Law S.A., (1989), "An Econometric Analysis of the Bank Credit Scoring Problem", Jour. of Econometrics, Vol. 40, pp.3-14.
3. Carter C. & Catlett J., (1987), "Assessing credit card applications using machine learning", IEEE Expert, Vol. 2, pp. 71-79.
4. Desai V.S., Convay D.G., Crook J.N. and Overstreet J.A., (1997), "Credit scoring models in the credit union environment using neural networks and genetic algorithms", IMA Journal of Mathematics Applied to Business & Industry, Vol. 8, pp. 323-346.
5. Fayyad V.M., Piatetsky-Shapiro G., Smyth P. and Uthurusamy R., (1996), Advances in Knowledge Discovery and Data Mining, AAAI & MIT Press.
6. Freund Y. and Schapire R., (1996), "Experiments with a New Boosting Algorithm", In Proc. Int. Conf. on Machine Learning, ICML'96 (Morgan Kaufmann), Bari, Italy, July 3-6, '96, pp. 148-156.
7. Herbert Simon, (1991), "Artificial Intelligence: Where Has It Been and Where is it Going?", IEEE Trans. Knowledge & Data Eng., Vol. 3, No. 2, pp. 128-136.
8. Kent Martin J., (1997), "An Exact Probability Metric for Decision Tree Splitting and Stopping", Machine Learning, 28, pp. 258-291.
9. Furnkranz Johannes, (1997), "Pruning Algorithms for Rule Learning", Machine Learning, 27, pp. 139-171.

10. Lewis E.M., (1992), An Intoduction to Credit Scoring" Athena Press, San Rafael, CA, USA.
11. Thomas L.C., (2000), "A survey of credit and behavioural scoring: forecasting financial risk of lending to consumers", International Journal of Forecasting, Vol. 16, pp. 149-172.
12. Martin Nelke, (2000), "Business Intelligence in Financial Engineering – A combined Approach to Customer Segmentation and Database Marketing", in Proc. of CoIL-2000, European Symposium on Computational Intelligence and Learning, 22-23 June 2000, Chios, Greece, pp. 171-177.
13. Mitchell T.M., (1997), Machine Learning, McGraw-Hill, New York.
14. Montserrat Guillen and Artis Manuel, (1992), "Count Data Models for a credit scoring system", Third Meeting on the European Series in Quantitative Economics and Econometrics on "Econometrics of Duration, Count and Transition Models", Paris, Dec.10-11, 1992, pp.1-9.
15. Nomikos I., Dounias G. and Vemmos K., (1999), "Comparison of Alternative Criteria for the Evaluation of Machine Learning in the Medical Diagnosis of Stroke", in Proc. 3rd Int. Data Analysis Symp., Aachen, Germany, pp.63-66 .
16. Langley P. and Simon H., (1995), "Applications of Machine Learning and Rule Induction", Communications of the ACM, 38 (11), pp.55-64.
17. Quinlan J.R., (1996), "Boosting First Order Learning" in Conf. on Algorithmic Learning Theory, 7th Int. Workshop ALT'96, Sydney, Australia, Oct. '96, Proc. pub. by Springer Verlag, pp.1-11.
18. Quinlan J.R., (1996), "Boosting, Bagging, and C4.5", Proc. of the 13th National Conf. on AI, Aug. 4-8, '96, Portland, Oregon, USA, pp. 725-730, AAAI Press.
19. Quinlan J.R., (1986), "Induction of Decision Trees", Machine Learning, Vol. 1, pp. 81-106.
20. Quinlan J.R., (1987), "Simplifying decision trees" International Journal of Man-Machine Studies, Vol. 12, pp.221-234.
21. Quinlan J.R., (1993), C4.5: Programs for Machine Learning, San Mateo: Morgan Kaufmann.
22. Weiss S.M. and Indurkhya N., (1998), Predictive Data Mining: A Practical Guide, Morgan Kaufmann.
23. Schaffer C., (1993), "Selecting a classification method by cross-validation", Machine Learning, Vol. 13, pp. 135-143.
24. Schapire R., (1990), "The strategy of weak learnability", Machine Learning, Vol.5, No. 2, pp. 197-227.
25. Steenackers A. and Goovaerts M.J., (1994), "A Credit Scoring Model for Personal Loans", Insurance: Mathematics and Economics, Vol. 8, pp. 31-34.
26. Thomas L.C., (1998), "Methodologies for classifying applicants for credit", in Hand D.J. and Jacka S.D. (Eds.), Statistics in Finance, Arnold, UK, pp.83-103.
27. Fayyad U. M. and Irani K. B., (1992), "Technical Note: On the Handling of Continuous-Valued Attributes in Decision Tree Generation", Machine Learning, Vol. 8, p.87-102.

This page is intentionally left blank

BANK ASSET LIABILITY MANAGEMENT TECHNIQUES: AN OVERVIEW

K. KOSMIDOU AND C. ZOPOUNIDIS

Technical University of CreteDept. of Production Engineering and Management, Financial Engineering Laboratory, University Campus, 73100 Chania, Greece

E-mail: kosmidou@ergasya.tuc.gr , kostas@ergasya.tuc.gr

The uncertainty that prevails in the financial and investment environment has prompted banks and other financial institutions to seek out greater efficiency in the management of their assets and liabilities. Today's asset management decisions create tomorrow's problems as well as tomorrow's opportunities. This need has led to studies concerning the optimal balance among profitability, risk, liquidity and other uncertainties. The present paper makes a brief overview of the bank ALM techniques that have been developed and used over the last 20 years.

1 Introduction

Financial engineering involves the design, development, and implementation of innovative financial instruments and processes, and the formulation of creative solutions to problems in finance [21]. During the last two decades there has been a tremendous increase in the use of operations research techniques. Among others, financial engineers have been heavily involved in risk management, in assessing the types of risk of different securities, in identifying and measuring them, and finally in developing systems for transforming high risk investment instruments to low risk ones. Risk management provides the most efficient way of managing risk through sophisticated quantitative and optimization models, such as asset/liability management model (ALM).

In asset/liability management (ALM) the exposure to various risks is minimized by holding the appropriate combination of assets and liabilities in order to meet the firm's objectives. More precisely, allocating assets lies at the heart of a strategic risk management system. In addition, liability streams and their uncertainty, institutional constraints and policies, taxes, transaction costs and the like are important features in real financial planning. Application areas include pension plans, insurance companies, banks, university endowments and other leveraged institutions, wealthy and ordinary individuals. These investors possess future liabilities and goals. They must make investment decisions while considering the use of their funds, that is, investing for a purpose. Risks must be measured in the context of the entire organization's or individual's financial situation. Asset investment decisions are combined with liability choices in order to maximize the investor's wealth over time.

The growing internationalization, the globalization of financial markets and the introduction of complex products have increased volatility and risks. The great and fast availability of all kinds of different information due to the development towards an "information society" has eliminated any delays between the occurrence of an event and the impact on the markets. Consideration of uncertainties is critical in financial planning. Investors often seek to develop long-term strategies that hedge against uncertainties.

The present paper focuses on the study of bank asset liability management. Many are the reasons that lead us to study bank asset liability management, as an application of ALM. Firstly, bank asset / liability management has always been of concern to bank managers, but in the last years and especially today its importance has grown more and more. The development of information technology has led to such an increasing public awareness that the bank's performance, its politics and its management are closely monitored by the press and the bank's competitors, shareholders and customers and thereby highly affect the bank's public standing.

The increasing competition in the national and international banking markets, the changeover towards the monetary union and the new technological innovations herald major changes in the banking environment and challenge all banks to make timely preparations in order to enter into the new competitive monetary and financial environment.

All the above drove banks to seek out greater efficiency in the management of their assets and liabilities. Thus, the central problem of ALM revolves around the bank's balance sheet and the main question that arises is: What should be the composition of a bank's assets and liabilities on average given the corresponding returns and costs, in order to achieve certain goals, such as maximization of the bank's gross revenues?

It is well known that finding an appropriate balance between profitability, risk and liquidity considerations is one of the main problems in ALM. The optimal balance between these factors cannot be found without considering important interactions that exist between the structure of a bank's liabilities and capital and the composition of its assets.

Bank asset / liability management is defined as the simultaneous planning of all asset and liability positions on the bank's balance sheet under consideration of the different banking and bank management objectives and legal, managerial and market constraints. Banks are looking to maximize profit and minimize risk.

In this paper we make a brief overview of bank ALM techniques, as an application of ALM. This overview is traced by classifying the models in two main categories. Finally, the concluding remarks are discussed.

2 Bank ALM Techniques

Asset and liability management models can be deterministic or stochastic. Deterministic models use linear programming, assume particular realizations for random events, and are computationally tractable for large problems. The banking industry has accepted these models as useful normative tools [11]. Stochastic models, however, including the use of chance-constrained programming, dynamic programming, sequential decision theory, and linear programming under uncertainty, presented computational difficulties.

The theoretical approach of these models is outlined in the following section, whereas the mathematical programming formulation is described in the Appendix.

2.1 Deterministic models

Looking to the past, we find the first mathematical models in the field of bank management. The deterministic linear programming model of Chambers and Charnes [8] is the pioneer on asset and liability management. Chambers and Charnes were concerned with formulating, exploring and interpreting the uses and constructs which may be derived from a mathematical programming model which expresses more realistically than past efforts the actual conditions of current operations. Their model corresponds to the problem of determining an optimal portfolio for an individual bank over several time periods in accordance with requirements laid down by bank examiners which are interpreted as defining limits within which the level of risk associated with the return on the portfolio is an acceptable one.

Cohen and Hammer [11] Robertson [46] Lifson and Blackman [33], Fielitz and Loeffler [20] are successful applications of Chambers and Charnes'model. Even though these models have differed in their treatment of disaggregation, uncertainty and dynamic considerations, they all have in common the fact that they are specified to optimize a single objective profit function subject to the relevant linear constraints.

Eatman and Sealey [18] developed a multiobjective linear programming model for commercial bank balance sheet management. The objectives used in their paper are based on the profitability and solvency. The profitability of a bank is measured by its profit function. Since the primary goals of bank managers, other than profitability, are stated in terms of liquidity and risk, measures of liquidity and risk would seem to reflect the bank's solvency objective. There are many measures of liquidity and risk that could be employed, just as there are many measures used by different banks and regulatory authorities. Eatman and Sealey measured liquidity and risk by the capital-adequacy (CA) ratio and the risk-asset to capital (RA) ratio respectively. The capital-adequacy ratio is a comprehensive measure of the bank's liquidity and risk because both asset and liability composition are considered when determining the value of the ratio. Since liquidity diminishes and risk increases as the CA ratio increases, banks can maximize liquidity and minimize risk by minimizing the CA ratio. The other objective reflecting the bank's solvency is the risk-asset to capital (RA) ratio. Using the RA ratio as a risk measure, the bank is assumed to incur greater risk as the RA ratio increases. Therefore, in order to minimize risk, the RA ratio is minimized. The constraints considered in the model of Eatman and Sealey are policy and managerial.

Apart from Eatman and Sealey, Giokas and Vassiloglou [23] developed a multiobjective programming for bank assets and liabilities management. They supported that apart from attempting to maximize revenues, management tries to minimize risks involved in the allocation of the bank's capital, as well as to fulfill other goals of the bank, such as retaining its market share, increasing the size of its deposits and loans, e.t.c. Conventional linear programming is unable to deal with this kind of problem, as it can only handle a single goal in the objective function. Goal programming is the most widely used approach in the field of multiple criteria decision making that enables the decision maker to incorporate easily numerous variations of constraints and goals.

2.2 *Stochastic models*

Apart from the deterministic models, several stochastic models have been attempted since the 1970's. These models, in their majority, originate from the portfolio selection theory of Markowitz [34] and they are known as *static mean-variance methods*. According to this approach the risk is measured by the variance in a single period planning horizon, the returns are normally distributed and the bank managers use risk-averse utility functions. In this case, the value of an asset depends not only on the expectation and variance of its return but also on the covariance of its return with the returns of all other existing and potential investments. Pyle [45] applied Markowitz's theory in his static model where a bank selects the asset and liability levels it wishes to hold throughout the period. He considers only the risk of the portfolio and not other possible uncertainties. The model omits trading activity, matching assets and liabilities, transactions costs, and other similar features. A more sophisticated approach was that of Brodt [6], who adapted Markowitz's theory and presented an efficient dynamic balance sheet management plan that maximizes profits for a given amount of risk over a multiperiod planning horizon. His two-period, linear model included uncertainty and based on the portfolio selection theory of Markowitz, he tried to build the efficient frontier between the function of expected profits and the linear one of its deviations. Instead of the variance, he used the mean absolute deviation or the semi-absolute deviation that is taken by varying the value of the upper or lower bound of one of the two functions.

Charnes and Thore [10], Charnes and Littlechild [9] developed *chance constrained programming models*. These models express future deposits and loan repayments as joint, normally distributed random variables, and replace the capital adequacy formula by chance-constraints on meeting withdrawal claims. These approaches lead to a computationally feasible scheme for realistic situations. Pogue and Bussard [44] have formulated a 12-period chance constrained model in which the only uncertain quantity is the future cash requirement. The major weakness is that the chance-constrained procedure cannot handle a differential penalty for either varying magnitudes of constraint violations or different types of constraints.

In 1969, Wolf proposed the *sequential decision theoretic approach* that employs sequential decision analysis to find an optimal solution through the use of implicit enumeration. This technique does not find an optimal solution to problems with a time horizon beyond one period, because it is necessary to enumerate all possible portfolio strategies for periods preceding the present decision point in order to guarantee optimality. In order to explain this drawback, Wolf makes the assertion that the solution to a one-period model would be equivalent to a solution provided by solving a n-period model. This approach ignores the problem of synchronizing the maturities of assets and liabilities. Bradley and Crane [3] have developed a stochastic decision tree model that has many of the desirable features essential to an operational bank portfolio model. The Bradley-Crane model depends upon the development of economic scenarios that are intended to include the set of all possible outcomes. The scenarios may be viewed as a tree diagram for which each element (economic condition) in each path has a set of cash flows and interest rates. The problem is formulated as a linear program, whose objective is the maximization of expected terminal wealth of the firm and the constraints refer to the cash flow, the inventory balancing, the capital loss and the class composition. To overcome computational difficulties, they reformulated the asset and liability problem and developed a general programming decomposition algorithm that minimizes the computational difficulties.

Another approach to stochastic modeling is *dynamic programming*. The approach dates to the work of Samuelson [49], Merton [35] and others. The main objective of this approach is to form a state space for the driving variables at each time period. Instead of discerning the scenarios, stochastic control perplexes the state space. Either dynamic programming algorithms or finite element algorithms are available for solving the problem. Merton [35] in his paper explores two classes of reasons why optimal endowment investment policy and expenditure policy can vary significantly among universities. This is done by relating the present value of the liability payments to the driving economic variables. The analysis suggests that managers and others who judge the prudence and performance of policies by comparisons across institutions should take account of differences in both the mix of activities of the institutions and the capitalized values of their no endowment sources of cash flows. Eppen and Fama [19] modeled two and three asset problems. The basic idea is to set up the optimization problem under uncertainty as a stochastic control model using a popular control policy. This model reallocates the portfolio in the end of each period such that the asset proportions meet the specified targets. The continuous sample space is represented via a discrete approximation. The discrete approximation offers a wider range of application and is easy to implement. These models are dynamic and account for the inherent uncertainty of the problem.

An alternative approach in considering stochastic models, is the *stochastic linear programming with simple recourse* (SLPSR), also called linear programming under uncertainty (LPUU). This technique explicitly characterizes each realization of the random variables by a constraint with a limited number of possible outcomes and time periods. Cohen and Thore [12] viewed their one-period model more as a tool for sensitivity analysis than a normative decision tool. Crane [13] on the other hand, modulated the model to a two-period one. The computational intractability and the perceptions of the formulation precluded consideration of problems other than those that were limited both in terms of time periods and in the number of variables and realizations. Booth [2] applied this formulation by limiting the number of possible realizations and the number of variables considered, in order to incorporate two time periods. Kallberg et al. [28] have formulated a firm's short term financial planning problem as a stochastic linear programming with simple recourse model where forecasted cash requirements are discrete random variables. The main goal of their paper was to minimize costs of the various sources of funds employed plus the expected penalty costs due to the constraint violations over the four quarter horizon. They concluded that even with symmetric penalty costs and distributions the mean model is significantly inferior to the stochastic linear programming formulation. Kusy and Ziemba [31] employed a multiperiod stochastic linear program with simple recourse to model the management of assets and liabilities in banking while maintaining computational feasibility. Their model tends to maximize the net present value of bank profits minus the expected penalty costs for infeasibility and includes the essential institutional, legal, financial and bank-related policy considerations and their uncertainties. It was developed for the Vancouver City Savings Credit Union for a 5-year planning period. The results indicate that ALM is theoretically and operationally superior to a corresponding deterministic linear programming model and that the effort required for the implementation of ALM and its computational requirements are comparable to those of the deterministic model. Moreover, the qualitative and quantitative characteristics of the solutions are sensitive to the model's stochastic elements, such as the asymmetry of cash flow distributions. This model had 1) Multiperiodicity incorporating changing yield spreads across time, transaction costs associated with selling assets prior to maturity, and the synchronization of cash flows across time by matching maturity of assets with expected cash outflows; 2) simultaneous considerations of assets and liabilities to satisfy accounting principles and match the liquidity of assets and liabilities; 3) transaction costs incorporating brokerage fees and other expenses incurred in buying and selling securities; 4) uncertainty of cash flows incorporating the uncertainty inherent in the depositors' withdrawal claims and deposits that ensures that the asset portfolio gives the bank the capacity to meet these claims; 5) the incorporation of uncertain interest rates into the decision-making process to avoid lending and borrowing decisions that may ultimately be detrimental to the financial well-being of the bank; and 6) legal and policy constraints appropriate to the bank's operating environment. The Kusy and Ziemba model did not contain end effects, nor was it truly dynamic since it was solved two periods at a time in rolling fashion. The scenarios were high, low, and average returns that were independent over time.

Another application of the multistage stochastic programming is Russell-Yasuda Kasai model [7], that aims at maximizing the long term wealth of the firm while producing high income returns. This model builds on this previous research to make a large scale dynamic model with possibly dependent scenarios, end effects, and all the relevant institutional and policy constraints of Yasuda Kasai's business enterprise. The multistage stochastic linear program, used by Carino et al., incorporates Yasuda Kasai's asset and liability mix over a five-year horizon followed by an infinite horizon steady-state end-effects period. The objective is to maximize expected long-run profits less expected penalty costs from constraint violations over the infinite horizon. The constraints represent the institutional, cash flow, legal, tax and other limitations on the asset and liability mix over time.

Based on one or more decision rules, it is possible to create an ALM model for optimizing the setting of decision rules or even to create a *scenario analysis*. These optimization problems are relatively small, but they often result in non-convex models and it is difficult to identify the global optimal solution. Examples of optimizing decision rules are Falcon Asset Liability Management [38] and Towers Perrin's Opt: Link System [37]. In general, scenario analysis is defined as a single deterministic realization of all uncertainties over the planning horizon. The process constructs, mainly, scenarios that represent the universe of possible outcomes [16, 24]. The main idea is the construction of a representative set of scenarios that are both optimistic and pessimistic within a risk analysis framework. Such an effort was undertaken by Towers Perrin, one of the largest actuarial firms in the world, who employs a capital market scenario generation system, called CAP: Link. This was done in order to help its clients to understand the risks and opportunities relating to capital market investments. The system produces a representative set of individual simulations – typically 500-1000, starting with the interest rate component. Towers Perrin employs a version of the Brennan and Schwartz [5] two factor interest rate model. The other submodels are driven by the interest rates and other economic factors. Towers Perrin has implemented the system in over 14 countries in Europe, Asia, and North America.

Derwa [16], Robinson [47] and Grubmann [24] report successful implementations of *simulation models* developed for various financial institutions. Derwa, for example, used a computer model, operating now at Société Générale de Banque, to improve management decision making in banks. The model was conceived as a form of a decision tree which made it possible to proceed step by step and examine the factors converging on the essential objectives of the bank. Derwa concludes that the problems raised by introducing models into management are much more difficult to solve than the technical ones connected with mathematics or date processing.

Mulvey and Vladimirou [41] used *dynamic generalized network programs* for financial planning problems under uncertainty. They developed a model in the framework of multiscenario generalized network that captures essential features of various discrete time financial decision problems and represented the uncertainty by a set of discrete scenarios of the uncertain quantities. However, these models have a small size and are not able to solve practical sized problems. Mulvey and Crowder [39] and Dantzig and Glynn [15] used the methods of sampling and cluster analysis respectively to limit the required number of scenarios to capture uncertainty and maintain computational tractability of the resulting stochastic programs.

3 Conclusions

The main purpose of the present paper was to provide a brief outline of the bank ALM techniques in the banking industry.

The banking business has recently become more sophisticated due to technological expansion, economic development, creation of financial institutions and increased competition. Moreover, the mergers and acquisitions that have taken place the last years create large groups of banking institutions. The success of a bank depends mainly on the quality of its asset and liability management, since the latter deals with the efficient management of sources and uses of bank funds concentrating on profitability, liquidity, capital adequacy and risk factors.

It is obvious that in the last two decades modern finance has developed into a complex mathematically challenging field. Various and complicated risks exist in financial markets. For banks, interest rate risk is at the core of their business and managing it successfully is crucial to whether or not they remain profitable. Therefore, it has been essential the creation of the department of financial risk management within the banks.

Despite the approaches described in this paper, little academic work has been done so far to develop a model for the management of assets and liabilities in the Greek banking industry. Based on the above we conclude that the quality of asset liability management in the Greek banking system has become significant as a resource of competitive advantage. Therefore, the development of new technological approaches in bank asset liability management in Greece is worth further research.

References

1. Beatty, Anne, "Assessing the use of derivatives as part of a risk-management strategy", *Journal of Accounting and Economics* 26, (1999), p. 353-357.
2. Booth, G.G., "Programming Bank Portfolios under Uncertainty: An Extension", *Journal of Bank Research* 2, (1972), p. 28-40.
3. Bradley, S.P., and Crane, D.B., "A Dynamic Model for Bond Portfolio Management", *Management Science* 19, (1972), p. 139-151.
4. Breiman, L., "Optimal gambling system for favorable games", *Proceedings of the 4th Berkeley Symposium on Mathematical Statistics and Probability* 1, (1961), p.63-68.
5. Brennan, M.J. and E.S. Schwartz, An equilibrium model of bond pricing and a test of market efficiency. *Journal of Financial and Quantitative Analysis* 17, (1982), p.75-100.
6. Brodt, A.I., "Dynamic Balance Sheet Management Model for a Canadian Chartered Bank", *Journal of Banking and Finance* 2(3), (1978), p. 221-241.
7. Carino, D.R., Kent, T., Muyers, D.H., Stacy, C., Sylvanus, M., Turner, A.L., Watanabe, K., and Ziemba, W.T., "The Russell-Yasuda Kasai Model: An Asset/Liability Model for a Japanese Insurance Company Using Multistage Stochastic Programming", *Interfaces* 24, (1994), p.29-49.
8. Chambers, D., and Charnes, A., "Inter-Temporal Analysis and Optimization of Bank Portfolios", Management Science 7, (1961), p. 393-410.
9. Charnes, A., and Littlechild, S.C., "Intertemporal Bank Asset Choice with Stochastic Dependence", Systems Research Memorandum no.188, The Technological Institute, Nortwestern University, Evanston, Illinois, (1968).
10. Charnes, A., and Thore, S., "Planning for Liquidity in Financial Institution: The Chance Constrained Method", *Journal of Finance* 21(4), (1966), p.649-674.
11. Cohen , K.J. and Hammer, F.S., "Linear Programming and Optimal Bank Asset Management Decisions", *Journal of Finance* 22, (1967), p. 42-61.
12. Cohen, K.J., and Thore, S., "Programming Bank Portfolios under Uncertainty", *Journal of Bank Research* 2, (1970), p.28-40.
13. Crane, B., "A Stochastic Programming Model for Commercial Bank Bond Portfolio Management", *Journal of Financial Quantitative Analysis* 6, (1971), p.955-976.
14. Crane, B., Knoop, F., and Pettigrew, W., "An Application of Management Science to Bank Borrowing Strategies", Interfaces 8(1), Part 2, (1977), p. 70-81.
15. Dantzig, B., and Glynn, P., "Parallel Processors for Planning under Uncertainty", *Annals of Operations Research* 22, (1990), p. 1-21.

16. Dantzig, G.B., and Infanger, G., "Multi-stage stochastic linear programs for portfolio optimization", *Annals of Operations Research* 45, (1993), p. 59-76.
17. Derwa, L., "Computer Models: Aids to management at Societe Generale de Banque", *Journal of Bank Research* 4(3), (1972), p. 212-224.
18. Eatman, L., and Sealey Jr., "A Multi-objective Linear Programming Model for Commercial bank Balance Sheet Management", Journal of Bank Research 9, (1979), p.227-236.
19. Eppen, G.D., and Fama, E.F., "Three Asset Cash Balance and Dynamic Portfolio Problems", *Management Science* 17, (1971), p. 311-319.
20. Fielitz, D., and Loeffler, A., "A Linear Programming Model for Commercial Bank Liquidity Management", *Financial Management* 8(3), (1979), p. 44-50
21. Finnetry, D., "Financial Engineering in corporate finance: An overview", *Financial Management* 17, (1988), p. 14-33.
22. Fortson, C., and Dince, R., "An Application of Goal Progamming to Management of a Country Bank", *Journal of Bank Research* 7, (1977), p. 311-319.
23. Giokas, D., and Vassiloglou, M., "A Goal Programming Model for Bank Assets and Liabilities", *European Journal of Operations Research* 50, (1991), p.48-60.
24. Glynn, P.W., and Iglehart, D.L., "Importance sampling for stochastic simulations", Management science 35, (1989), p.1367-1391.
25. Grubmann, N., "BESMOD: A Strategic Balance Sheet Simulation Model", *European Journal of Operations Research* 30, (1987), p. 30-34.
26. Guay, R. W., "The Impact of derivatives on firm risk: An empirical examination of new derivative users", *Journal of Accounting and Economics* 26, (1999), p.319-351.
27. Hakansson, H. and Ziemba, W., *Capital growth theory*. In Finance, R.A. Jarrow, V. Maksimovic and W. T. Ziemba, eds., (1995), p.123-144.
28. Kallberg, J.G., White, R.W., and Ziemba, W.T., "Short Term Financial Planning under Uncertainty", *Management Science* 28, (1982), p.670-682.
29. Kelly, J., "A new interpretation of information rate", *Bell system Technology Journal* 35, (1956), p.917-926.
30. Kischka, P., "Bestimmung optimaler Portfolios bei Ungewibheit", *Mathematical Systems in Economics* 97, (1984).
31. Kusy, I. M., Ziemba, T. W., "A Bank Asset and Liability Management model", *Operations Research*, vol. 34, no.3, May-June 1986, (1986), p. 356-376.
32. Langen, D., *Strategic Bank Asset Liability Management,* European University Studies, (1989).
33. Lifson, K.A. and Blackman, B.R., "Simulation and Optimization Models for Asset Deployment and Funds Sources Balancing Profit Liquidity and Growth", *Journal of Bank Research* 4(3), (1973), p. 239-255.

34. Markowitz, H.M., *Portfolio Selection, Efficient Diversification of Investments*, John Wiley and Sons, New York, (1959).
35. Merton, R.C., "Lifetime portfolio selection under certainty: the continuous time case", *Review of Economics and Statistics* 3, (1969), p.373-413
36. Merton, R.C., *Continuous-Time Finance.* Blackwell Publishers, (1990).
37. Mulvey, J.M., "Generating scenarios for the Towers Perrin investment system", *Interfaces* 26, (1996), p.1-15.
38. Mulvey, J.M., Correnti, S., and Lummis, J., *Total integrated risk management: insurance elements*, Princeton University Report, SOR-97-2, (1997).
39. Mulvey, J.M., and Crowder, H.P., "Cluster Analysis: An Application of Lagrangian Relaxation", *Management Science* 25, 4, (1979), p. 329-340 .
40. Mulvey, J.M., and Vladimirou, H., "Stochastic Network Oprimization Models of Investment Planning", *Annals of Operations Research* 20, (1989), p.187-217.
41. Mulvey, J.M., and Vladimirou, H., "Stochastic Network Programming for Financial Planning Problems", *Management Science* 38, (1992), p.1642-1663.
42. Oguzsoy C.B., Güven S., "Bank Asset and Liability Management under uncertainty", *European Journal of Operational Research* 102, (1997), p.575-600.
43. Pallaschke, D., and Vollmer, K., "A Partial Approach for the Optimization of Balance Sheet Structures under Special Risk-Functions", *Proceedings of the 3rd Symposium on Money, Banking and Insurance*, (1985), p. 699-710.
44. Pogue, G.A., and Bussard, R.N., "A Linear Programming Model for Short Term Financial Planning under Uncertainty", *Sloan Management Review* 13, (1972), p.69-98.
45. Pyle, D.H., "On the Theory of Financial Intermediation", *Journal of Finance* 26, (1971), p. 737-746.
46. Robetson, M., "A Bank Asset Management Model", in: S. Eilon and T.R. Fowkes (eds.), Applications of Management Science in Banking and Finance, Gower Press, Epping, Essex, (1972), p.149-158.
47. Robinson, R.S., "BANKMOD: An Interactive Simulation Aid for Bank Financial Planning", *Journal of Bank Research* 4(3), (1973), p.212-224.
48. Rotando, M. and Thorp, E., "The Kelly criterion and the stock market", American Mathematical Monthly (December), (1992), p.992-1032.
49. Samuelson, P., "Lifetime portfolio selection by dynamic stochastic programming", *Review of Economics and Statistics* (August), (1969), p.239-246.
50. Tayi, G. and Leonard, P., "Bank Balance Sheet Management: An Alternative Multi-Objective Model", *Journal of The Operational Research Society* 39, (1988), p. 401-410.
51. Wolf, C.R., "A Model for Selecting Commercial Bank Government Security Portfolios", *Rev. Econ. Stat.* 1, (1969), p.40-52.

52. Ziemba, W.T., Mulvey, J.M., *Worldwide Asset and Liability Modeling*, Cambridge University Press, (1998).
53. Zopounidis, C., *Analysis and Management of Financial Risks – Multicriteria Approaches*, Edition "Kleidarithmos" (in Greek), (1998).

APPENDIX

1 Linear programming formulation of assets and liabilities management

X_i : Mean balance of asset i
Y_j : Mean balance of asset j

Assuming that r_j is the unit revenue of asset i (in real terms) and c_j is the unit cost of liability j, the objective function is

$$\textit{Maximize } Z = \sum r_i X_i - \sum c_j Y_j$$

Where Z is the difference between the bank's interest income and interest expense, i.e. its revenues ignoring operational expense. The objective function is maximized under a set of constraints.

2 Linear goal programming model

The problem of a bank's assets and liabilities management can be formulated as the following goal programming model:

Determine $X = (x_1, x_2, \dots, x_j, \dots, x_n)$ that minimizes $Z = f(d_i^+, d_i^-)$ under *rigid constraints:*

$$\sum_{j=1}^{n} c_{mj} x_j \le \theta_m \quad \text{for } m=1,\dots,M \tag{1}$$

goals:

$$\sum_{j=1}^{n} a_{ij} x_j = b_i + d_i^+ - d_i^- \quad \text{for } i=1,\dots,I \tag{2}$$

$$x_j, d_i^+, d_i^- \ge 0 \tag{3}$$

where

- x_j is the mean balance of asset or liability j (structural variables)
- a_{ij} is the technological coefficient attached to x_j in goal i
- θ_m is the available amount of resource m

- b_i is the target value for goal i
- d_i^+, d_i^- are the positive and negative deviations from the target value of goal i

The rigid constraints (1) reflect the availability limitations of resources *m* and correspond to the constraints in the conventional linear programming model. The goals (2) represent the objectives set by management, with the right hand side of each goal consisting of the target value b_i and the positive/negative deviation d_i^+, d_i^- from it.

The difference in formulation between rigid constraints and goals can be handled in a number of ways. In the sequential linear goal programming model which was applied at a bank, these constraints are transformed to the same form as the goals. Thus, (1) becomes

$$\sum_{j=1}^{n} c_{mj} x_j = \theta_m + d_m^+ - d_m^- \quad \text{for } m=1,\ldots,M \tag{4}$$

The achievement function (objective function) has the following form:

$$Minimize\ Z = \left\{ \begin{array}{l} P_1\left[\sum_{m=1}^{M} W_{1m}(d_m^+, d_m^-)\right], P_2\left[\sum_{i=1}^{I} W_{2i}(d_i^+, d_i^-\right], \ldots, \\ P\left[\sum_{i=1}^{I} W_{\ i}(d_i^+, d_i^-)\right] \end{array} \right\}$$

Where

- P_φ are the priority levels, with $P_1 \geq P_2 \geq P \ \geq P_{\ +1}$
- $W_{\varphi i}$ is the linear weighting function of the deviation variables of constraint i at priority level $\varphi \leq i+1$, i.e. the number of priority levels is less than or equal to the number of goals plus 1, since all the rigid constraints appear at the first priority level.

3 Stochastic Linear Program with simple recourse

The general n-stage (SLPSR) model is

$$\text{maximise} \quad c_1 x_1 - E_{\xi_1}\left\{ \min \left[\begin{array}{l} q_1^+ y_1^+ + q_1^- y_1^- + \ldots + \\ \min \left[\begin{array}{l} c_n x_n + E_{\xi_n/\xi_{n-1}\cdots\xi_1} + \\ \{ \ \min \ \left[q_n^+ y_n^+ + q_n^- y_n^-\right] \ \} \end{array} \right] \cdots \end{array} \right] \right\}$$

Subject to

$$\sum_{j=1}^{i} T_{ij}x_i + Iy_i^+ - Iy_i^- = \xi_i,\ i=1,\ldots,n$$

The objective function implies the maximization of net present value of monthly profits minus expected penalty costs for constraint violations.

The approximation procedure aggregates $x_2,\ldots,x_n$ with x_1 and $\xi_2,\ldots,\xi_n$ with ξ_1. Thus, one chooses $x=(x_1,\ldots,x_n)'$ in stage one, observes $\xi=(\xi_1,\ldots,\xi_n)'$ at the end of stage one, and these steps together determine $(y^+, y^-) = [(y_1^+, y_1^-),\ldots,(y_n^+, y_n^-)]$ in stage two. This approach yields a feasible procedure for the true dynamic model (1) that is computationally feasible for large problems and incorporates partial dynamic aspects, since penalty costs for periods 2,…,n are considered in the choice of $x_1,\ldots,x_n$. Aggregating all future period decision variables into x_1 would make the first period decision function as if all future period decisions were the same regardless of the scenario. The decision maker is primarily interested in the immediate revision of the bank's assets and liabilities.

Author Index

FUZZY SETS IN MANAGEMENT, ECONOMICS AND MARKETING

The rapid changes that have taken place globally on the economic, social and business fronts characterized the 20th century. The magnitude of these changes has formed an extremely complex and unpredictable decision-making framework, which is difficult to model through traditional approaches. The main purpose of this book is to present the most recent advances in the development of innovative techniques for managing the uncertainty that prevails in the global economic and management environments. These techniques originate mainly from fuzzy sets theory. However, the book also explores the integration of fuzzy sets with other decision support and modeling disciplines, such as multicriteria decision aid, neural networks, genetic algorithms, machine learning, chaos theory, etc. The presentation of the advances in these fields and their real world applications adds a new perspective to the broad fields of management science and economics.

ISBN 981-02-4753-2
9 789810 247539
www.worldscientific.com
4813 hc