USING SIMULTANEOUS EQUATIONS MODELS TO ANALYZE THE CAUSES OF CORRUPTION AND ITS IMPLICATIONS

TUDOREL ANDREI[*]
BOGDAN OANCEA[**]

Abstract
For a country that is in the process of integration into EU structures, reducing the corruption is a good sign for attracting foreign investment and developing the economic environment. The paper estimates the parameters of a simultaneous equation model based on data sets obtained at a sample of employees in public administration. Statistical sample consist in 407 people and the maximum allowable error was estimated at ± 2.5%. For the effective development of the statistical questionnaire we identified major themes of public administration that are directly related to the problem of corruption: managing the institution, the civil service, transparency in the system, the decentralization process, causes and effects of corruption and the quality of the reform in the public administration. Based on the questionnaire we defined primary and secondary variables that have been used to define the model with simultaneous equations. For the variables in the model they have been divided into endogenous and exogenous.

Keywords: *simultaneous equations model, two stages least square method, corruption, public health system, Hausman statistics*

1. Introduction

During the transition, the level of corruption in Romania stood at the highest values among the countries of Eastern Europe. The reasons for this are various. The high level of corruption has led to significant social and economic losses: the reduction of foreign investment, reduction of production capacities in industry, construction, agriculture etc. through fraudulent privatization and liquidation of them, inefficient use of funds from the state and local communities.

There are a number of applications of simultaneous equations models for the analysis of social phenomena in the literature, while the parameters are estimated using data series that are obtained through the application of statistical surveys.

The major problems that arise in the definition and use of simultaneous equations models that uses the data sets obtained from a sample are related to two aspects: the definition of endogenous variables list, and the definition instrumental variables list that are used in parameter estimation by the two stages least squares method. Each of the two issues has an important role in obtaining conclusive results.

2. The simultaneous equations model for corruption analysis

To define the simultaneous equations model we considered two categories of variables: the endogenous variables that are specified in the model by the vectorial variable $\mathbf{y_i}$ and the exogenous variables included in the model by the variable $\mathbf{x_i}$. Under these conditions the simultaneous equations model is defined in the structural form as:

$$\mathbf{B}\mathbf{y}_i + \mathbf{C}\mathbf{x}_i = \boldsymbol{\varepsilon}_i \quad [1]$$

[**]Professor, "Nicolae Titluescu" University, Bucharest, (e-mail: oanceab@ie.ase ro).

where the residuum vector follows a normal distribution $\boldsymbol{\varepsilon}_i \rightarrow N(\mathbf{0},\boldsymbol{\Omega}), i = 1,...,G,$ and the matrix $\Omega = (\sigma_{ij})_{i,j=1,\ ,G}$.

To define this model, first, the following problem should be solved: the separation of the variables in the model in endogenous and exogenous. Moreover, among these variables can be identified causal relationships.

For a correct division of the variables in endogenous and exogenous variables, we will consider the analysis of the causal relationship that exists between different variables. In applying statistical tests to be taken into account that data series are recorded at the level of a statistical sample and are not recorded values for a variable for a certain period of time.

Solving the second problem is important for correct estimation of the parameters. This decision must be taken when the parameters of the model with simultaneous equations is estimated by the method of two-stage least squares (TSLS). Please note that in literature there is no uniform approach in choosing the list of instrumental variables in simultaneous equations models used to analyze issues related to corruption (Bai and Wei, 2000) (Kaufmann et al., 1999).

3. Aplication

In the following we present the application of the model with simultaneous equations for the analysis of some important issues in the public health system. The model thus defined can be used to analyze particular aspects of the system such as the quality of the reform of the system, the size of non-academic behavior in the public health system, government policy on health characteristics, education level of population for preventing or agravation of a disease, etc.

To estimate the parameters we used data sets obtained from a representative sample from Bucharest. The volume of the sample was 407 people, and the survey results are guaranteed with a probability of 95%, given that there was a representation error of ± 2.5%. For the data collection process we used a two-stage sampling plan. The first step was the health units in Bucharest (hospitals, health centers, clinics) that have been treated as primary sampling units. The second step was the doctors in each primary sampling units.

To define the endogenous and exogenous variables and the equations of the model we considered the following hypotheses:

- To define the variable that quantify the quality of the public health system reform it must be considered that the medical personnel opinion on this issue is in relation to quality of the public health system financing, the reform measures taken at the medical institutions, the quality of decentralization in the health system, drug procurement system characteristics and quality of employment and promotion system for staff in public health system;
- The effects of the reform process in this system are observable in the first period by increasing public health expenditure, the improvement of the quality of the national health care programs undertaken by the ministry, the ministry's decision to increase transparency at the medical units level, etc.
- The reform process must support measures to improve the health education of the population;
- The results of the reform process are perceived at the population level by reducing the corruption in the public health system, improving the medical care, etc.

The model is defined by the following equations:

$$RSS = f_1(CF, CSE, TMS, PDS, ESP, \mathbf{VP}) + \varepsilon_1$$

$$TMS = f_2(PDS, RSS, COR, \mathbf{VP}) + \varepsilon_2$$

$$ESP = f_3(PCS, PAC, ACS, DPE, UDP) + \varepsilon_3$$

$$COR = f_4(RSS, CSE, TMS, SCP, GSM, \mathbf{VP}) + \varepsilon_4$$

Within these linear models we used the following variables: the quality of the public health system reform (RSS), the quality of the factors influencing the achievement of a qualitative medical act (CF) and rating system of the institution and employees (CSE); the transparency of the ministry in making decisions about the ongoing reform (TMS); the ministry's policy in the area (PDS); the health education of the population (ESP), the interviewed person characteristics, including gender of the person, age and category of staff defined in the model on the basis of the vector variable VP; transparency of decisions from the public health system (TMS), the corruption at the national level (COR), the health education of the population (ESP), the frequency of application, at the end of a treatment period for a new medical examination (PCS), the extent to which people give enough importance to their health (PAC), factors related to the accessibility of citizens to primary, secondary and tertiary health care (ACS), the usefulness of developing health education and prevention programs for the population (UDP), the evaluation system of the quality of the medical services (CSE), changing the executives based on the political criteria (SCP) and medical staff satisfaction (GSM).

The simultaneous equations model variables are divided into endogenous and exogenous, as follows:

- Endogenous variables: RSS, TMS, ESP and COR;
- Exogenous variables: CF, CSE, PDS, GEN, ANI, PER, PCS, PAC, ACS, DPE, UDP, SCP, GSM and CSE.

In the following parameters for the four equations are estimated by two methods: the method of least squares (OLS) and two stages least square method (TSLS). The results are presented in Tables 1 and 2.

In the following we present the most important variables used in the model:

- RSS is a variable defined to measure the opinion on the quality of health care reform process in the public health system. To define we envisaged six components: the financing of the system, the procurement of the drugs, the decentralization process in the health system, the employment and promotion system of the medical personnel with secondary and higher education and the reform measures implemented in the medical units. For measuring the opinion of physicians on each of the six questions we used a scale with five values: 1 - very poor, 2, 3, 4, 5-very good. This variable is an aggregate variable that is defined on the basis of the six primary variables;
- CF is an aggregate variable used to measure the quality of the factors that contribute to a qualitative medical act in the public health units. To measure the values of these characteristics we used a scale with five values that are defined as: 1-very poor, 2, 3, 4, 5 - very good. The aggregate variable is calculated as the arithmetic mean of five primary variables defined directly on the basis of the questionnaire;
- CSE is an aggregate variable that is calculated as the average of three primary variables. It is used to estimate the quality of the assessment system of health services rendered to beneficiaries. To define the primary variables we used a scale with four values: 1 - unsatisfactory, 2, 3, 4 - very good.
- TMS is an aggregate variable used to assess transparency in ministry decision-making about the reform process. It is calculated as the arithmetic average of the two primary variables defined on the questionnaire. The measurement scale used has four values: 1 - unsatisfactory, 2, 3, 4 - very good;
- PDS is an aggregate variable used to assess the quality of government health policy in terms of volume of public health expenditure, the quality of national health programs run by the ministry and the transparent use of funds for compensated and free drugs in the primary care. The range is 1 - unsatisfactory, 2, 3, 4 - very good. This is calculated as an arithmetic average of the three primary variables;

- ESP is a variable used to measure the aggregate level of population health education and disease prevention or progression. The range is 1 - most people do not give importance to prevent the emergence or worsening of a disease, 2, 3, 4, 5 - most people consider this issue very important. It is calculated as the arithmetic average of the two primary variables;
- COR is an aggregate variable used to measure the corruption at the national level according to the medical staff with higher education. The range is 1 - there is no corruption, 2, 3, 4, 5 - there is a widespread corruption. It is calculated as the arithmetic average of the five variables defined directly from the questionnaire.
- PCS quantifies the extent to which patients that received a medical treatment calls for a new specialist advice. The range is 1 a small part of them, 2, 3, 4, 5 - with few exceptions, all patients.
- PAC is a variable used for an overall assessment of the extent to which people give enough importance to their health. The range is 1 - do not give enough importance to their health, 2, 3, 4, 5 – gives a great importance to health. It is determined based on the arithmetic average of two primary variables;
- ACS is a variable that measures the degree of accessibility of citizens to the primary, secondary and tertiary health care system. The range is 1 - poor accessibility, 2, 3, 4, 5 - highest availability to the medical act. It is calculated by means of three primary variables;
- DPE is an aggregate variable used to assess the overall contribution of public institutions to develop health education programs and prevention of disease among the population. The range is 1 - unsatisfactory, 2, 3, 4 - very good. The aggregate variable is calculated as an average of four primary variables;
- UDP is a primary variable used to assess the usefulness of health education programs and prevention among the population. The range is 1 - not useful, 2, 3, 4, 5 - totally useful;
- SCP is a primary variable used to assess to what extent the political level influences the changing of the management personnel based on political criteria. The range is 1 – there are no changes in the leadership based on political criteria, 2, 3, 4 – changing the personnel based on political criteria is a current practice;
- GSM is an aggregate variable defined to evaluate medical staff satisfaction. The range is 1 - not satisfied at all, 2, 3, 4, 5 - completely satisfied. It is calculated as the average of five primary variables that are defined directly on the questionnaire.

For each variable defined above we can computed a number of indicators to characterize the central tendency, dispersion and the asymmetry of the distribution. In this case, we used questionnaires with valid responses to questions used to define the underlying primary variables that define the aggregate variable.

To estimate the parameters of the four regression models we used only statistical questionnaires that have valid answers to all questions underlying the definition of the aggregate variables in the regression equation.

For a successful application of the method of two stages least squares in (TSLS) in the application it must be defined a list of instrumental variables. They must meet a number of conditions (Andrew and Bourbonnais 2008) to obtain suitable results.

For the second stage an important role is played by the definition of the list of instrumental variables. In this context the exogeneity of the variables of the model is analyzed. An important tool in this approach is the Hausman test (Hausman, 1978). It seeks to check the effectiveness and consistency of estimators. In this regard the following two hypotheses are defined.

The first is the case when the list of instrumental variables is correctly specified. The estimator of the parameter $\boldsymbol{\beta}$ obtained by the OLS, denoted by $\hat{\boldsymbol{\beta}}_0$, is effectively and consistent. In this case the explanatory variables of regression model $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u}$ are not correlated with residual variables, so $H_0 : \text{cov}(\mathbf{u}, \mathbf{X}) = \mathbf{0}$.

In the second case, the list of instrumental variables is not correctly specified. The estimator for the parameter $\boldsymbol{\beta}$ obtained by the OLS, denoted by $\hat{\boldsymbol{\beta}}_1$, is effective and inconsistent. Residual variables are correlated with one or more explanatory variables, so $H_1 : \text{cov}(\mathbf{u}, \mathbf{X}) \neq \mathbf{0}$.

The difference between the two estimators is $\hat{\mathbf{d}} = \hat{\boldsymbol{\beta}}_1 - \hat{\boldsymbol{\beta}}_0$, and the Hausman test statistics is:

$$H = \hat{\mathbf{d}}'(\text{var}(\hat{\boldsymbol{\beta}}_1) - \text{var}(\hat{\boldsymbol{\beta}}_0))^{-1}\hat{\mathbf{d}} \rightarrow \chi^2(r)$$

where r is the number of endogenous varibles from the list of explicative varibles, so of the variables $X_i, i = 1,...,r$ that verify the following $\text{cov}(u, X_i) \neq 0$.

If the statistics value is greater than the tabulated value, then we reject the null hypothesis, considering in this case that the second estimator gives more appropriate results.

Table 1. The parameters of the equations estimated by OLS

Explicative variable	Depedant variable RSS		Depedant variable TMS		Depedant variable ESP		Depedant variable COR	
	parameters	t-Student statistics	parameters	t-Student statistics	parameters	t-Student statistics	parameters	t-Student statistics
CF	0,309	8.784 (0,000)*						
CSE	0,017	0,406 (0,635)					0,029	0,415 (0,679)
TMS	0,041	1,010 (0,313)					0,125	1,939 (0,053)
PDS	0,393	6,746 (0,000)	0,758	11,359 (0,000)				
ESP	0,141	4,024 (0,000)						
GEN	- 0,027	0,539 (0,590)	0,043	0,645 (0,520)			0,061	1,418 (0,157)
ANI	0,047	2.074 (0,039)	0,018	0,623 (0,533)			0,155	3,195 (0,002)
PER	- 0,042	1.649 (0,100)	0,013	0,396 (0,692)			0,320	8,021 (0,000)
RSS			0,131	2,468 (0,014)			-0,028	-0,354 (0,725)
COR			0,045	1,412 (0,159)				
PCS					0,084	2,200 (0,028)		
PAC					0,520	11,188 (0,000)		
ACS					0,040	1,123 (0,262)		
DPE					0,204	4,513 (0,000)		
UDP					0,009	0,373 (0,709)		

SCP							0,186	3,043 (0,002)
GSM							0,758	8,357 (0,000)
F	1296,9 (0.00)*		563.8 (0.00)		429.9 (0.00)		566.1 (0.00)	

* the values for α are given in brackets

Table 2. The parameters of the equations estimated by TSLS

Explicative variable	Depedant variable RSS		Depedant variable TMS		Depedant variable ESP		Depedant variable COR	
	parameters	t-Student statistics	parameters	t-Student statistics	parameters	t-Student statistics	parameters	t-Student statistics
CF	0,463	3,408 (0,001)						
CSE								
TMS	-0,058	-0,572 (0,568)					0,369	1,459 (0,001)
PDS	0,549	1,836 (0,067)	1,672	3,111 (0,0020)				
ESP	0,178	2,536 (0,012)						
GEN							1,802	1,738 (0,083)
ANI	-0,094	-0,736 (0,462)					-1,085	-2,272 (0,024)
PER							0,062	-0,132 (0,895)
RSS			-0,372	-0,971 (0,332)				
COR			0,024	0,332 (0,740)				
PCS					-0,079	-0,648 (0,517)		
PAC					1,025	2,228 (0,026)		
ACS					-0,132	-0,557 (0,578)		
DPE					0,021	0,151 (0,880)		
UDP								
SCP							0,612	2,607 (0,009)
GSM							0,384	1,226 (0,221)
Variable list	CF, CSE, PDS, ANI, PER, PCS, PAC, ACS, DPE, UDP, SCP, GSM, CSE, COR, FLC, RCO, GSM, FMM		CF, CSE, COR, FLC, RCO, GSM, FMM		CF, CSE, PDS, ANI, PER, UDP		RSS, CF, CSE, PDS, ESP, PCS, DPE, FMM, ACS	

* the values for α are given in brackets

4. Conclusions

To increase the efficiency of the public health services is necessary to implement measures that will lead to major changes in the public health system. The new philosophy of operation of the health system will cause a reduction in corruption in the system and increase the quality of medical care.

To identify important aspects of health system operation an important tool is the statistical questionnaire. Its application to medical personnel will identify positive and negative aspects of the system.

The simultaneous equations model is used to analyze some aspects of functioning of the public health system. To define the model we used four endogenous variables and thirteen exogenous variables. The variables used in the model are mostly aggregated variables that are calculated as the average primary variables. They are defined based on statistical questions in the questionnaire taking into account the measurement scales used for each question in the questionnaire. For writing the four equations we start from a set of assumptions. The four equations are used for the assertion of some observations regarding the quality of the system reform and its implications for the quality of the medical services provided to citizens, identifying the characteristics of government health policy, evaluating the effect of public health education in preventing and worsening of a disease, assessing the extent of non-academic behavior in the public health system and its implications for the reform process and the quality of medical services provided to citizens.

For the identification of the four equations we used variables regarding the system behavior, the degree of satisfaction of physicians and physicians' personal characteristics. The results highlight the different opinions of doctors in relation to personal characteristics considered. In three of the four models the parameters that correspond to these variables are different from zero. To analyze the variable used to measure the levels of the health education we used a regression model with a series of explanatory variables related to the attention that people give to health (PAC), the practice of patients seeking care and new investigations at the end treatment period (PCS), the extent to which institutions or organizations are involved in developing health education programs in the population (DPE), population accessibility to health services (ACS) and the usefulness of these types of programs (UDP). To define the equation that describes the influence of corruption and non-academic conduct we considered the results of descriptive analysis of data series on corruption levels and intensity of factors acting to reduce it.

Aknowledgement

This work was supported by CNCSIS – UEFISCDI, project number PNII – IDEI code 1814/2008, financing contract no. 763/2009

References

1. Andrei T., Matei A., Oancea B., (2009), Simultaneous Equations Models Used in the Study of Some Issues Related to the Corruption and Performance of Services in the Public Health System, *Theoretical and Applied Economics,* 1(530), pp. 3-18.
2. Andrei T., Bourbonnais, R.,(2008) "Econometrie", *Editura Economica*, Bucureşti.
3. Andrei, T., Matei, A., Rosca, I.,(2008), "Corupția - o abordare socio-economică", *Editura Economică,* Bucureşti.
4. Bai, Chong-En, Wei, S-J., (2000), *The Quality of the Bureaucracy and Capital Account Policies,* World Bank Working Paper No. 2575.
5. Baltagi, B.H., (2008), "Econometrics", *4th Edition, Springer.*
6. Bardhan P.,(1997), "Corruption and Development: A Review of Issues", *Journal of Economic Literature*, Vol.35, no 3.

7. Emerson, M., P., (2002), "Corruption and industrial dualism in less developed countries, The Journal of International Trade & Economic Development", pp. 63-76.
8. Hausman, J.,(1978) „Specific Tests in Econometrics", *Econometrica*, 46.
9. Kaufmann, D., Kraay, A., Zoido-Lobaton, P., (1999), Gouvernance Matters, World Bank Working Paper No. 2195.
10. Lambsdorff J. G., (1999), "Transparency International Corruption Index. Responding to the Challenges of Corruption", *Act of the International Conference*, Milan, 19-20 November 1999, pp. 257-277.
11. McCabe, D.L., and Trevino, L.K., (1993), "Academic dishonesty: Honor codes and other contextual influences", *Journal of Higher Education*, 65, pp. 520-538.
12. Mauro P., (1995), "Corruption and Growth", *Quarterly Journal of Economics*,Vol. 110, no 3.
13. Mauro P., (1998), "La corruption: causes, conséquences et recommendations", *Finance et developpement*, volume 35, no 1, Mars.
14. Schleifer, A., & Vishny, R.,(1993), "Corruption", *Quarterly Journal of Economics*, 59, pp. 599-617.