AN ANALYSIS OF THE CHARACTERISTICS OF PUBLIC HEALTH SYSTEM AT REGIONAL LEVEL USING PANEL DATA

CATALINA LILIANA ANDREI[*]
BOGDAN OANCEA[**]
TUDOREL ANDREI[***]

Abstract

Reforming the public health system is a complex and long process, involving different categories of people. To accelerate the process of integration into the European Union, Romania is currently implementing strategies and programs aimed to increase the quality of public services. In the medical field a series of measures have been undertaken aimed at accelerating the decentralization process and optimize the activities of medical institutions. During the transition period a series of measures have been taken to decentralization and privatization of health services. However, currently we are witnessing a fragmentation of the system, which stressed the inequality in the distribution of medical personnel and reduced people's access to certain types of medical services. Please note that the number of doctors per capita in rural areas is only 20% of the urban area. Another major shortcoming of the system is linked to the financing system and its correlation with the strategies of decentralization. Frequently, decentralization has emerged as a way of placing the central tasks in the task of local government. Using panel data from developing regions we highlight a number of implications of the decentralization process.

Keywords: *panel data, regression models, public health care system, decentralization, regional analysis*

1. Introduction

Decentralization of public services is a process that takes place in all European countries, representing a means of increasing the efficiency of public services. In all cases, this process is based on the principle of subsidiarity. In Romania, along with political changes at the end of 1989, concrete actions have been taken to decentralize the public services. The process was difficult and long, considering the fact that during the communist period there was a hyper-centralized management of public services.

In the literature, there is a series of works that measure the impact of decentralization on the quality of health services offered to the population. Among the most recent studies we can mentioned in this respect the studies for Spain (Cantarero and Pascual, 2008, Cantarero, 2005) and Italy (Giannoni and Hitiris, 2002).

This paper will try to measure the effects of decentralization on the quality of medical services in Romania and compare results with those of the Spain, presented in (Cantarero and Pascual, 2008). The model presented here attempt to determine whether decentralization changed substantially in quality of life. The results should be viewed with reservation because the process of decentralization is still being implemented.

2. Characteristics of the public health system

In the following we present a number of features of public health system related to the evolution of the number of employees and the average wage. The indices that we've computed are aimed to analyze the evolution of the system in the last two election cycles, and a comparison of this system with other sectors. Table 1 presents some characteristics of the dynamics of health care

[***] Professor, “Academy of Economic Studies”, Bucharest , (e-mail: andreitudorel@yahoo.com).

personnel and social welfare and other sectors. Tables 2 and 3 show the ratio of nominal wage in the health sector compared with other sectors.

Table 1. The changes in the number of personnel on electoral cycles

	Total change in absolute values (people)						
	Total	Agriculture	Industry	Construction	Commerce	Health	Public administration and defence
2000-2004	-154189	-40045	-131828	6385	29360	1164	6544
2004-2007	416482	-15267	-126308	83102	200937	37045	43492
	Average annual rate in each period (%)						
2000-2004	0.84	-7.68	-1.81	0.50	1.27	0.09	1.09
2004-2007	3.01	-5.03	-2.48	7.94	10.15	3.88	8.62

Data Source: own calculations based on data from INSSE, Bucharest, 2008

Table 2. Average nominal wages in the health system compared with other sectors of the economy

	2000	2001	2002	2003	2004	2005	2006	2007	2008
Total	0.83	0.87	0.84	0.85	0.87	0.91	0.95	0.91	0.95
Agriculture	1.99	1.95	1.87	2.02	1.90	2.37	2.66	2.79	2.32
Industry	0.79	0.84	0.83	0.85	0.87	0.92	1.00	0.97	0.99
Construction	0.95	1.00	0.98	0.97	0.99	1.08	1.16	1.08	1.12
Education	0.86	0.91	0.84	0.87	0.80	0.82	0.77	0.81	0.89
Public administration	0.58	0.63	0.62	0.60	0.62	0.58	0.52	0.47	0.60

Table 3. Average hourly earnings

	Total	Agriculture	Industry	Construction	Education	Public administration	Public health
Average hourly wage (lei/hour)	7.80	5.52	8.29	6.43	8.43	12.63	7.42
Ratio Health other sectors	1.05	0.74	1.12	0.87	1.14	1.70	1.00

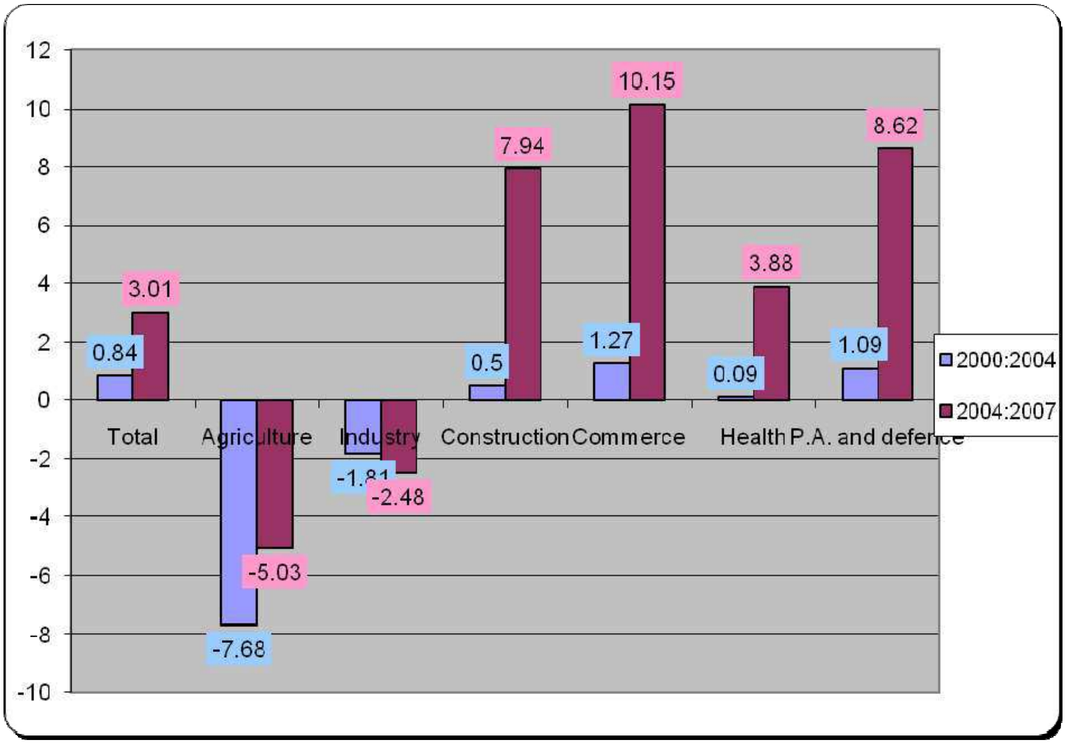

Figure 1. The annual average growth rhythms during the two election cycles

The above results show the following:

- The annual average growth in the health sector and social welfare in both periods was higher than the annual average growth in the economy.
- Personnel gains were significantly higher in public administration and defense in relation to that of health.
- Wages in this sector are below the national average salary and much lower than in administration and education. Thus, the average wage in this sector accounts for the year 2008 to only 60% of the average wage in the public administration.
- Average hourly earnings in the health system are at the level of the Average hourly earnings in the whole economy, but far below that of public administration and defense.
- The average real wages in the public health and social care system throughout the period 2000-2008 was lower than that of the public administration and defense (see Figure 2). The difference was reduced significantly in September 2008.

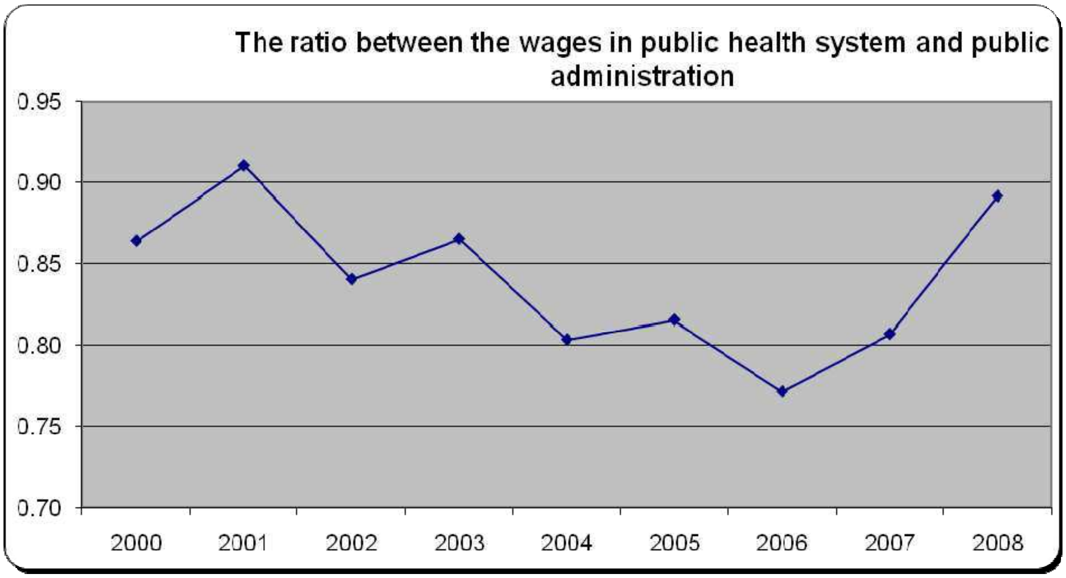

Figure 2. The evolution of ration between the real wages (1999 prices) in public health and social assistance and public administration in 2000-2004

3. Overview of the panel data models

Panel data are defined in relation to two dimensions. Thus, in this representation a variable is registered for each statistical unit in the population and for each time period from a given horizon. These types of data, take into account simultaneously both temporal dimension and the spatial dimension. There are a lot of examples of panel data that can be used in the analysis of economic and social regional level. In this paper we use panel data to highlight the effects of decentralization on some demographic characteristics that characterize the quality of life: average life expectancy and infant mortality.

We present a first example of using a panel data in economic analysis. In this paper we will estimate a model that will be used to analyze the impact of decentralization on public health system. We will use panel data.

This example is the assessment of investment issues in developing regions. In this approach, two aspects are important: defining the list of variables used in assessing the evolution of investment issues and their impact on regional economic and social development and defining the time horizon the series data are recorded. Thus, to highlight the factors that have allowed foreign investments in the developing regions in the last 20 years we mention the following categories of variables: the potential of the development area (number of enterprises with foreign capital participation per 100,000 inhabitants, the share of population employed in secondary sector, the share of population employed in agriculture, the share of population employed in services), foreign investment and fixed capital formation, labor market characteristics of the developing region (labor market pressures measured by each development region's unemployment rate, the efficient use of labor measured by the average labor productivity growth in the region, etc.) governmental decisions that are entered into the model by dummy variables which marks the emergence of a government act regulating the labor force (laws, retirement, layoff etc.), the demand at the development region level (the region's population share in total population, population density, etc.) living facilities in the developing region (the region's infrastructure, education system characteristics, the quality of the regional network of hospitals, etc.) features of the privatization process in the region (the number of units privatized by region and by year and the number of privatizations with foreign capital, etc.).

For the variables used in the model there are data series with double dimension: regional and temporal. Data series are available at regional level in publications edited by the National Institute of Statistics.

Each series of panel data recorded at the regions level is defined by:

$$x_{it}, i = 1,...,8, t = 1,...,T$$

The data series used in the analysis are usually recorded in quarters or years. These are taken directly from official statistics and are obtained by performing statistical calculations.

If the variable Y is considered endogenous and exogenous variables are $X_1,..., X_p$, the panel data model is defined based on the relation below:

$$A_p(L)y_{it} = b_{0it} + C'_{it}X_{it} + \varepsilon_{it}, i = 1,..., R, t = 1,...,T \quad [1]$$

where $A_p(L) = 1 - a_{i1}L - ... - a_{ip}L^p$ is a polinomial of p degree, C_{it} is a column vector with a value equal to the number of exogenous variables, X_{it} is the vector of exogenous variables, and ε_{it} are the residual variables of the model that are realizations of white noises. In the above relationship R represents the number of statistical units and T the number of time units (years, quarters or months).

Relative to the introduction of the endogenous variables a panel data model can be:

1.static, in this case $A_p(L) = 1$.

2.dynamic, such as the model defined by relationship [1].

Usually, in the above introduced model it is considered that its parameters are constant over time. Under these conditions the static form of the model is defined via the linear relation:

$$y_{it} = b_{0i} + C'X_{it} + \varepsilon_{it}, i = 1,..., R, t = 1,...,T \quad [2]$$

To estimate the parameters we considered some particular cases of the above model. The data series used to estimate the parameters are of the form:

$$(y_{it}, x_{1it},..., x_{mit}, i = 1,..., R, t = 1,...,T).$$

4. The advantages of the panel data models

We present in the following several advantages of panel data models.

- **Panel data include a variety of information.**

Panel data series are composed of a large number of values. At the level of statistical regions there are recorded various statistical information. The main issues raised by the use of these data are related by the usage of the same methodology of calculation and elimination of slope or level breaks. For these reasons, for a great variation in the data series we encourage using only data from 1998. It should be noted that the series of data that will be used to estimate parameters of econometric models can be annual or quarterly.

- **Since panel data have a double index (individual and temporal), they allow to analyze the dynamics and homogeneity of the statistical units.**

At the same time we must take into account that the heterogeneity of the statistical units has two components: an observable one, which is evidenced in the regression model by C_{it} parameters corresponding to the explanatory variables in the model, and another unobservable one that is not controlled based on recorded factors.

For example, using panel data we can reveal the effects of decentralization of public health care system on the quality of life. In this analysis we must consider that the life level in a region is determined by two factors: observable factors through a series of data recorded in official statistics, and unobservable factors such as the cultural model of the population of each region, etc.

Depending on the assumptions made on unobservable heterogeneity we can define the following types of panel data models:

- The model with common constant which is defined on the basis of the following relationship:

$$y_{it} = a + c_1 x_{1it} + \ldots + c_m x_{mit} + \varepsilon_{it}, i = 1,\ldots,R, t = 1,\ldots,T \quad [3]$$

Basically this is a classic regression model estimated using data series defined without taking into account the sharing of statistical units into groups. In this case, each set of data includes a number of $R \cdot T$ values.

- Fixed effect model is defined via the linear application**:**

$$y_{it} = a + a_i + c_1 x_{1it} + \ldots + c_m x_{mit} + u_{it}, i = 1,\ldots,R, t = 1,\ldots,T \quad [4]$$

The term a_i is called the individual specific effect and highlights the endogenous characteristic value that is determined by those factors acting locally.

- The variable effect model is represented by linear application:

$$y_{it} = a + c_1 x_{1it} + \ldots + c_m x_{mit} + (v_i + u_{it}), i = 1,\ldots,R, t = 1,\ldots,T \quad [5]$$

where v_i is a random variable of zero mean and standard deviation σ_v.

• **The form of the panel data allows the analysis of variance on three components: inter-individual factors, inter-temporal factors and intra-individual-temporal factors.**

In the following table we consider the statistical data for a variable (e.g. the number of of employees in agriculture) in developing regions of Romania for the period 1998-2008.

Table 4. Panel data for variable X

	Region								Average on years
	1	2	3	4	5	6	7	8	
1998	x_{11}	x_{12}	x_{13}	x_{14}	x_{15}	x_{16}	x_{17}	x_{18}	$x_{1\bullet}$
	…	…	…	…	…	…	…	…	…
2008	x_{111}	x_{112}	x_{113}	x_{114}	x_{115}	x_{116}	x_{117}	x_{118}	$x_{11\bullet}$
Average on regions	$x_{\bullet 1}$	$x_{\bullet 2}$	$x_{\bullet 3}$	$x_{\bullet 4}$	$x_{\bullet 5}$	$x_{\bullet 6}$	$x_{\bullet 7}$	$x_{\bullet 8}$	$x_{\bullet\bullet}$

For data in the table above we calculated three types of average: in each region, on each year and at national level throughout the whole period. In these circumstances, the total dispersion of the data series is broken down into three components:

Total variance = inter-individual variance + variance inter-temporal + intra-individual-temporal variance

The variances occurring in the above relationship are calculated as follows:

- inter-individual variance is calculated on the basis of the averages at regional development level;
- inter-temporal variance is based on averages for each period of time;
- intra-individual-temporal variance is based on reporting each value in the series with respect to the the average level of the region, and annual average level of the average of all values (for all statistical units in the entire time horizon).

If the panel data are defined by $(x_{it}, i = 1,...,R, t = 1,...,T)$, then the variance decomposition is represented by the following relationship:

$$\sum_{i=1}^{R}\sum_{t=1}^{T}(x_{it} - \bar{x}_{\bullet\bullet})^2 = T\sum_{i=1}^{R}(\bar{x}_{i\bullet} - \bar{x}_{\bullet\bullet})^2 + R\sum_{t=1}^{T}(\bar{x}_{\bullet t} - \bar{x}_{\bullet\bullet})^2 + \sum_{i=1}^{R}\sum_{t=1}^{T}(x_{ij} - \bar{x}_{i\bullet} - \bar{x}_{\bullet t} + \bar{x}_{\bullet\bullet})^2 \quad [6]$$

- **In general, panel data consist of a large number of values.**

Under these conditions we witness a rise in the number of degrees of freedom and improved quality of tests used to estimate the parameters and to verify the statistical assumptions.

The main disadvantage of using panel data is related to the effects of the errors propagated by them. In this sense we must take into account Huber's comment which showed that 3% error in the panel data causes significant changes in the estimates. For this reason it must be have developed techniques to detect and eliminate possible aberrant values in the data series. In the event of a aberrant value in the data series, it is recommended either eliminating or correcting it by the interpolation operation.

Not a few times, from statistical data sources several statistical records are missing, but these can be approximated by applying the interpolation operator. In relation to lack of data from the panel data we can identify two categories of panel data: balanced data sets, if panel data are registered for all statistical units and for all time periods, and unbalanced data sets case in which on certain positions the appropriate values are not recorded.

5. Application

To estimate the impact of decentralization of public health, on some of the demographic indicators we used data sets recorded for the statistical indicators of the eight regions of economic development for the period 1998-2005.

To analyze the decentralization process in Romania, we define the following model:

$$\log(H_{it}) = a + b \cdot PIB\_R_{it} + c \cdot \log(M\_P_{it}) + d \cdot P\_P_{it} + \varepsilon_{it},\ i = \overline{1,8},\ t = \overline{1998,2005}$$

In this model we used the following variables: H_{it} a global indicator of performance used to characterize the public health system in a region, for a year. To assess this issue we appeal to the infant mortality rate, PIB_R per capita in the region, P_P the number of beds per 1,000 inhabitants and M_P the number of doctors per 1,000 inhabitants.

To estimate the parameters of the model we used statistics from the eight development regions of Romania. The data are annual and have been reported for the period 1998-2008.

Parameter estimation was done for the classical model, the fixed effect model and the variable effects model. For estimation we used the least squares method (OLS) and two stage least square method (Baltagi, 2008). The results are presented in Table 5.

For the model with fixed effects we test the null hypothesis according to which specific effects are negligible in the regions. In this case, we use a Fisher-type test. Statistical values calculated for the case in which the parameters are estimated by OLS and TSLS are higher than F values determined from the distribution table, which signals that the specific effects of the regions are significant. The results are obvious, if we consider that there are large disparities between the eight development regions in terms of economic and social development.

Table 5. Characteristics of the model with dependent variable $\log(RMI)$

	Clasic Model		Fixed effects		Random effects	
	LS	TSLS	LS	TSLS	LS	TSLS
C (coeficient) t-Statistic	0.7619 (1.84)	0.8310 (1.69)	1.7222 (3.46)	1.3773 (2.10)	1.0985 (2.61)	0.8768 (1.83)
Log(PIB_R) (coeficient) t-Statistic	-0.3424 (-5.47)	- 0.3314 (-4.34)	-0.3311 (-4.83)	- 0.3477 (-4.53)	-0.3118 (-4.98)	- 0.3554 (-4.80)
P_SSA (coeficient) t-Statistic	0.8620 (1.42)	0.5591 (0.86)	0.1015 (0.06)	2.6871 (0.78)	0.4989 (0.51)	- 0.0294 (-0.03)
log(M_P) (coeficient) t-Statistic	-0.1873 (-1.84)	- 0.2405 (-2,16)	-0.6558 (-4.06)	- 0.6895 (-3.85)	-0.2059 (-2.11)	- 0.1697 (-1.55)
P_P (coeficient) t-Statistic	-0.0050 (-0.31)	- 0.0054 (-0.30)	-0.0683 (-3.20)	- 0.0763 (-3.24)	-0.0169 (-1.11)	- 0.0184 (-1.16)
R-squared	0.8270	0.8248	0.9116	0.9070	0.6404	0.6357
F-statistic and Prob (F)	60.94 (0.0000)		41.24 (0.0000)		22.71 (0.0000)	
F-statistic (Clasic Model vs Fixed effects) and Prob (F)	6.56 (0.000)	6.06 (0.000)				
Hausman statistic and Prob (Hausman)			15.443 (0.000)	26.004 (0.000)		

The results presented in the above table shows that the infant mortality is negatively correlated with variables PIB_R, M_P and P_P and P_SSA variable has no significant influence on it. To test the orthogonality of random effect and explanatory variables we used a Hausman test (Baltaga, 2008). The values determined for both statistics, entitled to use fixed effects models to estimate the impact of decentralization on the quality of life in developing regions, to the detriment of the other two models.

6. Conclusions

The usage of the panel data has a number of advantages mainly related to the availability of data sets to estimate the parameters. But bear in mind that relatively small errors in the statistical data can determine significant errors in the estimated parameters.

The results from the above table show that the infant mortality rate is negatively correlated with the variable that characterizes the region's economic development, the number of beds and number of doctors per 1,000 inhabitants. Note that the regions specific factors have a significant influence on child mortality. The estimated model shows a lack of positive effects of decentralization on the outputs of public health.

The results should be viewed with reservation for the following reasons: to estimate parameters of data sets we used data series during 1998-2008 time period when the economy and the public health system had major changes from one period to another, the health reform is being implemented; in general the effects of decentralization process is difficult to see in a short period of time.

A more detailed analysis recommends consideration of other important statistical variables to characterize the public health system and the decentralization of public administration. It is recommended to introduce some variables in the model that takes into account the cultural and economic development in each region.

To define the analysis model one could use the life expectancy as a dependent variable (endogenous) instead of the child mortality. We believe that the chosen version in this paper is more appropriate because the infant mortality rate is a characteristic more sensitive to changes in public health care system in relation to average life expectancy.

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

References

- Andrei, T., Matei, A., Stancu, I., Andrei, CL. (2009), Socioperformanța reformei sistemului public de sănătate, Editura Economică, București.
- Baltagi, B.H. (2008) Econometrics, 4th Edition, Springer, Berlin.
- Cantarero, D.and Pascual, Marta (2007) Analyzing the impact of fiscal decentralization on health outcomes: empirical evidence from Spain, *Applied Economics Letters*, 15, pp. 109-111.
- Cantarero, D. (2002) Decentralization and health care expenditure: the Spanish case, *Applied Economics Letters*, 12, pp. 963-966
- Giannoni, M., Hitiris, T. (2002) The regional impact of health care expenditure: the case of Italy, *Applied Economics,* 14, pp. 1829-1836