DECISION MAKING IN FINANCIAL MARKETS BY MEANS OF A MULTIVARIATE ORDERING PROCEDURE*

Luca Grilli [1], Massimo Alfonso Russo[1]

[1] Dipartimento di Scienze Economiche, Matematiche e Statistiche
Università degli Studi di Foggia
(e-mail: l.grilli@unifg.it, m.russo@unifg.it)

ABSTRACT: One of the main problems in managing multidimensional data for decision making is that it is not possible to define a complete ordering on multidimensional Euclidean spaces. In order to solve this problem the scientific community has devolped tecniques more and more sofisticated which can be collected under the name of Multivariate Statistics. Recently D'Esposito & Ragozini, 2004 have proposed an ordering procedure in which the "meaningful direction" is the "worst-best" one. The aim of this paper is to extend the approach in D'Esposito & Ragozini, 2004 by considering that, above all in financial applications, data are obtained by considering variables in different scales and, as we will show, this can lead to undesired results. In particular we show that, without an appropriate rescaling, data with a large range of variation are "overweighted" with respect to data with small range of variation.

KEYWORDS: Multivariate Data, Ordering procedures, Normalization.

1 Introduction

A decision maker is asked to perform decisions among many alternatives on the basis of multivariate data. An agent in a financial market face in making her decisions face this problem in fact she has to determine her portfolio of assets and for each asset she has a certain number of different variables to take into account. The problem is that, in general, the reality is multidimensional and it is not always possible to quantify phenomena by means of quantitative unidimensional "indicators".

The main problem in managing multidimensional data for decision making is that it is not possible to define a complete ordering on multidimensional Euclidean spaces. As a result, the scientific community has devolped tecniques more and more sofisticated which can be collected under the name of Multivariate Statistics.

The underling idea is that geometry can help multivariate data analysis; in fact data can be considered as a cloud of points in a multidimensional Euclidean space and the main question is: it is possible the interpretation of data on the clouds of points? In principle it seems difficult to answer since clouds of points come from data tables

*Research financially supported by Local research project of University of Foggia and National Research Project PRIN, co-financed by MIUR, Italian Ministery for Research, 2004-2006 Titled "Nonlinear models in economics and finance: interactions, complexity and forecasting".

and are not geometric objects; the construction is made on the basis of mathematical structures applied on data. The literature on this subject present very interesting results such as: Corrispondence Analysis, Principal Component Analysis, Structured Data Analysis and Inductive Data Analysis (complete reference on the subject in Roux & Rouanet, 2004).

In this paper we consider a different approach to the problem of decision making with multivariate data which answers the following question: It is possible to make a reduction of multidimensionality to unidimensionality? If this is the case the problem of ordering multivariate data is completly solved.

One way to answer this question is finding a "meaningful direction" in the data cloud and considering the projections of data on this "meaningful direction" in order to have unidimensional quantities.

In D'Esposito & Ragozini, 2004 the authors provide an ordering procedure in which the "meaningful direction" is the "worst-best" one. This approach seems to be very interesting since the "worst-best" direction is meaningful in the sense of information contained.

The aim of this paper is to extend this approach by considering a feature that is very important in order to have results that are not depend on the data set. The problem is that, above all in financial applications, data are obtained by considering variables in different scales and, as we will show, this can lead to results in which the range of variation of data influences the ordering procedure. In particular we show that, without an appropriate rescaling, data with a large range of variation are "overweighted" with respect to data with small range of variation.

In particular, in financial applications, the financial assets are rapresented by a certain numbers of financial indicators that are often in different scales. For istance, if we consider the market capitalization (in milion of euros) and EPS (in euros) in order to evaluate an asset, it is clear that a variation on few euros in the first variable has not the same effect of the same variation in the second variable.

2 The Ordering Procedure

The ordering procedure based on the "worst-best" direction for the projection requires the choice of the worst and best performance say the vectors x_{min} and x_{max}. This can be done in two different ways: the first is determine a priore the extreme performance (on the basis of an expert knowledge) and the second is to construct the extreme values with some procedure. The main issue is to determine the direction (a vector with norm equal to one) u on which the different values x_i are projected and so reduced to a single value. The authors define

$$u = \frac{x_{max} - x_{min}}{\| x_{max} - x_{min} \|},$$

that is the "worst-best" direction. In order to consider the point x_{min} as the origin of the axis each vector x_i is transformed in $\tilde{x}_i$ in the following way:

$$\tilde{x}_i = x_i - x_{min}.$$

The projection of the vector $\tilde{x}_i$ onto the vector u is the vector:

$$x_i^* = u \cdot \frac{< \tilde{x}_i, u >}{\| u \|} = u \cdot < \tilde{x}_i, u >,$$

where $< \cdot, \cdot >$ is the scalar product. The real number $< \tilde{x}_i, u >$ is the univariate indicator and is a sort of distance from the worst performance.

In the complete version of this paper we will give a graphical representation of this procedure.

This approach, however, do not consider an important feature that is often present in real data: the problem of different scale in which data are given. In the example given in D'Esposito & Ragozini, 2004 they consider bivariate data set consisting in the unemployment rate and the avarage bank deposit in order to evaluate ranking among territorial units. In this example the range of variation of the two parameters is very different. As a result the parameter with a large range of variation is overweighted with respet to the parameter with low range of variation. This problem leads to an ordering procedure in which the overweighted parameter is the most relevant one. A very simple way to confirm it is to compute the correlation among the single variables and the final score derived from them. In this case, in fact, the correlation is proportional to the range of variation of the variables.

The effect of the range of variation, that is a well-known bad indicator of variability, on the weight of the variable is absolutle not statistically justifiable.

In the next session we propose a possible solution to this problem.

3 The Problem of the Range of Variation

It is easy to show that, in the previous context, parameters with a large range of variation are weighted, a priori, more than parameters with a low range of variation and this influences negatively the interpretation of results. In order to solve this problem we propose a normalizing procedure on the data to be done before applying the ordering procedure defined above.

In our case it is enough to use (following Delvecchio, 1995) the normalizing procedure*:

$$\hat{x}_i = \frac{x_i - x_i^{min}}{x_i^{max} - x_i^{min}}.$$

This normalization has two important consequences: the range of variation of parameters is the same for all data ($[0,1]$) and the direction u has a fixed slope (in a the two dimensional case is exactly the diagonal of the $[0,1] \times [0,1]$ box).

The problem of the range of variation has an important application in the case of financial data. In fact, in order to evaluate different assets, financial agents has to make

*Clearly this is not the only normalizing procedure available, but any other choice do not modify the conclusions on the negative effect of the different range of variation of the variable to synthetize. The main problem is the necessity of homogenize data before being computed.

multidimensional ordering of data with parameters in very different range of variations and results can be strictly influenced by this fact. In Russo & Spada, 2005 it is shown how this problem is often underestimate and it leads to undesired results. They show that the relative weight of the single variables with respect to the aggregation of data is proportional to the range of variation of the variables and this is a very undesiderable result as indicated before.

In the complete version of this paper we will present some relevant (in the sense of effect on the decision procedure) applications to financial markets of such problem in order to avoid undesired effects in the evaluation process for decision makers.

Main References

BARNETT. 1976. The Ordering of Multivariate Data. *Journal of Royal Statistical Society*, **Ser. A, 139**, 318–354.

DELVECCHIO. 1995. *Scale di Misura e Indicatori Sociali*. Cacucci Editore.

D'ESPOSITO, MARIA ROSARIA, & RAGOZINI, GIANCARLO. 2004. Multivariate Ordering in Performance Analysis. *Atti della XLII riunione scientifica della Società Italiana di Statistica, Bari*, **Cleup, Padova**.

ROUX, BRIGITTE LE, & ROUANET, HENRY. 2004. *Geometric Data Analysis*. Dordrecht: Kluwer Academic Publishers.

RUSSO, & SPADA. 2005. Problemi statistici nella determinazione dell'indice di qualit (IGQ) per le cultivar di frumento duro. *Atti del XXXVI Convegno della Società Italiana di Agronomia*.