Abstract

Romania’s economic development requires the reconsideration of certain policies at the development regions level. Currently, there are large gaps between Romanian development regions. Moreover, certain socio-economic development regions are among the poorest from the EU. In the national statistics were recorded important data sets to characterize the economic and social phenomena at the development regions level. For their exploitation can be used panel data methods. The creation of certain coherent economic and social policies at the development regions level represents an immediate requirement for the reduction of the disparities between Romania and the EU average level.

Keywords: panel data, pooled regression model, fixed effects model, tourist demand, tourist supply.

JEL classification: C23, L83.

1. INTRODUCTION

Panel data are defined in relation with two dimensions. Thus, in this representation, a variable is registered at the level of each statistical unit of the population for each time period for a given time horizon. These types of data simultaneously take into account both the temporal dimension and the territorial one. Thus, may be presented numerous examples of panel data that can be used in the analysis of economic and social phenomena at regional level.
2. PANEL DATA MODELS APPLICATION IN THE REGIONAL ANALYSIS

If we consider the endogenous variable $Y$ and the exogenous variables $X_1, \ldots, X_p$, then the panel data model is defined as follows:

$$A_p(L)y_{it} = b_{0i} + C'X_{it} + \varepsilon_{it}, i = 1, \ldots, R, t = 1, \ldots, T$$  \hspace{1cm} (1)

Where $A_p(L) = 1 - a_1L - \ldots - a_pL^p$ is a polynomial of $p$ degree, $C_{it}$ is a column vector with a number of values equal to the one of the exogenous variables, $X_{it}$ is the exogenous variables vector, and $\varepsilon_{it}$ are the residuals, which are white noise achievements. We note by $R$ the number of statistical units and by $T$ the number of time units (years, quarters or months).

Taking into account of the manner of endogenous variables introduction, a panel data model can be:

- static, case in which $A_p(L) = 1$.
- dynamic, as the model defined by the first relation.

Usually, we consider that in the case of the previous model, their parameters are constant in time. In this case, the panel data model in the static form is defined by the following linear application:

$$y_{it} = b_{0i} + C'X_{it} + \varepsilon_{it}, i = 1, \ldots, R, t = 1, \ldots, T$$  \hspace{1cm} (2)

For the parameter estimation are considered certain particular cases of the above mentioned model. The data series used for the parameter estimation are presented as follows: $(y_{it}, x_{1it}, \ldots, x_{nit}, i = 1, \ldots, R, t = 1, \ldots, T)$.

Panel data contain a variety of information because they include a lot of values and they evidence a higher variability, mainly, in the statistical inter-units dimension. The major problems related to these data use are the application of the same calculus methodology and the elimination of the slope or level breaks. The data series used to estimate the models parameters can be annual or quarterly.

Due to the fact that panel data have double indexation (individual and temporal), they allow the analysis of the dynamics and homogeneity of the statistical units. In 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 econometric model by the $C_{it}$ parameters, which are corresponding to the exogenous variables of the model. This component is included in the model by the $C'X_{it}$ term. In this case, the panel data model parameters are independent of the statistical units or of the time;
- An unobservable one, which is not controlled on the basis of registered factors.

For example, using panel data we can evidence the fact that the size of the informal economy at regional level is determined by two categories of factors:

- Observable factors, represented by data series as: industrial production, agricultural production, the number of employees from the two sectors of activity, the number of high school and university graduates etc.;
- Unobservable factors, as the cultural model of the population of each region etc.
If we are not taking into account of the unobservable heterogeneity this fact can often lead to poor estimations of the regression model parameters defined above. There are three situations in which, if we are not taking into account of the unobservable heterogeneity, we obtain contradictory results:

- In the first case, the regression models defined at the level of the units groups differ by the intercept’s value, but they have the same slope value. The right estimated on the ensemble of population units has a slope with the same sign with the ones at the groups’ level but it significantly differs of them;

- In the second case, the differences are more significant because the estimated right on the ensemble of population tends to have a slope which doesn’t significantly differ to zero, while, at the groups’ level, the regression model defined for the two variables is practically insignificant;

- In the third case, while, at the groups’ level the variables dependence is a positive one, on the ensemble of population we obtain a negative slope for the linear regression model.

These results are obtained because the estimated overall regression (for example, the estimation based on data at national level without taking into account of the distribution on development regions) ignores the presence of unobservable heterogeneity, which represents an important factor, while the regression models estimated at the level of the statistical units groups (for example, which are taking into account that the data are obtained at the level of development regions) consider that the unobservable heterogeneity is due to the local factors that cannot be followed by registered data series. From this example it results the importance of taking into account the heterogeneity of statistical units in the parameters estimation. Depending on assumptions made on the unobservable heterogeneity we can define the following types of panel data models:

- **Pooled regression model** (the constant coefficients model), which is defined as follows:
  \[
  y_{it} = a + c_1 x_{it1} + ... + c_m x_{itm} + \varepsilon_{it}, i = 1,...,R, t = 1,...,T
  \]  
  Practically, this is a classical regression model estimated on the data series defined without taking into account on the split of the statistical units on groups. In this case, each data series includes \( R \cdot T \) values.

- **Fixed effects model**, defined by the linear application:
  \[
  y_{it} = a + a_i + c_1 x_{it1} + ... + c_m x_{itm} + u_i, i = 1,...,R, t = 1,...,T
  \]  
  The \( a_i \) term is called individual specific effect and it reveals the value of the endogenous characteristics, which is determined by those factors which are acting at local level.

- **Random effects model**, defined by the linear application:
  \[
  y_{it} = a + c_1 x_{it1} + ... + c_m x_{itm} + (v_i + u_i), i = 1,...,R, t = 1,...,T
  \]  
  where \( v_i \) is a random variable with zero mean and \( \sigma_v \) standard deviation.
3. THE APPLICATION OF PANEL DATA MODELS IN THE TOURIST SUPPLY AND DEMAND ANALYSIS

Information provided by National Institute of Statistics allows realizing analyses on key issues related to the Romanian tourist activity, both at national and regional level. The provided statistical data are registered at national, regional or at county level. Below is presented a panel data model for the analysis of tourist demand and supply using data series registered at development regions level which includes two equations defined as follows:

- For the analysis of the tourist supply, we express the total tourist activity, measured by the GDP from this sector of activity (GDP_T), according to three major factors: the region’s tourist capacity (x1), the density of public roads per 100 km² territory or of modernized public roads per 100 km² territory from the region (x2) and the development level of the region, characterized by regional GDP per capita (x3). Thus, we define the following model:

\[
GDP_T \text{it} = f(x_{1it}, x_{2it}, x_{3it}) + \varepsilon_{it} \quad (6)
\]

- For the analysis of the tourist demand, we express the total tourist activity (PIB_T) according to two factors: the average annual income of the region (x4) and the number of stayings overnight at region’s level (x5). The demand equation is defined using the following relationship:

\[
GDP_T \text{it} = f(x_{4it}, x_{5it}) + \varepsilon_{it} \quad (7)
\]

The supply and demand equations are represented using the following regression models:

\[
\log(GDP_T) = c_0 + a \log(x_2) + b \log(x_1) + c \log(x_3) + \varepsilon \quad (8)
\]

\[
\log(GDP_T) = c_0 + a \log(x_2) + b \log(x_1) + c \log(x_5) + \varepsilon \quad (9)
\]

Data series used to characterize the eight development regions were recorded during the period 2000-2006 and are expressed in 2000 prices.

- **Supply estimations**

  Based on the data series recorded for the four variables included in the (8) model, parameters were estimated using three types of panel data models: the constant coefficients model, the fixed effects model and the random effects model. For the first two methods the results are presented in equations (10) and (11). For the third method the results are not significant, and, therefore, were not presented.

  For the constant coefficients model, the obtained results are:

\[
\log(GDP_T) = -10.32 + 2.10 \log(x_2) + 0.43 \log(x_1) + 0.64 \log(x_3) \quad (10)
\]

- \(R^2 = 0.75, SSE = 2.574\). For the fixed effects model, the obtained results are:

\[
\log(GDP_T) = -10.32 + 2.10 \log(x_2) + 0.43 \log(x_1) + 0.64 \log(x_5) \quad (11)
\]

- \(R^2 = 0.94, SSE = 0.614\). In both cases, all parameters are statistically significant for a 1% significance level. By aggregating the regional specific effects for each historical province were obtained the results presented in Table 1.
The Identification of Economic Phenomena Trends at Regional Level

Table no. 1 The specific effects at the historical provinces level

<table>
<thead>
<tr>
<th>Historical Provinces</th>
<th>Development Region</th>
<th>The specific effects of the historical region</th>
</tr>
</thead>
<tbody>
<tr>
<td>Muntenia</td>
<td>SV, SM, B</td>
<td>2.772978</td>
</tr>
<tr>
<td>Moldova</td>
<td>NE, SE</td>
<td>-1.0851</td>
</tr>
<tr>
<td>Transilvania</td>
<td>V, NV, C</td>
<td>-1.68788</td>
</tr>
</tbody>
</table>

Source: Authors calculus

- Demand estimations

Based on the data series recorded for the three variables included in the (10) model, parameters were estimated using three types of panel data models: the constant coefficients model, the fixed effects model and the random effects model. For the first two methods the results are presented in equations (12) and (13). For the third method the results are not significant, and, therefore, were not presented.

For the constant coefficients model, the obtained results are:

\[
\log(GDP \_T) = -3.41 + 0.37 \log(x_3) + 0.74 \log(x_4)
\]

(12)

\[R^2 = 0.227, \text{ SSE} = 8.046.\]

For the fixed effects model, the obtained results are:

\[
\log(GDP \_T) = -1.70 + 0.29 \log(x_3) + 0.61 \log(x_4)
\]

(13)

\[R^2 = 0.944, \text{ SSE} = 0.584.\]

In both cases, all parameters are statistically significant for a 1% significance level.

Specific effects estimated for the tourist demand using the specific effects obtained for the historical provinces are presented in Table 2.

Table no. 2 The specific effects at the historical provinces level

<table>
<thead>
<tr>
<th>Historical Provinces</th>
<th>Development Region</th>
<th>The specific effects of the historical region</th>
</tr>
</thead>
<tbody>
<tr>
<td>Muntenia</td>
<td>SV, SM, B</td>
<td>0.14834</td>
</tr>
<tr>
<td>Moldova</td>
<td>NE, SE</td>
<td>-0.1011</td>
</tr>
<tr>
<td>Transilvania</td>
<td>V, NV, C</td>
<td>-0.0472</td>
</tr>
</tbody>
</table>

Source: Authors calculus

To determine whether specific effects at development regions level are statistically significant an F test is applied. The test’s statistics is expressed by the following relationship:

\[
F = \frac{(SSE_c - SSE_f)/(n-1)}{SSE_f/(nT - n - k)} 
\]

\[F(n-1,nT-n-k) \quad (14)\]

Where: \(SSE_c\) - sum of squared errors for the constant coefficients model, \(SSE_f\) - sum of squared errors for the fixed effects model, \(n\) - number of regions (equal to 8 in this case), \(T\) - number of years (equal to 7 in this case) and \(k\) - number of model’s explanatory variables (three for the first model and two for the second model). The values of the test’s statistics
for the two models are: 20.5 for tourist supply and 84.0 for tourist demand. In both cases is chosen the fixed effects model.

4. CONCLUSIONS

The above results highlight that in the case of the specific effects for the development regions estimated for both tourist supply and demand, the adequate model is the fixed effects model.

Acknowledgements
This work was supported by CNCSIS - UEFISCSU, project number PNII – IDEI code 1793/2008, financing contract no. 862/2009.

References