PREMISES FOR USING PRINCIPAL COMPONENT ANALYSIS IN STUDYING THE EFFICIENCY OF RENEWABLE ENERGY INVESTMENTS

Claudiu CICEA[1]
Corina MARINESCU[2]

ABSTRACT
This paper brings into attention a new method for analyzing renewable energy investments efficiency, a procedure called Principal Component Analysis. Authors aim to provide an adequate background for this method, before applying it to a set of data concerning renewable energy. The paper begins by describing the method, highlighting aspects that differentiate it from other method previously used to study the same issue. A brief literature review is included, as it was important to substantiate the role of Principal Component Analysis as a dimension-reduction tool and support to new indices development. The Principal Component Analysis terminology is shortly discussed and the first step for applying the method is undertaken: the study of correlation among variables. This assumption is validated by calculating Pearson's correlation coefficient and representing the correlation matrix. Different types of correlation are obtained according to the presented variation intervals. So, the Principal Component Analysis appliance is once more justified for developing efficiency indices.

KEYWORDS: *efficiency, investments, Pearson correlation coefficient, renewable energy, principal component analysis.*

JEL CLASSIFICATION: *C49, O13, Q29, Q43.*

1. INTRODUCTION

Researchers find the field of renewable energy appropriate for investigating issues related to: renewable energy sources (RES) exploitation, technology, electricity from RES, renewable energy production and generated emissions, renewable energy projects and so on. The development in the energy sector requires not only complex research but also financing. Significant investments are needed especially for RE projects, so concerns about the efficiency of these investments arise. If we look locally, the RE investments efficiency is assessed from an economic point of view, as every investor is interested in high returns on investments. Globally, things are different, as from a national economy view, not only the economic aspects are important, but also the social and environmental ones. So, the need to assess also the social and environmental efficiency of RE investments appears. One cannot apply the same methodology as for valuing economic efficiency of investments projects, based on cost-benefit analysis. An assessment that can include macroeconomic data is needed, so the idea of using an econometric approach was revealed. In a previous research, we managed to form three econometric models based on multiple linear regression applied to panel data. The relationships that underpin these models will be explained in this paper. They will also represent the base for starting the analysis of efficiency through a new method.
Taking into consideration all aspects regarding the use of panel data, the developed models offered

[1] The Bucharest University of Economic Studies, Romania, claudiu.cicea@man.ase.ro
[2] The Bucharest University of Economic Studies, Romania, corina_pirlogea@yahoo.com

important results and helped constructing indices of RE investments efficiency at macroeconomic level. Now, our purpose is to study the same phenomenon using a different approach, the one of Principal Component Analysis (PCA). In order to conduct the new analysis by using PCA, we need to establish the fundamentals of the method and to present the principal frame for its application.

2. PREVIOUS RESEARCHES ON PRINCIPAL COMPONENT ANALYSIS

As Krishnan (2013) explains it in his work, the appearance of PCA method is disputed between Pearson and Hotelling, but one fact is sure, that many researchers are using it when results from analytic procedures (for instance linear regression) become less reliable.

In this section of the paper we only aim to highlight the variety of fields suitable for the PCA appliance. PCA deals with variables describing different phenomenon or helping to construct new measures or indexes. For instance, Cahill and Sanchez (2001) give their contribution to the field of development measurement. They call their development index: The economic and social development index (ESDI) and use PCA technique for its quantification. The ESDI contains information gathered from 36 variables describing health and longevity, knowledge and communication and access to goods, as the authors explain. After obtaining ESDI values, they also realize a ranking of countries included in the analysis and compare it with rankings like the ones of HDI and Real GDP per capita.

Other trials of calculating composite indices are subject to fields like environmental science, air pollution (McNabola, Broderick & Gill, 2009; Chavent, Guegan, Kuentz, Patouille, & Saracco, 2007), finance, archeological science (Moropoulou & Polikreti, 2009), climate change, production, socio and economic development, food chemistry (Montero-Prado, Bentayeb & Nerín, 2013; Cheng, Qin, Guo, Hu & Wu, 2013). Some of these indices are:

- *River Water Quality Index* (WQI) (Ali, Ibrahim, Mengersen, Shitan & Juahir, 2013); it is calculated in Malaysia and before applying PCA method, it was calculated using six water quality variables.
- *Air quality index* (AQI); authors find this index “useful to assess the effects of air pollutants on human health in urban areas” (Kumar & Goyal, 2013), as air pollution represents a major problem in Delhi.
- *Stock market index* (Affleck & Troskie, 2001); the authors wanted to construct an index using ten securities quoted on a Stock Exchange (five securities from the coal sector and other five from the gold sector).

Other studies focus on water quality and use PCA for application examples in different countries. For instance, Gallo and Buciantti (2013) conduct an analysis for river water in Italy, using PCA for investigating the environmental and ecological characteristics of the river basin. Udayakumar, Abhilash & Ouseph (2009) investigate a related issue for a region in India. Their research results indicate that nitrate-nitrogen brought by the rivers in the region of Mangalore, represents a source of pollution to the ecosystem. In Turkey, another investigation reveals the importance of PCA in water quality evaluation; Mazlum, Ozer & Mazlum (1999) studied factors that caused variations in water quality of the Sakarya River.

3. PRINCIPAL COMPONENT ANALYSIS TECHNIQUE

Knowledge discovery often requires complex data analysis. Observations described by more than three variables are hard to be placed in a multi-dimensional space in order to be visualized, then analyzed. So, a method of reducing the number of variables without loosing information on observations is needed. This method appears to be the Principal Component Analysis and its first use is the one mentioned also by Fernandez (2003): “summarizing multivariate attributes by, two or three that can be displayed graphically with minimal loss of information”. These new variables are the so-called principal components.

As shown by other authors, PCA has also a contribution to identifying patterns in data (Smith, 2002). This tool practically finds the differences and also similarities in data of high dimensions.

Another important aspect of using PCA consists of overcoming the correlation problem of explanatory variables in regression analysis (Roberts & Martin, 2006).

The main step for finding the principal components of data consists of calculating the eigenvectors and eigenvalues of the data covariance matrix. In other words, by calculation, a system of axis in which the co-variance matrix is diagonal, should be obtained.

By eigenvalues we understand the "the amount of the variation explained by each PC" as Fernandez (2003) shows in his work. The same author synthesizes that eigenvectors refer to "the weights to compute the uncorrelated PC".

An eigenvector that contains the largest eigenvalue, represents the direction of the highest variation and indicates the first principal component. The direction of the second highest variation is captured by another eigenvector that contains high eigenvalues, but smaller than the first ones and so on. The second eigenvector indicates the second principal component.

Practically PCA identifies an orthogonal system of coordinates axes for utilized observations. Richardson (2009) affirms that "it is equivalent to obtaining the (least-squares) line of best fit through the plotted data". The first principal component of data represents a new axis, or the best direction of the largest variation in the data, as we already explained above. Atchley (2007) explains that "the principal components are linear combinations of the original variables weighted by their contribution to explaining the variance in a particular orthogonal dimension".

Once with PCA technique appliance, some important results are obtained. The most important ones are the contributions of each PC to total variance. For instance, Krishnan (2013) focused on creating a socioeconomic index by utilizing the PCA method. He managed to extract five PC from 26 variables characterizing socioeconomic issues, like: family and household, income, education, occupation, housing, and ethnicity; all together, these five factors explained 56% of the total variation. the proportion of total variation explained by the first PC was 16.3%, the proportion of total variation explained by the second PC was 14.7%, followed by proportions of total variation explained by the third PC: 9.2%, the fourth PC: 8.9%, and the fifth PC: 6.7%.

4. PCA IN THE CONTEXT OF RENEWABLE ENERGY INVESTMENTS EFFICIENCY

Pahor (2011) explains in two suggestive examples (one for measuring development and one for measuring quality) that in most cases, researchers are interested in extracting just one dimension from the original data. This dimension is seen as a new variable or indicator and helps in measuring the studied phenomenon. For instance, if one seeks to measure the development of a region, will have to analyze indicators like GDP, Literacy rate, Life Expectancy at Birth and others. By using PCA, it is possible to obtain a single indicator that could contain in certain proportion information from all studied indicators. The second example refers to several quality indicators of a controller from a factory. Using PCA for all those quality indicators, one can obtain a single index measuring quality.

Taking into consideration these two examples, we consider appropriate the application of PCA for analyzing the efficiency of renewable energy investments at macroeconomic level. The proposed analysis continues previous research (Pîrlogea, Popa & Frăsineanu, 2012) for countries in Europe, some of them members states of European Union. In Figure 1 are included all these countries after the level of investments in renewable energy in 2008. This level of investments was calculated for each country as the ratio between electricity production from renewable sources to total electricity production (Scandurra, 2012). So, the values obtained for investments are percentages here. For instance, for Ireland and Norway we can appreciate a high level of RE investments (stated on Y axes) supported by the high renewable electricity production in total production in these countries. The meaning of X axis is not an economic one; it contains only the countries alphabetically listed.

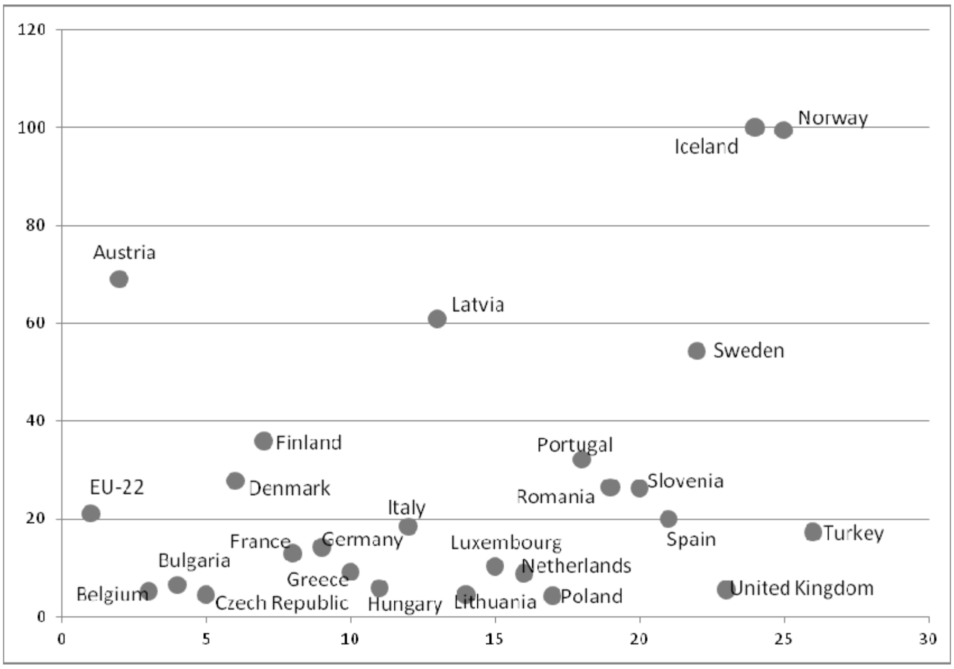

Figure 1. Countries included in the study and their level for renewable energy investments in 2008

Source: authors

The mentioned research proposed three econometric models based on linear regression, with the aim to study the efficiency of renewable energy investments at macroeconomic level. It is well known that analyzing the efficiency of investments from investors' perspective differs from analyzing the efficiency of investments seen at national economy level (Vasilescu et al., 2009). These differences consist in having a precisely methodology for the analysis at microeconomic level (using indicators that measure economic efficiency of investments), in the impossibility to apply the same methodology at macroeconomic level, in the possibility of studying other types of investments efficiency, by using macroeconomic indicators that highlight social and environmental benefits. So, it appears the need to find a new way to analyze the efficiency of investments (not only for investments in renewable energy production) at macroeconomic level. The method we used to study the efficiency of renewable energy investments at macroeconomic level and which results will be published in another work, is based on multiple linear regressions applied on panel data. Each regression highlights a type of investments efficiency that can be studied for each country included in the analysis. We will briefly present the three relationships that exist beyond each econometric model, in order to present the indicators that will be included also in the future principal component analysis. They are all macroeconomic indicators and represent the independent variables within each econometric model.

So, the relationships highlighted by each econometric model are:

$$\frac{C}{I} = \frac{G}{I} \times \frac{C}{E} \times \frac{E}{G} \tag{1}$$

$$\frac{G}{I} = \frac{C}{I} \times \frac{G}{E} \times \frac{E}{C} \tag{2}$$

$$\frac{H}{I} = \frac{G}{I} \times \frac{H}{E} \times \frac{E}{G} \tag{3}$$

Where:

C represents the CO_2 emissions from heat and electricity production;

I represents the investments in renewable energy production;

G represents GDP per capita;

H represents the Human Development Index;

E represents Gross internal energy consumption (1000 tons oil equivalent);

$\frac{C}{I}$ represents the index of environmental efficiency of investments in renewable energy;

$\frac{G}{I}$ represents the index of economic efficiency of investments in renewable energy;

$\frac{C}{E}$ represents the CO_2 intensity;

$\frac{E}{G}$ represents the energy intensity;

$\frac{G}{E}$ represents the economic rate of return on energy consumption;

$\frac{E}{C}$ represents a ratio between Gross internal energy consumption and CO_2 emissions from heat and electricity production;

$\frac{H}{I}$ represents the index of social efficiency of investments in renewable energy;

$\frac{H}{E}$ represents a ratio between Human Development Index and Gross internal energy consumption.

For estimating the three linear models corresponding to the above mentioned relationships, we used as a method of estimation Least Squares, so we had to validate hypothesis belonging to Gauss Markov model. These hypotheses refer to the model linearity, to independent variables and to errors. One of the hypotheses regarding independent variables consists in validating the absence of multicollinearity. Multicollinearity refers to multiple linear relationships between variables.

Multicollinearity cannot be ignored, but neither can be completely removed, most times being intrinsic (Yahoo Answers, 2007). Some authors say that only if is low we can ignore it, which could artificially stimulate the measurement of what is called “goodness of fit” (for instance those tests or factors that show the discrepancy between observed values and values expected under the model in question). Theoretically, the variables for which multicollinearity appears, should be changed with new ones, but the chances of not getting back multicollinearity are weak; therefore we adopted another strategy for “protecting” the results of estimates. Considering that one of the advantages of using panel data models is reducing multicollinearity (Meck, Rongping & Fanchen, 2008), we estimated the models’ coefficients and watched the standard error values that are associated to this coefficients. As long as they remain low, multicolinearity is also reduced.

We considered appropriate to give this explanation, because we intend to study the correlation of variables in each presented relationship. It does exist, we managed to work with it, for the reason mentioned above, without distorting results in the econometric panel data approach. Now, in the new approach of PCA, we will focus on this correlation, as it is the main concern of the method.

Practically correlation identifies simultaneous changes reflected by two or more variables. If the values of two variables tend to co-vary, then correlation measures to what extent is this happening.

So, the correlation matrix for the first relationship (1) (which contains variables influencing the index of environmental efficiency of investments in renewable energy) is presented in Table 2. All values included in the matrix are calculated by using Pearson correlation coefficient, which is similar to the classical linear correlation coefficient. Each value of the matrix is measuring the degree of linear correlation between two variables. As an interpretation, we followed the rule of Hinkle, Wiersma and Jurs (2003) and we constructed variation intervals (taking into account that Pearson's R can vary between -1 and 1) associated to different type of correlation. The variation intervals are shown in Table 1.

Table 1. Range of values for Pearson's correlation coefficient and their interpretation

No. crt.	Range of values	Interpretation
1	[-1, -0.90)	Very high negative correlation
2	[-0.90, -0.70)	High negative correlation
3	[-0.70, -0.50)	Moderate negative correlation
4	[-0.50, -0.30)	Low negative correlation
5	[-0.30, 0)	Little if any negative correlation
6	[0, 0.30)	Little if any positive correlation
7	[0.30, 0.50)	Low positive (negative) correlation
8	[0.50, 0.70)	Moderate positive correlation
9	[0.70, 0.90)	High positive correlation
10	[0.90, 1]	Very high positive correlation

Source: authors after Hinkle, Wiersma and Jurs (2003)

Table 2. Correlation matrix (Pearson) for the first relationship regarding the independent variables influencing the index of environmental efficiency of investments in renewable energy

Variables	*EG*	*CE*	*GI*
EG	1	0.473	-0.083
CE	0.473	1	-0.098
GI	-0.083	-0.098	1

Source: authors' calculation in XLSTAT 2012

EG represents a notation for energy intensity, *CE* represents a notation for CO_2 intensity, *GI* represents a notation for the index of economic efficiency of investments in renewable energy (calculated as a ratio between GDP per capita and investments in renewable energy). *EG*, *CE* and *GI* are independent variables from equation (1) presented above.

The significance of the values included in the matrix is the following:

- There is a low positive correlation between energy intensity and CO_2 intensity;

- There is a little negative correlation between energy intensity and the index of economic efficiency of investments in renewable energy, but also between CO_2 intensity and the index of economic efficiency of investments in renewable energy;
- It is normal to have value 1 between a variable and itself when calculating the correlation coefficient.

The correlation matrix for the second relationship (2) (which contains variables influencing the index of economic efficiency of investments in renewable energy) is presented in Table 3.

Table 3. Correlation matrix (Pearson)for the second relationship regarding the independent variables influencing the index of economic efficiency of investments in renewable energy

Variables	*CI*	*GE*	*EC*
CI	1	-0.262	-0.133
GE	-0.262	1	0.519
EC	-0.133	0.519	1

Source: authors' calculation in XLSTAT 2012

CI represents a notation for the index of environmental efficiency of investments in renewable energy, *GE* represents a notation for the economic rate of return on energy consumption, *EC* represents a notation for the ratio between Gross internal energy consumption and CO_2 emissions from heat and electricity production. *CI*, *GE* and *EC* are independent variables from equation (2) presented above.

The significance of the values included in the matrix is the following:

- The value of -0.262 indicates a little negative correlation between the economic rate of return on energy consumption and the index of environmental efficiency of investments in renewable energy;
- There is a little negative correlation between *EC* and *CI*;
- The value of 0.519 indicates moderate positive correlation between *EC* and *GE.*

The correlation matrix for the third relationship (3) (which contains variables influencing the index of social efficiency of investments in renewable energy) is presented in Table 4.

Table 4. Correlation matrix for the third relationship regarding the independent variables influencing the index of social efficiency of investments in renewable energy

Variables	*GI*	*HE*	*EG*
GI	1	0.038	-0.100
HE	0.038	1	-0.496
EG	-0.100	-0.496	1

Source: authors' calculation in XLSTAT 2012

GI represents a notation for the index of economic efficiency of investments in renewable energy, *HE* represents a notation for ratio between Human Development Index and Gross internal energy consumption, *EG* represents a notation for energy intensity. *GI*, *HE* and *EG* are independent variables from equation (3) presented above.

The significance of the values included in the matrix is the following:

- The value of -0.100 indicates a little negative correlation between the energy intensity and the index of economic efficiency of investments in renewable energy;
- There is a little positive correlation between *HE* and *GI*;
- The value of -0.496 indicates low negative correlation between *EG* and *HE.*

5. CONCLUSIONS

The analysis of investments' efficiency that we intend to develop by using PCA has a great degree of novelty, as no one before attempted to establish it by econometric means. So, it was necessary to highlight all important aspects regarding this technique, in order to apply it and investigate the investments' efficiency. The variables of interest in this analysis represent macroeconomic indicators describing the field of energy with focus on renewable energy. In the same time, associated in groups of three, they are integrated in relationships that focus on a certain type of investments' efficiency: economic, social and environmental. In a previous analysis, the correlation among these variables was avoided by using panel data, as we already explained. Now, the appliance of PCA accounts on the existence of correlation between variables. So, for each mentioned type of efficiency, a correlation matrix was developed. The correlation coefficients were calculated with Pearson's formula, offering information on the strength and direction of the linear associations between pairs of variables. Little, low or moderate correlation, either negative or positive was revealed by using Pearson's correlation coefficient. So, the assumption of PCA (the existence of correlation) is validated and the method can be applied. Even though the utilized variables are correlated, the principal components resulted from PCA are uncorrelated components. Taking all into consideration, by applying PCA on presented variables, three indices of investments' efficiency (social, environmental and economic) will be obtained. This will be subject to a future work.

ACKNOWLEDGMENT

This work was co-financed from the European Social Fund through Sectoral Operational Programme Human Resources Development 2007-2013; project number POSDRU/107/1.5/S/77213 „ Ph.D. for a career in interdisciplinary economic research at the European standards".

REFERENCES

Affleck, J.V., & Troskie, C.G. (2001). A principal component index subject to constraints. Retrieved from http://www.iassa.co.za/articles/014_sep1979_04.pdf

Ali, Z.M, Ibrahim, N.A., Mengersen, K., Shitan, M., & Juahir, H. (2013). The Langat River water quality index based on principal component analysis. AIP Conf. Proc. 1522, 1322 (2013); doi: 10.1063/1.4801283

Atchley, W.R. (2007). Introduction to Principal Components and Factor Analysis. Retrieved from ftp://statgen.ncsu.edu/pub/thorne/molevoclass/AtchleyOct19.pdf

Cahill, M.B., & Sanchez, N. (2001). Using Principal Components to Produce an Economic and Social Development Index: An Application to Latin America and the U.S. *Atlantic Econ. J.* 29(3): 311-329

Chavent, M., Guegan, H., Kuentz, V., Patouille, B., & Saracco, J. (2007). Pollution sources detection via principal component analysis and rotation, Proceedings of the XIIth International Symposium on Applied Stochastic Model and Data Analysis, Chania.

Cheng, H., Qin, Z.H., Guo, X.F., Hu, X.S., & Wu, J.H. (2013). Geographical origin identification of propolis using GC–MS and electronic nose combined with principal component analysis. *Food Research International*, *51(2)*, 813-822

Fernandez, G. (2003). Principal Component Analysis. Retrieved from http://www.cabnr.unr.edu/saito/classes/ers701/pca2.pdf

Gallo, M., & Buccianti, A. (2013). Weighted principal component analysis for compositional data: application example for the water chemistry of the Arno river (Tuscany, central Italy). *Environmetrics*, *24*, 269–277

Hinkle, D. E., Wiersma, W., & Jurs. S. G. (2003). Applied statistics for the behavioral sciences (5th ed.). Boston, MA: Houghton Mifflin Company

Krishnan, V. (2013). Constructing an Area-based Socioeconomic Index: A Principal Components Analysis Approach. Retrieved from http://www.cup.ualberta.ca/wp-content/uploads/2013/04/SEICUPWebsite_10April13.pdf

Kumar, A., & Goyal, P. (2013). Forecasting of Air Quality Index in Delhi Using Neural Network Based on Principal Component Analysis. *Pure and Applied Geophysics 170*(4), 711-722.

Mazlum, N., Ozer, A., & Mazlum, S. (1999). Interpretation of Water Quality Data by Principal Components Analysis. *Tr. J. of Engineering and Environmental Science, 23*, 19-26.

McNabola, A., Broderick, B.M, & Gill, L.W. (2009). A principal components analysis of the factors effecting personal exposure to air pollution in urban commuters in Dublin, Ireland. *Journal of Environmental Science and Health, Part A, 44*(12).

Meck, R., Rongping, M., & Fanchen, M. (2008). Technology and Innovation Management, Oldenbourg Wissenschaftsverlag GmbH, Munchen

Montero-Prado, P., Bentayeb, K., & Nerín, C. (2013). Pattern recognition of peach cultivars (Prunus persica L.) from their volatile components. *Food Chemistry*, *138*(1), 724-731.

Moropoulou, A., & Polikreti, K. (2009). Principal Component Analysis in monument conservation: Three application examples. *Journal of Cultural Heritage* 10, 73-81

Pahor, M. (2011). Seminar 6: PCA and Factor Analysis, retrieved from http://www.generalfiles.org/download/gs5959b3ech32i0/IMB_2011_Seminar6_PCA_FA.pdf.html

Pirlogea, C., Popa, I., & Frăsineanu, C. (2012). Macroeconomic Indicators Used to Study the Efficiency of Investments in Renewable Energy Field. Economia.Seria Management 15(2), 308-315.

Richardson, M. (2009). Principal Component Analysis, retrieved from http://people.maths.ox.ac.uk/richardsonm/SignalProcPCA.pdf

Roberts, S., & Martin, M.A. (2006). Using Supervised Principal Components Analysis to Assess Multiple Pollutant Effects. *Environmental Health Perspective 114*(12), 1877-1882.

Scandurra, G. (2012). Investments in Renewable Energies: evidence from a panel of countries. Retrieved March 12, 2012 from http://meetings.sis-statistica.org/index.php/sm/sm2012/paper/viewFile/2030/34

Smith, L.I. (2002). A tutorial on Principal Components Analysis, retrieved from http://www.cs.otago.ac.nz/cosc453/student_tutorials/principal_components.pdf

Udayakumar, P., Abhilash P.P., & Ouseph P.P. (2009). Assessment of water quality using principal component analysis--a case study of the Mangalore coastal region, India. Journal of environmental science & engineering 51(3), 179-186.

Vasilescu, I., Cicea, C., Dobrea, D., Busu, C., & Alexandru, Gh. (2009). Managementul investitiilor. Eficon Press, Bucharest

Yahoo Answers. (2007). Correlation between explanatory variables in a pooled regression ?, retrieved from http://answers.yahoo.com/question/index?qid=20070612090328AAcNmUo