DIDACTICS
OF
MATHEMATICS

10(14)

The Publishing House
of Wrocław University of Economics
Wrocław 2013

Reviewers cooperating with the journal
Giovanna Carcano, Igor Dubina, Salvatore Federico,
Marian Matłoka, Włodzimierz Odyniec, Anatol Pilawski,
Tadeusz Stanisz, Achille Vernizzi, Henryk Zawadzki

Copy-editing
Elżbieta Macauley, Tim Macauley

Proof reading
Barbara Cibis

Typesetting
Elżbieta Szlachcic

Cover design
Robert Mazurczyk

Front cover painting: W. Tank, Sower
(private collection)

This publication is available at: www. journal.ue.wroc.pl and www.ibuk.pl
Information of submitting and reviewing paper is available on
the Publishing House's website www.wydawnictwo.ue.wroc.pl

PL ISSN 1733-7941

The original version: printed

Printing: Printing House EXPOL,
P. Rybiński, J. Dąbek, sp.j. ul. Brzeska 4, 87-800 Włocławek

TABLE OF CONTENTS

DIDACTICS OF MATHEMATICS

No. 10 (14) 2013

SPATIAL ANALYSIS OF LEARNING RESULTS IN HIGH SCHOOL MATHEMATICS AND POLISH BY COUNTY

Wiktor Ejsmont, Janusz Łyko

Abstract. One way to assess the quality of the educational activities of schools is to analyze the educational value-added, with the help of which it is possible to measure the gain in students' knowledge that takes place at various stages of education. This is an objective measurement that takes into account the knowledge with which the student begins the next stage of learning. Access to data on the final results of tests at every stage of education enables the assessment of the quality of education in schools throughout Poland. The article aims to analyze these results and attempts to show the spatial dependence of the results obtained.

Keywords: educational value added, random effects model, taxonomy, development pattern, panel data.

DOI: 10.15611/dm.2013.10.02

1. Introduction

The advanced economies have been recently seeing fewer and fewer jobs that are directly connected with the manufacturing sector. Wages in the manufacturing sector are significantly lower than those in other sectors. The current transformations occurring in the labour market have resulted in the increased significance of white-collar workers who have been determining the economic success. This category includes a lot of workers, beginning with so-called call centre jobs, then architects, teachers, scientists and concluding with those employed in the financial sector. The benefits from a knowledge-based economy depend primarily on human capital, i.e. education, proficiency, the talent and skills of workers. Therefore various countries have been investing more and more in "people". This can be achieved

Wiktor Ejsmont, Janusz Łyko
Department of Mathematicsand Cybernetics, Wrocław University of Economics, Komandorska Street 118/120, 53-345 Wrocław, Poland.
E-mail: wiktor.ejsmont@gmail.com; janusz.lyko@ue.wroc.pl

by focusing on education and instruction, and these activities are now seen as increasingly more significant determining factors stimulating economic growth.

The issue of development and education is not brand new in the world literature. Among other researchers, Chabbott and Ramirez (2006) deal with it. They aim at explaining the benefits for the economy resulting from education. Research seeking to examine the influence of the education's quality on the economy is also broadly advanced. Mankiw, Romer and Weil (1992) extended the Solow model in their empirical analysis of economic growth. Their contribution consisted in including the process of human capital accumulation as regards school attendance. In a sense, the paper was a breakthrough because it was noticed that the addition of this factor to the model can totally change the understanding of what a growth process essentially is. It was demonstrated that the contribution of physical capital in income creation is nearly 0.31, whereas that of human capital is roughly 0.28. Bils and Klenow (2000) and Temple (2001) followed next in a similar vein.

This paper aims at investigating the relationship between the development of individual counties in Poland and the educational value added for students leaving high schools. Our research is empirical and makes use of two methods. The first one is the method put forward by Hellwig (1968) and used to measure the development of counties; the second one is based on the random effects model that was used to measure incremental knowledge.

2. Data description

The research covered 375 counties in Poland. The total number of counties is 379, however, the data from four counties were not complete and therefore they were excluded from the analysis. Data used to calculate the index of social and economic development were from 2010 and are available on the website of the Central Statistical Office. Due to the formal criteria when selecting variables, only measurable variables were chosen to calculate an index of counties' development. The selection process of variables to create the index of counties' development followed the literature, mostly Rosner (1999, 2002, 2007).

Table 1 presents the variables included into a development index. Preliminary analysis of empirical data covered more variables, however, some of them were excluded because of their inter-correlations and low variability. The thresholds applied for exclusion were 0.5 in the case of correlation and 0.1 in the case of variability.

Table 1. Variables selected to create the index of counties' development

Class of variables	Variable
Economic	Unemployment rate
	Wages
	Number of REGON entities per 10 thousand people
	Number of cars per 1 inhabitant
	Percentage of injured in accidents at work
	Budgetary income of self-government entities per 1 inhabitant
	Pre-working age population rate
	Investment expenditure by firms and gross fixed assets per 1 inhabitant
Access to the Internet at high schools	Number of students per 1 computer at high schools
	Percentage of high schools with computers
Libraries	Population per 1 public library
	Number of books at public libraries per 1 thousand people
	Readership of public libraries per 1 thousand people
	Lending transactions per 1 borrower
Environment	Emission of gas pollutants per 1 square kilometre
	Waste created per 1 square kilometre
	Industrial sewage treated per 100 square kilometres
Other	Number of deaths per 1 thousand people
	Number of community sports clubs per 1 thousand people

Source: authors' own.

The excluded variables that were strongly correlated with other variables or too stable, were e.g. net migration rate, participation rate, gross enrolment ratio, air pollutants emitted per 1 square kilometre, ratio of sewage requiring treatment, expenditure for education by local self-government entities per 1 inhabitant, deaths before age 1, joint participation rate of population at working and post-working age, rate of population with access to sewage treatment, gross fixed assets at enterprises per 1 inhabitant.

3. Results of students leaving high school

Table 2 presents the results of more than 95 per cent of students leaving Polish high schools in 2010, aggregated at a voivodeship level. The entire population could not be included because some students could not be identified in order to match their final examination scores at gymnasium and at high school.

Table 2. Average results of ‘gimnazjum’(secondary school) and maturity exams in basic Polish and basic mathematics, 2010

Voivodeship	Number of students	Number of counties	Polish		Mathematics	
			Average score in ‘gimnazjum’ exam (humanities)	Average score in ‘matura’ exam	Average score in ‘gimnazjum’ exam (science)	Average score in ‘matura’ exam
Dolnośląskie	14,194	29	75.92	63.99	62.87	66.47
Kujawsko--pomorskie	8,904	23	75.90	65.78	63.82	70.39
Lublin	15,301	24	75.66	61.01	60.26	64.81
Lubusz	4,931	14	74.88	63.03	60.28	67.14
Łódź	14,522	22	73.63	62.92	61.92	66.68
Małopolskie	19,910	22	77.22	64.67	63.52	67.67
Mazowieckie	30,621	42	77.23	62.52	64.42	68.41
Opolskie	5,021	12	75.22	63.80	61.77	66.39
Podkarpackie	13,231	25	75.92	62.26	60.65	66.04
Podlaskie	7,705	16	74.61	64.70	62.95	67.77
Pomeranian	10,114	19	75.18	66.49	64.92	69.49
Silesian	23,462	36	75.20	65.80	61.29	67.35
Świętokrzyskie	7,469	14	74.98	62.74	61.06	68.31
Warmińsko--mazurskie	7,403	21	74.37	63.03	63.27	68.09
Wielkopolskie	18,091	35	75.60	62.36	61.48	68.14
Zachodniopomorskie	8,109	21	75.21	61.86	60.69	66.26
Total	208,988	375	75.71	63.53	62.38	67.47

Source: own calculations based on data from the Central Examination Board in Warsaw (2010).

The analysis was conducted with two types of data. The first one includes the results in humanities (‘gimnazjum’ i.e. secondary school final exam) and in basic Polish (‘matura’ i.e. final high school exam), whereas the second one includes the results in science (‘gimnazjum’ exam) and in basic mathematics (‘matura’ exam). Taking into account the basic parts of examinations allows for minimizing the risk connected with side effects that do not depend on schools but have an impact on teaching effectiveness. These factors are principally additional private lessons increasingly attended by students. Therefore taking into account the obligatory (basic) examinations reduces the chance to unfairly evaluate a given school. The Public Opinion Research Centre estimates, based on their 2010 research, that nearly half of Polish parents pay for additional private lessons of their children. Private English lessons are taken by almost 22 per cent of those attending additional lessons, with mathematics ranked second – 17.3 per cent of all privately tutored.

4. Hellwig’s taxonomic measure of development

The variables under study allow us to classify counties with respect of their levels of social and economic development using the taxonomic measure of development put forward by Hellwig (1968); see also Nowak (1990). It is a widely accepted method based on calculating the distance of an object under study from a theoretical pattern of development. Using Hellwig’s development pattern we can order all counties P_j where $j \in \{1,\ldots, k = 375\}$. Each county is described by means of the diagnostic variables presented in Table 1. Each variable can be characterized as either a stimulant or anti stimulant. The description can be represented as a matrix

$$P = \begin{bmatrix} p_{1,1} & p_{1,2} & \cdots & p_{1,m} \\ p_{2,1} & p_{2,2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ p_{k,1} & p_{k,2} & \cdots & p_{k,m} \end{bmatrix} \tag{1}$$

where $p_{j,t}$ denotes the value of the t^{th} variable describing the j^{th} county. Next, the variables are standardized

$$z_{j,t} = \frac{p_{j,t} - \overline{p}_t}{S_t} \tag{2}$$

where $\overline{p}_t, S_t$ are mean and standard deviation, respectively.

A matrix of normalized variables is denoted by $Z = [z_{j,t}]_{k \times m}$. Let Z_t denote the t^{th} normalized variable. Then, a so-called development pattern is an abstract object P_0 with standardized coordinates $z_{0,1},...,z_{0,m}$, where $z_{0,t} = \max_j\{z_{j,t}\}$ if Z_t is a stimulant and $z_{0,t} = \min_j\{z_{j,t}\}$ if Z_t is anti-stimulant. Then, a distance of each county from the pattern is computed according to the formula

$$d_j = 1 - \frac{D_{j,0}}{D_0} \tag{3}$$

where

$$D_{j,0} = \sqrt{\sum_{t=1}^{m}(z_{j,t} - z_{0,t})^2},$$

$$D_0 = \bar{D}_0 + 2S_0,$$

$$\bar{D}_0 = \tfrac{1}{n_j}\sum_{t=1}^{m} D_{t,0},$$

$$S_0 = \sqrt{\tfrac{1}{n_j}\sum_{t=1}^{m}(D_{t,0} - \bar{D}_0)^2}.$$

In this way, combined indices are obtained for each county. The taxonomic measure takes values from the [0, 1] interval. The closer the values of a given county's attribute to the pattern, the higher the level of its development, while more distance from the pattern points to a lower level of the county's development. Typically, we introduce the three groups when classifying objects by means of development pattern:

- Class A – a high level of development, if $d_j > \bar{d} + 2S_d$.
- Class B – an average level of development, if

  $$\bar{d} - 2S_d < d_j \le \bar{d} + 2S_d.$$

- Class C – a low level of development, if $d_j \le \bar{d} - 2S_d$,

where $\bar{d}$ and S_d is the mean and standard deviation of d_j, respectively.

Table 3 shows the number of counties in respective development classes identified by means of Hellwig's the development measure depending on the voivodeship. The very good position of the Silesian voivodeship is worth emphasizing where nearly 33 per cent of counties are placed in class A of the highest development level and none are placed in the lowest class C.

Table 3. Number of counties in class A, B and C, by voivodeship

Voivodeship	A	B	C
Dolnośląskie	7	19	3
Kujawsko-pomorskie	1	19	3
Lublin	3	11	10
Lubusz		13	1
Łódź	3	17	2
Małopolskie	4	16	2
Mazowieckie	5	30	7
Opolskie	2	7	3
Podkarpackie	5	16	4
Podlaskie	2	12	2
Pomorskie	3	16	
Śląskie	12	23	1
Świętokrzyskie	2	7	5
Warmińsko-mazurskie	1	16	4
Wielkopolskie	3	32	
Zachodniopomorskie	3	15	3
Total	56	269	50

Source: own Excel and R-project calculations based on data from the Central Statistical Office (2010).

One can also detect that the lowest development is mostly characteristic for counties from poorest vvoivodeships such as Lublin, Świętokrzyskie, Warmińsko-mazurskie and Opolskie.

5. Random effects model

When building a random effects model, the following denotations will be used:

x_{ij} – 'gimnazjum' score of the i^{th} student in the j^{th} county (input),

y_{ij} – 'matura' score of the i^{th} student in the j^{th} county (output),

n_j – number of students in the j^{th} county,

n – number of all students, i.e. $n = n_1 + \ldots + n_k$,

k – number of counties , i.e. $j \in \{1,\ldots, k = 375\}$ (same as in the case of development measure),

$\bar{x}$ – average 'gimnazjum' score of all students,

$\bar{y}$ – average 'matura' score of all students,

$\bar{x}_j = \frac{1}{n_j}\sum_{i=1}^{n_j} x_{ij}$, $\bar{y}_j = \frac{1}{n_j}\sum_{i=1}^{n_j} y_{ij}$ – average 'gimnazjum' and 'matura' scores in the j^{th} county.

As a result, we get unbalanced panel data since the number n_j of observations for individual counties can vary. When modelling an inhomogeneous population, one has to introduce inhomogeneity into the model. As regards data under study, there may be various relations between output and input variables for respective counties. The econometric literature refers to this model as the unbalanced one-way error component model with random effects, e.g. Baltagi (2005). The random effects model is also known as a variance components model (VC), e.g. Maddala (2006). Wansbeek and Kapteyn (1982a, 1982b) first introduced this model. The model has the form:

$$y_{ij} = a + bx_{ij} + \xi_j + e_{ij} \tag{4}$$

where e_{ij} is a random variable following a normal distribution $N(0,\sigma^2)$, whereas ξ_j follows $N(0,\sigma_I^2)$. In addition, it is assumed that random components from different entities and different patients are uncorrelated and that individual random term ξ_j is uncorrelated with random term e_{ij}, i.e. $E(\xi_j, e_{is}) = 0$ for $j \neq s$.

It follows from the form of the model that ξ_j represents a deviation of the average score for the j^{th} county from the average score of the entire population. This average score in Figure 1 is shown as a dotted line while a solid line illustrates the average score of the entire population. If ξ_j is positive then one may argue that students from the j^{th} county improved their average score with respect to the average score of the entire population, or students from the whole of Poland, while a negative ξ_j indicates the deterioration of the j^{th} county's score compared to the average score of the population. Therefore the value of the parameter ξ_j is called the value added or operational effectiveness of the object under evaluation. The value e_{ij}, on the other hand, represents the deviation of the individual student's score from the average score of the j^{th} county.

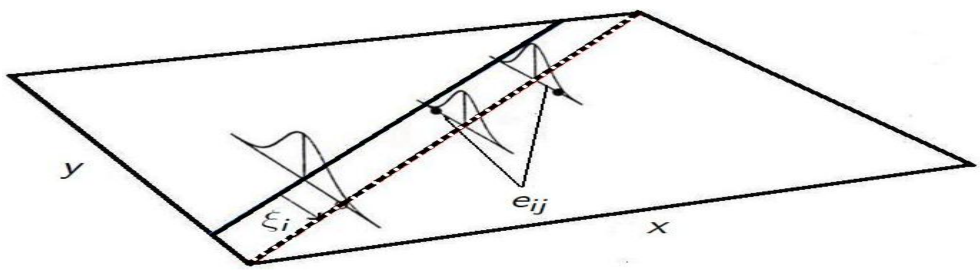

Fig. 1. The idea of measurement in a random effects model

Source: authors' own, based on Skrondal, Rabe-Hesketh (2008).

The above defined model is estimated by means of the maximum likelihood method (Aitkin, Longford 1986). The formulas for the estimates are given by Baltagi (2005) and Ejsmont (2009), where also a complete algorithm of estimating variance components σ^2 and σ_I^2 is provided. The obtained random effects are tested for significance by means of the Breusch-Pagan test (e.g. Hasio 1999; Baltagi 2005).

In order to estimate the value of ξ_j, one can use the mean squared error theorem (Jakubowski, Sztencel 2004, p. 135). Since both terms σ^2 and σ_I^2 are available before the model estimation, thus one can use them as the *a priori* information. Next, one determines a conditional distribution of random variable ξ_j given $\bar{y}_j$. The mean of the j^{th} county has the form:

$$\bar{y}_j = a + b\bar{x}_j + \xi_j + \bar{e}_j \tag{5}$$

and under appropriate assumptions it is distributed as

$$N\left(a + b\bar{x}_j, \sigma_I^2 + \frac{\sigma^2}{n_j}\right).$$

Since ξ_j follows $N(0,\sigma_I^2)$, thus the conditional distribution

$$f\left(\frac{\xi_j}{\bar{y}_j}\right)$$

is normal and has the form:

$$N\left(\rho n_j^*(\bar{y}_j - a - b\bar{x}_j),\ n_j^*(1-\rho)\frac{\sigma_I^2}{n_j}\right) \tag{6}$$

where

$$\rho = cor(y_{ij}, y_{pj}) = \frac{\sigma_I^2}{\sigma_I^2 + \sigma^2},$$

$$n_j^* = \frac{w_j}{1-\rho}$$

and

$$w_j = \frac{n_j\sigma^2}{\sigma^2 + n_j\sigma_I^2}.$$

Therefore, comparing the development of individual counties involves comparing the mean values from the conditional distribution, i.e.

$$e_j = \hat{\rho} n_j^*(\bar{y}_j - \hat{a} - \hat{b}\bar{x}_j) \tag{7}$$

where $\hat{\rho}, \hat{a}, \hat{b}$ are estimates of: correlation *ρ*, intercept *a* and slope *b*, respectively.

Table 4. Basic statistical characteristics of the random effects model

Characteristic	Subject	
	Polish	Mathematics
Pearson's correlation coefficient	0.505	0.714
Variance of the disturbance term – $\hat{\sigma}^2$	165.019	205.826
Between-counties variance – $\hat{\sigma}_I^2$	9.952	6.264
Normality – *p*-value	>0.01	>0.01
Coefficient $\hat{b}$	0.585	0.758
Coefficient $\hat{a}$	18.994	20.230
LM test – *p*-value	<0.01	<0.01

Source: own Excel and R-project calculations based on data from the Central Examination Board (2010).

Table 4 presents the main statistical characteristics of the estimated random effects models. The models obtained are well-fitted as regards the normality of residuals. The estimated values of the *LM* test indicate that σ_I^2 is statistically significant at a level of 0.01, therefore applying random effects connected with σ_I^2 is justified. 'Gimnazjum' and 'matura' exam scores are correlated in the case of humanities at the level exceeding 0.5 (correlations were calculated with all data, regardless of school). In the case of mathematics, the correlations are significantly stronger. The teaching of science also demonstrates a faster pace of knowledge increase indicated by slope coefficients $\hat{b}$. The variance estimating diversification within counties was considerably bigger for mathematics, i.e. teaching mathematics is more difficult than teaching Polish.

6. Conclusion

The paper aimed at dividing Poland into regions with respect to the level of social and economic development in order to compare them with their educational value added (EVA). We decided to classify counties into eight groups based on the development index (3) and thus to obtain a better illustration of the relation between EVA and the development than in the case of just three categories. Each group included roughly the same number of observations. The first group consisted of forty-six counties in increasing order, while the remaining groups consisted of forty-seven counties each, totalling 375 counties. In each group, the average EVA was calculated for humanities and science. The results are presented in Figure 2.

An increasing trend of the relation between EVA and the development index is evident in the case of mathematics, whereas the case of Polish is not so clear. The markedly best and worst educational results occur in two border groups, i.e. [0 – 324) and [1272 – 2545). This implies that the more developed regions educate best, while the least developed regions educate inadequately. Certainly such circumstances can follow from teachers' and parents' higher wages, especially as parents can afford to pay for additional private lessons for their children. Most students are classified in group B, i.e. 95 per cent of the total, except for two border groups, [0 – 324) and [1272 – 2545). This result may indicate a certain imperfection of development measure that is not as powerful as in the case of outlying observations when discriminating objects.

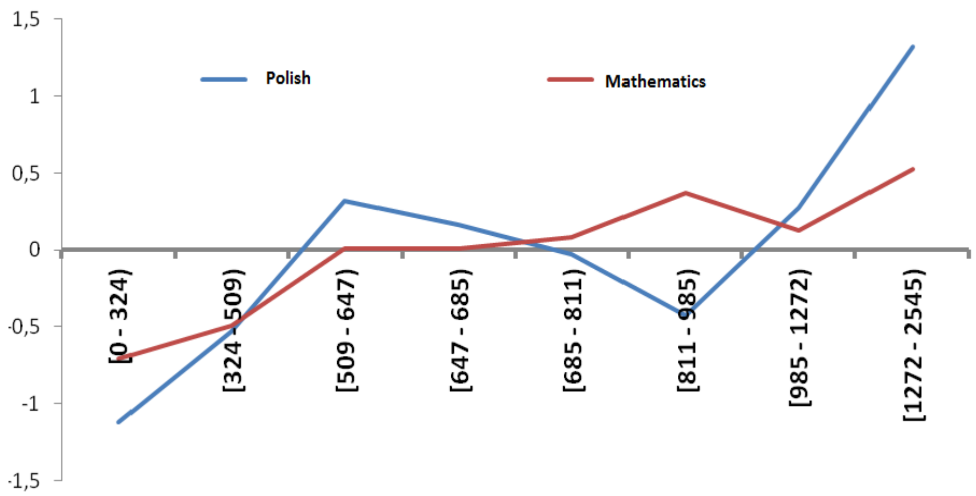

Fig. 2. Relation between mean EVA and grouped development index[1]

Source: own Excel calculations.

[1] Groups are left-bounded and right-unbounded; moreover, given numbers should be divided by 10,000; e.g. an interval [324 – 509) denotes actually [0.0324 – 0.0509).

As suggested by Figure 2, no clear relations are detected, especially as regards Polish. Thus, we deduce that teaching effects in Polish depend much less on the development level of counties from a medium developed group. This may result from the fact that in the case of similar levels of development of counties, students have very similar access to libraries and other cultural vehicles, therefore there is no significant variation. It is worth emphasizing here that four variables containing information about libraries were used to calculate the development index. Public libraries are mostly used by humanists, for that reason group B is not noticeably differentiated with respect to humanities. Moreover, both extreme groups, i.e. [0 – 324) and [1272 – 2545) deviate more from the remaining as regards humanities than mathematics.

References

Aitkin M., Longford N. (1986). *Statistical Modelling Issues in School Effectiveness Studies*. *Journal of the Royal Statistical Society. Vol*. 149. No. 1. Pp. 1-43.

Baltagi B.H. (2005). *Econometric Analysis of Panel Data*. John Wiley & Sons.

Bils M., Klenow P. (2000). *Does Schooling Cause Growth?* American Economic Review. Vol. 90. No 5. Pp. 1160-1183.

Chabbott C., Ramirez F. (2006). *Development and Education*. Handbook of the Sociology of Education. Part II. Pp. 163-187.

Ejsmont W. (2009). *Efektywność nauczania we wrocławskich liceach*. Didactics of Mathematics. No 5-6 (9-10). Wydawnictwo Uniwersytetu Ekonomicznego we Wrocławiu. Pp. 111-128.

Hellwig Z. (1968). *Zastosowanie metody taksonomicznej do typologicznego podziału krajów ze względu na poziom rozwoju i strukturę kwalifikowanych kadr*. Przegląd Statystyczny. No 4. Pp. 307-328.

Jakubowski J., Sztencel R. (2004). *Wstęp do teorii prawdopodobieństwa*. SCRIPT.

Maddala G.S. (2006). *Ekonometria*. PWN. Warszawa.

Mankiw N.G., Romer D., Weil D. (1992). *A Contribution to the Empirics of Economic Growth*. Quarterly Journal of Economics. No 107. Pp. 401-437.

Nowak E. (1990). *Metody taksonomiczne w klasyfikacji obiektów społeczno--gospodarczych*. PWE. Warszawa.

Rosner A. (1999). *Typologia wiejskich obszarów problemowych*. IRWiR PAN. Warszawa.

Rosner A. (2002). *Wiejskie obszary kumulacji barier rozwojowych*. IRWiR PAN. Warszawa.

Rosner A. (2007). *Zróżnicowanie poziomu rozwoju społeczno-gospodarczego obszarów wiejskich zróżnicowanie dynamiki przemian*. IRWiR PAN. Warszawa.

Skrondal A., Rabe-Hesketh S. (2008). *Multilevel and Longitudinal Modeling Using Stata*. College Station. Texas: Stata Press Publication – StataCorp LP.

Temple J. (2001). *Growth Effects of Education and Social Capital in the OECD Countries*. OECD Economic Studies. No 33. Pp. 57-101.

Wansbeek T.J., Kapteyn A. (1982). *A Class of Decompositions of the Variance–-Covariance Matrix of a Generalized Error Components Model*. Econometrica. No 50. Pp. 713-724.

Wansbeek T.J., Kapteyn A. (1982). *A Simple Way to Obtain the Spectral Decomposition of Variance Components Models for Balanced Data*. Communications in Statistics. No 11. Pp. 2105-2112.