



Munich Personal RePEc Archive

Corruption and Socioeconomics Determinants: Empirical Evidence of Twenty Nine Countries

George Halkos and Nickolaos Tzeremes

Department of Economics, University of Thessaly, Greece

3. January 2007

Online at <http://mpa.ub.uni-muenchen.de/2874/>
MPRA Paper No. 2874, posted 23. April 2007

Corruption and socioeconomics determinants: Empirical evidence of twenty nine countries

by

George Emm. Halkos and Nickolaos G. Tzeremes

*Department of Economics, University of Thessaly
Argonavton and Filellinon st., 38221, Volos, Greece*

Abstract

This paper measures the effect of different socioeconomic determinants on countries' transparency efficiency. Specifically, using Data Envelopment Analysis (DEA), the transparency efficiency of twenty nine countries is calculated. Then with the help of factor analysis we extract two factors from seven socioeconomic variables according to their communality of influence. Finally we set up a logistic regression using the efficiencies derived from DEA and the factors extracted from factor analysis. The results suggest that higher transparency efficiency appears in countries with cultural values of lower power distance, masculinity, uncertainty avoidance and lower individualism. Additionally, lower inflation rates and lower political and economical risks constitute to higher levels of countries' transparency efficiency while positive GDP growth doesn't ensure countries' transparency efficiency.

Keywords: Perceived transparency, business ethics, cultural dimensions, factor analysis, logistic regression, DEA

JEL codes: C10, C14, M20, Z19

1. Introduction

Current studies in economic development and international business suggest that corruption is a major “threat” for business operational efficiency and country’s economic development. Mauro (1995) suggests that corruption is a disincentive to investment, whereas Davids (1999) argues that the biggest threat of corruption lies upon its effect on misallocation of resources that disrupts economic development, the distortion of public policy and the degrading of integrity of the business system. Therefore, corruption forces multinational corporations to be careful in choosing the host countries for their foreign subsidiaries, because they are concerned of their increased operational costs and risks.

Among others Habib and Zurasawicki (2002) claim that foreign investors generally avoid corruption because it can create inefficiencies. Their analysis suggests that the different levels of corruption have a negative impact on foreign direct investment of both host and home country. In addition to this view, Tanzi and Davoodi (1997) emphasize that corruption may act as a tax on foreign direct investment, increasing considerably the operating costs of corporations and lower the public welfare.

Our paper is based on a fundamental assumption. Due to the fact that countries have a knowledge of corruption’s negative effect on economic development and business operation, we assume that every country tries to minimize the effect of corruption in different areas (like on government, tax, business, etc.) and in this way to maximise its transparency. Therefore, by using Data Envelopment Analysis CCR model the “transparency efficiency” of twenty nine countries is measured in terms of minimising corruption and maximizing their transparency.

Moreover, in contrast to other studies, this paper takes into account different cultural and economic variables and analyses their impact on countries' "transparency efficiency" simultaneously. For this reason factor analysis is used in order to group all the socioeconomic variables into two major factors according to their influence. Finally, the paper introduces a logistic regression using as depended variable the "transparency efficiencies" derived from our DEA analysis, (having 0 in case a country's transparency efficiency is below 70% and 1 otherwise) and as independent variables the two socioeconomic factors, as derived from our factor analysis, in order to emphasise the impact of socioeconomic determinants on country's perceived "transparency efficiency".

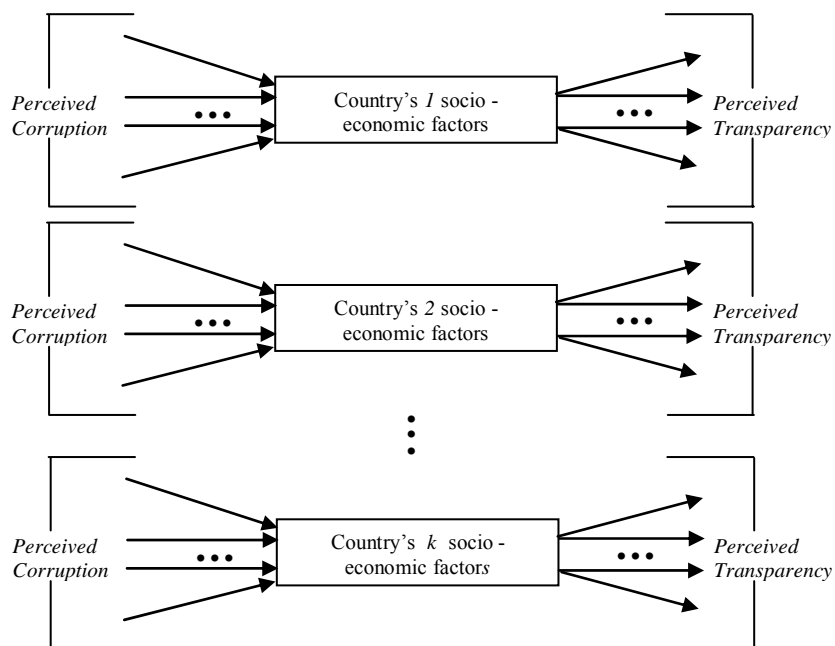
The structure of the paper is as follows. Section 2 describes the literature review, while section 3 analyses the data and methodology used in this research. Finally, section 4 presents and discusses the empirical results and section 5 concludes the paper.

2. Literature Review

Different studies in order to investigated socioeconomic factors related to corruption have used Corruption Index developed by Transparency International (TI), Hofstede's cultural dimensions and other economic variables such as GDP, inflation rates, unemployment, foreign direct investment and other macroeconomic variables (Davis and Ruhe 2003; Getz and Volkema 2001; Habib and Zurawicki 2002; Husted 1999; Robertson and Watson 2004). However, studies investigating the link between corruption and country's socioeconomic factors investigate separately those links in a hypotheses based form. Therefore, there is a "gap" in the literature investigating simultaneously those factors in addition to their effect on country's transparency.

Figure 1 illustrates the concept behind our research. Every country in our analysis is treated as an input-output system. In one hand the inputs of our research are related to the perceived corruption levels in different areas such as political parties, parliament / legislature, legal system / judiciary, tax revenue and business / private sector, whereas the output considered being the perceived transparency index (Transparency International, 2005). However, according to our conceptual input/output model the transparency efficiency of each country is characterised, predetermined and in fact imposed by the different socioeconomic unique characteristics of each country. Getz and Volkema (2001), analysing the socioeconomic factors associated with higher rates of perceived corruption, found that higher perceived corruption is positively associated with higher inflation and lower GDP rates, high masculinity levels in a culture, high power distance levels and higher collectivism and uncertainty levels. Similar results have been found in Davis and Ruhe (2003).

Figure 1: An input/ output conceptual model of transparency efficiency



The majority of the studies associated with the link between the social determinants and corruption have used Hofstede's cultural index. This paper uses the four cultural dimensions as introduced by Hofstede (1994). It also uses Hofstede's (1980a, p.25) definition of culture being 'the collective programming of the mind which distinguishes the members of one human group from another', which is regarded as the main determinant of the social aspect of corruption. In his research, Hofstede distributed more than 88,000 questionnaires to IBM's employees in forty different countries. Then based on a country level factor analysis, he classified the forty countries along four dimensions.

The first dimension is individualism/collectivism, with individualism characteristic defined as a social framework in which people are supposed to take care of themselves and of their immediate families; whereas collectivism is characterised by a social framework in which people distinguish between in-groups and out-groups, they expect their in-groups to look after them and in exchange for that they feel they owe absolute loyalty to it. The second dimension is power distance, defined as the extent to which a culture accepts the fact that power in institutions and organizations is distributed unequally. The third dimension (uncertainty avoidance) is defined as the extent to which a society feels threatened by uncertain and ambiguous situations and tries to avoid these situations by providing greater career stability, establishing more formal rules, not tolerating deviant ideas and behaviours, and believing in absolute truths and the attainment of expertise. Finally, the fourth dimension is masculinity/femininity, with masculinity defined as the extent to which the dominant values in a society are 'masculine'. According to Hofstede masculine values in a society are interrelated with assertiveness, acquisition of money, social isolation and ignorance of

quality of life. In contrast femininity is defined as the opposite of masculinity (Hofstede, 1980b, p.45-46).

3. Methodology

3.1 Data

This paper analyses twenty nine countries in terms of their socioeconomic factors influencing their transparency efficiency. For the calculation of transparency efficiency (see figure 1) the paper uses five inputs and one output. Table 1 illustrates the variables used for the calculation of transparency efficiency.

Table 1: Data description and data sources

<i>Variable Code</i>	<i>Variable Description</i>	<i>Source of variable</i>
POLP	Political Parties	Transparency International (2005)
PARLEG	Parliament / Legislature	Transparency International (2005)
LEGJUS	Legal system / Judiciary	Transparency International (2005)
TAXRE	Tax revenue	Transparency International (2005)
BUSPR	Business / private sector	Transparency International (2005)
TI	Corruption Transparency Index	Transparency International (2005)
PDI	Power Distance	Hofstede (1994)
IDV	Individualism/Collectivism	Hofstede (1994)
MAS	Masculinity/ Femininity	Hofstede (1994)
UAI	Uncertainty Avoidance	Hofstede (1994)
GDP	GDP % change 1995-2005	OECD (2005)
INFLA	Inflation % change 1995-2005	OECD (2005)
COUNTRISK	UNSTAD Country risk index	World Investment Report (2005)

The five inputs used are the levels of perceived corruption for political parties, parliament/ legislature, legal system/ judiciary, tax revenue and business/ private sector, taking values from one to five (1=no corrupted, 5= highly corrupted). The output used for this research is the perceived transparency index, taking prices from one to ten (1=less transparent or highly corrupted, 10=highly transparent or low corrupted). Moreover, for our factor analysis the four cultural indexes as introduced by Hofstede (1994) have been used in order to capture the social factors influencing country's transparency efficiency. Finally, different macroeconomic variables have been used in order to analyse the economic determinants of perceived transparency

efficiency, such as the percentage of GDP change for 1995-2005 and the percentage of inflation change for 1995-2005 (OECD, 2005). Moreover, the country risk index (World Investment Report, 2005) has been used in order to determine its relation with country's transparency efficiency. Country risk index ranges from 0% to 100% (0%=highly risk, 100%= no risk) indicating the investment risk associated with the host country. This index is associated with country's political and socio-economic stability.

3.2 Measuring Transparency Efficiency

DEA is widely acclaimed as a useful technique for measuring efficiency, including production possibilities, which are deemed to be one of the common interests of Operational Research and Management Science (Charnes *et al.*, 1994). It can be roughly defined as a nonparametric method of measuring the efficiency of a Decision Making Unit (DMU) with multiple inputs and/or multiple outputs. This is achieved by constructing a single 'virtual' output to a single 'virtual' input without pre-defining a production function. The terms DEA and the CCR model were first introduced in 1978 (Charnes *et al.*, 1978).

DEA is concerned with the efficiency of the individual unit, which can be defined as the Unit of Assessment (Thanassoulis, 2001) or the Decision Making Unit (*DMU*). DEA is used to measure the relative productivity of a DMU by comparing it with other homogeneous units transforming the same group of measurable positive inputs into the same types of measurable positive outputs. The input and output data as has been analysed above and illustrated in figure 1 can be expressed by matrices X and Y in (1) and (2), where x_{ij} refers to the i^{th} input data of DMU j , whereas y_{ij} is the i^{th} output of DMU j .

$$X = \begin{pmatrix} x_{11} & x_{12} & \cdots x_{1s} \\ x_{21} & x_{22} & \cdots x_{2s} \\ \vdots & & \\ x_{m1} & x_{m2} & \cdots x_{ms} \end{pmatrix} \quad (1)$$

$$Y = \begin{pmatrix} y_{11} & y_{12} & \cdots y_{1s} \\ y_{21} & y_{22} & \cdots y_{2s} \\ \vdots & & \\ y_{n1} & y_{n2} & \cdots y_{ns} \end{pmatrix} \quad (2)$$

This paper in order to measure country's transparency efficiency uses five inputs (POLP, PARLEG, LEGJUS, TAXRE, BUSPR) and one output (TI) by applying the CCR model (Charnes *et al.*, 1978).

The CCR model for the example of Figure 1 can be expressed by (3)-(6):

$$(FP_o)Max \quad \theta = \frac{u_1 y_{1o} + u_2 y_{2o} + \cdots + u_n y_{no}}{v_1 x_{1o} + v_2 x_{2o} + \cdots + v_m x_{mo}} \quad (3)$$

$$\text{Subject to: } \frac{u_1 y_{1j} + u_2 y_{2j} + \cdots + u_n y_{nj}}{v_1 x_{1j} + v_2 x_{2j} + \cdots + v_m x_{mj}} \leq 1 \quad (j = 1, \dots, s) \quad (4)$$

$$v_1, v_2, \dots, v_m \geq 0 \quad (5)$$

$$u_1, u_2, \dots, u_n \geq 0 \quad (6)$$

Given the data X and Y in (1) and (2), the CCR model measures the maximum efficiency of each DMU by solving the fractional programming (FP) problem in (3) where the input weights v_1, v_2, \dots, v_m and output weights u_1, u_2, \dots, u_n are variables to be obtained. o in (3) varies from 1 to s which means s optimisations for all s DMUs. Constraint 4 reveals that the ratio of 'virtual output' ($u_1 y_{1o} + u_2 y_{2o} + \cdots + u_n y_{no}$) to 'virtual input' ($v_1 x_{1o} + v_2 x_{2o} + \cdots + v_m x_{mo}$) cannot exceed 1 for each DMU, which conforms to the economic assumption that in production the output cannot be more than the input.

The above FP (3)-(6) is equivalent to the following linear programming (LP) formulation given in equations (7)-(11) (see e.g. Cooper *et al.*, 2000):

$$(LP_o)Max \quad \theta = u_1 y_{1o} + u_2 y_{2o} + \cdots + u_n y_{no} \quad (7)$$

Subject to:

$$v_1 x_{1o} + v_2 x_{2o} + \cdots + v_m x_{mo} = 1 \quad (8)$$

$$u_1 y_{1j} + u_2 y_{2j} + \cdots + u_n y_{nj} \leq v_1 x_{1j} + v_2 x_{2j} + \cdots + v_m x_{mj} \quad (j = 1, \cdots, s) \quad (9)$$

$$v_1, v_2, \cdots, v_m \geq 0 \quad (10)$$

$$u_1, u_2, \cdots, u_n \geq 0 \quad (11)$$

It is worth mentioning that the computation of the above DEA CCR model by transforming the FP model into the LP model has been of great significance for the rapid development and wide application of DEA. As a long-established mathematical method with various sophisticated computation methods and commercially available solution software, LP possesses inherent advantages that make the complicated computation both easier and more feasible.

3.3 Factor Analysis

After measuring the transparency efficiency using DEA CCR model, the paper uses factor analysis in order to group the socioeconomic variables (see table 1) into main factors according to their impact similarity and avoiding the problem of multicollinearity.

Specifically, if we have a p-indicator m-factor model then the basic factor analysis equation is given by:

$$X = \Phi \zeta + u \quad (12)$$

where X is a $p \times 1$ vector of variables, Φ is a $p \times m$ matrix of factor pattern loadings, ζ is an $m \times 1$ vector of unobservable factors and u is a $p \times 1$ vector of unique factors. It is assumed that the factors are not correlated with the error components. The correlation matrix R of the indicators is given by

$$E(XX') = E(\Phi\zeta\zeta'\Phi') + E(uu') \quad (13)$$

$$R = \Phi\Lambda\Phi' + \Omega$$

where Φ , Λ , Ω matrices are parameter matrices and where R is the correlation matrix of the observables, Φ is the correlation matrix of the factors and Ω is a diagonal matrix containing the unique variances. The diagonal of the R - Ω matrix gives the communalities. The off diagonal of the R matrix give the correlation among the indicators.

The correlation between the indicators and the factors is given as

$$E(X\zeta') = \Phi E(\zeta\zeta') + E(u\zeta')$$

$$A = \Phi\Lambda \quad (14)$$

where A is the correlation between indicators and factors. Rotations of the factor solution are the common type of constraints placed on the factor model for obtaining the unique solution. In our case we have followed the varimax rotation. The objective of this rotation is to determine the transformation matrix C in such a way as any given factor will have some variables loaded high on it and some loaded low on it. This may be achieved by maximizing the variance of the square loading across variables subject to the constraint that the communalities of each variable remain the same (Johnson and Wichern, 1998; Sharma, 1996).

The factor scores are calculated as:

$$\hat{F} = X\hat{B} \quad (15)$$

where \hat{F} is an $m \times n$ matrix of m factor scores for n indicators, X is an $n \times p$ matrix of observed variables and \hat{B} is a $p \times m$ matrix of estimated factor score coefficients. If we standardized our variables

$$\hat{F} = Z\hat{B} \Rightarrow \frac{1}{n}Z'\hat{F} = \frac{1}{n}Z'Z\hat{B} \Rightarrow \Phi = R\hat{B} \quad (16)$$

as $\frac{1}{n}(Z'Z) = R$ and $\frac{1}{n}Z'\hat{F} = \Phi$

Thus the estimated factor scores coefficient matrix is given as $\hat{B} = R^{-1}\Phi$ and the estimated factor scores by $\hat{F} = ZR^{-1}\Phi$.

The factor scores are extracted using the following expression

$$f_j = w_{j1}X_1 + w_{j2}X_2 + w_{j3}X_3 + \dots + w_{jp}X_p \quad (17)$$

where f_j is the score of the j common factor, w_{ji} are considered unknown and they are estimated using regression. In the Principal Components method applied here the scores are exactly calculated. Residuals are computed between observed and reproduced correlations.

If the common factors F and the specific factors u can be assumed normally distributed, then maximum likelihood estimates of the factor loadings and specific variances may be obtained. When F_j and u_j are jointly normal the likelihood is given by:

$$L(\mu, \Sigma) = (2\pi)^{-\frac{np}{2}} |\Sigma|^{-\frac{n}{2}} e^{-\frac{1}{2}tr[\Sigma^{-1}(\sum_{j=1}^n (X_j - \bar{X})(X_j - \bar{X})' + n(\bar{X} - \mu)(\bar{X} - \mu)']]} =$$

$$2\pi^{-\frac{(n-1)p}{2}} |\Sigma|^{-\frac{(n-1)}{2}} e^{-\frac{1}{2}tr[\Sigma^{-1}(\sum_{j=1}^n (X_j - \bar{X})(X_j - \bar{X})']]} (2\pi)^{-\frac{p}{2}} |\Sigma|^{-\frac{1}{2}} e^{-\frac{n}{2}(\bar{X} - \mu)' \Sigma^{-1} (\bar{X} - \mu)} \quad (18)$$

which depends on L and Ψ from the covariance matrix for the m common factor model of $\Sigma = LL' + \Psi$. The maximum likelihood estimates of \hat{L} and $\hat{\Psi}$ are obtained by

maximizing (18). The maximum likelihood estimators \hat{L} , $\hat{\Psi}$ and $\hat{\mu} = \bar{X}$ maximize (18) subject to $\hat{L}'\hat{\Psi}^{-1}\hat{L}$ being diagonal. The maximum likelihood estimates of the communalities are

$$\hat{h}_i^2 = l_{i1}^2 + l_{i2}^2 + \dots + l_{im}^2 \quad \text{for } i=1,2,\dots,p$$

The proportion (P_{var}) of the total sample variance to the j th factor is given by

$$P_{\text{var}} = \frac{\hat{l}_{i1}^2 + \hat{l}_{i2}^2 + \dots + \hat{l}_{im}^2}{s_{11} + s_{22} + \dots + s_{pp}}$$

A proof is provided in Johnson and Wichern (1998).

The elements of the residual matrix are much smaller for the residuals corresponding to maximum likelihood compared to those corresponding to principal components. Based on this, the ML approach is preferred.

The idea to perform a Factor Analysis using as method of extraction the Principal Components, came from the fact that according to previous research outcomes all those variables affect or explain partly the transparency efficiency. Additionally, the proposed variables are expected to present an increased correlation as a result of overlapping variation between them in terms of multicollinearity in a regression model setup. Researchers suggest the application of factor analysis in order to examine the structure of the overlapping variation between the predictors (Leeflang *et al.*, 2000) claiming that the only problem in this case remains the theoretical interpretation of the final components (Greene, 2000; Gurmu *et al.*, 1999).

Table 2 presents the factor loadings and specific variance contributions according to the Maximum Likelihood method of extraction in a Factor Analysis setup. Looking at Table 2 it can be seen that variables PDI, IDV and GDP define factor 1 (high loadings on factor 1, small or negligible loadings on factor 2); while variables UAI, INFLA and COUNTRISK define factor 2 (high loadings on factor 2, small or negligible loadings on factor 1). Variable 3 (MAS) is most closely aligned with factor 1, although it has aspects of the theory represented by factor 2. The communalities (0.434, 0.999, 0.131, 0.289, 0.371, 0.466, 0.999) being moderate indicate that the two factors account for an average percentage of the sample variance of each variable.

Table 2: Statistical Output of the proposed method

Variables	Estimated factor loadings		Rotated factor loadings		Communalities \hat{h}_i^2
	F ₁	F ₂	F ₁ *	F ₂ *	
PDI	0.598	-0.276	0.606	0.258	0.434
IDV	-0.792	0.610	-0.984	-0.176	0.999
MAS	-0.004	0.176	-0.133	0.115	0.131
UAI	0.519	0.140	0.245	0.479	0.289
GDP	0.500	-0.348	0.594	0.136	0.371
INFLA	0.595	0.334	0.152	0.665	0.466
COUNTRISK	-0.791	-0.610	-0.079	-0.996	0.999
Cumulative Percentage of Total sample Variance explained	35.494	51.270	25.646	51.270	
Kaiser-Meyer-Olkin	0.539				
Bartlett's test of Sphericity	61.688 (P=0.000)				

In an attempt to explain the results of our analysis we can conclude that there are clearly two different sets of independent variables in our sample. The first set consisting of PDI, IDV, MAS, and GDP is the set of variables we can group as the factor of socioeconomic stereotypes determining countries' perceived transparency efficiency. Similarly, the second set consists of UAI, INFLA and COUNTRISK

which is the set of variables we can group as the factor of socioeconomic ambiguity determining countries' perceived transparency efficiency. The two factors include three variables that describe country's economic adversity (Getz and Volkema, 2001) and four determinants of country's cultural values (Hofstede, 1994). Socioeconomic ambiguity which in turn may have a negative effect on countries' transparency can be appeared in counties with an environment of political and investment instability, with high levels of uncertainty avoidance and high inflation rates.

3.4 Logistic regression

Let us now use the logistic regression in formulating a model of explaining the transparencies with the extracted factors. First we define the distributional properties of the dependent variable, (for more details on the properties and applications of logistic regression see Halkos 2006; Kleinbaum 1994; Hosmer and Lemeshow 1989; Collett 1991; Kleinbaum *et al.* 1999; Hair *et al.* 1998; Sharma, 1996).

In our sample the first n_1 out of n observations have the characteristic under investigation (transparency efficiency $\geq 70\%$ – transparency efficiency $< 70\%$) and so $Y_1=Y_2=\dots=Y_{n_1}=1$ while the rest of the observations do not and so $Y_{n_1+1}=Y_{n_1+2}=\dots=Y_n=0$.

Instead of minimizing the squared deviations as in a multiple regression, logistic regression maximizes the likelihood that an event (E) will take place.

$$\ln \frac{\Pr(E)}{1 - \Pr(E)} = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k \quad (19)$$

or

$$\Pr(E) = \frac{1}{1 + \exp^{-(\beta_0 + \sum_{i=1}^k \beta_i X_{ij})}} \quad (20)$$

where P is the probability of having the characteristic under investigation given the independent variables X_1, X_2, \dots, X_k . Equation (20) models the log of the odds as a linear function of the independent variables and it is equivalent to a multiple regression equation with log of the odds as the dependent variable.

The logit form of the model is a transformation of the probability $\Pr(Y=1)$ that is defined as the natural log odds of the event $E(Y=1)$. That is

$$\text{logit} [\Pr(Y=1)] = \ln[\text{odds}(Y=1)] = \ln \left[\frac{\Pr(Y=1)}{1 - \Pr(Y=1)} \right] \quad (21)$$

In the general case, where the dichotomous response variable Y, denotes whether (Y=1) or not (Y=0) the characteristic under investigation (transparency efficiency $\geq 70\%$ – transparency efficiency $< 70\%$) is linked with the k regression variables $X=(X_1, X_2, \dots, X_k)$ via the logit equation, recall (18)

$$P(Y=1) = \frac{\exp \left\{ \beta_0 + \sum_{k=1}^K \beta_k X_k \right\}}{1 + \exp \left\{ \beta_0 + \sum_{k=1}^K \beta_k X_k \right\}} \quad (22)$$

This is equivalent to

$$\text{logit} \Pr(Y=1 | X) = \beta_0 + \sum_{k=1}^K \beta_k X_k \quad \text{due to (21).}$$

The regression coefficients β 's of the proposed logistic model quantifies the relationship of the independent variables to the dependent variable involving the parameter called the *Odds Ratio* (OR). As odds we define the ratio of the probability that implementation will take place divided by the probability that implementation will not take place.

That is

$$\text{Odds}(E | X_1, X_2, \dots, X_n) = \frac{\Pr(E)}{1 - \Pr(E)} \quad (23)$$

4. Empirical findings

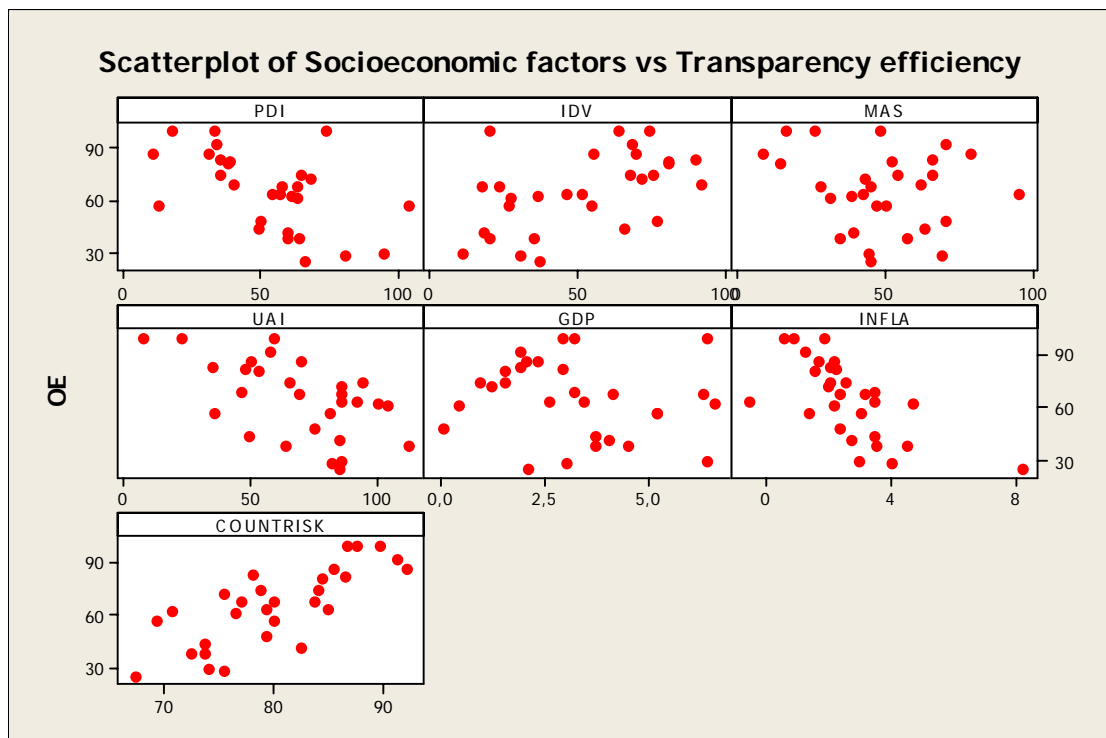
According to table 3 the results derived from our DEA analysis indicate that three out of the twenty nine countries (Denmark, Finland and Singapore) are transparent efficient. This means that these three countries have lower levels of perceived corruption in their political, legislative, legal, tax and business environment and higher transparency levels. Therefore, their transparency efficiency scores are equal to 100%. The countries with lower transparency efficiency and thus with higher levels of perceived corruption are reported to be Greece, Panama, Mexico and Turkey with transparency efficiency scores of 37.9%, 29.9%, 28.8% and 25.3% respectively. Looking at the results of our DEA analysis four of the EU countries located in the Mediterranean region (Spain, Portugal, Italy and Greece) have transparency efficiency scores below 70%, which is also the case for the USA and Japan.

Table 3: Transparency efficiency scores and country ranking

Country	Transparency Efficiency	Ranking	Country	Transparency Efficiency	Ranking
<i>Denmark</i>	100,00	1	<i>Japan</i>	63,83	14
<i>Finland</i>	100,00	1	<i>Spain</i>	63,04	15
<i>Singapore</i>	100,00	1	<i>Uruguay</i>	62,10	16
<i>Switzerland</i>	92,37	2	<i>Portugal</i>	61,59	17
<i>Austria</i>	86,96	3	<i>Israel</i>	56,55	18
<i>Norway</i>	85,98	4	<i>Malaysia</i>	56,52	19
<i>UK</i>	83,45	5	<i>Italy</i>	48,06	20
<i>Canada</i>	81,57	6	<i>South Africa</i>	43,77	21
<i>Netherlands</i>	80,85	7	<i>Korea</i>	41,61	22
<i>Germany</i>	74,93	8	<i>Thailand</i>	38,41	23
<i>Belgium</i>	74,03	9	<i>Greece</i>	37,90	24
<i>France</i>	71,86	10	<i>Panama</i>	29,91	25
<i>USA</i>	68,37	11	<i>Mexico</i>	28,77	26
<i>Chile</i>	68,11	12	<i>Turkey</i>	25,27	27
<i>Taiwan</i>	67,60	13			

Furthermore, figure 2 represents graphically the association of the seven socioeconomic factors used in our study with country levels of transparency efficiency. The results clearly indicate that higher levels of a country's transparency efficiency is positively associated with national cultures with social characteristics of lower power distance, higher individualism values, lower masculine values, lower uncertainty avoidance. Moreover, the association between higher levels of efficiency transparency and lower levels of inflation and country risk (the higher the *COUNTRISK* score the lower the investment risk) is also emphasised. However for the case of countries' GDP change the results are not conclusive.

Figure 2: Graphical representation of socioeconomic influence to country's transparency efficiency



In this stage of our analysis the results regarding the association of the socioeconomic factors with country's transparency efficiency are fully supported by the studies from Getz and Volkema (2001) and by Davis and Ruhe (2003), which in their research found a link between country's economic adversity, cultural

characteristics and perceived corruption. However, in our study the analysis goes further by introducing countries' transparency efficiency and by formulating two main factors of all the socioeconomic variables according to their communality of influence in a logistic regression analysis.

The idea of performing a regression analysis between a dependent variable and extracted factors is not a new one. Dunteman (1989) also suggests this process to cope with multicollinearity in a regression analysis model and it is also an indicated way to minimize the number of independent variables and maximize the degrees of freedom.

As our main interest is in terms of the main effects we have ignored interactions. Working with the two factors extracted most statistical significant variables we derive the logit form of the fitted model, which may be represented as

$$\text{logit} [\text{Pr}(Y=1)] = \beta_0 + \beta_1 \text{FACTOR 1} + \beta_2 \text{FACTOR 2} + \varepsilon_t$$

where Y denotes the dependent variable as 1 for countries with transparency efficiency scores of \geq to 70% and 0 for countries with transparency efficiencies less than 70%. The beta terms are the unknown coefficients to be estimated, and ε_t is the error term, assumed to be normally distributed with mean 0 and variance 1.

Specifically, the dependent variable is the answer to the question of the influence of transparency efficiency derived from the DEA application and adopting as a rule a level of efficiency greater than 70%. The results of the fitted models are presented in Table 4.

Table 4: Logistic Regression results

Dependent: Transparency				
Variables	Estimates	Odds Ratio	Estimates	Odds Ratio
Constant	-1.611 Wald [3.009] P-value (0.083)	0.200		
Factor 1	-2.210 Wald [5.337] P-value (0.021)	0.110	-1.558 [6.293] (0.012)	0.210
Factor 2	-3.161 Wald [5.565] P-value (0.018)	0.042	-2.254 [6.452] (0.011)	0.105
Cox and Snell R ²	0.555		0.504	
Nagelkerke R ²	0.756		0.673	
Hosmer Lemeshow	2.147 [0.976]		6.445 [0.597]	
Likelihood Ratio	15.000 [0.000]		19.840 [0.000]	

We may compute the difference $e^{\hat{\beta}_i} - 1$ which estimates the percentage change (increase or decrease) in the odds $\pi = \frac{\Pr(Y = 1)}{\Pr(Y = 0)}$ for every 1 unit in X_i holding all the other X 's fixed. The coefficient of factor 1 is $\hat{\beta}_1 = -2.210$, which implies that the relative risk of this particular variable is $e^{\hat{\beta}_1} = 0.110$ and the corresponding percentage change is $e^{\hat{\beta}_1} - 1 = -0.89$. This means that in relation to the socioeconomic stereotypes (factor 1) the transparency efficiency decreases by almost 90% ceteris paribus. In the case of the determinants of socioeconomic ambiguity (factor 2) the result is $\hat{\beta}_2 = -3.161$, which implies that the relative risk of this particular variable is $e^{\hat{\beta}_2} = 0.042$ and the corresponding percentage change is $e^{\hat{\beta}_2} - 1 = -0.958$. This means that in relation to factor 2 the odds of transparency efficiency decrease by almost 96% all other remaining fixed.

In case we run the model with no constant term then the coefficient of factor 1 is $\hat{\beta}_1 = -1.558$, which implies that the relative risk of this particular variable is

$e^{\hat{\beta}_1}=0.210$ and the corresponding percentage change is $e^{\hat{\beta}_1}-1=-0.79$. This means that in relation to factor 1 the transparency efficiency decreases by almost 79% *ceteris paribus*. In the case of factor 2 $\hat{\beta}_2=-2.254$, which implies that the relative risk of this particular variable is $e^{\hat{\beta}_2}=0.105$ and the corresponding percentage change is $e^{\hat{\beta}_2}-1=-0.895$. This means that in relation to factor 2 the odds of transparency efficiency decrease by almost 0.89% all other remaining fixed.

The individual statistical significance of the β estimates is presented by the Wald (Chi-square). The significance levels of the individual statistical tests (i.e. the P-values) are presented in parentheses and correspond to $\text{Pr}>\text{Chi-square}$. Note that both factors are statistically significant for statistical levels of 0.05 and 0.1 while the constant term is statistically significant for 0.1. Running the logistic regression without the constant term then both factors are statistically significant at $\alpha=0.05$ and $\alpha=0.1$. The model certainly fits the data well and provides evidence that the economical interpretation of the logit model.

To assess the model fit we compare the log likelihood statistic ($-2 \log \hat{L}$) for the fitted model with the explanatory variables with this value that corresponds to the reduced model (the one only with intercept). The likelihood ratio statistic for comparing the two models is given by the difference

$$\text{LR} = (-2 \log \hat{L}_R) - (-2 \log \hat{L}_F) = 15$$

where the subscripts R and F correspond to the Reduced and Full model respectively. That is, in our case the overall significance of the model is $X^2=15.000$ (or 19.840 in the case with no constant) with a significance level of $P=0.000$. Based on this value

we can reject H_0 (where $H_0: \beta_0 = \beta_1 = \beta_2 = 0$) and conclude that at least one of the β coefficients is different from zero.

Finally, the Hosmer and Lemeshow value equals to 2.147 (with significance equal to 0.976). In the case with no constant the results are 6.445 (with P-value equal to 0.597). The non-significant X^2 value indicates a good model fit in the correspondence of the actual and predicted values of the dependent variable.

The results of our logit model partially support the theory for the case of factor 1. The negative effect of high PDI and MAS cultural characteristics on countries' transparency efficiency are supported by the theory (Getz and Volkema, 2001; Davis and Ruhe, 2003). However, we found that higher values of IDV cultural characteristics have also a negative impact on countries' transparency efficiency. Furthermore, a GDP change doesn't ensure countries' higher transparency efficiency levels. Finally, the results for factor 2 are fully support the theory. It seems that countries with lower levels of UAI, INFLA and COUNTRISK ensure higher levels of countries' transparency efficiency. These results support the study by Davis and Ruhe (2003) and constitute to the fact that corruption results to the countries with an environment of political and economic risks.

5. Conclusion

This study investigates the link between countries' perceived transparency and eight socioeconomic determinants. The Hofstede's cultural dimensions and three economic variables have been used in order to justify their influence on countries' ability to reduce corruption. For the first time this paper tries to measure this ability by introducing the term transparency efficiency. By using DEA methodology we measure the perceived transparency levels of twenty nine countries. Furthermore, using factor analysis we separate the seven socioeconomic determinants into two main factors according to their communality of influence. Finally, logistic regression has been used in order to clarify the way these two factors influence countries' transparency efficiency. The results indicate that in relation to the factors socioeconomic stereotypes (factor 1) the odds of transparency efficiency decreases by almost 90% all others remaining fixed while in relation to the factors of socioeconomic ambiguity (factor 2) the odds of transparency efficiency decreases by almost 96% all others remaining fixed. In case with no constant term the results indicate that in relation to the factors socioeconomic stereotypes (factor 1) the odds of transparency efficiency decreases by almost 79% all other remaining fixed while in relation to the factors of socioeconomic ambiguity (factor 2) the odds of transparency efficiency decreases by almost 89% all other remaining fixed.

The results indicate that countries with cultural characteristics of lower power distance, lower masculinity values, lower uncertainty avoidance and lower values of individualism constitute to higher levels of countries transparency efficiency. Additionally, lower inflation rates and lower political and economical risks constitute to higher levels of countries' transparency efficiency while GDP growth doesn't ensure countries' transparency efficiency.

REFERENCES

- Charnes, A., Cooper, W. W. and Rhodes, E. (1978), Measuring the Efficiency of Decision Making Units, *European Journal of Operational Research*, Vol. 2, pp. 429-444.
- Charnes, A., Cooper, W. W., Lewin, A. Y. and Seiford, L. M., (1994), *Data Envelopment Analysis: Theory, Methodology and Application*, Kluwer Academic Publishers, Boston/Dordrecht/London.
- Cooper, W.W., Seiford, L. M. and Tone, K. (2000), *Data Envelopment Analysis: A Comprehensive Text with Models, Applications, References and DEA-Solver Software*, Kluwer Academic Publishers: Boston
- Dauids, M. (1999), Global standards, Local Problems, *Journal of Business Strategy*, 38-43.
- Davis, H. J. and Ruhe, A., J, (2003), Perceptions of country corruption: Antecedents and outcomes, *Journal of Business Ethics*, 43, 275-288.
- Dunteman, G.H. (1989), *Principal Components Analysis*, Sage.
- Farrell, M. J. (1957), The Measurement of Productive Efficiency, *Journal of Royal Statistical Society A*, Vol. 120, pp. 253-281.
- Getz, A., K. and Volkema, R., J. (2001), Culture, perceived corruption and economics: a model of predictors and outcomes, *Business & Society*, 40,1, 7-30.
- Greene, W. (2000), *Econometric Analysis*, Prentice Hall International.
- Gurmu, S., P. Rilstone and S. Stern (1999), 'Semiparametric estimation of count regression models. ', *Journal of Econometrics*, 88(1), 123-150.
- Habib, M. and Zurawicki, L. (2002), Corruption and Foreign Direct Investment, *International Journal of Business Studies*, 33, 2, 291-307.
- Hair, J. F, Anderson R. E, Tatham, R. L and Black, W. C. (1998). *Multivariate data analysis*, Prentice Hall, Fifth Edition
- Halkos G. E. (2006), *Econometrics : Theory and Practice*, Giourdas Publications.
- Hofstede, G. (1980a), *Culture's consequences: international differences in work related values*, Sage: Beverly Hills, CA.
- Hofstede, G. (1980b), Motivation, leadership and organization: do American theories apply abroad?, *Organizational Dynamics*, 9, 1, 42-63.
- Hofstede, G. (1994), *Cultures and Organizations: Software of the mind*, Harper Collins Business.

Hosmer D., W and Lemeshow S. (1989). Applied logistic regression. John Wiley and Sons, New York

Husted, B. (1999), Wealth, culture and corruption, *Journal of International Business Studies*, 30, 2, 339-360.

Johnson, R.A. and Wichern, D.W. (1998), *Applied Multivariate Statistical Analysis*, Fourth Edition, Prentice Hall.

Kleinbaum D., G. and Kupper L.,L., Muller KE. and Nizam A. (1999). Applied regression analysis and other multivariate techniques, Duxbury, Third Edition.

Kleinbaum D.G. (1994). Logistic regression: A self learning text, Springer-Verlag, New York.

Leeflang, P., Wittink, D., Wedel, M. and Naert, P. (2000), 'Building models for marketing decisions', Kluwer Academic Publishers.

Mauro, P. (1995), Corruption and Growth, *Quarterly Journal of Economics*, 109, 681-712.

OECD (2005), *OECD in figures 2005 edition*, Statistics on the member states countries, OECD Observer.

Robertson, C., J. and Watson, A. (2004), Corruption and Change: the impact of foreign direct investment, *Strategic Management Journal*, 25, 4, 385-396.

Sharma S. 1996. Applied multivariate techniques, John Wiley and Sons, New York.

Stolp, C. (1990), Strengths and Weaknesses of Data Envelopment Analysis: An Urban and Regional Perspective, *Computers, Environment and Urban Systems*, Vol. 14, No. 2, pp. 103-116.

Tanzi, V. and Davoodi, H., (1997), Corruption, Public Investment and Growth, *International Monetary Fund*, Working paper, 97/139.

Thanassoulis, E. (2001), *Introduction to Theory and Application of Data Envelopment Analysis*, Kluwer Academic Publishers, Norwell.

Transparency International, (2005), *Corruption Perception Index*, Retrieved from: www.transparency.org.

World Investment Report (2005), *Transnational Corporations and Internationalization of R&D*, United Nations, New York and Geneva.