

Capital Asset Pricing Model Using Regression Methods for Interval-Valued Variables *

F. Gioia

Department of Management Studies and Quantitative
Methods, University of Naples Parthenope
gioia@uniparthenope.it

Abstract

It is known that the Capital Asset Pricing Model (CAPM) provides an expression which relates the expected return of an asset to its systematic risk. In a decision making problem involving financial data however, we have to take in account the uncertainty given by the imprecision and the incompleteness of the information. Uncertainty in the data may be treated by considering, rather than a single value, the interval of values in which the data may fall. The extension of the CAPM to the case in which the returns of any considered asset are interval-valued variables (IntervalCAPM) has been introduced by Gioia (2009). The methodology makes use of the interval regression method Iregr presented in Gioia and Lauro (2005), but other different methods are already present in the literature and are described in the present work. A contribution of this work is the comparison of those interval regression methods on sets of real data. The algorithms of the considered methodologies have been implemented in MATLAB and the numerical results are compared to one another highlighting the good advantages for the method Iregr. Interval regression methods showed to be useful in the application of CAPM to interval financial data. However, as shown in this work, the interval regression method Iregr is more suitable in the framework of the Interval-CAPM, with respect to the other interval regression methods presented in the present work.

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

AMS subject classifications: 65-05, 91G70

1 Introduction

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 many times we do not know the exact value of the return of

*Submitted: May 5, 2015; Revised: November 12, 2015, February 4, 2016, and April 12, 2016 ; Accepted: May 6, 2016.

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 its “mean behavior” not only but also its “variability”. Variability is one of the most remarkable factors in economics and finance. In the literature there are some works on confidence-based interval forecasting methods to predict variability [6], [7] but poor quality is the major problem of these methods. In order to improve interval forecasting quality, He and Hu (2007) [10] use interval data, rather than point data, to get OLS lower and upper bound forecasts to predict variability in the stock markets. Their results suggest that their use of interval data can significantly improve the accuracy of OLS variability forecast. Accuracy of variability of interval forecasts may be measured by accuracy ratio as suggested in [13]. Further, He and Hu (2008) [11] provide evidence that higher forecasting accuracy can be reached if interval data are combined by the interval algebra tools. It is known that the Capital Asset Pricing Model (CAPM) [22] provides an expression which relates the expected return of an asset to its systematic risk [14]. The CAPM to the case in which the returns of any considered asset and the market portfolio are *interval-valued* variables (IntervalCAPM) has been introduced in [8]. The purpose of the present work is first to describe some regression methodologies for interval-valued variables already present in the literature, and second to compare those different methodologies in the framework of the Interval CAPM. A real financial case is analyzed, the numerical results using the different methodologies are compared one another. Interval regression has been treated in the literature. In [3], [4], [16] and [21], the authors derive the results from some classical regression methods which minimize criteria which are different of the least squares one. An alternative methodology, is proposed by [17] 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 [19], [1], [20], [15], [2], 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, the *slope* and *intercept* functions, when the observed values vary in their own interval of values.

In Section 2 below, the CAPM for single-valued and interval-valued variables is described. Section 3 is devoted to the methodological description of the following regression methods for interval-valued variables: CM [3], MinMax [4], CRM [16] and Iregr [9]. In Section 4 the Iregr in the framework of the CAPM is analysed pointing out the motivations of using the Interval CAPM methodology. A real financial case is analyzed. The algorithms of each considered method have been implemented in MATLAB and tested numerically. The interval time series of the asset **ABBOTT** (*Abbot Laboratories*), and of the **SP500** (*Standard and Poor’s 500 Composite*) index have been considered. The downloaded data refer to real-valued variables; we have artificially transformed these variables into interval-valued ones by applying a perturbation using a uniform distribution $U(0, 0.01)$.

The numerical results of the different methodologies have been compared. Further-

more, in the framework of the IntervalCAPM, the interval regression method Iregr, with respect to the other interval regression methods ([3], [4], [16]), shows some good advantages well described in the present work.

2 The Capital Asset Pricing Model (CAPM)

2.1 Classical CAPM

Since it is known the meaning of financial market (trade of financial products) the problem of an actor (investor) in this market is to buy a portfolio of assets which gives the highest return given a level of risk.

The common sense suggests that risky investments will generally yield higher returns than investments free of risk. With the development of the Capital Asset Pricing Model (CAPM) [22] it is possible to quantify risk and the reward for bearing it [14].

Let us suppose to have p securities denoted as S_j ($j = 1, \dots, p$) and let us suppose that t_0 is the instant in which an investor decides to buy a portfolio of those securities. The investor will diversify his portfolio according to the expected returns of the assets at a future instant of time t_1 . Instant by instant the prices of the assets are determinate according to the law of supply and demand; this occurs at both considered instants with the difference that at time t_0 the prices are known, while at instant t_1 are expected (according to the behavior of the investors at time t_0).

Let us indicate with x_j the portion of total investment funds devoted to this security. Thus

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

$\mathbf{x} = (x_1, x_2, \dots, x_p)^T$ is the portfolio of the considered consumer and X is the set of possible portfolios.

At time t_1 the returns of the p securities are assumed to be $n \times 1$ random variables denoted by: \mathbf{R}_j ($j = 1, \dots, n$) (n different states of the world are contemplated), represented as columns of the following $n \times p$ matrix:

$$\mathbf{R}^T = \begin{pmatrix} R_{11} & R_{12} & \cdots & R_{1p} \\ R_{21} & R_{22} & \cdots & R_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ R_{n1} & R_{n2} & \cdots & R_{np} \end{pmatrix} \quad (2.1)$$

Let us suppose a discrete probability distribution $\mathbf{p} = (p_1, p_2, \dots, p_n)$ on the outcomes of the random variable \mathbf{R}_j , ($j = 1, \dots, p$). The expected return of the j^{th} security is:

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

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

The expected return and the variance of a portfolio \mathbf{x} can be written as:

$$E(\mathbf{x}) = \sum_{j=1}^n E(\mathbf{R}_j) \cdot x_j \quad (2.2)$$

$$\sigma^2(\mathbf{x}) = \sum_{j=1}^n x_j \cdot \sum_{i=1}^n x_i \sigma(\mathbf{R}_i, \mathbf{R}_j) \quad (2.3)$$

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.

Markowitz, in [18] stated the investor's portfolio selection problem in terms of expected return and variance of return. In particular he argued that investors would optimally hold a mean-variance efficient portfolio, that is, a portfolio with the highest expected return for a given level of variance.

Since the variance of a portfolio \mathbf{x} is regarded as the risk of investment, the (conditional) best investment is one with the minimum variance (2.3) subject to a given return E_0 . This leads to the following quadratic programming problem:

$$\min_{\mathbf{x}} \sigma^2(\mathbf{x}) \quad (2.4)$$

so that

$$\begin{aligned} E(\mathbf{x}) &= E_0 \\ \sum_{j=1}^n x_j &= 1 \\ x_i &\geq 0 \end{aligned}$$

The Capital Asset Pricing Model (CAPM) [14] deduce from an individual decision procedure (Markowitz), a decision criterion which is shared by all agents. The additional hypothesis introduced to derive the model is that the market is in *equilibrium*, i.e., that the sum of the quantities of all assets available in the market is equal to the sum of the quantities of all assets required. The model provides an expression which relates the expected return of an asset to its risk as will be shown by the theorem herewith reported.

Let us define the *market portfolio* as the vector $\mathbf{A} = (A_1, A_2, \dots, A_n)$ in which A_j is the aggregate quantity of the j^{th} risky asset available in the economy. In other words the market portfolio is a portfolio consisting of a weighted sum of every asset in the market, with weights in the proportions that they exist in the market, with the necessary assumption that these assets are infinitely divisible.

In analogy with (2.2) the expected return on the market portfolio is defined as:

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

The following Theorem holds true:

Theorem 1 *If the market is in equilibrium, for each asset \mathbf{R}_j then*

$$E(\mathbf{R}_j) - r_f = (E(\mathbf{A}) - r_f) \cdot \beta_j \quad (2.5)$$

where r_f is the risk-free rate and the factor of proportionality β_j has the following expression:

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

having indicated with $\sigma(\mathbf{A}, j)$ the covariance between the return of the market portfolio and the return of the j^{th} asset, and with $\sigma^2(\mathbf{A})$ the variance of the market portfolio.

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 model (2.5) may have different uses:

1. given the *expected returns* $E[\mathbf{R}_j]$ and $E[\mathbf{A}]$ of asset S_j and the market portfolio respectively, and given the risk free rate r_f , calculate the corresponding β_j ;
2. given β_j of S_j and the risk free rate r_f , calculate the *expected return* $E[\mathbf{R}_j]$;
3. given the risk-free rate r_f , use an estimation method to determine the β_j of a particular risky asset and therefore compute the expected return $E[\mathbf{R}_j]$.

Let us analyze point 3 in more details. The CAPM is a single-period model [5]; hence (2.5) do not have a time dimension. Although time-series of returns are readily available and one can use familiar estimation methods to determine the β of a particular risky asset.

Let \mathbf{Z}_j and \mathbf{Z}_m be the vectors of the excess returns ¹ of the asset S_j and the market portfolio respectively in different periods of time, those excess returns can be described using the excess-return single-index *marked model*:

$$\mathbf{Z}_j = \alpha + \beta \mathbf{Z}_m + \mathbf{e}_j, \tag{2.6}$$

where \mathbf{e}_j is the vector of disturbances. The following hypothesis concerning time-independence hold:

$$\begin{aligned} E(\mathbf{e}_j) &= 0 \\ E(\mathbf{e}_j \mathbf{e}_j^T) &= \Sigma \\ E(\mathbf{Z}_m) &= \mu_m, \\ E[(\mathbf{Z}_m - \mu_m)^2] &= \mu_m^2 \\ \sigma(\mathbf{Z}_m, \mathbf{e}_j) &= \Sigma \end{aligned}$$

It is known from classical theory that estimators for β and α are the OLS (Ordinary Least Square) estimators; from the estimation of β , 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.

¹ $\mathbf{Z}_j = \mathbf{R}_j - r_f$ and $\mathbf{Z}_m = \mathbf{A} - r_f$

2.2 The Interval CAPM

As mentioned previously, intervals may be useful for representing *uncertainty* in financial data. In the case of interval returns an investor should ask the following questions: if the returns *fluctuate* from their fixed values, how do the perturbations reflect on the risk? In which interval that risk ranges? In which interval the estimate of the actual return of the considered asset falls? The CAPM, to the case in which the returns of any considered asset are *interval-valued* variables (IntervalCAPM), has been introduced in [8] and is here below described.

Let us suppose for example that the return of the security S_j ($j = 1, \dots, n$) not only varies with time but can be represented by an *interval* of values when each state occurs. Thus \mathbf{R}_j ($j = 1, \dots, n$) are assumed to be *interval-valued* variables denoted by: \mathbf{R}_j^I ($j = 1, \dots, n$), and represented as columns in the following interval matrix:

$$(\mathbf{R}^T)^I = \begin{pmatrix} [\underline{R}_{11}, \bar{R}_{11}] & [\underline{R}_{12}, \bar{R}_{12}] & \cdots & [\underline{R}_{1p}, \bar{R}_{1p}] \\ [\underline{R}_{21}, \bar{R}_{21}] & [\underline{R}_{22}, \bar{R}_{22}] & \cdots & [\underline{R}_{2p}, \bar{R}_{2p}] \\ \vdots & \vdots & \ddots & \vdots \\ [\underline{R}_{n1}, \bar{R}_{n1}] & [\underline{R}_{n2}, \bar{R}_{n2}] & \cdots & [\underline{R}_{np}, \bar{R}_{np}] \end{pmatrix} \quad (2.7)$$

where: $[\underline{R}_{ij}, \bar{R}_{ij}]$ is the interval in which the return rate of security S_j “falls” when the i^{th} state occurs.

The aim of the Interval CAPM is to extend the model (2.6) to the case of interval returns.

More in details: when an interval data matrix of returns (2.7) is given, also the excess returns of the j^{th} title and of the market portfolio are described by the following *interval-valued* variables respectively:

$$\mathbf{Z}_j^I = (Z_{jt} = [\underline{Z}_{jt}, \bar{Z}_{jt}]), \quad t = 1, \dots, n \quad (2.8)$$

$$\mathbf{Z}_m^I = (Z_{mt} = [\underline{Z}_{mt}, \bar{Z}_{mt}]), \quad t = 1, \dots, n \quad (2.9)$$

In this special case the regression equation (2.6) must be treated by means of an *interval regression method*; in the following section four different methodologies, already present in the literature are described: *Centre based Method*, *MinMax Method*, *Centre and Range based Method*, *Optimal interval of solution based method (Iregr)*.

The Interval CAPM introduced in [8] in particular apply the latter method *Iregr* (section 4.4) which make extensively use of the interval algebra tools combined with some optimization techniques for treating equation (2.6) as an interval *interval regression equation*:

$$\mathbf{Z}_j^I = \hat{\alpha}_j^I + \hat{\beta}_j^I \mathbf{Z}_m^I + \mathbf{E}^I \quad (2.10)$$

The methodology computes the *interval* $\hat{\beta}_j^I$ and the *interval* $\hat{\alpha}_j^I$ of the interval regression line (2.10) which are both well interpretable:

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} , $i = 1, \dots, k$ ranges in its own interval of values.

2. Remarking that for real-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 $E(\mathbf{R}_j)$ of security S_j for each $R_{ij} \in [\underline{R}_{ij}, \bar{R}_{ij}]$.

The estimation of an interval Z_{jt} is given by:

$$Z_{jt} = \hat{\alpha}_j^I + \hat{\beta}_j^I Z_{mt}$$

where the operations are interval algebra operations.

In the next section it could be of help to discuss in detail the cited interval regression methods in order to highlight the advantages in using the regression method *Iregr*.

3 Interval Regression

In this section different methods for fitting a linear regression of interval-valued variables will be presented. Two different approaches, which have the common task of forecasting a dependent interval-valued variable versus one or more independent interval-valued variables, are explained. The first approach, followed by [3], [4], [16], consists in forecasting the dependent interval-valued variable using some regression functions with *scalar* parameters. The authors derive the results from some classical regression methods which minimize criteria different from the least squares one. The second approach, our methodology [9], follows the alternative approach of forecasting the dependent interval-valued variable computing also the *interval parameters*, i.e., defining the *interval regression line*; the methodology uses optimization techniques for computing, rather than the interval solution, 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 interval of variation. It exists in [17] an attempt to compute the interval parameters of the regression line by means of the interval algebra tools. It has been proved, however, that this approach leads to the resolution of a linear interval system which has a very complicated structure and that has no convex solutions. Let us consider a dependent and p independent interval-valued variables:

$\mathbf{Y}^I = (Y_i = [\underline{y}_i, \bar{y}_i])$, $\mathbf{X}_j^I = (X_{ij} = [\underline{x}_{ij}, \bar{x}_{ij}])$, $i = 1, \dots, n$, $j = 1, \dots, p$. Let $E = \{e_1, \dots, e_m\}$ be a set of examples described by the $p + 1$ interval-valued variables $\mathbf{Y}^I, \mathbf{X}_1^I, \dots, \mathbf{X}_p^I$. Each example $e_i \in E$ is represented by the interval vector $\mathbf{Z}_i^I = (X_{i1}, \dots, X_{ip}, Y_i)$ where X_{ij} and Y_i are the observed values of \mathbf{X}_j^I and \mathbf{Y}^I respectively.

3.1 Centre Based Method

Here below the Billard and Diday (2000) method (CM) will be presented. The authors consider the following two linear relationship:

$$\underline{y}_i = \beta_0 + \beta_1 \underline{x}_{i1} + \dots + \beta_p \underline{x}_{ip} + \varepsilon_i \tag{3.1}$$

$$\bar{y}_i = \beta_0 + \beta_1 \bar{x}_{i1} + \dots + \beta_p \bar{x}_{ip} + \bar{\varepsilon}_i \tag{3.2}$$

and minimize the *sum of squares deviations*:

$$C_1 = \sum_{i=1}^n (\varepsilon_i + \bar{\varepsilon}_i)^2 \tag{3.3}$$

which represent the sum of the squares of the lower plus the upper bound errors. It can be proved that, the estimates of $\beta_0, \beta_1, \dots, \beta_p$ that minimize (3.3) are the solution of the following system of $p + 1$ normal equations written in matrix form:

$$\hat{\boldsymbol{\beta}} = (\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_p)^T = (\mathbf{A})^{-1} \mathbf{b} \quad (3.4)$$

where \mathbf{A} is a $(p + 1) \times (p + 1)$ matrix and \mathbf{b} is a $(p + 1) \times 1$ vector here below described:

$$\mathbf{A} = \begin{pmatrix} n & \sum_i x_{i1}^c & \cdots & \sum_i x_{ip}^c \\ \sum_i x_{i1}^c & \sum_i (x_{i1}^c)^2 & \cdots & \sum_i x_{ip}^c x_{i1}^c \\ \vdots & \vdots & \ddots & \vdots \\ \sum_i x_{ip}^c & \sum_i x_{i1}^c x_{ip}^c & \cdots & \sum_i (x_{ip}^c)^2 \end{pmatrix}$$

and

$$\mathbf{b} = \left(\sum_i y_i^c, \sum_i y_i^c x_{i1}^c, \dots, \sum_i y_i^c x_{ip}^c \right)^T$$

Given a new example e , described by $\mathbf{Z}^I = (X_1, \dots, X_p, Y)$ where $X_j = [\underline{x}_j, \bar{x}_j]$ ($j = 1, \dots, p$), the prediction $\hat{Y} = [\hat{y}, \hat{\bar{y}}]$ of the interval $Y = [y, \bar{y}]$ will be computed as follow:

$$\hat{y} = (\underline{\mathbf{x}})^T \hat{\boldsymbol{\beta}} \text{ and } \hat{\bar{y}} = (\bar{\mathbf{x}})^T \hat{\boldsymbol{\beta}}$$

where $\underline{\mathbf{x}}^T = (1, \underline{x}_1, \dots, \underline{x}_p)$ and $\bar{\mathbf{x}}^T = (1, \bar{x}_1, \dots, \bar{x}_p)$.

3.2 MinMax Based Method

Here below the Billard and Diday (2002) method (MinMax) will be presented. The idea is to suppose the *independence* between lower and upper bounds of the considered intervals, and to take into account two different linear equations:

$$\underline{y}_i = \beta_0^L + \beta_1^L \underline{x}_{i1} + \dots + \beta_p^L \underline{x}_{ip} + \varepsilon_i \quad (3.5)$$

$$\bar{y}_i = \beta_0^U + \beta_1^U \bar{x}_{i1} + \dots + \beta_p^U \bar{x}_{ip} + \bar{\varepsilon}_i. \quad (3.6)$$

The criterion to be minimized is:

$$C_2 = \sum_{i=1}^n (\varepsilon_i)^2 + \sum_{i=1}^n (\bar{\varepsilon}_i)^2 \quad (3.7)$$

which represents the sum of the lower bound square errors plus the sum of the upper bound square errors, considering *independent* vectors of the parameters to predict the bounds of the intervals. It can be proved that, the estimates of $\beta_0^L, \beta_1^L, \dots, \beta_p^L$ and $\beta_0^U, \beta_1^U, \dots, \beta_p^U$ that minimize (3.7) are the solution of the following system of $2(p + 1)$ normal equations written in matrix form:

$$\hat{\boldsymbol{\beta}} = (\hat{\beta}_0^L, \hat{\beta}_1^L, \dots, \hat{\beta}_p^L, \hat{\beta}_0^U, \hat{\beta}_1^U, \dots, \hat{\beta}_p^U)^T = (\mathbf{A})^{-1} \mathbf{b} \quad (3.8)$$

where \mathbf{A} is a $2(p + 1) \times 2(p + 1)$ matrix and \mathbf{b} is a $2(p + 1) \times 1$ vector here below described:

$$\mathbf{A} = \begin{pmatrix} n & \sum_i \underline{x}_{i1} & \cdots & \sum_i \underline{x}_{ip} & 0 & \cdots & 0 \\ \sum_i \underline{x}_{i1} & \sum_i (\underline{x}_{i1})^2 & \cdots & \sum_i \underline{x}_{ip}\underline{x}_{i1} & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \sum_i \underline{x}_{ip} & \sum_i \underline{x}_{i1}\underline{x}_{ip} & \cdots & \sum_i (\underline{x}_{ip})^2 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 & n & \cdots & \sum_i \bar{x}_{ip} \\ 0 & 0 & \cdots & 0 & \sum_i \bar{x}_{i1} & \cdots & \sum_i \bar{x}_{ip}\bar{x}_{i1} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & 0 & \sum_i \bar{x}_{i1}\bar{x}_{ip} & \cdots & \sum_i (\bar{x}_{ip})^2 \end{pmatrix}$$

and

$$\mathbf{b} = \left(\sum_i \underline{y}_i, \sum_i \underline{y}_i \underline{x}_{i1}, \dots, \sum_i \underline{y}_i \underline{x}_{ip}, \sum_i \bar{y}_i, \sum_i \bar{y}_i \bar{x}_{i1}, \dots, \sum_i \bar{y}_i \bar{x}_{ip} \right)^T$$

Given a new example e , described by $\mathbf{Z}^I = (X_1, \dots, X_p, Y)$ where $X_j = [\underline{x}_j, \bar{x}_j]$ ($j = 1, \dots, p$), the prediction $\hat{Y} = [\hat{y}, \hat{\bar{y}}]$ of the interval $Y = [y, \bar{y}]$ will be computed as follow:

$$\hat{y} = (\mathbf{x})^T \hat{\boldsymbol{\beta}}^L \text{ and } \hat{\bar{y}} = (\bar{\mathbf{x}})^T \hat{\boldsymbol{\beta}}^U$$

where $\mathbf{x}^T = (1, \underline{x}_1, \dots, \underline{x}_p)$, $\bar{\mathbf{x}}^T = (1, \bar{x}_1, \dots, \bar{x}_p)$, $\hat{\boldsymbol{\beta}}^L = (\hat{\beta}_0^L, \hat{\beta}_1^L, \dots, \hat{\beta}_p^L)^T$ and $\hat{\boldsymbol{\beta}}^U = (\hat{\beta}_0^U, \hat{\beta}_1^U, \dots, \hat{\beta}_p^U)^T$.

Remark

It is important to remark that in [3], [4] it is assumed that, for an interval valued function $Y^I = f(X_1^I, X_2^I, \dots, X_n^I)$, the lower bound of Y^I depends only on the lower bounds of $X_1^I, X_2^I, \dots, X_n^I$, and the upper bound of Y^I depends only on the upper bounds of $X_1^I, X_2^I, \dots, X_n^I$. These assumptions can be misleading at both the theoretical and practical levels as shown in [11] where is reported that, using least-squares to forecast lower and upper bounds separately for the SP 500, in a considerable number of cases the predicted upper bound results less than the predicted lower bound.

3.3 Centre and Range Based Method

In the CM and MinMax methods we have seen that the estimate of the parameters β involves only the extreme values of the intervals according to the criterion considered. In [12] results suggest that both the midpoint and interval computing methods can generate significantly higher accuracy in variability forecast than the OLS lower and upper bound method. The (CRM) method, proposed by Neto and de Carvalho (2007), estimates the β taking into account both the centre and the radius of each considered interval in order to improve the model prediction performance.

Let us indicate with

$x_{ij}^c = (\underline{x}_{ij} + \bar{x}_{ij})/2$, $x_{ij}^r = (\bar{x}_{ij} - \underline{x}_{ij})/2$ the centre and the radius of the i^{th} element of the j^{th} interval-valued variable \mathbf{X}_j^I . Let us also indicate with $y_i^c = (\underline{y}_i + \bar{y}_i)/2$ and $y_i^r = (\bar{y}_i - \underline{y}_i)/2$ the centre and the radius of the i^{th} component of the dependent interval-valued \mathbf{Y}^I respectively. The authors consider the following linear relationships:

$$y_i^c = \beta_0^c + \beta_1^c x_{i1}^c + \dots + \beta_p^c x_{ip}^c + \varepsilon_i^c \quad (3.9)$$

$$y_i^r = \beta_0^r + \beta_1^r x_{i1}^r + \dots + \beta_p^r x_{ip}^r + \varepsilon_i^r \quad (3.10)$$

The criterion to be minimized is:

$$C_3 = \sum_{i=1}^n ((\varepsilon_i^c)^2 + (\varepsilon_i^r)^2) \quad (3.11)$$

which represents the sum of the *mid-points* square error plus the *ranges* square error of the intervals.

It can be proved that, the estimates of $\beta_0^c, \beta_1^c, \dots, \beta_p^c$ and $\beta_0^r, \beta_1^r, \dots, \beta_p^r$ that minimize (3.11) are the solution of the following system of $2(p+1)$ *normal equations* written in matrix form:

$$\hat{\boldsymbol{\beta}} = (\hat{\beta}_0^c, \hat{\beta}_1^c, \dots, \hat{\beta}_p^c, \hat{\beta}_0^r, \hat{\beta}_1^r, \dots, \hat{\beta}_p^r)^T = (\mathbf{A})^{-1} \mathbf{b} \quad (3.12)$$

where \mathbf{A} is a $2(p+1) \times 2(p+1)$ matrix and \mathbf{b} is a $2(p+1) \times 1$ vector here below described:

$$\mathbf{A} = \begin{pmatrix} n & \sum_i x_{i1}^c & \dots & \sum_i x_{ip}^c & 0 & \dots & 0 \\ \sum_i x_{i1}^c & \sum_i (x_{i1}^c)^2 & \dots & \sum_i x_{ip}^c x_{i1}^c & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \sum_i x_{ip}^c & \sum_i x_{i1}^c x_{ip}^c & \dots & \sum_i (x_{ip}^c)^2 & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 & n & \dots & \sum_i x_{ip}^r \\ 0 & 0 & \dots & 0 & \sum_i x_{i1}^r & \dots & \sum_i x_{ip}^r x_{i1}^r \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots & 0 & \sum_i x_{i1}^r x_{ip}^r & \dots & \sum_i (x_{ip}^r)^2 \end{pmatrix}$$

and

$$\mathbf{b} = \left(\sum_i y_i^c, \sum_i y_i^c x_{i1}^c, \dots, \sum_i y_i^c x_{ip}^c, \sum_i y_i^r, \sum_i y_i^r x_{i1}^r, \dots, \sum_i y_i^r x_{ip}^r \right)^T$$

Let us consider again an example e , described by $\mathbf{Z}^I = (X_1, \dots, X_p, Y)$ where $X_j = [\underline{x}_j, \bar{x}_j]$ ($j = 1, \dots, p$). Let us indicate with $x_j^c = (\underline{x}_j + \bar{x}_j)/2$ and $x_j^r = (\bar{x}_j - \underline{x}_j)/2$ ($j = 1, \dots, p$) the centre and the radius of X_j . The value of an interval $Y = [\underline{y}, \bar{y}]$ of \mathbf{Y}^I will be predicted by $\hat{Y} = [\hat{y}, \hat{\bar{y}}]$ as follows:

$$\hat{y} = \hat{y}^c - \hat{y}^r \text{ and } \hat{\bar{y}} = \hat{y}^c + \hat{y}^r$$

where

$$\hat{y}^c = (\mathbf{x}^c)^T \hat{\boldsymbol{\beta}}^c, \hat{y}^r = (\mathbf{x}^r)^T \hat{\boldsymbol{\beta}}^r, (\mathbf{x}^c)^T = (1, x_{i1}^c, \dots, x_{ip}^c), (\mathbf{x}^r)^T = (1, x_{i1}^r, \dots, x_{ip}^r), \hat{\boldsymbol{\beta}}^c = (\hat{\beta}_0^c, \hat{\beta}_1^c, \dots, \hat{\beta}_p^c)^T \text{ and } \hat{\boldsymbol{\beta}}^r = (\hat{\beta}_0^r, \hat{\beta}_1^r, \dots, \hat{\beta}_p^r)^T.$$

3.4 Optimal Interval of Solution Based Method (Ireg)

We have seen, in the previous sections, that the prediction of an interval $\hat{Y} = [\hat{y}, \hat{\bar{y}}]$, if computed by CM, MinMax, or CRM 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. The aim of the present section is to describe the *interval regression method* (Iregr) [9]; the methodology makes extensively use of the *interval algebra* tools combined with some *optimization techniques* to consider the interval as a whole structure and to compute, rather than the interval solution, 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 interval of values. Iregr follows the approach of forecasting an interval-valued variable linearly dependent from one or several independent interval-valued variables, defining the *interval regression equation*, i.e., a regression relationship with *interval parameters*.

Let us first consider the case of simple interval regression, thus let us consider the following independent and dependent interval-valued variables respectively:

$$\mathbf{X}^I = (X_i = [\underline{x}_i, \bar{x}_i]), \quad i = 1, \dots, n$$

$$\mathbf{Y}^I = (Y_i = [\underline{y}_i, \bar{y}_i]), \quad i = 1, \dots, n$$

it is:

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

where \mathbf{E}^I is the error 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 $i = 1, \dots, n$.

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 $x_1, x_2, \dots, x_n, y_1, y_2, \dots, y_n$ take a particular value in their own interval of variation.

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

$$\hat{\beta}^I = \left\{ \hat{\beta}(\mathbf{X}, \mathbf{Y}) = \frac{\sum_{i=1}^n (x_i - x_m)(y_i - y_m)}{\sum_{i=1}^n (x_i - x_m)^2}, \quad / \mathbf{X} \in \mathbf{X}^I, \mathbf{Y} \in \mathbf{Y}^I \right\} \quad (3.13)$$

$$\hat{\alpha}^I = \left\{ \hat{\alpha}(\mathbf{X}, \mathbf{Y}) = y_m - \hat{\beta}x_m, \quad / \mathbf{X} \in \mathbf{X}^I, \mathbf{Y} \in \mathbf{Y}^I \right\} \quad (3.14)$$

where x_m and y_m , regarded as functions of $x_1, x_2, \dots, x_n, y_1, y_2, \dots, y_n$, are given by

$$x_m = \frac{1}{n} \sum_{i=1}^n x_i \quad ; \quad y_m = \frac{1}{n} \sum_{i=1}^n y_i$$

These sets may be computed numerically by solving some optimization problems; i.e., searching for the minimum and for the maximum of functions $\hat{\alpha}(\mathbf{X}, \mathbf{Y})$ and $\hat{\beta}(\mathbf{X}, \mathbf{Y})$ in (3.13) and (3.14).

These functions are both continuous² on a connected and compact set and this assures that sets (3.13) and (3.14) are the following closed intervals

$$\hat{\beta}^I = \left[\begin{array}{c} \min_{\substack{\mathbf{X} \in \mathbf{X}^I \\ \mathbf{Y} \in \mathbf{Y}^I}} \hat{\beta}(\mathbf{X}, \mathbf{Y}), \max_{\substack{\mathbf{X} \in \mathbf{X}^I \\ \mathbf{Y} \in \mathbf{Y}^I}} \hat{\beta}(\mathbf{X}, \mathbf{Y}) \end{array} \right] \quad (3.15)$$

$$\hat{\alpha}^I = \left[\begin{array}{c} \min_{\substack{\mathbf{X} \in \mathbf{X}^I \\ \mathbf{Y} \in \mathbf{Y}^I}} \hat{\alpha}(\mathbf{X}, \mathbf{Y}), \max_{\substack{\mathbf{X} \in \mathbf{X}^I \\ \mathbf{Y} \in \mathbf{Y}^I}} \hat{\alpha}(\mathbf{X}, \mathbf{Y}) \end{array} \right] \quad (3.16)$$

and may be interpreted as follows: *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

$$\mathbf{Y} = \hat{\alpha} + \hat{\beta}\mathbf{X} \quad (3.17)$$

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

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

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

Let us turn again to the case of multiple regression; i.e., let us consider a dependent and p independent interval-valued variables

$$\mathbf{Y}^I = (Y_i = [y_i, \bar{y}_i]), \mathbf{X}_j^I = (X_{ij} = [x_{ij}, \bar{x}_{ij}]), \quad i = 1, \dots, n, \quad j = 1, \dots, p. \text{ It is}$$

$$\mathbf{Y}^I = \beta_0^I + \beta_1^I \mathbf{X}_1^I + \beta_2^I \mathbf{X}_2^I + \dots + \beta_p^I \mathbf{X}_p^I + \mathbf{E}^I$$

where \mathbf{E}^I is the error interval component. Also in this case the aim is to take into account *all possible values* of the components $x_{ij} y_i$ each of which is in its interval of values $[x_{ij}, \bar{x}_{ij}], [y_i, \bar{y}_i]$ for $i = 1, \dots, n$.

The method computes the *set of hyperplanes* each of which realizes the best fit, in the Minimum Least Square sense, of a set of points in the space. A hyperplane is univocally determined each time the components take a particular value in their own interval of variation.

Let us indicate with: $\mathbf{X}^I = (\mathbf{1}, \mathbf{X}_1^I, \mathbf{X}_2^I, \dots, \mathbf{X}_p^I)$ the $n \times (p + 1)$ interval matrix in which the first column is the $n \times 1$ vector of ones, and the other $p - 1$ columns are the considered p independent interval-valued variables. The method consists in describing the following interval vector:

$$\begin{aligned} \hat{\beta}^I &= (\hat{\beta}_0^I, \hat{\beta}_1^I, \dots, \hat{\beta}_p^I) = \\ &= \{ \hat{\beta} = (\hat{\beta}_0(\mathbf{X}, \mathbf{Y}), \hat{\beta}_1(\mathbf{X}, \mathbf{Y}), \dots, \hat{\beta}_p(\mathbf{X}, \mathbf{Y})) = (X^T X)^{-1} X^T Y, \quad \mathbf{X} \in \mathbf{X}^I, \mathbf{Y} \in \mathbf{Y}^I \} \end{aligned}$$

which is the set of all vectors $\hat{\beta} = (\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_p)$ coefficients of some hyperplane:

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

$$Y = \hat{\beta}_0 + \hat{\beta}_1 \mathbf{X}_1 + \dots + \hat{\beta}_p \mathbf{X}_p$$

realizing the best fit, in the Minimum Least Square sense, of a set of points in the space generated by x_{ij} and y_i ($i = 1, \dots, n, j = 1, \dots, p$), each of which in its own interval of variation. Mathematically the following optimization problems are solved:

$$\hat{\beta}_j^I = \left[\begin{array}{c} \min_{\substack{\mathbf{X} \in \mathbf{X}^I \\ \mathbf{Y} \in \mathbf{Y}^I}} \hat{\beta}_j(\mathbf{X}, \mathbf{Y}), \max_{\substack{\mathbf{X} \in \mathbf{X}^I \\ \mathbf{Y} \in \mathbf{Y}^I}} \hat{\beta}_j(\mathbf{X}, \mathbf{Y}) \end{array} \right] \quad (3.18)$$

for each $j = 1, \dots, p$.

In order to analyse some special cases of Iregr on simulated data, let us recall the definition of *interval correlation coefficient* between two interval-valued variables [9]. It is known that the correlation between two single-valued variables \mathbf{X} and \mathbf{Y} it may be written as follow:

$$corr(\mathbf{X}, \mathbf{Y}) = \frac{\sigma(\mathbf{X}, \mathbf{Y})}{\sigma(\mathbf{x}_r)\sigma(\mathbf{x}_s)} \quad (3.19)$$

In the case in which the variables are interval-valued variables $\mathbf{X}^I, \mathbf{Y}^I$ the interval correlation coefficient is the set of all correlation coefficients that may be calculated when each component of the variables varies in its interval of variation:

$$\mathbf{Corr}(\mathbf{X}^I, \mathbf{Y}^I) = [corr(\mathbf{X}, \mathbf{Y}) \mid \mathbf{X} \in \mathbf{X}^I, \mathbf{Y} \in \mathbf{Y}^I]$$

the above set is computed numerically finding the extrema of the nonlinear multi-variable function (3.19) as follows:

$$\mathbf{Corr}(\mathbf{X}^I, \mathbf{Y}^I) = \left[\begin{array}{c} \min_{\substack{\mathbf{X} \in \mathbf{X}^I \\ \mathbf{Y} \in \mathbf{Y}^I}} corr(\mathbf{X}, \mathbf{Y}), \max_{\substack{\mathbf{X} \in \mathbf{X}^I \\ \mathbf{Y} \in \mathbf{Y}^I}} corr(\mathbf{X}, \mathbf{Y}) \end{array} \right] \quad (3.20)$$

Remark It is important to remark that Iregr is a method which is quite sensitive to the characteristic of the input data, in particular it works well when: i) there are no outliers; ii) the radii of the intervals reflect natural fluctuations excluding exceptional events; iii) the intervals are well separated.

Illustrative examples of (3.15), (3.16) and of (3.20).

Let us analyze some special cases of Iregr on simulated data (the interval slope, the interval intercept and the interval correlation will be indicated as Beta, Alpha and Corr respectively).

The simulated data, reported in Figure 1, is a cloud of rectangles rather dispersed and clearly not correlated. 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. In Figures 3,4 and 5, we analyze the interval regression

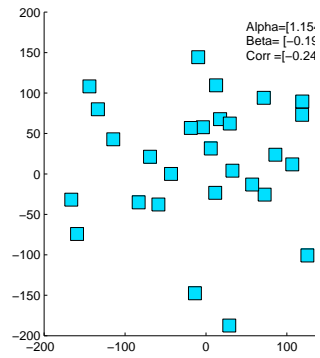


Figure 1: High dispersion

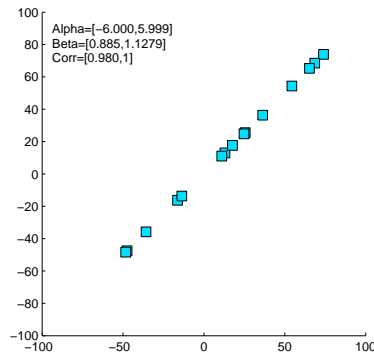


Figure 2: Max correlation

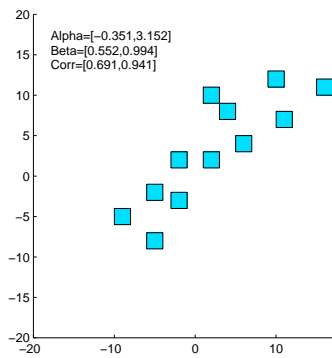


Figure 3: Cloud of rectangles

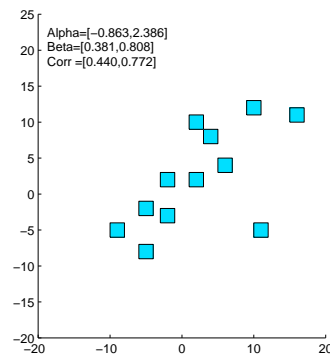


Figure 4: Shift of a rectangle

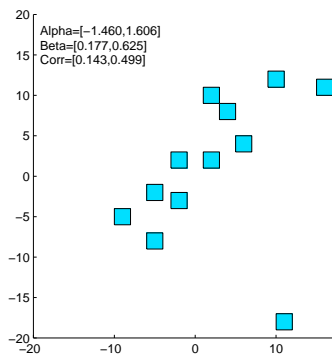


Figure 5: Shift of a rectangle

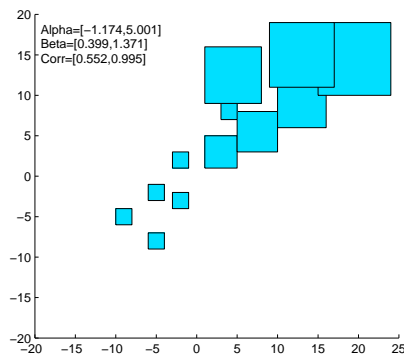


Figure 6: Different forms

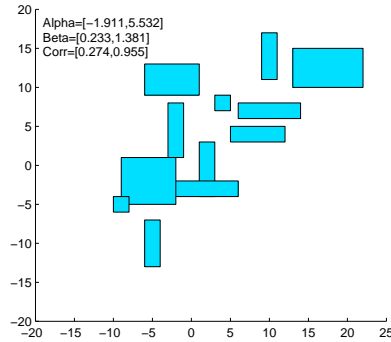


Figure 7: Different forms

coefficients and the interval correlation of some clouds of rectangles in which one rectangle is changing its position with respect to the second axis. In Figure 3 the regression coefficients $\text{Alpha}=[-0.351, 3.152]$, $\text{Beta}=[0.552, 0.994]$ and the correlation between the variables $\text{Corr}=[0.691, 0.941]$, are well in agreement with the position of the rectangles. Furthermore perturbing the data with an anomalous rectangle as in Figure 4 and 5, the interval correlations $\text{Corr}=[0.440, 0.772]$, $\text{Corr}=[0.143, 0.499]$ become intervals containing lower values with respect to the interval correlation in Figure 3. This is perfectly in agreement with the increasing variability of the considered cloud of the rectangles. Moreover, the slopes $\text{Beta}=[0.381, 0.808]$, $\text{Beta}=[0.177, 0.625]$, influenced by the anomalous rectangle, are intervals containing lower values but a bigger radius with respect to those reported in Figure 3. In Figures 6 and 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, $\text{Beta}=[0.399, 1.371]$, $\text{Beta}=[0.233, 1.381]$, present a bigger radius according to the increased variability of the input data.

Remark

It is important to remark that, in the special case of simple regression, the methods CM, MinMax and CRM may have some problem with the inversion of matrix \mathbf{A} . This problem is bypassed with Iregr, which computes $\hat{\alpha}^I$ and $\hat{\beta}^I$ without inverting any matrix but using explicit formulas as in (3.14), (3.13).

4 Iregr in the Framework of the Interval CAPM

Using the previous interval regression methods *Centre based Method*, *MinMax Method*, *Centre and Range based Method*, for the CAPM with interval data, an estimate of the interval expected return of an asset S_j has been computed, but no information is given for the slope and the intercept of the regression line; i.e., we don't know what happens to the *risk* and to the intercept of the regression line when the returns vary in their interval of variation. This is a drawback if we would like to answer to the previous introduced questions: if the returns *fluctuate* from their fixed values, how the

perturbations reflect on the *risk*? In what interval that risk ranges? 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 (2.10) which are both well interpretable: $\hat{\beta}_j^I$ and $\hat{\alpha}_j^I$ are the set of estimations of *all slopes* and *intercepts* of the considered security market line, when each return ranges in its own interval of values; an investor using the IntervalCAPM 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*, but also the “uncertainty” around those quantities. In the following subsection the aim is to apply and to compare one to another the different regression methods described before in the framework of the interval CAPM on financial data. Iregr appears to be particularly appropriate; the numerical results will be compared among them and the advantages in using Iregr, with respect to the other methods, will be pointed out.

4.1 A Real case

In the following examples the IntervalCAPM, using different regression methods, is applied to the asset **ABBOTT** (*Abbot Laboratories*), which belongs to the **SP500** (*Standard and Poor’s 500 Composite*) index. The downloaded data refer to real-valued variables; we have artificially transformed these variables into interval-valued ones by applying a perturbation using a uniform distribution $\mathbf{U}(0, 0.01)$, in order to analyze how a perturbation on the input data reflect on the final solution.

- **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 R_f).
2. The input real-valued time series are transformed into interval-valued one by a perturbation using an uniform distribution $\mathbf{U}(0, 0.01)$.
3. The IntervalCAPM, using different regression methods, is applied and the results are compared one to another.

- **The Output Parameters**

The estimations of α and β in (2.6) (indicated as *Alpha* and *Beta*) of the classical CAPM on the real-valued variables are showed in Table 1. The results of the IntervalCAPM, using CM, MinMax, CRM and Iregr interval regression methods are reported in the same table. A simple interval regression is computed thus *Alpha* and *Beta* are the first two components of (3.4), (3.8), (3.12) with respect to CM, MinMax and CRM respectively. It is important to remark that CM, MinMax and CRM do not produce interval results thus the pair of numbers shown in parenthesis are not necessarily ordered. The results of Iregr are instead *intervals*, in particular *Alpha* and *Beta* are (3.14) and (3.13) respectively.

In Table 2 we present the comparison among CM, MinMax, CRM and Iregr methods. The first column of the table, refers to the *observed* returns of the

Method	<i>Alpha</i>	<i>Beta</i>
CAPM	0	0.806
IntervalCAPM		
CM	-0.0210	0.5156
CR	(-0.0223,-0.0197)	(0.5350,0.5350)
CRN	(-0.0210,0.0058)	(0.5156,-0.1276)
Iregr	[-0.044,0.005]	[0.060,1.007]

Table 1: CAPM/IntervalCAPM

Asset: ABBOTT	Method: CM	Method: MinMax	Method: CRM	Method: Iregr
Y(1)=[0.005,0.014]	[-0.015,-0.012]	[-0.016,-0.011]	[-0.019,-0.008]	[-0.043,0.021]
Y(2)=[-0.023,-0.017]	[-0.040,-0.0362]	[-0.042,-0.034]	[-0.043,-0.032]	[-0.081,0.003]
Y(3)=[0,0.009]	[-0.052,-0.046]	[-0.055,-0.044]	[-0.054,-0.044]	[-0.106,0.002]
Y(4)=[-0.043,-0.039]	[-0.033,-0.026]	[-0.035,-0.024]	[-0.034,-0.025]	[-0.068,0.004]
Y(5)=[-0.053,-0.041]	[-0.079,-0.073]	[-0.082,-0.069]	[-0.081,-0.071]	[-0.157,0]
⋮	⋮	⋮	⋮	⋮
Y(49)=[-0.058,-0.046]	[-0.031,-0.028]	[-0.032,-0.027]	[-0.035,-0.024]	[-0.064,0.004]
GOF	[0.020,0.033]	[0.018,0.035]	[0.017,0.036]	[0,0.079]

Table 2: Forecasting asset **ABBOTT**

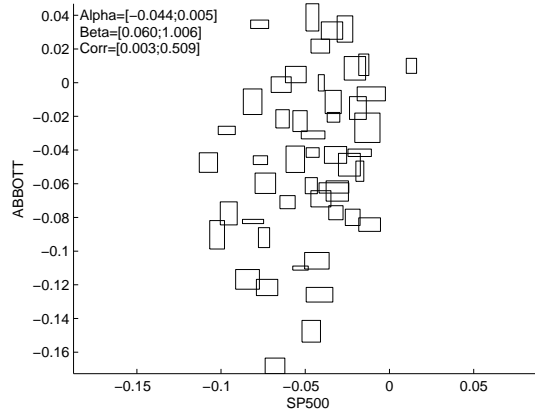
title **ABBOTT**, the other columns refer to their *estimate* by the described interval regression methods. GOF is the Goodness of Fit Index: given two interval vectors: $\mathbf{Y}^I = (Y_i = [\underline{y}_i, \bar{y}_i])$, and $\mathbf{X}^I = (X_i = [\underline{x}_i, \bar{x}_i])$ ($i = 1 \dots n$), the Goodness of Fit Index between \mathbf{X}^I and \mathbf{Y}^I is:

$$GOF = 1/n \sum_{i=1}^n |X_i - Y_i|$$

In Figure 8 the cloud of rectangles generated by the two considered interval time series is represented on a cartesian plane.

4.1.1 Analysis of the results

The methods CM, MinMax, CRM have the computational cost of the product $(\mathbf{A})^{-1}\mathbf{b}$ where \mathbf{A} and \mathbf{b} have been described in Sections 3.1, 3.2, 3.3. Iregr *minimize/maximize* each component of $(\mathbf{A})^{-1}\mathbf{b}$ over a compact and connected set thus it has a computational cost which is quite greater than that of CM, MinMax, CRM, anyway it shows some good advantages with respect to them: 1-Iregr computes the *interval* $\hat{\beta}_j^I$ (**Beta**) and the *interval* $\hat{\alpha}_j^I$ (**Alpha**) of the interval regression line (2.10) which are both well interpretable: $\hat{\beta}_j^I$ and $\hat{\alpha}_j^I$ are the *set* of estimations of all *slopes* and *intercepts* of (2.6), when each return ranges in its own interval of values. CM, MinMax, CRM produce just scalar slope and scalar intercept. 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

Figure 8: **SP500/ABBOTT**

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 real-valued security the CAPM states that the intercept **Alpha** in (2.6) 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 IntervalCAPM 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*. Furthermore, it has been observed that the IntervalCAPM with Iregr 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. 2- the actual return of asset **ABBOTT**, if computed by CM, MinMax, and CRM 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. Furthermore, while CM, MinMax and CRM compute the *interval solution*, i.e. the interval estimation of an interval return, Iregr computes the *interval of solutions* i.e. the interval of the estimations of each point in the chosen interval return. The solutions are always well interpretable. 3- As seen in Section 3.4, Iregr is a method which computes *all possible* regression lines varying each return in its own interval of variation. Thus, the estimated intervals reported in Table 2 (5th column) have a bigger radius with respect to those computed by CM, MinMax and CRM. The predictions computed by Iregr, taking into account also *extreme situations*, may be considered more reliable with respect to those computed by the other meth-

ods. It is important to point out that the quality of our predictions is strictly related to characteristic of the input data; in fact Iregr is quite sensitive to outliers. In the case of simple regression, it is always possible to compute the actual return of asset **ABBOTT** by means of Iregr. On the contrary the other methods may not apply if the matrix involved in the computation of the β is singular.

5 Conclusions

In the present work different regression methodologies for interval data have been described. A first contribution of this work is the comparison of the various regression methods, reported in the literature, on sets of real data. The algorithms of the considered methodologies have been implemented in MATLAB and the numerical results are compared to one another highlighting the following good advantages for the method Iregr:

1) in [3], [4], [16] and [21], the authors derive the results from some classical regression methods which minimize criteria different from the least squares one. Those methods, do not consider the interval as a whole structure or special kind of data, but reconstruct interval solution *ex post*. Iregr, instead, makes 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 interval of values.

2) With difference to the other regression methods that compute *scalar* regression coefficients, Iregr computes the *interval slope* and the *interval intercept* of the interval regression line which are the *set* of estimations of all *slopes* and *intercepts* generated when each data varies in its own interval of variation.

3) Iregr computes the interval of *possible solutions* of the considered problem when each quantity is in its own interval of variation.

Interval regression methods showed to be useful in practical cases, as for example, in the application of CAPM to interval financial data: the interval time series of the asset **ABBOTT** (*ABBOTT Laboratories*), and of the **SP500** (*Standard and Poor's 500 Composite*) index have been considered. The numerical results of the different methodologies are compared one another. The following new findings on the application of Iregr in the framework of the CAPM have been highlighted:

i) Iregr computes the *interval beta* $\hat{\beta}_j^I$ and the *interval alpha* $\hat{\alpha}_j^I$ of the security market line which are both well interpretable: $\hat{\beta}_j^I$ and $\hat{\alpha}_j^I$ are the *set* of estimations of all *slopes* and *intercepts* of (2.6), when each return ranges in its own interval of values. An investor using the IntervalCAPM 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*, but also the “uncertainty” around those quantities. CM, MinMax, CRM produce just scalar slope and scalar intercept.

ii) Considering the *interval predictions*, while CM, MinMax and CRM compute the *interval solution*, i.e. the interval estimation of an interval return, Iregr computes the *interval of solutions* i.e. the interval of the estimations of each point in the chosen interval return. The solutions are always well interpretable.

iii) In the case of simple interval regression, it is always possible to compute the estimate of the actual return of asset **ABBOTT** by means of Iregr. On the contrary the other methods may not apply if the matrix involved in the computation of the β

is singular.

Ireg is a method which is quite sensitive to the characteristic of the input data, in particular it works well when: there are no outliers; the radii of the intervals reflect natural fluctuations excluding exceptional events; the intervals are well separated. Thus, as a future perspective of research, it could be interesting to: i) use some *trimmed method* in order to handle outliers; ii) study the robustness of the method, introducing indexes regarding the ratio centre/radius.

Acknowledgements

I wish to thank Prof. V. Aversa and Prof. C. Lauro for their precious support and fruitful discussions during the development of this research work.

References

- [1] G. Alefeld and J. Herzberger. *Introduction to Interval Computations*. New York: Academic Press, 1996.
- [2] G. Alefeld and G. Mayer. Interval analysis: theory and applications. *Journal of Computational and Applied Mathematics*, 121:421–464, 2000.
- [3] L. Billard and E. Diday. Data analysis, classification and related methods. In H.-H. Bock and E. Diday, editors, *Regression Analysis for Interval-Valued Data*, Studies in Classification, Data Analysis, and Knowledge Organization, pages 103–124. Springer-Verlag, 2000.
- [4] L. Billard and E. Diday. Classification, clustering and data analysis. In K. Jajuga, A. Sokolowski, and H.-H. Bock, editors, *Symbolic Regression Analysis*, Studies in Classification, Data Analysis, and Knowledge Organization, pages 281–288. Springer-Verlag, 2002.
- [5] Y. Campbell, A. W. Lo, and A. C. MacKinlay. *The Econometrics of Financial Markets*. New Jersey: Princeton University Press, 1997.
- [6] C. Chatfield. Calculating interval forecasts. *Journal of Business and Economic Statistics*, 11:121–144, 1993.
- [7] C. Chatfield. Principles of forecasting: A handbook for researchers and practitioners. In J. S. Armstrong, editor, *Prediction Intervals for Time-Series Forecasting*, International Series in Operations Research and Management Science, pages 475–494. Springer US, 2001.
- [8] Gioia F. Portfolio selection model with interval data. *Statistica Applicata*, 21(1):43–75, 2009.
- [9] Gioia F. and Lauro C. Basic statistical methods for interval data. *Statistica Applicata*, 17:75–104, 2005.
- [10] L. He and C. Hu. Impacts of interval measurement on studies of economic variability - evidence from stock market variability forecasting. *Journal of Risk Finance*, 12:489–507, 2007.
- [11] L. He and C. Hu. Impacts of interval computing on stock market variability forecasting. *Journal of Computational Economics*, 33(3):263–276, 2008.

- [12] L. He and C. Hu. Midpoint method and accuracy of variability forecasting. *Journal of Empirical Economics*, 38:705–715, 2010.
- [13] C. Hu and L. He. An application of interval methods to the stock market forecasting. *Reliable Computing*, 13:423–434, 2007.
- [14] Eichberger J. and Harper IR. *Financial Economics*. Oxford University Press, 1997.
- [15] R. B. Kearfott and V. Kreinovich. *Applications of Interval Computations*. Kluwer Academic Publishers, 1996.
- [16] E. Lima Neto and F. de Carvalho. Centre and range method for fitting a linear regression model to symbolic interval data. *Computational Statistics and Data Analysis*, 52:1500–1515, 2007.
- [17] M. Marino and F. Palumbo. Interval arithmetic for the evaluation of imprecise data effects in least squares linear regression. *Statistica Applicata*, 14(3):277–289, 2002.
- [18] H. Markowitz. Portfolio selection. *The Journal of Finance*, VII(1):77–91, 1952.
- [19] R. E. Moore. *Interval Analysis*. Englewood Hills, N.J.: Prentice-Hall, 1996.
- [20] A. Neumaier. *Interval methods for Systems of Equations*. Cambridge University Press, 1990.
- [21] O. Rodriguez. *Classification et Modèles Lineaires en Analyse des Données Symboliques*. PhD thesis, Universite de Paris Dauphine IX, 2000.
- [22] W. F. Sharpe. Capital asset prices: A theory of market equilibrium under conditions of risk. *Journal of Finance*, 19(3):425–442, 1964.