JOURNAL OF APPLIED QUANTITATIVE METHODS

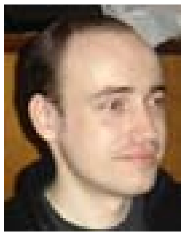
A MATHEMATICAL DETERMINISTIC APPROACH IN MODELLING NATIONAL ECONOMIC EVOLUTION

Marius GIUCLEA[1]
PhD, Assistant Professor
Department of Mathematics, University of Economics, Bucharest, Romania

E-mail: mgiuclea@yahoo.com

Ciprian Costin POPESCU[2]
PhD, University Assistant
Department of Mathematics, University of Economics, Bucharest, Romania

E-mail: cippx@yahoo.com



Abstract: *In this work a regression model based on total least squares approach is presented. An application of the theoretical results in estimating the tendency of some countries economic evolution is given.*

Key words: *Regression line; Total least squares; Economic indicators*

1. Introduction

On the basis of statistical data $(x_{1i}, x_{2i}, ..., x_{ni})$, $i = \overline{1, m}$, n-dimensional orthogonal regression (or total least squares-TLS), is equivalent to solving a problem in the general form:

$$\min_{(c_1, c_2, ..., c_n) \in \mathbf{R}^n} J(c_1, c_2, ..., c_n) = \sum_{i=1}^{m} \left[\text{distance}\left((x_{1i}, x_{2i}, ..., x_{ni}), H(c_1, c_2, ..., c_n; x_{1i}, x_{2i}, ..., x_{ni})\right)\right]^2$$

where H is a regression function (generally, a hyperplane). TLS approach has some interesting properties [3][3] which make it more appropriate than classical regression methods in many cases: for instance the sum of squared distances is independent of the choice of axis system but, the most important fact, this method can be applied when the order cause-effect is not very clear (in other words, any $x_1, x_2, ..., x_n$ can be chosen as dependent variable). In the following sections, consider $n = 3$, that is H is a line or a plane. Graphic generating procedures are available in Matlab for both cases [4]. The economic evolution through the instrumentality of three among most important indicators (GDP (I_1), Inflation (I_2) and Unemployment (I_3)-measured in the form of annual percent changes) is a process which can be very well described using TLS for $n = 3$ (remark that there isn't a direct causal relation between these indicators). In more details, at the given moments in time, I_1, I_2, I_3 receive (after measurements) some numerical values. These are the initial set of data and are considered a starting point by applying the TLS method a global image of economic evolution in the time intervals which are studied. Moreover, a mathematical relation between I_1, I_2, I_3 can be obtained, which allow us to forecast one of them if the others are easily predictable. In Section 3 this reasoning (mention that it is used to measure V4 group economic differences [3], too) is applied to 3 economics entities (Romania, Eurozone and The United States of America) on a period of eight years (2000-2007) and the results are figured as a regression line. The information regarding comparison at a single indicator level (L_1) can be obtained by measuring the slopes relative to every positive oriented space direction, and at level of group of two of them (L_2) by the angles between the regression line and the three coordinate planes (as basis plans). Taking into account levels 1 and 2 we receive a complex, combined image which takes us to L_3 (top level) which gives a three dimensional space image of the whole economic process. The article contains two main parts: the theoretical considerations and the application of the method of orthogonal regression in comparative study and graphic evolutive presentation of the selected economies.

2. A model of three-dimensional orthogonal regression

Consider the data $\left(x_1^{(i)}, x_2^{(i)}, y^{(i)}\right)^T \in \mathbf{R}^3, i = \overline{1,m}$. The linear model is given by:

$$y(x_1, x_2) = c_0 + c_1 x_1 + c_2 x_2$$

Ordinary least squares:

$$\min_{c_0, c_1, c_2} J(c_0, c_1, c_2) = \sum_{i=1}^{m} \left(y^{(i)} - c_0 - c_1 x_1^{(i)} - c_2 x_2^{(i)}\right)^2$$

This is equivalent with search the best fitting plane in $\mathbf{R}^3$,

$$\left\{(x_1, x_2, y)^T \in \mathbf{R}^3 \middle| y = c_0 + c_1 x_1 + c_2 x_2\right\}.$$

Consider the centroid

$$\left(\overline{x}_1, \overline{x}_2, \overline{y}\right) = \left(\frac{1}{m}\sum_{i=1}^{m} x_1^{(i)}, \frac{1}{m}\sum_{i=1}^{m} x_2^{(i)}, \frac{1}{m}\sum_{i=1}^{m} y_i\right).$$

One can observe that as

$$y^{(i)} = c_0 + c_1 x_i^{(i)} + c_2 x_2^{(i)}, \forall i = \overline{1,m}$$

then

$$\frac{\sum_{i=1}^{m} y^{(i)}}{m} = c_0 + c_1 \frac{\sum_{i=1}^{m} x_i^{(i)}}{m} + c_2 \frac{\sum_{i=1}^{m} x_2^{(i)}}{m}$$

that is

$$\overline{y} = c_0 + c_1 \overline{x}_1 + c_2 \overline{x}_2 .$$

Therefore the objective function can be written

$$J(c_0, c_1, c_2) = \sum_{i=1}^{m} \left[\left(y^{(i)} - \overline{y} \right) - c_1 \left(x_1^{(i)} - \overline{x}_1 \right) - c_2 \left(x_2^{(i)} - \overline{x}_2 \right) + \overline{y} - c_0 - c_1 \overline{x}_1 - c_2 \overline{x}_2 \right]^2 =$$

$$= \sum_{i=1}^{m} \left[\left(y^{(i)} - \overline{y} \right) - c_1 \left(x_1^{(i)} - \overline{x}_1 \right) - c_2 \left(x_2^{(i)} - \overline{x}_2 \right) \right]^2 +$$

$$+ 2 \sum_{i=1}^{m} \left[\left(y^{(i)} - \overline{y} \right) - c_1 \left(x_1^{(i)} - \overline{x}_1 \right) - c_2 \left(x_2^{(i)} - \overline{x}_2 \right) \right] \left(\overline{y} - c_1 \overline{x}_1 - c_2 \overline{x}_2 \right) + m \left(\overline{y} - c_0 - c_1 \overline{x}_1 - c_2 \overline{x}_2 \right)^2 =$$

$$= \sum_{i=1}^{m} \left[\left(y^{(i)} - \overline{y} \right) - c_0 - c_1 \left(x_1^{(i)} - \overline{x}_1 \right) - c_2 \left(x_2^{(i)} - \overline{x}_2 \right) \right]^2 +$$

$$+ m \left(\overline{y} - c_0 - c_1 \overline{x}_1 - c_2 \overline{x}_2 \right)^2 \geq \sum_{i=1}^{m} \left[\left(y^{(i)} - \overline{y} \right) - c_0 - c_1 \left(x_1^{(i)} - \overline{x}_1 \right) - c_2 \left(x_2^{(i)} - \overline{x}_2 \right) \right]^2$$

with equality if and only if $\overline{y} = c_0 + c_1 \overline{x}_1 + c_2 \overline{x}_2$. As a consequence the centroid belongs to the fitting plane.

Consider

$$\mathbf{x}_k = \left(x_k^{(1)}, x_k^{(2)}, \dots, x_k^{(m)} \right)^T, \mathbf{y} = \left(y^{(1)}, \dots, y^{(m)} \right)^T$$

and

$$A = \left(\mathbf{1} \middle| \mathbf{x}_1 \middle| \mathbf{x}_2 \right)$$

where

$$\mathbf{1} = (1, \dots, 1)^T \in \mathbf{R}^m, \mathbf{x}_1 = \left(x_1^{(1)}, x_1^{(2)}, \dots, x_1^{(m)} \right)^T \in \mathbf{R}^m, \ \mathbf{x}_2 = \left(x_2^{(1)}, x_2^{(2)}, \dots, x_2^{(m)} \right)^T \in \mathbf{R}^m$$

Thus

$$J(c_0, c_1, c_2) = \left\| y - c_0 \mathbf{1} - c_1 x_1 - c_2 x_2 \right\|^2 = \left\| \mathbf{y} - A\mathbf{c} \right\|_2^2$$

and the solution may be derived from the normal equations.

On the other hand, taking into account that

$$J(c_0, c_1, c_2) = \sum_{i=1}^{m} \left[\left(y^{(i)} - \overline{y} \right) - c_0 - c_1 \left(x_1^{(i)} - \overline{x}_1 \right) - c_2 \left(x_2^{(i)} - \overline{x}_2 \right) \right]^2$$

the problem is equivalent with finding $\mathbf{r}$ with $\|\mathbf{r}\|_2 = 1$ which minimize $\|B\mathbf{r}\|_2^2$, where $B = \left(\mathbf{x}_1 - \overline{\mathbf{x}}_1 / \mathbf{x}_2 - \overline{\mathbf{x}}_2 / \mathbf{y} - \overline{\mathbf{y}} \right)$ for $\overline{\mathbf{x}}_s = \left(\overline{x}_s, \dots, \overline{x}_s \right) \in \mathbf{R}^m, s = \overline{1,2}$ and $\overline{\mathbf{y}} = \left(\overline{y}, \dots, \overline{y} \right) \in \mathbf{R}^m$. The solution always exists [2].

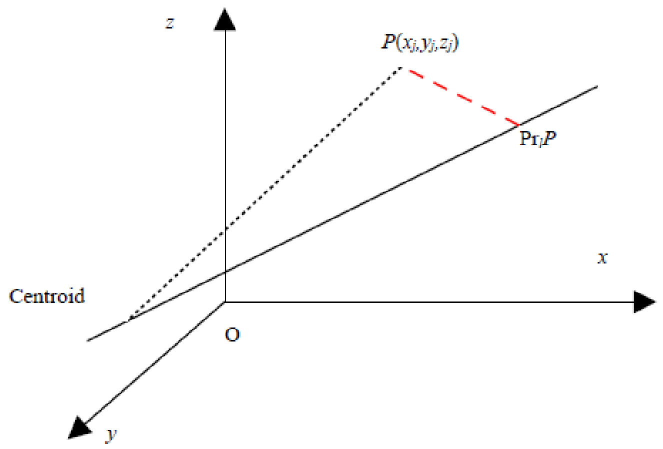


Figure 1.

Next, we study the same problem by substituting the previous plane with a regression line l. The parametric equations of l (with the property $(x_0, y_0, z_0) \in l$) in 3D space are:

$$x = x_0 + ta,\ y = y_0 + tb,\ z = z_0 + tc \text{ where } t \in \mathbf{R}.$$

On the basis that the centroid is on the regression line (as is derived in the Appendix), one may write the parametric equations of the line l in the form:

$$x = \overline{x} + ta,\ y = \overline{y} + tb,\ z = \overline{z} + tc.$$

Moreover, for a point $P(x_j, y_j, z_j)$ the following relation can be derived:

$$d^2(P,l) + d^2(\text{Centroid}, \Pr_l P) = d^2(\text{Centroid}, P) \text{ (Figure 1).}$$

Then

$$(x_j - \overline{x} - ta)^2 + (y_j - \overline{y} - tb)^2 + (z_j - \overline{z} - tc)^2 + t^2(a^2 + b^2 + c^2) = (x_j - \overline{x})^2 + (y_j - \overline{y})^2 + (z_j - \overline{z})^2$$

$$\Rightarrow (x_j - \overline{x})^2 + (y_j - \overline{y})^2 + (z_j - \overline{z})^2 + t^2(a^2 + b^2 + c^2) - 2at(x_j - \overline{x}) - 2bt(y_j - \overline{y}) - 2ct(z_j - \overline{z}) + + t^2(a^2 + b^2 + c^2) = (x_j - \overline{x})^2 + (y_j - \overline{y})^2 + (z_j - \overline{z})^2$$

$$\Rightarrow 2t^2(a^2 + b^2 + c^2) = 2at(x_j - \overline{x}) + 2bt(y_j - \overline{y}) + 2ct(z_j - \overline{z})$$

$$\Rightarrow t = \frac{a(x_j - \overline{x}) + b(y_j - \overline{y}) + c(z_j - \overline{z})}{a^2 + b^2 + c^2} \neq 0$$

(for points which are not projected in the centroid; for the others, the distance to the line is constant and has no influence in the minimization of the objective function).

Thus the squared distance from $P(x_j, y_j, z_j)$ to line l is given by:

$$\left[x_j-\bar{x}-a\frac{a(x_j-\bar{x})+b(y_j-\bar{y})+c(z_j-\bar{z})}{a^2+b^2+c^2}\right]^2$$
$$+\left[y_j-\bar{y}-b\frac{a(x_j-\bar{x})+b(y_j-\bar{y})+c(z_j-\bar{z})}{a^2+b^2+c^2}\right]^2+$$
$$+\left[z_j-\bar{z}-c\frac{a(x_j-\bar{x})+b(y_j-\bar{y})+c(z_j-\bar{z})}{a^2+b^2+c^2}\right]^2.$$

The sum which must be minimized becomes

$$\sum_{j=1}^{m}\left[x_j-\bar{x}-a\frac{a(x_j-\bar{x})+b(y_j-\bar{y})+c(z_j-\bar{z})}{a^2+b^2+c^2}\right]^2+$$
$$+\sum_{j=1}^{m}\left[y_j-\bar{y}-b\frac{a(x_j-\bar{x})+b(y_j-\bar{y})+c(z_j-\bar{z})}{a^2+b^2+c^2}\right]^2+$$
$$+\sum_{j=1}^{m}\left[z_j-\bar{z}-c\frac{a(x_j-\bar{x})+b(y_j-\bar{y})+c(z_j-\bar{z})}{a^2+b^2+c^2}\right]^2$$

which can be solved by mathematical computer programs.

Another method is to use spherical coordinates (Figure 2). As before, $C\in l$.

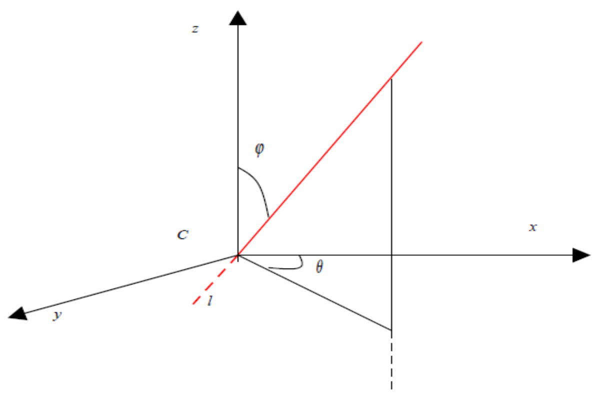


Figure 2.

First, a translation is made in such a manner that the origin of the rectangular coordinates becomes the centroid C. Consequently, a point $P(x,y,z)$ has the new coordinates $(x-\bar{x}, y-\bar{y}, z-\bar{z})$.

Let θ be the angle between the line l and the positive Cz -axis and φ the angle between the plane through l and orthogonal to xCy and the plane xCz. A rotation around Cz such that l comes in xCz is performed. The new position for the generic point is $P(x_1,y_1,z_1)$ where

$$x_1 = \left(x - \bar{x}\right)\cos\theta + \left(y - \bar{y}\right)\sin\theta,\ y_1 = -\left(x - \bar{x}\right)\sin\theta + \left(y - \bar{y}\right)\cos\theta,\ z_1 = z - \bar{z}$$

Now, make a rotation around Cy such that $l \equiv Cz$. It implies that P has the coordinates $\left(x_2, y_2, z_2\right)$ with

$$x_2 = -z_1\sin\varphi + x_1\cos\varphi,\ y_2 = y_1,\ z_2 = z_1\cos\varphi + x_1\sin\varphi$$

Thus

$$x_2 = -\left(z - \bar{z}\right)\sin\varphi + \left(x - \bar{x}\right)\cos\theta\cos\varphi + \left(y - \bar{y}\right)\sin\theta\cos\varphi,$$

$$y_2 = -\left(x - \bar{x}\right)\sin\theta + \left(y - \bar{y}\right)\cos\theta,$$

$$z_2 = \left(z - \bar{z}\right)\cos\varphi + \left(x - \bar{x}\right)\cos\theta\sin\varphi + \left(y - \bar{y}\right)\sin\theta\sin\varphi$$

The squared distance from $P\left(x_2, y_2, z_2\right)$ to l is $x_2^2 + y_2^2$. We have:

$$\frac{\partial x_2^2}{\partial\theta} = 2x_2\frac{\partial x_2}{\partial\theta} = 2\left[-\left(z - \bar{z}\right)\sin\varphi + \left(x - \bar{x}\right)\cos\theta\cos\varphi + \left(y - \bar{y}\right)\sin\theta\cos\varphi\right]\cdot$$
$$\cdot\left[-\left(x - \bar{x}\right)\sin\theta\cos\varphi + \left(y - \bar{y}\right)\cos\theta\cos\varphi\right],$$

$$\frac{\partial x_2^2}{\partial\varphi} = 2x_2\frac{\partial x_2}{\partial\varphi} = 2\left[-\left(z - \bar{z}\right)\sin\varphi + \left(x - \bar{x}\right)\cos\theta\cos\varphi + \left(y - \bar{y}\right)\sin\theta\cos\varphi\right]\cdot$$
$$\cdot\left[-\left(z - \bar{z}\right)\cos\varphi - \left(x - \bar{x}\right)\cos\theta\sin\varphi - \left(y - \bar{y}\right)\sin\theta\sin\varphi\right],$$

$$\frac{\partial y_2^2}{\partial\theta} = 2y_2\frac{\partial y_2}{\partial\theta} = 2\left[-\left(x - \bar{x}\right)\sin\theta + \left(y - \bar{y}\right)\cos\theta\right]\cdot\left[-\left(x - \bar{x}\right)\cos\theta - \left(y - \bar{y}\right)\sin\theta\right]$$

The sum of squared distances from the points $\left(x_i, y_i, z_i\right)_{i=\overline{1,m}}$ to the line l is given by $J\left(\theta,\varphi\right) = \sum_{i=1}^{m}\left[\left(x_i\right)_2^2 + \left(y_i\right)_2^2\right]$. After equal to zero the partial derivatives with respect to θ and φ obtain:

$$\sum_{i=1}^{m}\left[\left(-\left(z_i - \bar{z}\right)\sin\varphi + \left(x_i - \bar{x}\right)\cos\theta\cos\varphi + \left(y_i - \bar{y}\right)\sin\theta\cos\varphi\right)\cdot\right.$$
$$\cdot\left(-\left(x_i - \bar{x}\right)\sin\theta\cos\varphi + \left(y_i - \bar{y}\right)\cos\theta\cos\varphi\right) + \left(-\left(x_i - \bar{x}\right)\sin\theta + \left(y_i - \bar{y}\right)\cos\theta\right)\cdot$$
$$\left.\cdot\left(-\left(x_i - \bar{x}\right)\cos\theta - \left(y_i - \bar{y}\right)\sin\theta\right)\right] = 0$$

and

$$\sum_{i=1}^{m}\left[\left(-\left(z_i - \bar{z}\right)\sin\varphi + \left(x_i - \bar{x}\right)\cos\theta\cos\varphi + \left(y_i - \bar{y}\right)\sin\theta\cos\varphi\right)\cdot\right.$$
$$\left.\cdot\left(-\left(z_i - \bar{z}\right)\cos\varphi - \left(x_i - \bar{x}\right)\cos\theta\sin\varphi - \left(y_i - \bar{y}\right)\sin\theta\sin\varphi\right)\right] = 0$$

Taking into account the fundamental trigonometric formula for both angles and with computing technique of SWP for instance, the solution can be obtained. In addition, there are Matlab routines for the representation of both types of regression in 3D space: plane and line [3,4].

3. Case study: Eurozone countries, Romania and the United States of America

In this section a comparative study is developed for the period 2000-2007 with data regarding *GDP*, *inflation* and *unemployment* associated to: Eurozone, Romania and the US. The numerical values of the mentioned variables are given in the next tables.

Table 1. Euro zone countries

Eurozone (2009)	Adopted	Population
Austria	1999	8,316,487
Belgium	1999	10,666,866
Cyprus	2008	766,400
Finland	1999	5,289,128
France	1999	63,392,140
Germany	1999	82,314,906
Greece	2001	11,125,179
Ireland	1999	4,239,848
Italy	1999	59,131,287
Luxembourg	1999	476,200
Malta	2008	404,962
Netherlands	1999	16,471,968
Portugal	1999	10,599,095
Slovakia	2009	5,389,180
Slovenia	2007	2,013,597
Spain	1999	45,116,894

Source: European Central Bank, Map of euro area, March 2009

Table 2. Eurozone

Subject Descriptor	2000	2001	2002	2003	2004	2005	2006	2007
GDP	3.842	1.905	0.912	0.795	2.085	1.608	2.757	2.595
Inflation	2.171	2.368	2.292	2.073	2.149	2.191	2.183	2.148
Unemployment	8.092	7.800	8.200	8.675	8.825	8.575	8.746	7.427

Source: International Monetary Fund, World Economic Outlook Database, October 2008.
Note: measuring units: annual percent changes.

Table 3. Romania

Subject Descriptor	2000	2001	2002	2003	2004	2005	2006	2007
GDP*	2.149	5.745	5.120	5.224	8.455	4.180	7.855	6.042
Inflation*	45.667	34.468	22.537	15.274	11.881	9.025	6.552	4.840
Unemployment**	7.100	6.600	8.400	7.000	8.000	7.200	7.300	6.400

***Source:** International Monetary Fund, World Economic Outlook Database, October 2008.
****Source:** ILO (International Labour Organization) Bureau of Statistics, LABORSTA-database of labour statistics, March 2009.

Table 4. United States

Subject Descriptor	2000	2001	2002	2003	2004	2005	2006	2007
GDP	3.660	0.751	1.599	2.510	3.637	2.939	2.779	2.028
Inflation	3.367	2.817	1.596	2.298	2.668	3.375	3.226	2.858
Unemployment	3.967	4.742	5.783	5.992	5.542	5.067	4.608	4.642

Source: International Monetary Fund, World Economic Outlook Database, October 2008.

The graphical representations for variations of the considered variables with respect to time are given in Figure 3.

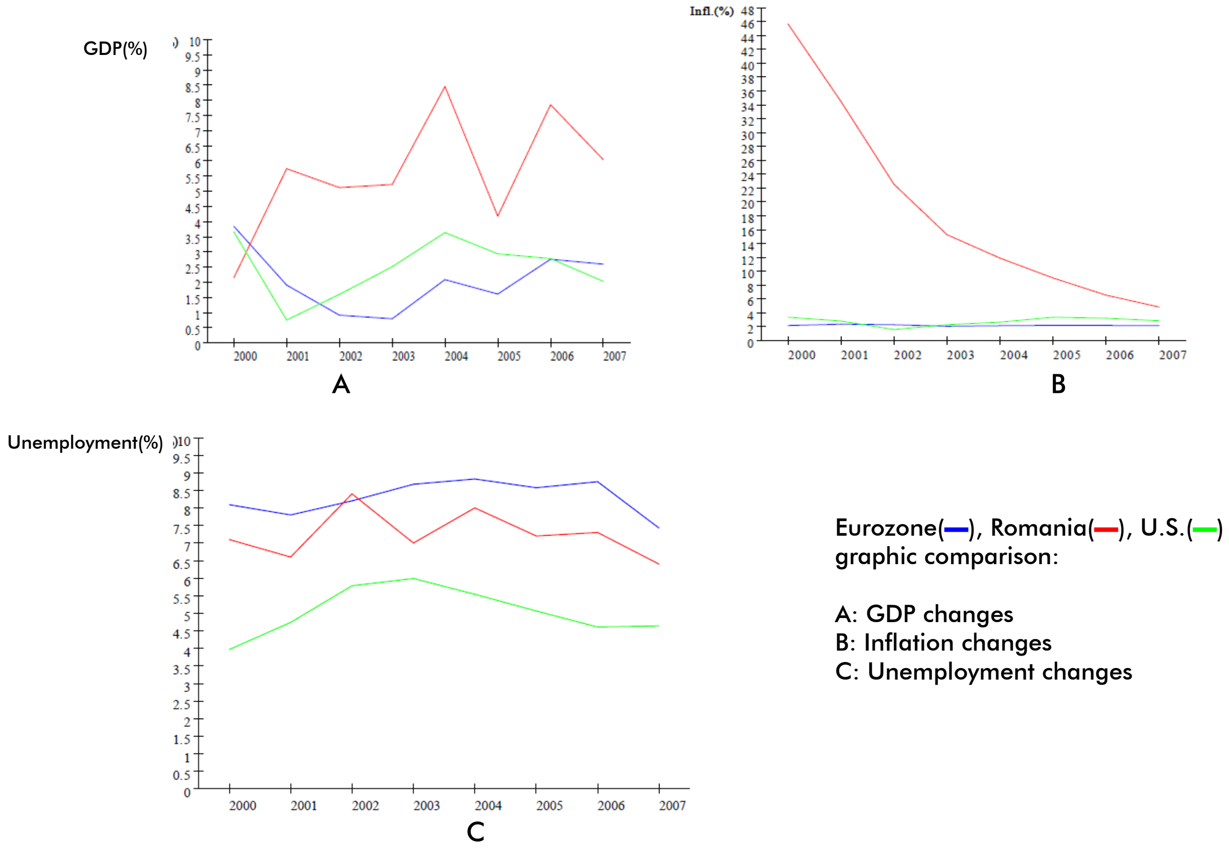


Figure 3

By using the developed method and some MATLAB algorithms [3,4], one can obtain the regression lines from Figure 4 (for Eurozone), Figure 5 (for Romania) and Figure 6 (for USA).

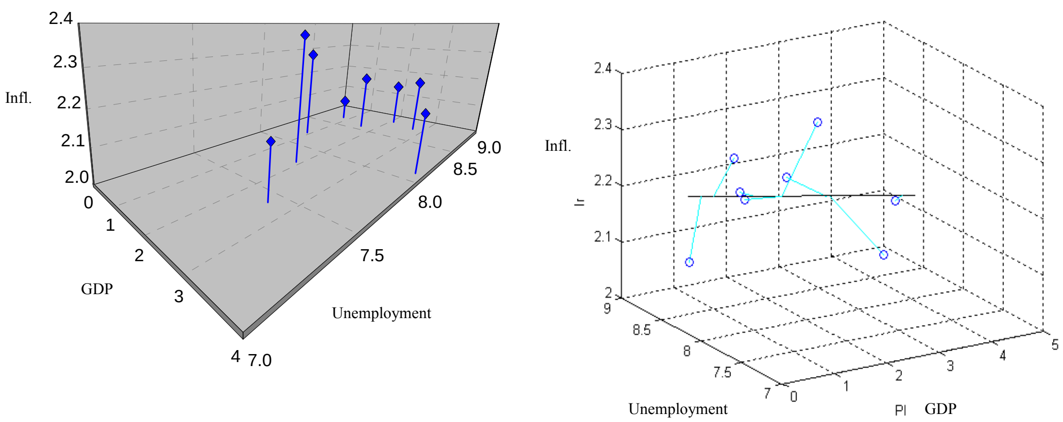


Figure 4. Eurozone – datapoints and regression line

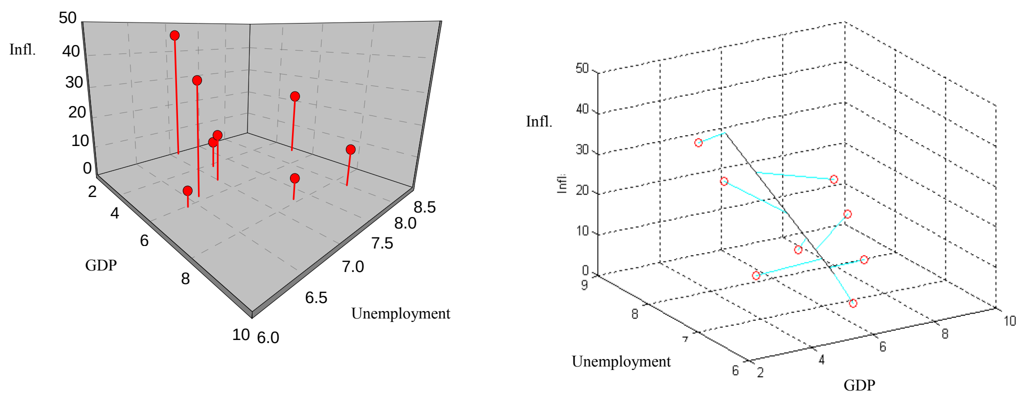


Figure 5. Romania – datapoints and regression line

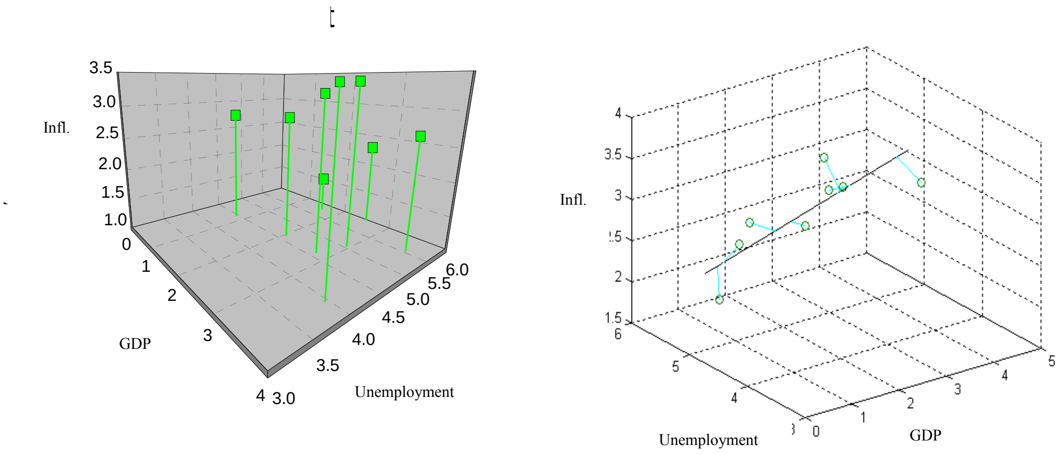


Figure 6. United States–datapoints and regression line

4. Conclusions

A deterministic method for mapping the linear evolution of an economy using three main indicators, each of them regarded as a time function, was discussed. The algorithm was applied to a cross-national comparative study regarding economic changes in a period of several years.

References

1. Arthanary, T. S. and Dodge, Y. **Mathematical programming in statistics**, John Wiley and Sons, New York, 1980
2. de Groen, P. **An introduction to total least squares**, Nieuw Archief voor Wiskunde, Vierde serie 14, 1996, pp. 237-253
3. Petras, I. and Podlubny, I. **State space description of national economies: The V4 countries**, Computational Statistics & Data Analysis 52, 2007, pp. 1223-1233
4. Petras, I. and Podlubny, I. **Orthogonal linear regression in 3D-space by using principal components analysis. A function for MATLAB,** MATLAB Central File Exchange, 2006

Appendix

In this section the following notations are used:

-The dot product (or the inner product) between three-dimensional vectors $\vec{a} = (a_1, a_2, a_3)$ and $\vec{b} = (b_1, b_2, b_3)$ is given by

$$\vec{a} \cdot \vec{b} = a_1 b_1 + a_2 b_2 + a_3 b_3$$

-The cross product is defined by

$$\vec{a} \times \vec{b} = (a_2 b_3 - a_3 b_2, a_3 b_1 - a_1 b_3, a_1 b_2 - a_2 b_1)$$

Let $A(x_0, y_0, z_0)$ be a point belonging to line l, $\vec{r}_0 = (x_0, y_0, z_0) = x_0 \vec{i} + y_0 \vec{j} + z_0 \vec{k}$ (position vector) and $\vec{v} = (a, b, c)$ (direction vector). Consider an arbitrary point $B(x, y, z) \in l$ and $\vec{r} = (x, y, z)$. Then there is a real number t such that $\overrightarrow{AB} = t\vec{v}$. Therefore one can write $\vec{r} = \vec{r}_0 + \overrightarrow{AB} = \vec{r}_0 + t\vec{v}$ (Fig. A). So the vector equation of the line is given by $\vec{r} = \vec{r}_0 + t\vec{v}$, $t \in \mathbf{R}$ (parameter). The vector equation is equivalent to $(x, y, z) = (x_0, y_0, z_0) + (ta, tb, tc)$, from which the following parametric equations are derived: $x = x_0 + at, y = y_0 + bt, z = z_0 + ct$, with $t \in \mathbf{R}$, or (if $a, b, c \neq 0$): $\dfrac{x - x_0}{a} = \dfrac{y - y_0}{b} = \dfrac{z - z_0}{c}$.

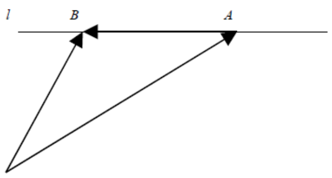


Figure A

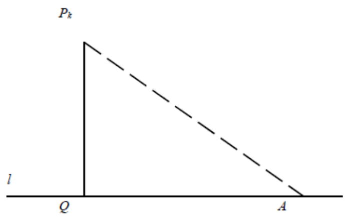


Figure B

Consider $P_k(x_k, y_k, z_k)$ a point in the three-dimensional space and the line l (defined as above). If θ is the angle between $\overrightarrow{AP_k}$ and $\vec{v}$ ($\theta = Q\hat{A}P_k$, Fig. B) it results the following relation between the cross product and $\sin\theta$:

$$\left\|\overrightarrow{AP_k} \times \vec{v}\right\| = \left\|\overrightarrow{AP_k}\right\| \cdot \left\|\vec{v}\right\| \cdot \sin\theta .$$

Thus, the formula for the distance from the point P_k to the line l is given by:

$$P_k Q = \left\|\overrightarrow{P_k Q}\right\| = \left\|\overrightarrow{AP_k}\right\| \sin\theta = \left\|\overrightarrow{AP_k}\right\| \cdot \frac{\left\|\overrightarrow{AP_k} \times \vec{v}\right\|}{\left\|\overrightarrow{AP_k}\right\| \cdot \left\|\vec{v}\right\|} = \frac{\left\|\overrightarrow{AP_k} \times \vec{v}\right\|}{\left\|\vec{v}\right\|} .$$

Consequently,

$$P_k Q = \frac{\left\|(x_k - x_0, y_k - y_0, z_k - z_0) \times (a, b, c)\right\|}{\sqrt{a^2 + b^2 + c^2}} =$$

$$= \frac{\left\|(y_k - y_0)c - (z_k - z_0)b, (z_k - z_0)a - (x_k - x_0)c, (x_k - x_0)b - (y_k - y_0)a\right\|}{\sqrt{a^2 + b^2 + c^2}} =$$

$$= \sqrt{\frac{[(y_k - y_0)c - (z_k - z_0)b]^2 + [(z_k - z_0)a - (x_k - x_0)c]^2 + [(x_k - x_0)b - (y_k - y_0)a]^2}{a^2 + b^2 + c^2}}$$

The sum of squared distances from points $P_k(x_k, y_k, z_k), k = \overline{1,m}$ to l is:

$$\frac{1}{a^2+b^2+c^2}\sum_{k=1}^{m}\left\{[(y_k-y_0)c-(z_k-z_0)b]^2+[(z_k-z_0)a-(x_k-x_0)c]^2+[(x_k-x_0)b-(y_k-y_0)a]^2\right\}$$

If one set the first derivatives w.r.t. x_0, y_0, z_0 equal to zero, obtain:

$$\sum_{k=1}^{m}\{c[(z_k-z_0)a-(x_k-x_0)c]-b[(x_k-x_0)b-(y_k-y_0)a]\}=0$$
$$\sum_{k=1}^{m}\{-c[(y_k-y_0)c-(z_k-z_0)b]+a[(x_k-x_0)b-(y_k-y_0)a]\}=0$$
$$\sum_{k=1}^{m}\{b[(y_k-y_0)c-(z_k-z_0)b]-a[(z_k-z_0)a-(x_k-x_0)c]\}=0$$

which lead to

$$c\left[(\overline{z}-z_0)a-(\overline{x}-x_0)c\right]-b\left[(\overline{x}-x_0)b-(\overline{y}-y_0)a\right]=0$$
$$-c\left[(\overline{y}-y_0)c-(\overline{z}-z_0)b\right]+a\left[(\overline{x}-x_0)b-(\overline{y}-y_0)a\right]=0$$
$$b\left[(\overline{y}-y_0)c-(\overline{z}-z_0)b\right]-a\left[(\overline{z}-z_0)a-(\overline{x}-x_0)c\right]=0$$

or

$$-(\overline{x}-x_0)(b^2+c^2)+(\overline{y}-y_0)ab+(\overline{z}-z_0)ac=0$$
$$ab(\overline{x}-x_0)-(\overline{y}-y_0)(a^2+c^2)+(\overline{z}-z_0)bc=0$$
$$(\overline{x}-x_0)ac+(\overline{y}-y_0)bc-(\overline{z}-z_0)(a^2+b^2)=0$$

Consider the determinant:

$$\Delta=\begin{vmatrix}-b^2-c^2 & ab & ac\\ ab & -a^2-c^2 & bc\\ ac & bc & -a^2-b^2\end{vmatrix}$$

As (Column 1)= $\frac{-b}{a}\cdot$(Column 2)+ $\frac{-c}{a}\cdot$(Column 3) it result that $\Delta=0$.

Consider

$$\Delta=\begin{vmatrix}-b^2-c^2 & ab\\ ab & -a^2-c^2\end{vmatrix}=c^2(a^2+b^2+c^2)$$

$$\Delta_1=\begin{vmatrix}-ac(\overline{z}-z_0) & ab\\ -bc(\overline{z}-z_0) & -a^2-c^2\end{vmatrix}=ac(a^2+b^2+c^2)(\overline{z}-z_0)$$

$$\Delta_2=\begin{vmatrix}-b^2-c^2 & -ac(\overline{z}-z_0)\\ ab & -bc(\overline{z}-z_0)\end{vmatrix}=bc(a^2+b^2+c^2)(\overline{z}-z_0)$$

Thus $\overline{x}-x_0=\frac{\Delta_1}{\Delta}=\frac{a}{c}(\overline{z}-z_0)$ and $\overline{y}-y_0=\frac{\Delta_2}{\Delta}=\frac{b}{c}(\overline{z}-z_0)$. Hence $\frac{\overline{x}-x_0}{a}=\frac{\overline{y}-y_0}{b}=\frac{\overline{z}-z_0}{c}$ and it's obvious that the centroid belonging to the fitting line.

---

[1]**Dr. Marius Giuclea** graduated in Mathematics from University of Bucharest, Romania in 1994 and obtained his MSc degree in Mathematics in 1995.
Between 1994 and 2001 he worked as scientific researcher at National Institute for Research and Development in Microtechnologies and from 2001 as lecturer at Academy of Economic Studies.

In 2004 he was awarded a Ph.D. in applications of intelligent techniques in dynamic systems control by Institute of Mathematics, Bucharest.
His research interests include intelligent techniques and their applications in modelling and control of dynamic systems. He is author of 3 books and about 35 publications in scientific journals and international conference proceedings.

[2] **Dr. Ciprian Popescu** is author of 2 books and about 21 research papers in scientific journals, international and national conferences.

[3]Codification of references:

[1]	Arthanary, T. S. and Dodge, Y. **Mathematical programming in statistics**, John Wiley and Sons, New York, 1980
[2]	de Groen, P. **An introduction to total least squares**, Nieuw Archief voor Wiskunde, Vierde serie 14, 1996, pp. 237-253
[3]	Petras, I. and Podlubny, I. **State space description of national economies: The V4 countries**, Computational Statistics & Data Analysis 52, 2007, pp. 1223-1233
[4]	Petras, I. and Podlubny, I. **Orthogonal linear regression in 3D-space by using principal components analysis. A function for MATLAB,** MATLAB Central File Exchange, 2006