PORTFOLIO SELECTION MODELS WITH INTERVAL DATA

Federica Gioia

Università degli Studi di Napoli Parthenope, Italia
gioia@uniparthenope.it

Abstract

Financial data are often affected by uncertainty: imprecision, incompleteness etc. Therefore in a decision making problem, we should be able to process uncertain information.

The uncertainty in the data may be treated by considering, rather than a single value, the interval of values in which the data may fall. Purpose of the present work is to extend the Markowitz's portfolio selection model and the CAPM to the case in which the returns of any considered asset or the market portfolio are interval-valued variables. We aim at enforcing the power of decision of a classical decision-making method analyzing the interval of solutions when each quantity varies in its own interval of values. The algorithm of the introduced portfolio selection models have been developed and tested numerically when the returns of risky assets over time are described by real interval time series. The numerical results are well interpretable.

Keywords: Interval algebra, interval-valued variables, interval financial returns, interval portfolio.

1. INTRODUCTION

The mathematical modelling of real life must account in the majority of cases of "errors" both in the data and in the solution. These errors may be classified as:

- *measurement errors*: the measured value of a physical quantity $\boldsymbol{x}$ (e.g. the temperature) may be different of the"exact value" of the quantity;
- *computation errors*: due to the finite precision of computers the numerical results are distorted by roundoff errors;
- *errors due to uncertainty in the data*: frequently there is uncertainty associated with the data used in the computation as far as their value cannot be set precisely. For example, the actual construction cost of a facility may be known only in terms of a range of costs. Furthermore, the inflation rate, the cost of money may be known in the general case with uncertainty.

Tools which turned to be particularly adequate for the analysis of data when dealing with inexact, uncertain or vague knowledge, are *fuzzy set*, *rough set* theory and *interval mathematics*.
Fuzzy set and rough set capture two distinct aspects of imperfection in knowledge: vagueness and indiscernibility (RADIZKOWSKA-KERRE 2002), (WALCZAC-MASSART 1999), (YAO 1998).
Interval mathematics neither changes the "nature" of the sets nor defines new parameters to be taken into account, but uses sets themselves (e.g. intervals) instead of real numbers.

A form of interval algebra appeared for the first time in the literature in (BURKILL 1924), (YOUNG 1931); then in (SUNAGA 1958). Modern developments of such an algebra were started by R.E. Moore (MOORE 1966). Main results may be found in (ALEFELD-HERZBERGER 1983), (NEUMAIER 1990), (KEARFOTT-KREINOVICH 1996), (ALEFELD-MAYER 2000). Beyond its main application in optimization theory, interval algebra becomes more and more applied in domains like: economics, statistics, engineering etc.. Nowadays, many numerical aspects are solved thanks to the modern powerful computers and to innovative and efficient numerical algorithms. In the recent, we noticed an upsurge of scientific contributions published in specialized reviews and new software dedicated to the treatment of data in different application domains (ALEFELD-HERZBERGER 1983), (ISHIBUCHI-TANAKA 1990), (KEARFOTT-KREINOVICH 1996), (CHINNECK-RAMADAN 2000), (HICKEY et al. 2001).
Financial data are often affected by uncertainty: imprecision, incompleteness etc. Therefore in a decision making problem, we should be able to process uncertain information.
For example, in real portfolio selection models, many times we do not know the exact value of the return of an asset in the i^{th} state of the world but we know, at best, the *interval* of its possible values.
Intervals may be useful for representing *uncertainty* in financial data or, by converse, it may be useful to *construct* intervals from scalar financial data, for analyzing the uncertainty in the solution of real financial problems.
Moreover, interval data may be considered when it is of interest to analyze a phenomenon in a given interval of time, daily, monthly, etc., with the aim of evaluating not only its 'mean behavior' but also its 'variation'. In this case it is assumed that the variations are uniformly distributed between the minimum and the maximum value. The intervals are representative of both the *location* and the *size* of that phenomenon. The data are intervals or hypercube depending on the dimen-

sion of the problem. In the case of a multidimensional problem it is possible, by the interval data, to analyze also of the *shape* of that problem.
Purpose of the present work is to extend the *Markowitz's portfolio selection model* (MARKOWITZ 2003) and the CAPM (SHARPE 1964) to the case in which the returns of any considered asset or the market portfolio are *interval-valued* variables; in this case the data are described by intervals and rectangles. The basic idea is to revisit computational formulas and related mathematical results for using *intervals* instead of real numbers.
Methodologies for portfolio selection with some uncertainty in the data are proposed in (TANAGA et al. 2000), (INUIGUCHI-RAMIK 2000), (INUIGUCHI-TANINO 2000), (CARLSSON et al. 2002), (LAI et al. 2002), (GIOVE et al. 2006). Those methods handle rather than intervals, fuzzy probabilities, possibility distributions, regret functions. An approach which deals with a linear programming problem with interval objective function with interval coefficients, is proposed in (INUIGUCHI-SAKAWA 1995).
In section 2 of the present work, some notations, definitions, and propositions of the interval mathematics are presented.
In section 3 the Markowitz's portfolio selection model with interval data is introduced.
In section 4 the CAPM with interval returns is proposed. Three different cases will be treated:

1. known the interval expected return of an asset, compute an estimate of the corresponding interval beta;

2. known the interval beta of an asset, compute the corresponding interval expected return;

3. compute an estimate of the interval beta of an asset given available time series for the interval excess returns of a security and the market portfolio respectively.

In section 4 the interval CAPM is introduced also for the case in which the return of any asset is a single-valued variable while the market portfolio is described by an interval vector, i.e., we only know the intervals in which fall the aggregate quantities of any considered risky asset available in the economy.
Section 6 is devoted to the numerical examples. The algorithms of the introduced portfolio selection models, **IMSM** and **ICAPM**, have been implemented in MATLAB and tested numerically when the returns of risky assets over time

are described by real interval time series. The numerical results, analyzed and discussed in this section, are well interpretable.

2. DEFINITIONS NOTATIONS AND BASIC FACTS

Extensions of number systems involving ordered pairs of numbers from the given system are commonplace. The rational numbers are essentially ordered pairs of integers; complex numbers are ordered pairs of real numbers; in each case arithmetic operations are defined with rules for computing the components of a pair resulting from an arithmetic operation on a pair of pairs (MOORE 1966).
An interval $[a,b]$ with $a \leq b$, is defined as the set of real numbers between a and b:

$$[a,b] = \{x/a \leq x \leq b\}$$

Degenerate intervals of the form $[a,a] = a$, are named *thin* intervals. The symbols will be used in the common sense of set theory. For example by $[a,b] \subset [c,d]$ we mean that interval $[a,b]$ is included as a set in the interval $[c,d]$. Furthermore it is $[a,b] = [c,d] \Longleftrightarrow a = c, b = d$.
Let $\boldsymbol{I}$ be the set of closed intervals. Thus $I \in \boldsymbol{I}$ then $I = [a,b]$ for some $a \leq b$. Let us introduce an arithmetic on the elements of $\boldsymbol{I}$. The arithmetic will be an extension of real arithmetic.
If $\diamond$ is one of the symbols $+, -, \cdot, /$, we define arithmetic operations on intervals by:

$$[a,b] \diamond [c,d] = \{x \diamond y / a \leq x \leq b, c \leq y \leq d\} \tag{2.1}$$

except that we do not define $[a,b]/[c,d]$ if $0 \in [c,d]$.
Let us write an equivalent set of definitions in terms of formulas for the endpoints of resultant intervals:

$$\begin{array}{lcl} [a,b] & + & [c,d] = [a+c, b+d] \\ [a,b] & - & [c,d] = [a-d, b-c] \\ [a,b] & \cdot & [c,d] = [min(ac,ad,bc,bd), max(ac,ad,bc,bd)] \\ \multicolumn{3}{l}{\text{if } 0 \in [c,d] \text{ then}} \\ [a,b] & / & [c,d] = [a,b] \cdot [1/d, 1/c] \end{array} \tag{2.2}$$

Properties of the interval algebra may be found in (MOORE 1966), (ALEFELD-HERZBERGER 1983), (KEARFOTT-KREINOVICH 1996), (NEUMAIER 1990),

(ALEFELD-MAYER 2000).

Proposition 1 *If $f(x_1,x_2,...,x_n)$ is a real rational function in which each variable x_i occurs only once and at the first power, then the corresponding interval expression $f(X_1,X_2...,X_n)$ will compute the actual range of values of f:*

$$f(X_1,X_2...,X_n) = \{y/y = f(x_1,x_2,...,x_n, x_i \in X_i, i = 1,...n\}$$

An interval matrix, that will be indicated as X^I, is a matrix in which the elements are *intervals*. Mathematically:

Definition 1 *An $n \times m$ interval matrix X^I is the following set of matrices:*

$$\boldsymbol{X}^I = \{\boldsymbol{X}/\underline{\boldsymbol{X}} \leq \boldsymbol{X} \leq \bar{\boldsymbol{X}}\}$$

where $\underline{\boldsymbol{X}}$ and $\bar{\boldsymbol{X}}$ are $n \times m$ numerical matrices which verify:

$$\underline{\boldsymbol{X}} \leq \bar{\boldsymbol{X}}$$

The inequalities are understood to be componentwise.

3. MARKOWITZ'S INTERVAL-PORTFOLIO SELECTION MODEL

3.1 CLASSIC MARKOWITZ'S MODEL

Let us give a brief description of the classic Markowitz's model. Assume that there are n securities denoted as $S_j\,(j = 1,\cdots,n)$; let us indicate with R_j the return of the security S_j and with x_j the portion of total investment funds devoted to this security. Thus

$$\sum_{j=1}^{n} x_j = 1$$

The vector $\boldsymbol{x} = (x_1,x_2,...,x_n)^T$ is the portfolio of the considered consumer.
In the real setting one can seldom obtain the return rate $R_j\,(j = 1,\cdots,n)$ without any uncertainty, furthermore since returns vary from time to time, they are assumed to be random variables and will be denoted by: $\boldsymbol{R}_j\,(j = 1,\cdots,n)$, as far as are vectors and not real numbers.
Let us consider the following data set in which the returns of the n securities

$S_1, S_2, ..., S_n$ are given for k different states of the world. Thus the j^{th} return $\boldsymbol{R}_j$ is a random variable represented as the j^{th} column of the following $k \times n$ matrix:

$$\boldsymbol{R}^T = \begin{pmatrix} R_{11} & R_{12} & \cdots & R_{1n} \\ R_{21} & R_{22} & \cdots & R_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ R_{k1} & R_{k2} & \cdots & R_{kn} \end{pmatrix}$$

Let us indicate with $\boldsymbol{p} = (p_1, p_2, ..., p_k)$ a discrete probability distribution on the outcomes of the random variable: $\boldsymbol{R}_j$, $(j = 1, \ldots, k)$. The vector of the expected returns of the n assets over the k states is:

$$\boldsymbol{R}^0 = \left[R_1^0, R_2^0, \cdots ..., R_n^0\right]$$

this may be written as:

$$\boldsymbol{R}^0 = \left[\sum_{i=1}^{k} p_i R_{i1}, \cdots ..., \sum_{i=1}^{k} p_i R_{in}\right]^T$$

The covariance matrix $\mathbf{Q}$, associated to the returns matrix $\boldsymbol{R}^T$, has on the r^{th} row and on the s^{th} column, the covariance between $\boldsymbol{R}_r$ and $\boldsymbol{R}_s$:

$$q_{rs}{}^2 = \sum_{i=1}^{k} p_h (R_{hr} - R_r{}^0)(R_{hs} - R_s{}^0) \qquad (r, s = 1, \ldots, n)$$

Associated with a portfolio $\boldsymbol{x}$, the state-contingent wealth vector is defined as:

$$\boldsymbol{W}^T(\boldsymbol{x}) = \boldsymbol{x}^T \boldsymbol{R}$$

and it is the vector of the returns of the portfolio $\boldsymbol{x}$ in each state of the world. The expected return and the variance of a portfolio $\boldsymbol{x}$ can be written as:

$$E(\boldsymbol{x}) = E(\boldsymbol{x}^T \boldsymbol{R}) = \boldsymbol{x}^T E(\boldsymbol{R}) = \boldsymbol{x}^T \boldsymbol{R}^0 \tag{3.1}$$

$$\sigma(\boldsymbol{x}) = \sigma(\boldsymbol{x}^T \boldsymbol{R}) = \boldsymbol{x}^T \boldsymbol{Q} \boldsymbol{x} \tag{3.2}$$

Thus the portfolio's expected return is simply the weighted average of the expected returns of its component securities, a portfolio variance is a more complicated concept, it depends on more than just the variances of the component securities. Since the variance of a portfolio $\boldsymbol{x}$ is regarded as the risk of investment, the (conditional) best investment is one with the minimum variance (3.2) subject to a given return (3.1). This leads to the following quadratic programming problem:

$$\min_{\boldsymbol{x}} \boldsymbol{x}^T \boldsymbol{Q} \boldsymbol{x} \tag{3.3}$$

so that

$$\begin{aligned} \boldsymbol{x}^T \boldsymbol{R}^0 &= E \\ \sum_{j=1}^{n} x_j &= 1 \\ x_i &\geq 0 \end{aligned}$$

3.2 MARKOWIT'S MODEL WITH INTERVAL DATA

Financial data are often affected by uncertainty, imprecision or incompleteness. Therefore in a decision making problem, we should be able to process uncertain and/or incomplete information.
For example, in real portfolio selection models, many times we do not know the exact value of the return of an asset in the i^{th} state of the world but we know, at best, the *interval* of its possible values.
Let us suppose for example that the return of the security S_j $(j = 1, \cdots, n)$ not only varies with time but can be represented by an *interval* of values when each state occurs. Thus $\boldsymbol{R}_j$ $(j = 1, \cdots, n)$ are assumed to be *interval-valued* variables denoted by: $\boldsymbol{R}_j{}^I$ $(j = 1, \cdots, n)$, and represented as columns in the following interval matrix:

$$(\boldsymbol{R}^T)^I = \begin{pmatrix} [\underline{R}_{11}, \bar{R}_{11}] & [\underline{R}_{12}, \bar{R}_{12}] & \cdots & [\underline{R}_{1n}, \bar{R}_{1n}] \\ [\underline{R}_{21}, \bar{R}_{21}] & [\underline{R}_{22}, \bar{R}_{22}] & \cdots & [\underline{R}_{2n}, \bar{R}_{2n}] \\ \vdots & \vdots & \ddots & \vdots \\ [\underline{R}_{k1}, \bar{R}_{k1}] & [\underline{R}_{k2}, \bar{R}_{k2}] & \cdots & [\underline{R}_{kn}, \bar{R}_{kn}] \end{pmatrix}$$

where: $\left[\underline{R}_{ij}, \bar{R}_{ij}\right]$ is the interval in which the return rate of security S_j "falls" when the i^{th} state occurs. The vector of the *expected returns* over the k periods is now an interval vector defined as follow (GIOIA-LAURO 2005):

$$(\boldsymbol{R}^I)^0 = (\left[\underline{R}_1^0, \bar{R}_1^0\right], \left[\underline{R}_2^0, \bar{R}_2^0\right], \cdots, \left[\underline{R}_n^0, \bar{R}_n^0\right])^T$$

where

$$\left[\underline{\boldsymbol{R}}_j^0, \bar{\boldsymbol{R}}_j^0\right] = \left[\sum_{i=1}^{k} p_i \underline{R}_{ij}\,,\, \sum_{i=1}^{k} p_i \bar{R}_{ij}\right]$$

More precisely the j^{th} *interval expected return* $\left[\underline{\boldsymbol{R}}_j^0, \bar{\boldsymbol{R}}_j^0\right]$ is an interval with lower and upper bounds equal to the expected returns of the lower and upper bounds respectively of the intervals:

$$\left[\underline{R}_{1j}, \bar{R}_{1j}\right] \left[\underline{R}_{2j}, \bar{R}_{2j}\right] \left[\underline{R}_{kj}, \bar{R}_{kj}\right] \tag{3.4}$$

associated to the security S_j; it represents the set of all and only the expected returns of k elements each of which is chosen in a different interval of (3.4).
Also the covariance matrix is of interval type and is defined as follows:

$$\boldsymbol{Q}^I = \begin{pmatrix} \left[\underline{q}_{11}^2, \bar{q}_{11}^2\right] & \left[\underline{q}_{12}^2, \bar{q}_{12}^2\right] & \cdots & \left[\underline{q}_{1k}^2, \bar{q}_{1k}^2\right] \\ \left[\underline{q}_{21}^2, \bar{q}_{21}^2\right] & \left[\underline{q}_{22}^2, \bar{q}_{22}^2\right] & \cdots & \left[\underline{q}_{2k}^2, \bar{q}_{2k}^2\right] \\ \vdots & \vdots & \ddots & \vdots \\ \left[\underline{q}_{k1}^2, \bar{q}_{k1}^2\right] & \left[\underline{q}_{k2}^2, \bar{q}_{k2}^2\right] & \cdots & \left[\underline{q}_{kk}^2, \bar{q}_{kk}^2\right] \end{pmatrix}$$

where $[\underline{q}_{ij}^2, \bar{q}_{ij}^2]$ is the interval covariance[1] between securities S_i and S_j; it is defined as follow (GIOIA-LAURO 2005):

$$\left[\underline{q}_{rs}^2, \bar{q}_{rs}^2\right] = \left[\min_{\substack{\boldsymbol{R}_r \in \boldsymbol{R}_r{}^I \\ \boldsymbol{R}_s \in \boldsymbol{R}_s{}^I}} g(\boldsymbol{R}_r\,, \boldsymbol{R}_s)\,,\ \max_{\substack{\boldsymbol{R}_r \in \boldsymbol{R}_r{}^I \\ \boldsymbol{R}_s \in \boldsymbol{R}_s{}^I}} g(\boldsymbol{R}_r\,, \boldsymbol{R}_s) \right]$$

where:

$$g(\boldsymbol{R}_r\,, \boldsymbol{R}_s) = \sum_{h=1}^{k} p_h \cdot (R_{hr} - \sum_{l=1}^{k} p_l \cdot R_{lr})(R_{hs} - \sum_{l=1}^{k} p_l \cdot R_{ls})$$

It can be seen (GIOIA-LAURO 2005) that $[\underline{q}_{rs}^2, \bar{q}_{rs}^2]$ is the interval of only and all the covariances that may be computed when each component of the considered

[1] *By* $\boldsymbol{R}_r \in R_r^I$ *we will refer to a real vector* $\boldsymbol{R}_r$ *having each component in the corresponding interval component of the interval vector* $\boldsymbol{R}_r^I$*.*

vectors ranges in its interval of values.
Analogously to the classical case, the interval return $\boldsymbol{W}(\boldsymbol{x})^I$ associated to the portfolio $\boldsymbol{x}$, may be defined as follow:

$$W(\boldsymbol{x})^I = (\boldsymbol{R}^T)^I \cdot \boldsymbol{x} \tag{3.5}$$

with the difference that now the product $(\boldsymbol{R}^T)^I \cdot \boldsymbol{x}$ in (3.5) is an interval product as defined in (2.2). Thus, in the case in which an interval data matrix $(\boldsymbol{R}^T)^I$ of returns is given, the return $\boldsymbol{W}(\boldsymbol{x})^I$ of a portfolio $\boldsymbol{x}$ is itself an interval-valued variable with an *interval variance* and an *interval expected return*.
We have just seen, in section 3, that the (conditional) best investment is one with the minimum variance subject to a given expected return; this leads to a quadratic programming problem (*QP*) which requires that specific values for the coefficients of the model (3.3) must be chosen; in particular, specific values for the covariances q_{rs}^2 lead to specific values of the coefficients of the objective function to be optimized.
When an interval data matrix $(\boldsymbol{R}^T)^I$ of returns is given, the values of the coefficients in (3.3) are known only approximately, in fact the interval of values in which they fall is given. In this special case it could be interesting to face the "set" of solutions of the optimization problem (3.3) when the coefficients of the objective function vary in their own interval of values. Furthermore, we are searching for the range of optimum solutions, that could be returned by a *QP* model, with various settings of the uncertain coefficients each of which belonging to its interval of values.
We refer to the problem of finding the two extreme solutions and the associated coefficient settings as Quadratic Programming with Interval Coefficients .
Mathematically the following problem must to be solved:

$$\min_{\boldsymbol{R} \in \boldsymbol{R}^I} G(\boldsymbol{R}) \tag{3.6}$$

where:

$$G(\boldsymbol{R}) = \begin{cases} \min_{\boldsymbol{x}} \boldsymbol{x}^T \boldsymbol{Q} \boldsymbol{x} \\ \boldsymbol{x}^T \boldsymbol{R}^0 = E \\ \sum_{j=1}^n x_j = 1 \\ x_i \geq 0 \end{cases}$$

We refer to problem (3.6) as: *Interval Markowitz's Selection Model*.

Hence, when the data are of interval type some uncertainty has to be handled both in the data and in the solution of the problem to be solved. Supposing a uniform distribution on each interval, all the elements of any interval have the same probability of occurrence. Thus for an investor who does not know which will be the return of the j^{th} security at the i^{th} state, could be remarkable to take into account all possible values of that return in an *interval* of values, and to minimize the risk of investment for each value of the returns each of which is in its own interval of values i.e., the investor could be interested in computing the "range of risk" and the two portfolio which achieves the "worst" optimum and the "best" optimum.

Remark

It is important to observe that the classical Markowitz's model postulate that the covariance matrix be definite positive (rather than semidefinite positive) to ensure the necessary linear independence. This hypothesis is *still preserved* under interval arithmetics: effectively the introduced interval covariance matrix $\boldsymbol{Q}^{\boldsymbol{I}}$ may contain scalar matrices which are *not* scalar *variance/covariance* matrices and which are not definite positive and even not semidefinite positive.

In truth, the interval covariance matrix is introduced just to face the range of variation of each '*variance/covariance*' as each return varies in its own interval of variation, but it is *not* used in its 'interval form'. The methodology developed in the paper solves the optimization problem (3.3) *varying* the *scalar variance/covariance* matrix $\boldsymbol{Q}$ in the *interval variance/covariance* $\boldsymbol{Q}^{\boldsymbol{I}}$: $\boldsymbol{Q} \in \boldsymbol{Q}^{\boldsymbol{I}}$. Thus only the semidefinite/definite positive matrices in $\boldsymbol{Q}^{\boldsymbol{I}}$ are involved in the computation; by this way the hypothesis of the classical Markowitz's model is preserved.

4. INTERVAL CAPM

4.1 CLASSIC CAPM

Investors are risk averse, so they will choose to hold a portfolio of securities to take advantage of the benefits of diversification. Therefore, when they are deciding whether or not to invest in a particular stock, they want to know how the stock will contribute to the risk and expected return of their portfolios. The Capital Asset Pricing Model (CAPM) (SHARPE 1964), (EICHBERGER-HARPER 1997) provides an expression which relates the expected return on an

asset to its systematic risk.
In section 3 we have assumed n assets $S_1, S_2, ..., S_n$ and we have indicated with $x_1, x_2, ..., x_n$ respectively the portions of total investment devoted to those securities. We have indicated with $\boldsymbol{R}^T$ (3) the matrix in which the j^{th} column $\boldsymbol{R}_j$ is the random variable representing the returns of the j^{th} asset over k considered states of the world.
Let us recall some of the previously introduced definitions without using the symbolism of the algebra matrix in order to face the formulas explicitly.
Given a probability distribution for the returns of the N considered securities, the expected return of the j^{th} security is:

$$E(\boldsymbol{R}_j) = \sum_{i=1}^{k} p_i R_{ij}$$

The wealth of a portfolio has be defined as the following vector:

$$W(\boldsymbol{x})^t = (W_1(\boldsymbol{x}), W_2(\boldsymbol{x}), \cdots, W_k(\boldsymbol{x}))$$

where

$$W_i(\boldsymbol{x}) = \sum_{j=1}^{n} R_{ij} \cdot x_j$$

is the wealth of the portfolio when the i^{th} state occurs.
The expected return $E(\boldsymbol{x})$ derived from a portfolio $\boldsymbol{x}$ equals the sum of the expected pay-off from the individual assets weighted by the quantities of the assets held in portfolio:

$$E(\boldsymbol{x}) = \sum_{j=1}^{n} E(\boldsymbol{R}_j) \cdot x_j$$

It is important at this point to introduce a "particular" portfolio named the *market portfolio*.
The market portfolio is a vector $\mathbf{A} = (A_1, A_2, \cdots, A_n)$ in which A_j is the aggregate quantity of the j^{th} risky asset available in the economy.
If we denote with q_t the price by which consumers may trade freely the j^{th} asset, the sum:

$$W_0 = \sum_{j=1}^{n} q_j \cdot A_j$$

represents the *market value* of the market portfolio and it is the value of the aggregate endowments available in the economy.
The wealth of the market portfolio is defined by the following vector:

$$W(\boldsymbol{A})^t = (W_1(\boldsymbol{A}), W_2(\boldsymbol{A}), \cdots, W_k(\boldsymbol{A}))$$

where:

$$W_i(\mathbf{A}) = \sum_{j=1}^{n} R_{ij} \cdot A_j$$

is the wealth of the market portfolio at the i^{th} state. The expected return on the market portfolio is defined as:

$$E(\mathbf{A}) = \sum_{j=1}^{n} E(\boldsymbol{R}_j) \cdot A_j$$

and it follows directly from the definition of expected return of a random variable:

$$E(\mathbf{A}) = \sum_{i=1}^{k} p_i \cdot W_i(\mathbf{A}) = \sum_{j=1}^{k} p_i \cdot \left(\sum_{j=1}^{n} R_{ij} \cdot A_j \right) = \sum_{j=1}^{n} \sum_{i=1}^{k} p_i \cdot R_{ij} \cdot A_j = \sum_{j=1}^{n} E(\boldsymbol{R}_j) \cdot A_j$$

Using the finance theorist's preferred mode of operation, and measuring asset returns as pay-offs per unit invested and asset quantities in units of expenditure, the CAPM equation states that, in equilibrium, the difference between the expected rate of return on each risky asset and the riskless rate of return is proportional to the difference between the expected rate of return on the market portfolio and the riskless rate of return.
The relationship, known as Security Market Line equation, is expressed as follow:

$$E(\boldsymbol{R}_j) = R_f + (E(\boldsymbol{A}) - R_f) \cdot \beta_j \tag{4.1}$$

R_f is the risk-free rate; the factor of proportionality β_j has the following expression:

$$\beta_j = \frac{\sigma(\boldsymbol{A}, j)}{\sigma^2(\boldsymbol{A})}$$

where $\sigma(\boldsymbol{A}, j)$ is the covariance between the return of the market portfolio and the return of the j^{th} asset; $\sigma^2(\mathbf{A})$ is the variance of the market portfolio.
If the covariance between the return of the j^{th} asset and the market portfolio is greater than the variance of the market portfolio the risk premium required by the

market in equilibrium will exceed that required on the entire portfolio of risky assets.

The CAPM is a single-period model (CAMPBELL et al. 1997); hence (4.1) do not have a time dimension. Although time-series of returns are readily available and one can use familiar estimation methods to determine the *Beta* of a particular risky asset. Let Z_j and Z_m represent the excess returns for asset S_j and the market portfolio respectively:

$$Z_j \equiv R_j - R_f, Z_m \equiv W(A) - R_f$$

Z_{jt} and Z_{mt} will represent the described excess returns in a fixed time period t:

$$Z_{jt} \equiv R_{jt} - R_f, Z_{mt} \equiv W_t(A) - R_f$$

Let us define $\boldsymbol{Z}_t$ as an $n \times 1$ vector of excess returns for n assets at a time period t. For these n assets, the excess returns can be described using the excess-return single-index *marked model*:

$$\boldsymbol{Z}_t = \alpha + \beta Z_{mt} + \boldsymbol{e}_t \tag{4.2}$$

with the following hypothesis concerning time-independence:

$$\begin{aligned} E(\boldsymbol{e}_t) &= 0 \\ E(\boldsymbol{e}_t \boldsymbol{e}_t^t) &= \Sigma \\ E(Z_{mt}) &= \mu_m, \\ E[(Z_{mt} - \mu_m)^2] &= \mu_m^2 \\ Cov(Z_{mt}, \boldsymbol{e}_t) &= \Sigma \end{aligned}$$

β is the $n \times 1$ vector of betas, α and $\boldsymbol{e}_t$ are $n \times 1$ vectors of asset return intercepts, and disturbances respectively.
There is a close relationship between the single-index market model and the CAPM, that is: the beta from Sharpe's derivation of equilibrium prices is essentially the same beta that can be obtained doing a least-squares regression against the data. Beta is the slope of the regression line; alpha, the vertical intercept, indicates how much better the fund did than CAPM predicted (or maybe more typically, a negative alpha tells you how much worse it did).
It is known from classical theory that estimators for beta and alpha are the OLS

(Ordinary Least Square) estimators; from the estimation of beta, it is possible to compute the risk premium required by the market and compare it with the one implicit in the actual asset price. Purchasing an asset with an actual risk premium exceeding the one predicted by the CAPM, and selling assets with CAPM risk premiums that exceed the actual one, is the common decision rule for investors in financial markets. This fact makes the CAPM a useful instrument for the analysis of asset prices in financial markets.

4.2 CAPM WITH INTERVAL DATA

In this section an interval CAPM approach is introduced. The task is to extend the Security Market Line equation in the case in which some involved quantities are not known precisely, but the interval in which they fall is given.
Let us suppose that the matrix of the returns of the n securities over the k considered states is an interval matrix (3).
This means that some involved quantities in the CAPM equation (4.1) are also intervals:

- the expected return of the j^{th} security over the k states: $[\underline{E}(\boldsymbol{R}_j), \bar{E}(\boldsymbol{R}_j)]$
- the beta of the j^{th} security: $[\underline{\beta}_j, \bar{\beta}_j]$
- the expected return of the market portfolio: $[\underline{E}(\boldsymbol{A}), \bar{E}(\boldsymbol{A})]$

The risk-free rate R_f is supposed to be a known real number, instead the expected return of the market portfolio is supposed to be a known real interval $[\underline{E}(\boldsymbol{A}), \bar{E}(\boldsymbol{A})]$.
Three different situations will be analyzed:

1. given the *interval expected return*: $[\underline{E}(\boldsymbol{R}_j), \bar{E}(\boldsymbol{R}_j)]$ on asset j, calculate an estimation of the corresponding interval Beta: $[\underline{\beta}_j, \bar{\beta}_j]$
2. given the *interval beta*: $[\underline{\beta}_j, \bar{\beta}_j]$ on asset j calculate an estimation of the corresponding interval expected return: $[\underline{E}(\boldsymbol{R}_j), \bar{E}(\boldsymbol{R}_j)]$
3. given available time-series for *interval excess returns* of security S_j and market portfolio respectively, use an interval estimation method to determine the *interval Beta* of a particular risky asset.

case 1

Let us suppose to know the interval in which the expected return of the j^{th} asset "falls": $[\underline{E}(\boldsymbol{R}_j), \bar{E}(\boldsymbol{R}_j)]$.

The task is to *range* the interval beta: $[\underline{\beta}_j, \bar{\beta}_j]$, that is the set of all β_j each of which corresponds, according to (4.1), to an $E(\boldsymbol{R}_j)$ in $[\underline{E}(\boldsymbol{R}_j), \bar{E}(\boldsymbol{R}_j)]$.

Let as consider the function:

$$f : \boldsymbol{R} \in \boldsymbol{R}^I \to f(\boldsymbol{R}) = \frac{E(\boldsymbol{R}_j) - R_f}{E(\boldsymbol{A}) - R_f}$$

f is continuous[2] on a compact and connected set thus the interval $[\underline{\beta}_j, \bar{\beta}_j]$ is the image of f:

$$\begin{aligned} [\underline{\beta}_j, \bar{\beta}_j] &= \left\{ f(\boldsymbol{R}) \,/\, \boldsymbol{R} \in \boldsymbol{R}^I \right\} = \\ &= \left\{ f(R_{ij}) / R_{ij} \in [\underline{R}_{ij}, \bar{R}_{ij}], \quad i = 1, \cdots, k,\ j = 1, \cdots, n \right\} \end{aligned}$$

Refer to worse, function f is a real rational function in which the variables $R_{ij}, (j = 1, \cdots, n, i = 1, \cdots, k)$ appear more than one time, then for proposition 1, the corresponding interval expression:

$f([\underline{E}(\boldsymbol{R}_j), \bar{E}(\boldsymbol{R}_j)], [\underline{E}(\boldsymbol{A}), \bar{E}(\boldsymbol{A})])$ will not compute the actual range of values of f but provide an estimate of it, in fact the following inclusion may be assured:

$$\begin{aligned} f \quad & ([\underline{E}(\boldsymbol{R}_j), \bar{E}(\boldsymbol{R}_j)], [\underline{E}(\boldsymbol{A}), \bar{E}(\boldsymbol{A})]) = \frac{[\underline{E}(\boldsymbol{R}_j), \bar{E}(\boldsymbol{R}_j)] - R_f}{[\underline{E}(\boldsymbol{A}), \bar{E}(\boldsymbol{A})] - R_f} \supset \qquad (4.3) \\ \supset \quad & \left\{ f(R_{ij}) / R_{ij} \in [\underline{R}_{ij}, \bar{R}_{ij}], \quad i = 1, \cdots, k,\ j = 1, \cdots, n \right\} \\ = \quad & \left[\underline{\beta}_j, \bar{\beta}_j \right] \end{aligned}$$

It is important to remark that the operations in (4.3) are *interval algebra* operations as described in (2.2) [3].

Thus by substituting in (4.1) the known intervals $[\underline{E}(\boldsymbol{R}_j), \bar{E}(\boldsymbol{R}_j)]$ and $[\underline{E}(\boldsymbol{A}), \bar{E}(\boldsymbol{A}$ and combining them by using the interval algebra instruments, it is possible to

---

[2] *E(A) is supposed to be not the risk-free rate.*

[3] *Considering intervals which represent small perturbations of the data, the denominator in (4.3) is supposed to be an interval not containing the zero.*

compute an *inclusion* (*the interval solution*) of the interval of betas, by means of (1), corresponding to different possible values of R_{ij}, $(i = 1,\dots,n, j = 1,\dots,k)$ each of which in its interval of values (*the interval of solutions*).

case 2

The task is to range the interval expected return: $[\underline{E}(\boldsymbol{R}_j),\bar{E}(\boldsymbol{R}_j)]$ that is the set of all $E(\boldsymbol{R}_j)$ each of which corresponds, according to eq.(4.1), to an β_j in $[\underline{\beta}_j,\bar{\beta}_j]$. Let us consider the following function:

$$g:\boldsymbol{R}\in\boldsymbol{R}^I\to g(\boldsymbol{R})=R_f+(E(\mathbf{A})-R_f)\cdot\beta_j$$

g is continuous [4] on a compact and connected set, thus the interval $[\underline{\boldsymbol{E}}(R_j),\bar{\boldsymbol{E}}(R_j)]$ is the image of g:

$$\begin{aligned}[\underline{E}(\boldsymbol{R}_j),\bar{E}(\boldsymbol{R}_j)] &= \left\{g(\boldsymbol{R})\,/\,\boldsymbol{R}\in\boldsymbol{R}^I\right\}= \\ &= \left\{g(R_{ij})/R_{ij}\in[\underline{R}_{ij},\bar{R}_{ij}],\, i=1,\cdots,k,\, j=1,\cdots,n\right\}\end{aligned}$$

It is important to remark that both $E(\boldsymbol{A})$ and β_j are functions of R_{ij}, thus g is a rational function in which the variables from which it depends appear more than one time, therefore the corresponding interval expression will not compute the actual range of values of g but will provide an estimate of it; the following inclusion is true:

$$\begin{aligned} g\left([\underline{\beta}_j,\bar{\beta}_j],[\underline{E}(\boldsymbol{A}),\bar{E}(\boldsymbol{A})],\right) &= R_f+([\underline{E}(\boldsymbol{A}),\bar{E}(\boldsymbol{A})]-R_f)\cdot[\underline{\beta}_j,\bar{\beta}_j]\subset \qquad (4.4)\\ &\subset \left\{g(R_{ij})/R_{ij}\in[\underline{R}_{ij},\bar{R}_{ij}]\right\}\\ &= [\underline{E}(\boldsymbol{R}_j),\bar{E}(\boldsymbol{R}_j)] \qquad (4.5)\end{aligned}$$

The operations in (4.4) are interval algebra operations as described in (2.2).

Also in this case, by substituting in (4.1) the known intervals $[\underline{\beta}_j,\bar{\beta}_j]$ and $[\underline{E}(\boldsymbol{A}),\bar{E}(\boldsymbol{A})]$, and combining them by using the interval algebra instruments, it is possible to compute an *inclusion* (*the interval solution*) of the interval of values of $E(\boldsymbol{R}_j)$, by means of (1), corresponding to different possible values of

[4] *It is supposed that $\sigma^2(\mathbf{A})\neq 0$.*

R_{ij}, $(i = 1,\ldots,n, j = 1,\ldots,k)$ each of which in its interval of values (*the interval of solutions*).

case 3

The task of the present section is to regress *interval excess returns* of security S_j on the *interval market risk premium*, in order to estimate the interval beta of the interval-valued security S_j.

Let us recall some basic concepts about interval regression (**Iregr**) (GIOIA-LAURO 2005).

For sake of simplicity let us indicate with $\boldsymbol{X^I}$ and $\boldsymbol{Y^I}$ the independent and the dependent interval-valued variables respectively, which assume the following interval values for k considered states of the world:

$$\boldsymbol{X^I} = (X_i = [\underline{x}_i, \bar{x}_i]), \quad i = 1,\cdots,k$$

$$\boldsymbol{Y^I} = \left(Y_i = [\underline{y}_i, \bar{y}_i]\right), \quad i = 1,\cdots,k$$

it is:

$$\boldsymbol{Y^I} = \alpha^I + \beta^I \boldsymbol{X^I} + \boldsymbol{E^I}$$

where E^I is the erratic interval component.

The aim is to take into account *all possible values* of the components x_i y_i each of which is in its interval of values $[\underline{x}_i, \bar{x}_i]$, $[\underline{y}_i, \bar{y}_i]$ for $i = 1,\cdots k$.

Thus making regression between two interval-valued variables means to compute the *set of regression lines* each of which realizes the best fit, in the Minimum Least Square sense, of a set of points in the plane. This set of points is univocally determined each time the components $x_1, x_2, \cdots, x_k$, $y_1, y_2, \cdots, y_k$ take a particular value in their own interval of variation.

Mathematically computing the interval regression line between two interval-valued variables $\boldsymbol{X^I}$ and $\boldsymbol{Y^I}$ is equivalent to compute the following two sets:

$$\hat{\beta}^I = \left\{ \hat{\beta}(\boldsymbol{X},\boldsymbol{Y}) = \frac{\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})}{\sum_{i=1}^{n}(x_i - \overline{x})^2}, \quad /\boldsymbol{X} \in \boldsymbol{X^I}, \boldsymbol{Y} \in \boldsymbol{Y^I} \right\} \tag{4.6}$$

$$\hat{\alpha}^I = \left\{ \hat{\alpha}(\boldsymbol{X},\boldsymbol{Y}) = \bar{y} - \hat{\beta}\bar{x}, \quad /\boldsymbol{X} \in \boldsymbol{X^I}, \boldsymbol{Y} \in \boldsymbol{Y^I} \right\} \tag{4.7}$$

where and $\bar{x}$ and $\bar{y}$, regarded as functions of $x_1, x_2, \cdots, x_k$, $y_1, y_2, \cdots, y_k$, are given by:

$$\bar{x} = \frac{1}{k}\sum_{i=1}^{k} x_i \quad ; \quad \bar{y} = \frac{1}{k}\sum_{i=1}^{k} y_i$$

These sets may be computed numerically by solving some optimization problems; i.e., searching for the minimum and for the maximum of functions $\hat{\alpha}(\boldsymbol{X},\boldsymbol{Y})$ and $\hat{\beta}(\boldsymbol{X},\boldsymbol{Y})$ in (4.6) and (4.7).
These functions are both continuous[5] on a connected and compact set and this assures that sets (4.6) and (4.7) are the following closed intervals:

$$\hat{\beta}^I = \left[\min_{\substack{\boldsymbol{X} \in \boldsymbol{X}^I \\ \boldsymbol{Y} \in \boldsymbol{Y}^I}} \hat{\beta}(\boldsymbol{X},\boldsymbol{Y}), \max_{\substack{\boldsymbol{X} \in \boldsymbol{X}^I \\ \boldsymbol{Y} \in \boldsymbol{Y}^I}} \hat{\beta}(\boldsymbol{X},\boldsymbol{Y}) \right] \tag{4.8}$$

$$\hat{\alpha}^I = \left[\min_{\substack{\boldsymbol{X} \in \boldsymbol{X}^I \\ \boldsymbol{Y} \in \boldsymbol{Y}^I}} \hat{\alpha}(\boldsymbol{X},\boldsymbol{Y}), \max_{\substack{\boldsymbol{X} \in \boldsymbol{X}^I \\ \boldsymbol{Y} \in \boldsymbol{Y}^I}} \hat{\alpha}(\boldsymbol{X},\boldsymbol{Y}) \right] \tag{4.9}$$

and may be interpreted as follow:
chosen an intercept $\hat{\alpha}$ in the interval $\hat{\alpha}^I$ it exists a slope $\hat{\beta}$ in the interval $\hat{\beta}^I$ so that the regression line:

$$y = \hat{\alpha} + \hat{\beta}x \tag{4.10}$$

is the unique line that realizes the best fit, in the of Minimum Least Square sense, of a given set of points $(x_1, y_1), (x_2, y_2), \cdots, (x_n, y_n)$ in the plane $(x_i \in X_i, y_i \in Y_i, \quad i = 1, \cdots, n)$.

The prediction of an interval $Y = [\underline{y}, \bar{y}]$ of $\boldsymbol{Y^I}$ will be computed as follow:

$$Y = \hat{\alpha}^I + \hat{\beta}^I X$$

[5] *The quantity $\sum_{i=1}^{n}(x_i - \bar{x})^2$ is nil only in the case in which: $x_1 = x_2 = \ldots = x_n$. This is in contradiction with the classic hypothesis that at least two different observations must be available in the experiment.*

Recalling that in our special case the independent and dependent variables are the following *interval-valued* variables respectively:

$$\boldsymbol{Z}_j{}^I = (Z_{jt} = [\underline{Z}_{jt}, \bar{Z}_{jt}]), \quad t = 1, \cdots, k \tag{4.11}$$
$$\boldsymbol{Z}_m{}^I = (Z_{mt} = [\underline{Z}_{mt}, \bar{Z}_{mt}]), \quad t = 1, \cdots, k \tag{4.12}$$

The *interval regression equation* is:

$$\boldsymbol{Z}_j^I = \hat{\alpha}_j{}^I + \hat{\beta}_j{}^I \boldsymbol{Z}_m^I + \boldsymbol{E}^I$$

Every time each return R_{ij} takes a particular value in its own interval of variation $[\underline{R}_{ij}, \bar{R}_{ij}]$, a set of points in the (Z_j, Z_m) plane is univocally determined; the slope and the intercept of the regression line, which realizes the ”best” fit of that set of points, are elements belonging to the intervals $\hat{\beta}_j{}^I$ and $\hat{\alpha}_j{}^I$ respectively. Therefore:

1. the interval $\hat{\beta}_j{}^I$ is the *set of all beta* of security S_j when each return R_{ij}, $i = 1, \cdots, k$ ranges in its own interval of values.
2. Remarking that for single-valued security S_j the CAPM states that the intercept α_j in (3.2) is zero, we can interpret the interval $\hat{\alpha}_j{}^I$ as the *set of all ”errors”* that we may do using the CAPM for predicting the expected return $E(\boldsymbol{R}_j)$ of security S_j for each $R_{ij} \in [\underline{R}_{ij}, \bar{R}_{ij}]$.

In order to show the good agreement between the proposed method and the input interval data, we analyze some special cases here below (the interval slope, the interval intercept and the interval correlation will be indicated as Beta, Alpha and Corr respectively).

A cloud of rectangles rather dispersed and clearly not correlated is reported in Figure 1. The application of the proposed methodology produces Alpha=[1.154,18.680], Beta=[-0.192,0.026] and Corr=[-0.243,0.034] which are well in agreement with the position of the rectangles and that clearly confirm the poor correlation of the analyzed data .

On the contrary rectangles which are visually strongly correlated are reported in Figure 2. The application of the method produces a correlation $Corr = [0.980, 1]$ which confirms this strong correlation and the regression coefficient Alpha=[-6.000,5.999] and Beta=[0.885,1.1279] which are well in agreement with the position of the rectangles.

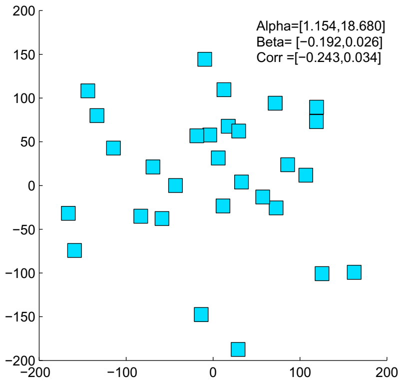

Fig. 1: Iregr:high dispersion.

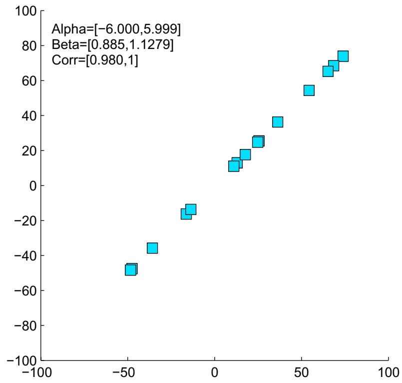

Fig. 2: Iregr: max correlation.

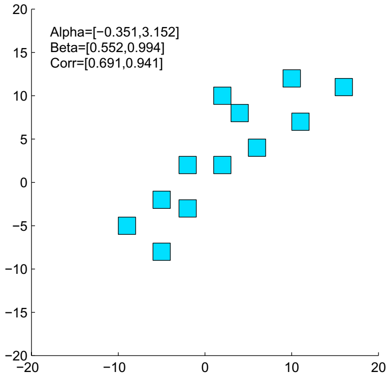

Fig. 3: Iregr: cloud of rectangles.

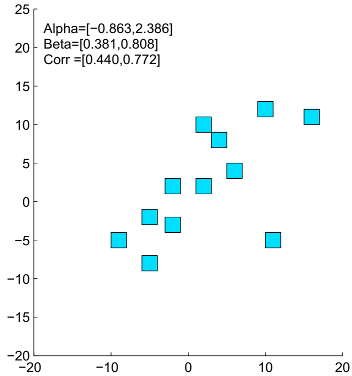

Fig. 4: Iregr: shift of a rectangle.

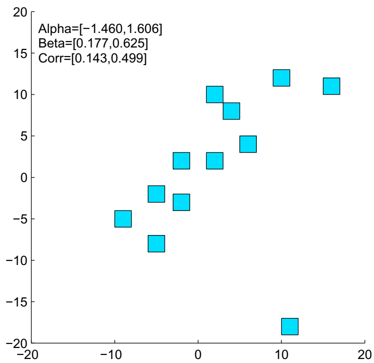

Fig. 5: Iregr: shift of a rectangl.

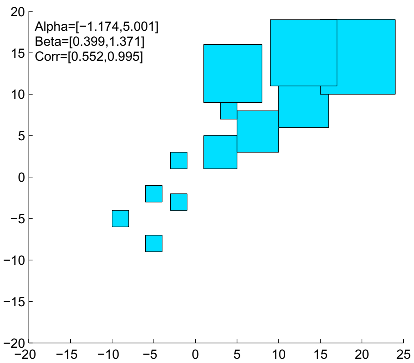

Fig. 6: Iregr: different forms.

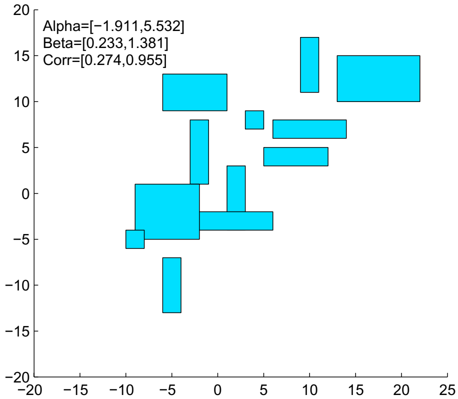

Fig.7: Iregr: different forms

In Figures 3,4,5, we analyze the interval regression coefficients and the interval correlation of some clouds of rectangles in which one rectangle is changing its position with respect to the second axis. Also in this case the regression coefficients Alpha=[-0.351,3.152], Beta=[0.552,0.994] and the correlation between the variables Corr=[0.691,0.941], are well in agreement with the position of the rectangles presented in Figure 3. Furthermore perturbing the data with an anomalous rectangle as in Figure 4,5, the interval correlations Corr=[0.440,0.772], Corr=[0.143,0.499] becomes intervals containing lower values with respect to the interval correlation in Figure 3. Moreover, the intercepts and the slopes Alpha=[-0.863,2.386], Beta=[0.381,0.808], Alpha=[-1.460,1.606], Beta=[0.177,0.625], "attracted" by the anomalous rectangle, are intervals containing higher values with r

creasing "variability" of the considered cloud of the rectangles.

In Figures 6,7 the regression coefficients and the interval correlation are computed for two clouds of rectangles having different forms with respect to those presented in Figure 3. In both situations the computed intervals, Alpha=[-1.174,5.001], Beta=[0.399,1.371], Alpha=[-1.911,5.532], Beta=[0.233,1.381], are intervals which present a bigger radius according to the visible higher variability of the problem.

Interval regression has been treated in the literature. In (BILLARD-DIDAY 2000), (BILLARD-DIDAY 2002), (LIMA NETA-DE CARVALHO 2007) and (RODRIGUEZ 2000), the authors derive the results from some classical regression methods which minimize criteria different from the least squares one. An alternative methodology, is proposed by (MARINO-PALUMBO 1952) with an approach which is typical for handling imprecise data, taking into account the center and the radius of each considered interval and the relations between these two quantities. Those methods, do not consider the interval as a whole structure or special kind of data, but reconstruct interval solution ex post.

Here we make extensively use of the interval algebra tools combined with some optimization techniques to consider the interval as a whole structure and to compute the *interval of solutions*, which is the interval containing all possible values assumed by a considered function, in this special case the *beta* and *alpha* functions, when the observed values vary in their own interval of values.

4.3 CAPM WITH INTERVAL MARKET PORTFOLIO

As last case suppose that the return of each security in each state of the world is not an interval but a real number. However let us suppose that the aggregate quantity A_l of the $l^{th}(l = 1, \cdots, n)$ risky asset available in the economy it is not known precisely and the range in which its value falls is given. In this special case the market portfolio is the following interval vector:

$$\boldsymbol{A}^I = \left([\underline{A}_1, \bar{A}_1], [\underline{A}_2, \bar{A}_2], \cdots, [\underline{A}_n, \bar{A}_n]\right)$$

Considering the security market line (4.1), now the involved interval quantities are:

- the expected return of the market portfolio: $[\underline{E}(\boldsymbol{A}), \bar{E}(\boldsymbol{A})]$
- the beta on asset $[\underline{\beta}_j, \bar{\beta}_j]$

The expected return $E(\boldsymbol{R}_j)$ and the risk-free rate R_f are real numbers.
Let us suppose that the expected return $E(\boldsymbol{R}_j)$ of the j^{th} security is known; we want to compute the interval $[\underline{\beta}_j, \bar{\beta}_j]$ that is: the set of all β_j on asset j, according to eq. (4.1), corresponding to different values of the market portfolio when the l^{th} component A_l ranges in its own interval of values $[\underline{A}_l, \bar{A}_l]$, $(l = 1, \cdots, n)$.
Considering the function:

$$h : \boldsymbol{A} \in \boldsymbol{A}^I \rightarrow \beta_j = \frac{E(\boldsymbol{R}_j) - R_f}{E(\boldsymbol{A}) - R_f}$$

h is continuous on a compact and connected set but, with difference with cases 1 and 2 of section (4.2), each variable A_l, $(l = 1, \cdots, n)$ from which h depends compares once and at the first power; thus in this special case proposition (1) applies so the interval expression:

$$\begin{aligned} h([\underline{E}(\boldsymbol{A}), \bar{E}(\boldsymbol{A})]) &= \frac{E(\boldsymbol{R}_j) - R_f}{[\underline{E}(\boldsymbol{A}), \bar{E}(\boldsymbol{A})] - R_f} = \\ &= \{h(\boldsymbol{A}) / \boldsymbol{A} \in \boldsymbol{A}^I\} = \\ &= \{h(A_1, A_2, \cdots, A_n) / A_l \in \left[\underline{A}_l, \bar{A}_l\right] l = 1, \cdots, n\} \end{aligned}$$

computes the *actual range* of function h that is: the interval $[\underline{\beta}_j, \bar{\beta}_j]$ for which we are looking for.

Refer to worse instead, in the case in which the interval $[\underline{\beta}_j, \bar{\beta}_j]$ is given, for similar considerations with respect to ***cases 1,2*** of previous section, only an interval containing the real number $E(\boldsymbol{R}_j)$ may be computed.

Remark

The CAPM is a pricing methodology, and as such it is supposed to provide a pricing functional for several assets (which is in practice never exactly correct, but should be a workable approximation) that at least in the limit satisfies the no-arbitrage condition. In the 'interval case' the *no-arbitrage condition* is the more preserved the more the computed *interval solution* approximate the *interval of solutions*.

5. FURTHER REMARKS

5.1 FUZZINESS, PROBABILITY AND INTERVALS

Any reasoning which considers not exact data cannot be itself vague: for example, the logic used for treating fuzzy sets is not fuzzy. The difference among fuzziness, probability, and intervals is in the proposition:

$$x \in A$$

1. In the usual logic the above proposition is defined according to the definition of set A, thus it is known that (not if) for each element x, the proposition is true or false but not both things.

2. In the probability theory, the proposition $x \in A$ is still defined, but its trueness is related to an event.

3. In fuzzy logic the trueness of the proposition is accompanied by a function with values in $[0,1]$ whose meaning depends on the context in which the logic is used.

4. Considering interval data $x \in \{a\}$ is substituted with $x \in [\underline{a}, \bar{a}]$, as well as $x \cong a$ is substituted with $|x-a| < \varepsilon$ if an approximation epsilon may be accepted.

The comparison of items 1,2,3,4 outside the epistemological context appears as a purely intellectual speculation, while it is apparent that the four points provide different information, each one valid for the specific suitable use.

5.2 UTILIZATION OF AN INTERVAL METHOD

The reason why an investor should use an *interval method* of decision making is the following: when the returns of n securities in k states of the world are not known precisely, a classical method of decision making imposes to *set* those returns to be numbers: R_{ij}, $(i = 1, \cdots, n, j = 1, \cdots, k)$. This is evidently an artifice: the number R_{ij} will be an *approximation* of the real value of the return of the j^{th} security in the i^{th} state of the world. Using, for example, the classical Markowitz's portfolio selection model, the investor knows the *minimum risk* that he will have in holding a portfolio with a given expected return, but he also knows that, that *minimum risk* is only indicative (a good approximation) of its true value.

The investor should ask the following questions:
if the returns *fluctuate* form their fixed values, how the perturbations reflect on the *minimum risk* ? In what interval that minimum risk ranges? Which are the *worst optimum* and the *best optimum*, i.e., which are the *portfolios* corresponding to the 'best minimum risk' and to the 'worst minimum risk' respectively?
It could be useful for an investor to know the *minimum risk* that he may have, *for each* value of the returns each of which in its own interval of values. The investor should want to know "how good" and "how bad" it could be the minimum risk when he does not know precisely "how good" and "how bad" are the returns of each security in each state of the world.

Interval data are an efficient instrument for preserving all the information of the phenomenon under study as both the mean value of the data and the variation from that mean value are taken into account. It must be considered, however, that small fluctuations in the data are favorable while big ones increase the uncertainty in the decision making process.
Moreover an investor using the interval CAPM computes the intervals ***Beta*** and ***Alpha*** having the possibility to know, not only the scalar risk and the scalar intercept, but also the 'uncertainty' around those quantities.
The interval methodologies introduced in the present paper, should be regarded as methods which may give some *additional* information to that provided by the classical methods. In fact they embrace the corresponding classical methods as special cases. More precisely, we aim at enforcing the power of decision of a classical method of decision-making analyzing the *interval of solutions* when each quantity varies in its own interval of values.

6. NUMERICAL RESULTS

In this section some results of the **IMSM** and **ICAPM** are presented. The data have been downloaded from the following webpages [6]:

http://it.finance.yahoo.com
http://www.dse.unibo.it/pastorel/emf1.htm

The algorithms are implemented in MATLAB.

6.1 INTERVAL MARKOWITZ'S

Daily time series of the returns of the following four risky assets, each of which belonging to the **SPMIB** index, have been considered: ***Eni, Telecom, Generali, Unicredit***.

The corresponding *interval* time series are constructed considering minimum and maximum value of the returns in each week of the considered period of time. Two different portfolios are constructed considering interval returns time series from *2003* to *2008*.
The first portfolio is a *non-financial* one composed by ***Eni*** and ***Telecom*** assets; the second portfolio is a *financial* one composed by ***Generali*** and ***Unicredit*** assets. Thus, two different numerical examples are reported. Let us give herewith a brief description of each of them.

- **The input parameters**

  1. **Portfolio 1**: in the first example we consider a 286×2 interval input data matrix $(\mathbf{R}^T)^I$ in which the interval returns of the 2 risky assets: ***Eni*** and ***Telecom*** from *January 2003* to *December 2008* are reported.
  2. **Portfolio 2**: in the second example we consider a 286×2 interval input data matrix $(\mathbf{R}^T)^I$ in which the interval returns of the 2 risky assets: ***Generali*** and ***Unicredit*** are reported with respect to the same period of time of Portfolio 1.

- **The method**

  1. For each portfolio the classical Markowitz's (CM) selection model is applied on the center scalar data matrix $\mathbf{RC}^T$. Fixed the portfolio expected return, the Markowitz's portfolio and its risk are computed.

[6] *I wish to thank Dr. C. Drago for his precious help in constructing the used interval financial time series, and for his useful comments.*

2. The **IMSM** algorithm is applied. For each portfolio the algorithm computes the range of optimum solutions that could be returned by a *Quadratic Programming problem*, with various settings of the uncertain quantities R_{ij} each of which belonging to its interval of values; i.e., it computes ”the worst” optimum and ”the best” optimum of the interval Markowitz's selection model described in section 3.

   It is important to point out that the expected return of the optimization problem (3.6) is *dynamic*, i.e., it depends on the returns of the assets thus it varies each time the returns vary in their own interval of variation: $\boldsymbol{R}^T \in (\boldsymbol{R}^T)^I$.

- **The output parameters**

  vmin: ”worst” optimum (worst variance)
  vmax: ”best” optimum (best variance)
  xmin: portfolio corresponding to ***vmin***.
  xmax: portfolio corresponding to ***vmax***.

The numerical results are reported in Table 2 and Table 2.

Tab. 1: Classical Markowitz (CM).

Portfolio	**E**	***v***	***x***
Eni-Tel	0.00080	0.00856	(0.32553 , 0.67447)
Gen-Uni	0.00027	0.00596	(0.6 9443 , 0.30557)

Tab. 2: Interval Markowitz (IMSM).

Portfolio	***vmin***	***vmax***	***xmin***	***xmax***
Eni-Tel	0.00087	0.02283	(0.98309, 0.01691)	(0.97788 ,0.02211)
Gen-Uni	0 .00051	0.01901	(0.88329, 0.11671)	(0.93767 ,0.06232)

6.1.2 ANALYSIS OF THE RESULTS

Given a scalar return matrix, following the classical Markowitz's selection model, an investor knows the *minimum risk* that he will have in holding a portfolio with a given expected return. He also knows that the returns of the risky assets involved in his portfolio are approximatively calculated, thus he may ask "how much" he may be wrong in taking his investment decision. In other words it could be useful for the investor to know how perturbations on the input data (returns) reflect on for the investor to know how perturbations on the input data (returns) reflect on the final solution (risk). The interval method (**IMSM**) may be used to assess the range of risk: $\mathbf{v}^I = [\boldsymbol{vmin}, \boldsymbol{vmax}]$ and the two portfolios: ***xmin*** and ***xmax*** which achieves the "worst" optimum ***vmin*** and the "best" optimum ***vmax*** when each return varies in its interval of variation.

With respect to the considered interval time serie, an investor applies the introduced **IMSM** to know that: investing in the first portfolio his *minimum risk* will range from 0.00087 to 0.02283; investing in the second one it will range from 0.00051 to 0.01901. Thus the **IMSM** gives some *additional* information to that provided by the classical method while it computes the *interval of solutions* when each quantity varies in its own interval of values.

6.2 INTERVAL CAPM

In the following examples the interval ***Beta*** and the interval ***Alpha*** of the asset ***abbott*** (*Abbot Laboratories*), which belongs to the ***SP500*** (*Standard and Poor's 500 Composite*) index, is estimated using the interval CAPM approach (**ICAPM**) described in section 4 (***case 3***).
The downloaded data refer to single-valued variables; we have artificially transformed these variables into interval-valued ones by applying a perturbation using a uniform distribution $\mathbf{U}(0, 0.01)$.

- **The input parameters**

  In this example we consider a 49×2 interval matrix, in which monthly time series of the interval excess returns of the ***abbott*** asset and of the ***SP500*** index are reported for the period: January 1991- February 1995.

- **The method**

  1. First of all the classical CAPM is applied on the input scalar matrix of the *excess returns* of the considered time series (with respect to a given risk-free rate *rf*).
  2. The input single-valued time series are transformed into an interval-valued one by a perturbation using an uniform distribution $\mathbf{U}(0, 0.01)$.
  3. The **ICAPM** algorithm is applied. The range of all ***Betas*** and of all ***Alphas*** is computed solving an *interval regression problem* [Gioia-Lauro 2005], that is: the set of all slopes and all intercepts of the *security market line* is computed for the case in which the excess returns are not known precisely but range in an interval of values.

- **The output parameters**

  Alpha = [***Alpham,AlphaM***]
  Beta = [***Betam,BetaM***]

  where:

  Alpham: the minimum intercept.
  AlphaM: the maximum intercept.
  Betam: the minimum slope.
  Betam: the maximum slope.

The ***Alpha*** and ***Beta*** of the classical CAPM on the single-valued variables are also reported.

The numerical results are illustrated in Table 3. In Figure 8 the cloud of rectangles generated by the two considered interval time series is represented on a cartesian plane. No numerical example are reported with respect to ***case 1*** and ***case 2*** considering that only few interval operations are involved.

Tab. 3: CAPM/ICAPM.

Method	*Alpha*	*Beta*
CAPM	0	0.806
ICAPM	[-0.044,0.005]	[0.060,1.007]

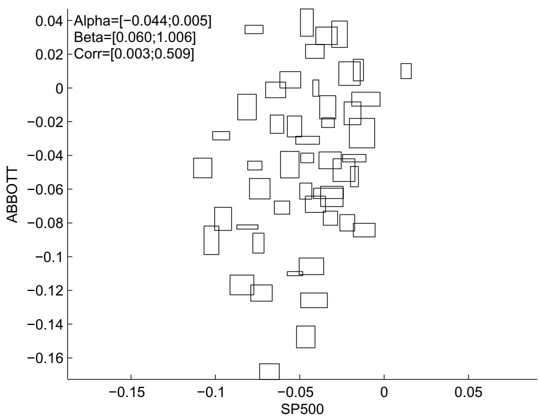

Fig. 8: Interval CAPM (ICAPM)

6.2.1 ANALYSIS OF THE RESULTS

Given time series of the scalar excess returns of a risky asset and of the market portfolio respectively, the ***Beta*** and the ***Alpha*** of the security market line may be estimated using a classical regression model. The question is: are those estimations reliable if the excess returns fluctuate in an interval of values?
In the considered numerical example, ***Beta*** $= [0.060, 1.006]$ is in accordance with the cloud of rectangles represented in Figure 8, considering that the interval correlation is $[0.003, 0.509]$. The interval ***Beta*** is also well interpretable considering that it does not contain the zero; an investor knows that, even if the returns *fluctuate* around their fixed values, the ***Beta*** is always positive and it ranges from 0.060 to 1.006.

Furthermore, remarking that for a single-valued security the CAPM states that the intercept ***Alpha*** in (4.2) is zero, the interval ***Alpha***, which is an interval around zero, is interpretable as the set of all "errors" that we may do using the CAPM for predicting the expected return of the considered risky security. It is remarkable that ***Alpha*** $= [-0.044, 0.0005]$ ranges very closely to zero, namely does not contain elements with absolute value significantly different from zero. For the considered numerical example the interval CAPM approach may be a good way of prediction.
The reliability of the method and consequently the estimate of the intervals ***Alpha*** and ***Beta***, is significative when the radii of the intervals reflect natural fluctuations excluding exceptional events.

In general, it has been observed that the **ICAPM** works well when the rectangles are *well separated*; if they are not, i.e., the rectangles are included one into another, the interval ***Beta*** is an interval containing the zero, and so not well interpretable, considering that 'everything may happen' for the slope of the regression lines of a 'circular cloud' of rectangles. It could be of interest to investigate this aspect by means of simulations, in order to study the robustness of the interval method, introducing indexes regarding the ratio centre/radius.

7. CONCLUSIONS

Methodologies for portfolio selection with some uncertainty in the data are proposed in (TANAGA et al. 2000), (INUIGUCHI-RAMIK 2000), (INUIGUCHI-TANINO 2000), (CARLSSON et al. 2002), (LAI et al. 2002), (GIOVE et al. 2006). Those methods handle rather than intervals, fuzzy probabilities, possibility distributions, regret functions. An approach which deals with a linear pro-

gramming problem with interval objective function with interval coefficients, is proposed in (INUIGUCHI-SAKAWA 1995).
In this paper the interval is considered as a *special kind of data* able to describe both the *position* and the *variability* of what we are observing. Thus we make extensively use of the interval algebra tools combined with some optimization techniques to consider the interval as a *whole structure* and to compute the *interval of solutions*, which is the interval containing all possible values assumed by a considered function when the observed values vary in their own range of variation. The presented numerical results show that the methods adopted in this paper features the input data fairly well. In the framework of the interval CAPM, the interval regression method **Iregr** (GIOIA-LAURO 2005), with respect to other interval regression methods CM, CR, and CRM (BILLARD-DIDAY 2000), (BILLARD-DIDAY 2002), (LIMA NETA-DE CARVALHO 2007), shows some good advantages: the actual return of an asset, if computed by CM, CR, and CRN methods, is an interval which is calculated *ex post*; i.e., it is constructed from the results of some classical regression methods which minimize criteria different from the least squares one. On the contrary, applying **Iregr**, an inclusion of the *interval of solutions* is computed; i.e., the computed *interval actual return* includes all possible actual returns of the considered asset varying each return in its own interval of variation. The solutions are always well interpretable. With difference to the other regression methods, **Iregr** computes the *interval slope* ${\hat{\beta}_j}^I$ and the *interval intercept* ${\hat{\alpha}_j}^I$ of the interval regression line which are both well interpretable: ${\hat{\beta}_j}^I$ and ${\hat{\alpha}_j}^I$ are the *set of all betas and alphas* of the considered security market line, when each return ranges in its own interval of values. An investor using the interval CAPM and computing ${\hat{\beta}_j}^I$ and ${\hat{\alpha}_j}^I$ of a given asset has the possibility to know, not only the scalar *risk* and the scalar *intercept* of the regression line, but also the 'uncertainty' around those quantities.

As a future prospective of research, it could be interesting to use the *Principal Component Analysis* with interval data (GIOIA-LAURO 2006) to compute the interval portfolio of the Markowitz's model. With respect to the CAPM with interval data instead, it could be interesting the extension of the described methodology to *histogram data* or to interval data when a distribution different from the uniform one is supposed.

ACKNOWLEDGEMENTS

First of all I wish to thank Prof. V. Aversa and Prof. C.N. Lauro for their valuable support, fruitful discussions, and intelligent suggestions during the development of this research work. I am grateful to Prof. M. LiCalzi for his precious remarks and suggestions on a previous version of this paper.

REFERENCES

ALEFELD, G., HERZBERGER, J. (1983). *Introduction for interval computations,* New York: Accademic Press.

ALEFELD, G., MAYER, G. (2000). Interval analysis: theory and applications, *Journal of computation and Applied Mathematics,* **121**: 421–464.

BILLARD, L., DIDAY, E. (2000). Regression Analysis for Interval-Valued Data, *Data Analysis, Classification and Related Methods* (eds.H.-H. Bock and E. Diday), Springer: 103–124.

BILLARD, L., DIDAY, E. (2002). Symbolic regression Analysis, *Data Analysis, Classification and Clustering Methods Heidelberg,* (eds.Krzysztof Jajuga et al), Springer-Verlag.

BURKILL, JC. (1924). Functions of Intervals, *Proceedings of the London Mathematical Society* **22**: 375–446.

CAMPBELL, Y.,LO, AW., MACKINLAY, AC. (1997). *The Econometrics of Financial Markets*, New Jersey: Princeton University Press.

CARLSSON, C., FULLÉR, R., MAJLENDER, P., (2002). A possibilistic approach to selecting portfolios with highest utility score, *Fuzzy Sets and Systems* **131**: 13–21.

CHINNECK, JW., RAMADAN, K., (2000). Linear programming with interval coefficients, *Journal of the Operational Research Society* **51**: 209–220.

EICHBERGER, J., HARPER, IR., (1997). *Financial Economics,* Oxford University Press.

GIOIA, F., LAURO, C. (2005). Basic statistical methods for interval data, *Statistica Applicata* **17**: 75–104.

GIOIA, F., LAURO, C. (2006). Principal Component Analysis with Interval Data, *special issue of Computational Statistics Journal on "Statistical Analysis of Interval Data"* **21**: 343–363.

GIOVE, S., FUNARI, S., NARDELLI, C., (2006). An interval portfolio selection problem based on regret function, *European Journal of Operational Research* **170**: 253–264.

HICKEY, T., JU, Q., VAN EMDEN, MH., (2001). Interval arithmetic: From principles to implementation, *Journal of the ACM* **48**(5): 1038–1068.

INUIGUCHI, M., RAMêK, J., (2000). Possibilistic linear programming: a brief review of fuzzy mathematical programming and comparison with stochastic programming in portfolio selection problem, *Fuzzy sets and systems* **111**: 3–28.

INUIGUCHI, M., SAKAWA, M., (1995). Minimax regret solution to linear programming problems with interval objective function, *European Journal of Operational Research* **86**: 526–536.

INUIGUCHI, M., TANINO, T., (2000). Portfolio selection under independent possibilistic information, *Fuzzy sets and systems* **115**: 83–92.

ISHIBUCHI, H., TANAKA, H., (1990). Multiobjective programming in optimization of the interval objective function, *European Journal of Operational Research* **48**: 219–225.

KEARFOTT, V., KREINOVICH, RB., (1996). *Applications Of Interval Computations*, Kluwer Academic Publishers.

LAI, KK., WANG, SY., XU, JP., ZHU, SS., FANG, Y., (2002). A Class of Linear Interval Programming Problems and its Application to Portfolio Selection, *IEEE Transactions on Fuzzy Systems* **10**: 698–704.

LIMA NETO, E., DE CARVALHO, F., (2007). Centre and Range method for fitting a linear regression model to symbolic interval data, *Computational Statistic and Data Analysis* **52**: 1500–1515.

MARINO, M., PALUMBO, F., (1952). Interval Arithmetic for the evaluation of imprecise data effects in least squares linear regression, *Statistica Applicata* **3**.

MARKOWITZ, H., (2003). Portfolio selection, *The Journal of Finance* **VII**(1).

MOORE, RE., (1966). Interval Analysis, *Series in Automatic Computation*, Inc. Englewood Cliff: N.J Prentice-Hall.

NEUMAIER, A., (1990). *Interval methods for Systems of Equations*, Cambridge University Press.

RADIZKOWSKA, AM., KERRE, EE., (2002). A comparative study of fuzzy rough sets, *Fuzzy sets and systems* **126**:137–155.

RODRIGUEZ, O., (2000). Classification et Modeles Lineaires en Analyse des Donnes Symboliques, *Doctoral Thesis*, Universite de Paris Dauphine IX.

SHARPE, WF., (1964). Capital asset prices: A theory of market equilibrium under conditions of risk, *Journal of Finance* **19**(3):425–442.

SUNAGA, T., (1958). *Theory of an Interval Algebra and its Application to Numerical Analysis*, Tokyo: Gaukutsu Bunken Fukeyu-kai.

TANAGA, H., GUO, P., TURKSEN, B., (2000). Portfolio selection based on fuzzy probabilities and possibility distributions, *Fuzzy sets and systems* **111**(3):387–397.

WALCZAC, B., MASSART, DL., (1999). Rough sets theory, *Chemometrics and intelligent laboratory systems* **47**:1–16.

YAO, YY., (1998). A comparative study of fuzzy sets and rough sets, *Information Sciences* **109**:227–242.

YOUNG, R.C., (1931). The algebra of many-valued quanties, *Math. Ann* **104**:260–290.

MODELLI DI SELEZIONE DEL PORTAFOGLIO PER DATI DI INTERVALLO

Riassunto

I dati finanziari sono spesso incompleti, imprecisi o in generale soggetti ad incertezza.

L'incertezza nei dati può essere trattata considerando, piuttosto che uno scalare, l'intervallo dei valori a cui il dato appartiene. L'obiettivo del presente lavoro è di estendere il problema di selezione del portafoglio di Markowitz ed il Capital Asset Pricing Model, nel caso in cui i rendimenti dei titoli considerati o del portafoglio di mercato siano descritti da variabili di intervallo. Le metodologie introdotte calcolano l'intervallo delle soluzioni quando ciascuna quantità varia nel proprio intervallo di valori. Gli algoritmi relativi sono stati implementati in MATLAB e testati numericamente su serie storiche reali ad intervallo. I risultati numerici sono ben interpretabili.