Covariance Structure Analysis of Regional Development Data: An Application to Municipality Development Assessment[†]

Dario Cziráky[a,‡] , Jakša Puljiz[b], Krešimir Jurlin[b], Sanja Maleković[b], Mario Polić[b]
[a]*Department for Resource Economics, IMO*
[b]*Department for International Economic Relations, IMO*
IMO, 2 Farkaša-Vukotinovića, 10000 Zagreb, Croatia

Abstract

This paper proposes a multivariate statistical approach based on covariance structure analysis for assessment of the regional development level with an application to development ranking of 545 Croatian municipalities. Municipality-level data ware collected on economic, structural, and demographic dimensions and preliminary factor and principal component analysis were computed to analyse empirical groupings of the variables. Next, confirmatory factor analytic models were estimated with the maximum likelihood technique and subsequently their implied structure was formally tested. Testing was extended to a joint model including all three dimensions (economic, structural and demographic) and their covariance structure was modelled with a recursive structural equation model. Finally, scores were estimated for latent variables thereby allowing (i) estimation of the latent development level of the territorial units, (ii) ranking of all units on an interval scale in respect to their latent development level, and (iii) selection of a given percentage of units for inclusion into special state-care subsidy programme.

JEL Classification: R1, C3, C1

Keywords: Regional development; Covariance structure analysis; Multivariate methods; Factor analysis; Structural equation modelling; Latent variables

1. Introduction

Regional development assessment is a methodologically challenging and policy relevant issue. Aside from purely academic investigation into geo-economic and social patterns and

[†] Paper presented at 42nd Congress of the European Regional Science Association, Dortmund, Germany, August 27 – 31, 2002.
[‡] Corresponding author. Tel.: +358 1 48 26 522; fax: +358 1 48 28 361
E-mail address: dario@irmo.hr (D. Cziráky).

groupings of regional units, there is an important policy requirement for estimating the level of regional development for the purpose of development classification and funding considerations. The Structural Funds of the European Union provide one such example where the level of economic development (approximated by the GDP per capita), in principle, determines the inclusion or exclusion of particular European regions into the regional financial schemes allocations.

In this paper we present a multivariate statistical framework for assessment of the regional development level developed for the purpose of ranking 545 Croatian municipalities. The statistical model was needed to identify municipalities lacking in development. A given percentage of the most underdeveloped municipalities was planned to be subsequently included into a regional funding scheme financed from the national budget. This paper presents the results from the second phase of the project "Criteria for the Development Level Assessment of the Areas Lagging in Development" that was carried out by the IMO for the Croatian Ministry of Public Works (Maleković, 2001). The purpose of the project was to provide an analytical base for evaluation of the development level of the Croatian territorial units (municipalities) with an aim of widening the span of territorial units which are currrently receiving state support under the "Law on Areas of Specific Governmental Concern". The results of the analysis were intended to serve as the basis for changing the approach to defining and supporting the development of areas of special state concern. Unlike the former approach, whose sole criterion has been whether an area was war-affected (i.e., under Serbian occupation in the 1991-1995 war), the new approach defines *economic*, *structural*, and *demographic* criteria (dimensions), as well as a combination of indicators and choice procedures with an aim of obtaining a better quality development evaluation.

There are some important aspects that have substantially influenced our approach. First, the unit of analysis is municipality and that fact has resulted in some major problems, mostly regarding the availability and a quality of the data. Second is the policy relevance and a political sensitivity of the whole project. Namely, in the situation when direct result of the project is a list of municipalities eligible to enter "Areas of Special State Concern" which results in their privileged status concerning state subsidies, tax deductions, etc., the proposed solution needed to be maximally transparent and unambiguous, and a space left for political manipulations has to be minimised.

There are several possible approaches to assess the development level of territorial units, most often some form of classification and data reduction techniques is employed. Soares, *et al.* (2002) suggested a combination of factor and cluster analysis (see Everitt, 1993) and

provided an example of a regional classification for Portugal. Lipshitz and Raveh (1998; 1994) proposed the use of a co-plot technique for the study of regional disparities. Multidimensional scaling techniques (Borg and Groenen, 1997), metric scaling (Weller and Romney, 1990) and correspondence analysis (Greenacre, 1993; Greenacre and Blasius, 1994; Blasius and Greenacre, 1998) can be also used to investigate clustering and grouping of territorial units. Most of these methods minimise some metric or not metric criteria in respect to given variables thereby allowing proximity groupings of units and/or variables. However, it is often the case that the applied model selection criteria are to a large degree arbitrary and subject to "fine-tuning" and data mining. At best such grouping techniques offer broad geographical picture of similarity clusters, i.e., territorial units that are more similar among themselves than with the rest of the units. Moreover, clustering and grouping techniques do not offer justification for exclusion decision, namely, if an "underdeveloped" cluster is defined so to include more units that can be funded from the subsidiary funds, there are no grounds for exclusion of some members of that cluster as they do not posses a unique "development score".

In cases such as ours clear universal criteria and transparent models are needed. Furthermore, it is often necessary to estimate the underlying (i.e., latent) development level for each territorial unit, not merely classify them into separate clusters and then substantively interpret these clusters as more or less developed.

To address these problems we propose an inferential multivariate statistical methodology framework within the general class of covariance structure analysis to estimate the regional development level of territorial units. The formal analysis is preceded by extensive descriptive analysis and data screening including principal component and factor analysis methods. We then develop a structural equation model with latent variables and subsequently compute scores of the underlying latent variables thereby achieving three important goals of the project assignment: (i) estimation of the latent development level of the territorial units, (ii) ranking of all units on an interval scale in respect to their latent development level, and (iii) enable selection of a given percentage of units for inclusion into special state-care subsidy programme. The paper is organised as follows. In the second part we give brief description of the variables and present the results of preliminary descriptive analysis. The third part explains econometric methodology and model building strategy and subsequently presents the estimation results. In the fourth part we compute the underlying development scores and rank the territorial units.

2. Regional development data

The data for the analysis came from several Croatian sources such as the Ministry of finance, the Health Insurance Institute, and the Statistical Bureau. The collected data is on the municipality level and presents lowest aggregation level available in Croatia. We were able to collect data on 11 indicators, roughly grouped into *economic*, *structural* and *demographic* dimensions. These broad categories are defined on substantive grounds and are later analysed by the means of exploratory and confirmatory factor analysis to check their empirical similarity patterns. The economic indicators generally included income-type variables, namely income per capita, share of population earning income and municipality (direct) income. The structural indicators were essentially employment measures such as employment percentage (we used the 2000 data), unemployment percentage (using 2001 data) and social aid per capita. Data from different years were used due to availability of more reliable employment indicators for the year 2000. Demographic indicators included the age index (defined as the number of people older then 65 divided by the number of people younger then 20), density (defined as the number of inhabitants per square kilometre), vitality (number of live births per 100 births), distance (time in minutes needed to reach the County centre by car) and population trend (total municipality population in 2001 over total population in 1991). The codes and brief description are shown in Table 1. We also include the LISREL notation (see Jöreskog, *et al.*, 2000) that will be used in later analysis in the third column.

Table 1
Codes and descriptions of the variables

Code	Description	LISREL notation
INC_PC	Income per capita (in thousands HRK)	y_1
POP_INC	Population share making income (%)	y_2
MUN_INC	Municipality income per capita (in thousands HRK)	y_3
EMP_00	Employment in 2000 (%)	y_4
UNEMP_01	Unemployment in 2001 (%)	y_5
SOC_AID	Social aid per capita (in thousands HRK)	y_6
AGEINDEX	Age index	x_1
DENSITY	Density (inhabitants per km^2)	x_2
VITALITY	Vitality index	x_3
DISTANCE	Distance	x_4
POPTREND	Population trend	x_5

The initial data screening (Table 2 & Fig. 1) showed high degree of skewness and excess kurtosis in all variables (the density plots and kernel estimates were created with PcGive 9.1, Hendry and Doornik, 1999). An informal look at the standard deviations (column two of table

2) indicates relatively comparable variances of most variables with noted exception of the DENSITY variable. DENSITY has higher variance because it is expressed in the original metric without rescaling while most other variables were expressed either in thousands or in percentages. We left the population density variable in people per square kilometre units because we could find no meaningful rescaling that could be justified on substantive grounds. Consequently, we note that DENSITY has greater relative variance then any other variable in the analysis. The issue of removing differences in variances across variables through standardisation is a rather debated one. Some authors base their entire analysis on standardised variables (e.g., Soares *et al.*, 2002) which amounts to analysing correlation matrices instead of covariance matrices in all subsequent econometric models (see Gerbing and Anderson, 1984).

Table 2
Univariate summary statistics for continuous variables

Variable	Mean	St. Dev.	Skewness	Kurtosis	Minimum	Maximum
INC_PC	10.214	3.298	0.260	− 0.482	2.744	21.367
POP_INC	47.862	7.462	− 0.279	− 0.391	25.949	66.746
MUN_INC	1.043	1.077	2.892	10.958	0.034	7.827
EMP_00	53.864	10.482	− 0.827	1.477	5.625	79.084
UNEMP_01	24.825	14.382	1.614	3.944	0.696	95.152
SOC_AID	0.095	0.111	4.062	21.643	0.015	0.839
AGEINDEX	107.197	76.186	7.296	83.904	24.408	1164.286
DENSITY	98.076	220.497	10.150	130.166	1.361	3372.907
VITALITY	82.434	39.643	1.462	4.133	10.811	308.929
DISTANCE	31.561	27.960	3.143	13.227	2.700	210.000
POPTREND	− 8.790	19.391	− 0.297	3.203	− 91.328	73.476

However, our methodology is mainly based on the analysis of covariance rather then correlation structures thus we preserve the original metrics of the variables (at least up to the point of substantive interpretability).

We note that certain multivariate techniques yield identical results regardless of which covariance matrix is analysed, however in general the standards errors and overall fit statistics can be wrong if the correlation matrix is analysed in place of the covariance matrix (see Cudek, 1989; Jöreskog, 2001: 209-214). We proceed with formal univariate and multivariate normality tests (D'Agostino, 1986; Doornik and Hansen, 1994; Mardia, 1980). The results of the normality tests are shown in Table 3.

It is clear that the visible univariate deviations from the Guassian density in Fig. 1 are statistically strongly significant for all variables. AGEINDEX and DENSITY have particularly large chi-square values (strongly rejecting the null hypothesis that the variable is

Gaussian or normally distributed). For more information on normality tests see also D'Agustino (1970; 1971), Bowman and Shenton (1975), Shenton and Bowman (1977) and Belanger and D'Agostino (1990).

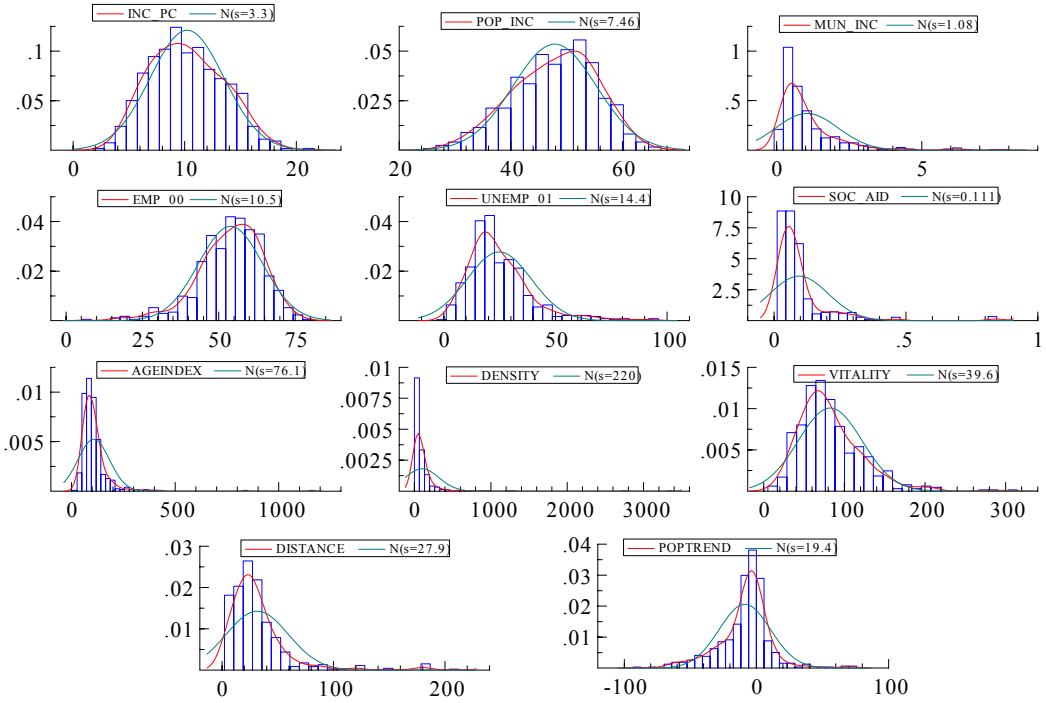


Figure 1. Empirical density (Gaussian kernel estimate): original data

Table 3
Tests of univariate normality*

	Skewness		Kurtosis		Skewness and Kurtosis	
Variable	Z-Score	P-Value	Z-Score	P-Value	X^2	P-Value
INC_PC	2.484	0.013	– 2.985	0.003	15.084	0.001
POP_INC	– 2.669	0.008	– 2.260	0.024	12.229	0.002
MUN_INC	27.635	0.000	10.598	0.000	876.035	0.000
EMP_00	– 7.901	0.000	4.446	0.000	82.199	0.000
UNEMP_01	15.427	0.000	7.429	0.000	293.200	0.000
SOC_AID	38.822	0.000	12.440	0.000	1661.866	0.000
AGEINDEX	69.731	0.000	15.261	0.000	5095.250	0.000
DENSITY	97.004	0.000	15.954	0.000	9664.310	0.000
VITALITY	13.976	0.000	7.580	0.000	252.776	0.000
DISTANCE	30.035	0.000	11.135	0.000	1026.094	0.000
POPTREND	– 2.839	0.005	6.766	0.000	53.847	0.000

* The normality tests were computed with PRELIS 2 (Jöreskog and Sörbom, 1996).

In addition to univariate tests we also compute the multivariate normality test (see Mardia, 1980) which produced the chi-square of 16462.229 with multivariate skewness of 274.269 (z-score 123.181) and multivariate kurtosis of 500.210 (z-score 35.896) which strongly rejects multivariate normality. These are important findings because we intend to use inferential

techniques which are rather sensitive to departures from normality. Consequently, in section 3.3 we will attempt to normalise these variables using monotonic transformation techniques.

3. Econometric methodology

3.1. Multivariate modelling

Multivariate methods used in regional development research generally fall into variable classification (e.g., factor analysis) or classification of cases techniques (e.g. cluster analysis, Q-factor analysis). The two classes of techniques could be also combined. For example, Soares, *et al*. (2002) first perform factor analysis on the variables and then they cluster the cases (i.e., territorial units) using cluster analysis on the original variables as well as on the factor scores. This approach of searching for general patterns of similarities also underlines other similar space-proximity methods such as the co-plot technique (Lipshitz and Raveh 1998; 1994), multidimensional scaling (Borg and Groenen, 1997), metric scaling (Weller and Romney, 1990) and correspondence analysis (Greenacre, 1993; Greenacre and Blasius, 1994; Blasius and Greenacre, 1998). However, nether of these techniques allows estimation of the development level of territorial units on a single scale (preferably interval) nor do they allow ranking of all analysed units in respect to some uniquely defined development criteria. The trouble is that these criteria are at the heart of the problem we need to solve in the first place. Consequently, we need to design an alternative methodological framework within which we can achieve given objectives of comparative development evaluation and ranking of all units.

Our proposed solution is to model the covariance structure of the municipality socio-economic and demographic data within the class of general structural equation models with latent variables (Jöreskog, 1973; Hayduk, 1987, 1996; Bollen, 1989; Jöreskog *et al.* 2000). It can be easily shown that factor analysis, errors-in-variables models, classical econometric simultaneous models and several other model types are all special cases of the general linear structural equation model with latent variables (LISREL).

In our approach we propose to start from exploratory techniques (e.g., principal component analysis) and then combine these results with theoretically-driven modelling strategies that utilise substantive insights from the economic and social theory.

Within the LISREL class of (sub)models we initially wish to mention one special case that appears particularly appealing for our methodological objectives, namely the second-order factor analysis (Buntig, *et al*., 1987; Gerbing, *et al*., 1994; Kaplan and Elliott, 1997a, 1997b; Mulaik, and Quartetti, 1997). The second-order factor analysis assumes two layers of latent

variables which can nicely correspond to our initially assumed economic, structural and demographic factors (first layer) and the overall regional development (second layer, i.e., second-order factor). The underlying covariance structure of such model would imply that there is one common dimension (regional development) that can be measured by separate types of development dimensions (economic, structural and demographic) which are themselves latent variables measured by the observed development indicators (e.g., our municipality variables from Table 1). If such model would fit the data well we could indeed argue that we modelled a single "regional development" level, and by computing factor scores for the second-order factor (Lawley and Maxwell, 1971; Jöreskog, 2000) we would immediately have an indicator that would satisfy all of our project objectives. Unfortunately, second-order factor models rarely fit in practice and are highly unlikely to be applicable to economic data which, by theoretical assumption, include a wealth of causal (both recursive and non-recursive) relationships among variables. Despite these obvious shortcomings many applied researchers rely on assumptions very similar to those behind second-order factor models by sweeping them under the carpet of non-inferential and informal techniques. It is often assumed, without any testing, that the underlying factors are orthogonal and preliminary factor analysis solutions are frequently accepted without confirmatory testing.

Our approach, on the contrary, is inferential and it emphasises model testing and evaluation. Initially, we perform exploratory analysis but in subsequent stages of data analysis we formally test the insights gained from the exploratory analysis. For this purpose we first perform principal component factor analysis and then test each implied dimension (factor) for specification using maximum likelihood confirmatory factor analysis. Finally, we develop a recursive structural equation model with latent variables that includes more complex relationships among the analysed variables.

3.2. Factor analysis

First we performed principal component analysis extracting 11 principal components. Three components had eigenvalues above one contributing 66% of the variance (see Anderson, 1984). The Cattell's scree plot (Fig. 2) levels out after the third eigenvalue and variance contributions diminish after the third component (see Tacq, 1998).

Table 4

Principal component analysis: Eigenvalues and variance extraction

	Extraction sums of squared loadings			Rotation sums of squared loadings		
Component	Total	Variance %	Cumulative %	Total	Variance %	Cumulative %
1	4.014	36.487	36.487	2.450	22.274	22.274
2	1.993	18.121	54.608	1.055	9.590	31.864
3	1.272	11.560	66.168	1.048	9.529	41.392
4	0.959	8.718	74.886	1.048	9.526	50.919
5	0.733	6.664	81.550	1.046	9.506	60.425
6	0.651	5.914	87.464	1.021	9.282	69.707
7	0.483	4.392	91.856	1.008	9.165	78.873
8	0.339	3.086	94.941	1.007	9.157	88.029
9	0.316	2.870	97.812	1.007	9.153	97.182
10	0.167	1.518	99.329	0.185	1.683	98.866
11	0.074	0.671	100.000	0.125	1.134	100.000

Keeping also in mind that we conjectured about three dimensions, i.e., factors (economic, structural and demographic) we computed factor analysis retaining three factors.

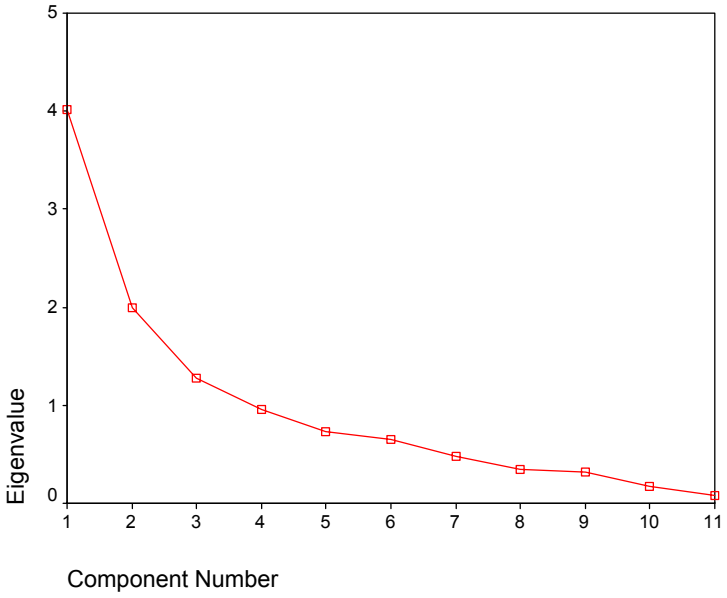


Figure 2. Cattell's scree test

In addition, on the basis of principal component solution and insignificant correlations with other variables we dropped DISTANCE variable from further analysis. Un-rotated and rotated loadings (for the remaining ten variables) are shown in Table 5.

The first factor loads highly on POP_INC, EMP_00, UNEMP_01 and SOC_AID which corresponds to our postulated *structural* dimension. The exception is POP_INC which loads ambiguously on two factors.

Table 5
Factor loadings structure

	Component loadings			Rotated component loadings*			Extracted
Variable	1	2	3	1	2	3	communalities
INC_PC	0.740	–0.138	0.556	0.197	0.113	0.908	0.875
POP_INC	0.801	–0.413	0.123	0.582	–0.149	0.683	0.828
MUN_INC	0.704	–0.125	0.480	0.219	0.111	0.825	0.742
EMP_00	0.849	–0.255	–0.141	0.759	0.005	0.478	0.805
UNEMP_01	–0.832	0.234	0.297	–0.846	–0.013	–0.347	0.836
SOC_AID	–0.719	0.054	0.425	–0.812	–0.145	–0.139	0.700
AGEINDEX	–0.286	–0.883	–0.050	0.000	–0.930	–0.014	0.864
DENSITY	0.479	0.664	–0.086	0.273	0.772	0.082	0.678
VITALITY	0.096	0.851	0.262	–0.271	0.851	0.062	0.802
POPTREND	0.606	0.526	–0.287	0.525	0.670	0.041	0.726
Variance total	4.303	2.549	1.004	2.800	2.703	2.353	–
Variance %	43.030	25.490	10.036	28.000	27.027	23.530	–
Cumulative %	43.030	68.520	78.557	28.000	55.027	78.557	–

* Verimax rotation.

The *demographic* dimension (AGEINDEX, DENSITY and VITALITY) seem well captured with the second factor. The third factor then appears to account for the *economic* dimension and includes INC_POP, POP_INC and MUN_INC with high loadings. However, EMP_00 appears to also load on this factor. The general impression from this three-factor solution is that there does not appear to be a "simple structure" in the data. There are several complex loadings, i.e., several variables appear to be indicators of more then one latent variable. Also, on theoretical grounds it is highly unlikely that these three underlying dimensions are uncorrelated in the population thus the above factor solution can, at best, serve as a starting point for more detailed analysis in which these indicative and partly ambiguous findings could be statistically tested.

3.3. Confirmatory maximum likelihood factor analysis

Formal testing of factor structures is most conveniently done in the confirmatory factor analysis framework using the maximum likelihood technique. Although all of our variables are continuous we have found that they are not normally distributed. Normality is, however, important in maximum likelihood estimation based on multivariate normal likelihood. We proceed by transforming the variables closer to the Gaussian distribution and this way try to avoid potential problems with the analysis of non-normal variables (see Babakus, *et al.*, 1987; Curran, *et al.*, 1996; West, *et al.*, 1995). For this purpose we apply the *normal scores* technique (Jöreskog *et al.*, 2000, Jöreskog, 1999). The technique can be summarised as

follows. Given a sample of N observations on the j^{th} variable, $\mathbf{x}_j = \{x_{j1}, x_{j2}, \ldots, x_{jN}\}$, the normal scores transformation is computed in the following way. First define a vector of k distinct sample values, $\mathbf{x}_j^k = \{x_{j1}', x_{j2}', \ldots, x_{jk}'\}$ where $k \leq N$ thus $\mathbf{x}^k \subseteq \mathbf{x}$. Let f_i be the frequency of occurrence of the value x_{ji} in $\mathbf{x}_j$ so that $f_{ji} \geq 1$ and. Then normal scores x_{ji}^{NS} are computed as $x_{ji}^{NS} = (N/f_{ji})\{\phi(\alpha_{j,i-1}) - \phi(\alpha_{ji})\}$ where ϕ is the standard Gaussian density function, α is defined as

$$\alpha_{ji} = \begin{cases} -\infty, & i = 0 \\ \Phi^{-1}\left(N^{-1}\sum_{t=1}^{i} f_{jt}\right), & i = 1, 2, \ldots, k-1, \\ \infty, & i = k \end{cases}$$

and Φ^{-1} is the inverse of the standard Gaussian distribution function. The normal scores are further scaled to have the same mean and variance as the original variables.

The transformation of the variables resulted in insignificant normality chi-square statistics. The mean and variances remained unaltered, though some values became negative (note the minimums) due to rescaling (Table 6).

Table 6
Univariate Summary Statistics for Continuous Variables

Variable	Mean	St. Dev.	Skewness	Kurtosis	Minimum	Maximum
INC_PC	10.214	3.298	0.000	− 0.007	− 0.316	20.745
POP_INC	47.862	7.462	0.000	− 0.007	24.037	71.687
MUN_INC	1.043	1.077	0.000	− 0.007	− 2.395	4.480
EMP_00	53.864	10.482	0.000	− 0.007	20.399	87.330
UNEMP_01	24.825	14.382	0.000	− 0.007	− 21.091	70.742
SOC_AID	0.095	0.111	0.003	− 0.096	− 0.186	0.387
AGEINDEX	107.197	76.186	0.000	− 0.007	− 136.043	350.438
DENSITY	98.076	220.497	0.000	− 0.007	− 605.911	802.063
VITALITY	82.434	39.643	0.000	− 0.007	− 44.137	209.005
DISTANCE	31.561	27.960	0.000	− 0.035	− 51.982	120.885
POPTREND	− 8.790	19.391	0.000	− 0.007	− 70.700	53.119

Fig. 3 shows close fit between empirical densities of all variables and the theoretical Gaussian curve. Formal normality tests (Table 7) no longer reject univariate normality null for any of the variables. Having transformed the data we perform maximum-likelihood based tests for the number of factors (Table 8) which reject simple multi-factor solutions up to 6 factors. This is another indication that the covariance structure of these data is too complex to be explained by simple multi-factor solutions.

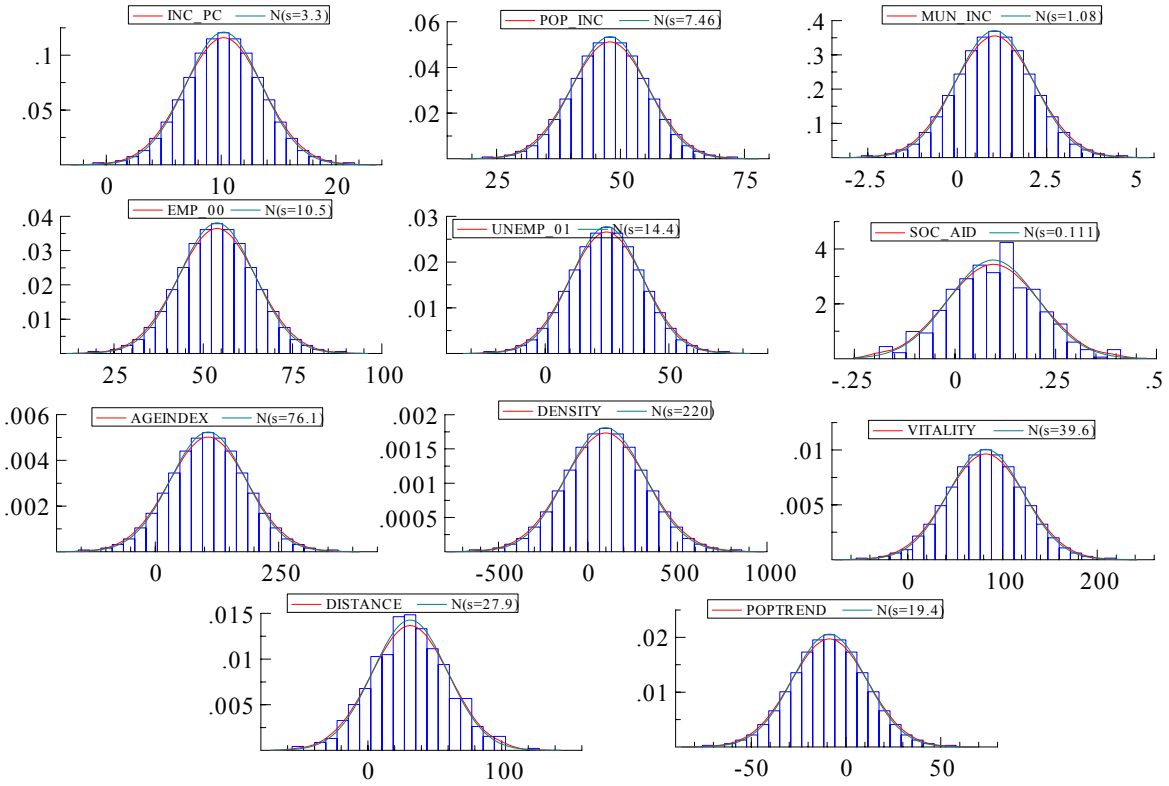


Figure 3. Empirical density (Gaussian kernel estimate): normalised data

Table 7

Test of Univariate Normality for Continuous Variables: Normalised data

	Skewness		Kurtosis		Skewness and Kurtosis	
Variable	Z-Score	P-Value	Z-Score	P-Value	Chi-Square	P-Value
INC_PC	0.000	1.000	0.065	0.948	0.004	0.998
POP_INC	0.000	1.000	0.065	0.948	0.004	0.998
MUN_INC	0.000	1.000	0.065	0.948	0.004	0.998
EMP_00	0.000	1.000	0.065	0.948	0.004	0.998
UNEMP_01	0.000	1.000	0.065	0.948	0.004	0.998
SOC_AID	0.030	0.976	– 0.388	0.698	0.151	0.927
AGEINDEX	0.000	1.000	0.065	0.948	0.004	0.998
DENSITY	0.000	1.000	0.065	0.948	0.004	0.998
VITALITY	0.000	1.000	0.065	0.948	0.004	0.998
DISTANCE	– 0.004	0.997	– 0.074	0.941	0.005	0.997
POPTREND	0.000	1.000	0.065	0.948	0.004	0.998

It is therefore likely that the underlying factors have complex rather then simple structure and it is also likely that they are correlated (for more details on modelling strategies and testing for number of factors see Bollen, 2001 and Bai and Ng, 2002).

Table 8
Maximum likelihood decision table for the number of factors

Factors	X^2	d.f.	P	X^2 difference	d.f.	PD	RMSEA
0	4081.480	55	0.000				0.367
1	2066.900	44	0.000	2014.590	11	0.000	0.290
2	886.070	34	0.000	1180.830	10	0.000	0.214
3	460.080	25	0.000	425.980	9	0.000	0.179
4	276.740	17	0.000	183.350	8	0.000	0.167
5	71.260	10	0.000	205.470	7	0.000	0.106
6	30.260	4	0.000	41.000	6	0.000	0.110

RMSEA = root mean square error of approximation

In the following analysis we test separate measurement models for three implied dimensions starting from the indicative factor analysis results. Our purpose here is to statistically evaluate validity of single dimensions separately.

The confirmatory measurement models as well as later structural models are estimated within the class of general linear structural equation models (Jöreskog, 1973; Hayduk, 1987; Bollen, 1989; Hayduk, 1996; Jöreskog, *et al.*, 2000). Denoting the latent endogenous variables by $\boldsymbol{\eta}$ and latent exogenous variables by $\boldsymbol{\xi}$ and their respective observed indicators by $\mathbf{y}$ and $\mathbf{x}$, the structural part of the model is given by

$$\boldsymbol{\eta} = \mathbf{B}\boldsymbol{\eta} + \boldsymbol{\Gamma}\boldsymbol{\xi} + \boldsymbol{\zeta} \tag{1}$$

The measurement models are given in form of standard factor analytic models as

$$\mathbf{y} = \boldsymbol{\Lambda}_y\boldsymbol{\eta} + \boldsymbol{\epsilon}, \tag{2}$$

for latent endogenous and

$$\mathbf{x} = \boldsymbol{\Lambda}_x\boldsymbol{\xi} + \boldsymbol{\delta}, \tag{3}$$

for latent exogenous variables. Using Jöreskog's LISREL notation we also define the following second moment matrices: $E(\boldsymbol{\xi}\boldsymbol{\xi}^T) = \boldsymbol{\Phi}$, $E(\boldsymbol{\zeta}\boldsymbol{\zeta}^T) = \boldsymbol{\Psi}$, $E(\boldsymbol{\epsilon}\boldsymbol{\epsilon}^T) = \boldsymbol{\Theta}_\epsilon$, $E(\boldsymbol{\delta}\boldsymbol{\delta}^T) = \boldsymbol{\Theta}_\delta$, and $E(\boldsymbol{\epsilon}\boldsymbol{\delta}^T) = \boldsymbol{\Theta}_{\delta\epsilon}$. The covariance matrix implied by the model is comprised from three separate covariance matrices, covariance matrix of the observed indicators of the latent endogenous variables

$$\begin{aligned}\boldsymbol{\Sigma}_{yy} &= E(\mathbf{y}\mathbf{y}^T) \\ &= E[(\boldsymbol{\Lambda}_y\boldsymbol{\eta} + \boldsymbol{\epsilon})(\boldsymbol{\Lambda}_y\boldsymbol{\eta} + \boldsymbol{\epsilon})^T] \\ &= \boldsymbol{\Lambda}_y E(\boldsymbol{\eta}\boldsymbol{\eta}^T)\boldsymbol{\Lambda}_y^T + \boldsymbol{\Theta}_\epsilon \\ &= \boldsymbol{\Lambda}_y(\mathbf{I} - \mathbf{B})^{-1}(\boldsymbol{\Gamma}\boldsymbol{\Phi}\boldsymbol{\Gamma}^T + \boldsymbol{\Psi})[(\mathbf{I} - \mathbf{B})^{-1}]^T\boldsymbol{\Lambda}_y^T + \boldsymbol{\Theta}_\epsilon,\end{aligned} \tag{4}$$

the covariances between the indicators of latent endogenous and indicators of latent exogenous variables

(5) $$\begin{aligned}\boldsymbol{\Sigma}_{yx} &= E(\mathbf{y}\mathbf{x}^{T}) \\ &= E[(\boldsymbol{\Lambda}_{y}\boldsymbol{\eta} + \boldsymbol{\epsilon})(\boldsymbol{\Lambda}_{x}\boldsymbol{\xi} + \boldsymbol{\delta})^{T}] \\ &= \boldsymbol{\Lambda}_{y}E(\boldsymbol{\eta}\boldsymbol{\xi}^{T})\boldsymbol{\Lambda}_{x}^{T} + \boldsymbol{\Theta}_{\delta\epsilon}^{T} \\ &= \boldsymbol{\Lambda}_{y}(\mathbf{I} - \mathbf{B})^{-1}\boldsymbol{\Gamma}\boldsymbol{\Phi}\boldsymbol{\Lambda}_{x}^{T} + \boldsymbol{\Theta}_{\delta\epsilon}^{T},\end{aligned}$$

and finally, the covariance matrix of the indicators of the latent exogenous variables

(6) $$\begin{aligned}\boldsymbol{\Sigma}_{xx} &= E(\mathbf{x}\mathbf{x}^{T}) \\ &= E[(\boldsymbol{\Lambda}_{x}\boldsymbol{\xi} + \boldsymbol{\delta})(\boldsymbol{\Lambda}_{x}\boldsymbol{\xi} + \boldsymbol{\delta})^{T}] \\ &= \boldsymbol{\Lambda}_{x}E(\boldsymbol{\xi}\boldsymbol{\xi}^{T})\boldsymbol{\Lambda}_{x}^{T} + \boldsymbol{\Theta}_{\delta} \\ &= \boldsymbol{\Lambda}_{x}\boldsymbol{\Phi}\boldsymbol{\Lambda}_{x}^{T} + \boldsymbol{\Theta}_{\delta}.\end{aligned}$$

Arranging the above three matrices together (noting that the lower left block is just a transpose of the upper right block we get the joint covariance matrix implied by the model, i.e.,

(7) $$\boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_{yy} & \boldsymbol{\Sigma}_{yx} \\ \boldsymbol{\Sigma}_{xy} & \boldsymbol{\Sigma}_{xx} \end{pmatrix}.$$

Using Eq. 4-7 the implied covariance matrix can be written in terms of model parameters as

(8) $$\boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Lambda}_{y}(\mathbf{I}-\mathbf{B})^{-1}(\boldsymbol{\Gamma}\boldsymbol{\Phi}\boldsymbol{\Gamma}^{T} + \boldsymbol{\Psi})[(\mathbf{I}-\mathbf{B})^{-1}]^{T}\boldsymbol{\Lambda}_{y}^{T} + \boldsymbol{\Theta}_{\varepsilon} & \boldsymbol{\Lambda}_{y}(\mathbf{I}-\mathbf{B})^{-1}\boldsymbol{\Gamma}\boldsymbol{\Phi}\boldsymbol{\Lambda}_{x}^{T} + \boldsymbol{\Theta}_{\delta\varepsilon}^{T} \\ \boldsymbol{\Lambda}_{x}\boldsymbol{\Phi}\boldsymbol{\Gamma}^{T}[(\mathbf{I}-\mathbf{B})^{-1}]^{T}\boldsymbol{\Lambda}_{y}^{T} + \boldsymbol{\Theta}_{\varepsilon\delta} & \boldsymbol{\Lambda}_{x}\boldsymbol{\Phi}\boldsymbol{\Lambda}_{x}^{T} + \boldsymbol{\Theta}_{\delta} \end{pmatrix}.$$

The maximum likelihood estimates of the model parameters, given the model is identified, can be obtained by numerical maximisation of the multivariate Gaussian log-likelihood function

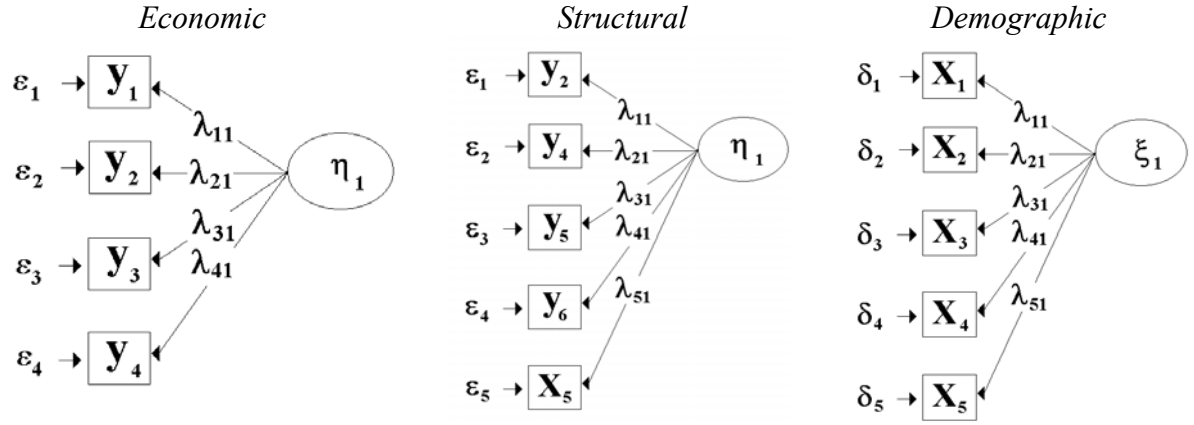
(9) $$F = \ln|\boldsymbol{\Sigma}| + tr\{\mathbf{S}\boldsymbol{\Sigma}^{-1}\} - \ln|\mathbf{S}| - (p+q),$$

where p and q are the numbers of the observed indicators of latent endogenous and latent exogenous variables, respectively.

The model building approach, given such a complex structure (measurement and structural parts), can take several routes and there is no agreement in the literature of the unique best approach (see Bollen, 2001). An approach of initially fitting separate measurement models and then, in the second stage, estimating a pooled model (with all measurement models together) is most appropriate for our purposes as it simultaneously allows testing of the underlying latent structures in each initially conjectured dimension (economic, structural and

demographic). Table 9 shows conceptual path diagrams for the three hypothetical latent dimensions including complex loadings suggested from the factor solution in Table 5. Note that due to simplicity we use the LISREL notation defined in Table 1.

Table 9
Measurement models for the development dimensions



The "economic" measurement model is given in matrix notation as

(10) $$\mathbf{y}_1 = \mathbf{\Lambda}_y \boldsymbol{\eta}_1 + \boldsymbol{\epsilon},$$

or equivalently, in terms of scalar elements of the matrices as

(11) $$\begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{pmatrix} = \begin{pmatrix} \lambda_{11}^{(y)} \\ \lambda_{21}^{(y)} \\ \lambda_{31}^{(y)} \\ \lambda_{41}^{(y)} \end{pmatrix} (\eta_1) + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \varepsilon_4 \end{pmatrix}.$$

The model (11) includes three economic criteria variables (y_1, y_2 and y_3), however y_4 (EMP_00) was also included because it had an ambiguous loading on this factor (see Table 5). The maximum likelihood estimates of the coefficients, their accompanying standard errors and the chi-square for overall fit are given in Table 10 (we term the model "M_1"). The chi-square is 76.18 with 2 degrees of freedom which does not indicate good fit. In an attempt to improve the model, based on the largest modification index (see Sörbom, 1989) we re-estimated the model with the error covariance between y_2 and y_4 (POP_INC and EMP_00) set free (for a discussion on the meaning of error covariances see Gerbing and Anderson, 1984). In terms of the model notation this amounts to freeing the (4, 2) element of the $\mathbf{\Theta}_\epsilon$ residual covariance matrix, which is thus no longer constrained to be diagonal. Therefore, we estimate this modified model ("M_2") with the $\mathbf{\Theta}_\epsilon$ matrix specified as

(12) $$\Theta_{\varepsilon} = \begin{pmatrix} \theta_{11}^{(\varepsilon)} & & & \\ 0 & \theta_{22}^{(\varepsilon)} & & \\ 0 & 0 & \theta_{33}^{(\varepsilon)} & \\ 0 & \theta_{42}^{(\varepsilon)} & 0 & \theta_{44}^{(\varepsilon)} \end{pmatrix}.$$

The estimation results are shown in Table 10. It can be seen that the chi-square dropped to 4.6 with 1 degree of freedom, which is no longer significant, thus the modified model M_2 now fits the data.

Table 10
Maximum likelihood estimates (*Economic* model)

	M_1		M_2	
Parameter	Estimate	(S.E.)	Estimate	(S.E.)
λ_{11}	0.71	(0.04)	0.94	(0.04)
λ_{21}	0.99	(0.03)	0.75	(0.04)
λ_{31}	0.45	(0.04)	0.57	(0.04)
λ_{41}	0.82	(0.04)	0.57	(0.04)
θ_{11}	0.49	(0.03)	0.12	(0.05)
θ_{22}	0.03	(0.03)	0.44	(0.04)
θ_{33}	0.80	(0.05)	0.67	(0.05)
θ_{44}	0.33	(0.03)	0.67	(0.05)
θ_{42}	–	–	0.38	(0.04)
X^2	76.18		4.60	
d.f.	2		1	

A similar procedure is used to assess the measurement model for the second hypothesised dimension (structural). Building a measurement model with three structural indicators (y_4, y_5, y_6) and also including y_2 and x_5 which had moderate loadings on this factor in the exploratory factor analysis we get

(13) $$\mathbf{y}_2 = \mathbf{\Lambda}_y \boldsymbol{\eta}_2 + \boldsymbol{\epsilon},$$

or in full matrix notation

(14) $$\begin{pmatrix} y_2 \\ y_4 \\ y_5 \\ y_6 \\ x_5 \end{pmatrix} = \begin{pmatrix} \lambda_{11}^{(y)} \\ \lambda_{21}^{(y)} \\ \lambda_{31}^{(y)} \\ \lambda_{41}^{(y)} \\ \lambda_{51}^{(y)} \end{pmatrix} (\eta_2) + \begin{pmatrix} \varepsilon_2 \\ \varepsilon_4 \\ \varepsilon_5 \\ \varepsilon_6 \\ \varepsilon_5 \end{pmatrix}.$$

The inclusion of y_2 (POP_INC) makes sense from the economic point of view, while the inclusion of x_5 (POPTREND) is less clear. Its inclusion was based on its moderately high loading on the structural factor (Table 5) and now it can be formally tested. Table 11 gives estimates for model M_1 and M_2, the later one excluding x_5. Though neither model has good fit, the chi-square difference between the two models is 176.34 which strongly rejects M_2 in favour of M_1, that is, the inclusion of the ambiguous variable (POPTREND) into the structural factor is rejected.

Table 11
Maximum likelihood estimates (*Structural* model)

	M_1		M_2		M_3	
Parameter	Estimate	(S.E.)	Estimate	(S.E.)	Estimate	(S.E.)
λ_{11}	0.81	(0.04)	0.81	(0.04)	0.79	(0.04)
λ_{21}	0.99	(0.03)	1.01	(0.03)	1.03	(0.03)
λ_{31}	–0.86	(0.03)	–0.84	(0.04)	–0.82	(0.04)
λ_{41}	–0.49	(0.04)	–0.47	(0.04)	–0.47	(0.04)
λ_{51}	0.39	(0.04)	–	–	–	–
θ_{11}	0.34	(0.02)	0.35	(0.02)	0.38	(0.03)
θ_{22}	0.02	(0.02)	–0.01	(0.02)	–0.06	(0.02)
θ_{33}	0.27	(0.02)	0.29	(0.02)	0.32	(0.02)
θ_{44}	0.76	(0.05)	0.78	(0.05)	0.78	(0.05)
θ_{55}	0.85	(0.05)	–	–	–	–
θ_{43}	–	–	–	–	0.22	(0.02)
X^2	312.51		136.17		13.21	
d.f.	5		2		1	

The fit of the model, however, is still not good enough, and we again modify it on the bases of the largest modification index by freeing the (4, 3) element of the $\mathbf{\Theta}_\epsilon$ matrix, i.e.,

$$\Theta_\varepsilon = \begin{pmatrix} \theta_{11}^{(\varepsilon)} & & & & \\ 0 & \theta_{22}^{(\varepsilon)} & & & \\ 0 & 0 & \theta_{33}^{(\varepsilon)} & & \\ 0 & 0 & \theta_{43}^{(\varepsilon)} & \theta_{44}^{(\varepsilon)} & \\ 0 & 0 & 0 & 0 & \theta_{55}^{(\varepsilon)} \end{pmatrix}, \tag{15}$$

which results in model M_3 with the chi-square of 13.21 with 1 degree of freedom indicating acceptable fit of the model.

Finally, we estimate the confirmatory model for the demographic dimension noting that this was the only "simple-structured" factor, i.e., without any ambiguous loadings. The model is specified as

$$\mathbf{x} = \mathbf{\Lambda}_x\boldsymbol{\xi} + \boldsymbol{\delta}, \tag{16}$$

where the **x-ξ** notation indicates that this factor is treated as exogenous. This assumption will be clarified in the context of the full LISREL model in section 3.4. Remembering that we dropped the DISTANCE variable from further analysis on the basis of its low loading in principal component analysis and insignificant correlations with other variables, we are now in position to formally test its exclusion from the demographic measurement model. The model that includes the DISTANCE variable (x_4) is given as

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{pmatrix} = \begin{pmatrix} \lambda_{11}^{(x)} \\ \lambda_{21}^{(x)} \\ \lambda_{31}^{(x)} \\ \lambda_{41}^{(x)} \\ \lambda_{51}^{(x)} \end{pmatrix} (\xi_1) + \begin{pmatrix} \delta_1 \\ \delta_2 \\ \delta_3 \\ \delta_4 \\ \delta_5 \end{pmatrix}. \tag{17}$$

We term this model M_1. Model M_2 sets λ_{41} to zero. The estimation results are shown in Table 12. M_1 has a chi-square of 189.96 with 5 degrees of freedom and M_2 has a chi-square of 97.89 with 2 degrees of freedom thus the chi-square difference is 92.07 which is highly significant. Therefore, we can reject the inclusion of x_4 into the model.

Table 12
Maximum likelihood estimates (*Demographic* model)

	M_1		M_2		M_3	
Parameter	Estimate	(S.E.)	Estimate	(S.E.)	Estimate	(S.E.)
λ_{11}	0.98	(0.03)	1.00	(0.03)	1.06	(0.04)
λ_{21}	− 0.65	(0.04)	− 0.64	(0.04)	− 0.59	(0.04)
λ_{31}	− 0.79	(0.04)	− 0.78	(0.04)	− 0.73	(0.04)
λ_{41}	0.30	(0.04)	–	–	–	–
λ_{51}	− 0.57	(0.04)	− 0.56	(0.04)	− 0.53	(0.04)
θ_{11}	0.04	(0.03)	0.00	(0.03)	− 0.13	(0.05)
θ_{22}	0.57	(0.04)	0.60	(0.04)	0.65	(0.04)
θ_{33}	0.38	(0.03)	0.39	(0.03)	0.46	(0.04)
θ_{44}	0.91	(0.06)	–	–	–	–
θ_{55}	0.68	(0.04)	0.69	(0.04)	0.72	(0.04)
θ_{52}	–	–	–	–	0.28	(0.03)
X^2	189.96		97.89		5.27	
d.f.	5		2		1	

Once again, the fit can be improved by adding an error covariance between POPTREND (x_5) and AGEINDEX (x_2) where the $\mathbf{\Theta}_\delta$ matrix is specified by freeing the (5, 2) element of

(18) $$\Theta_{\delta} = \begin{pmatrix} \theta_{11}^{(\delta)} & & & & \\ 0 & \theta_{22}^{(\delta)} & & & \\ 0 & 0 & \theta_{33}^{(\delta)} & & \\ 0 & 0 & 0 & \theta_{44}^{(\delta)} & \\ 0 & \theta_{52}^{(\delta)} & 0 & 0 & \theta_{55}^{(\delta)} \end{pmatrix}.$$

The modified model (M_3) had chi-square of 5.27 with 1 degree of freedom, which now indicates acceptably good fit.

3.4. The structural equation model

Having estimated the three measurement models separately, we now estimate a joint model that includes all three dimensions simultaneously. We conducted some preliminary confirmatory analysis, using chi-square difference approach in ML estimation of restricted and unrestricted models, to test for the significance of correlations between factors finding that these are highly significant (detailed results are omitted for brevity).

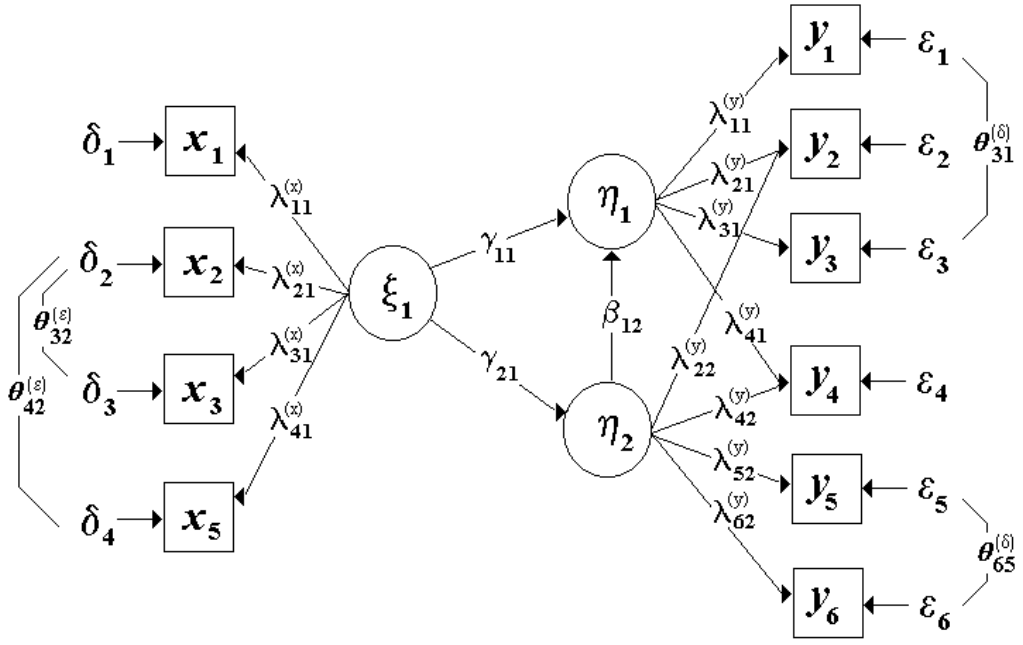


Figure 5. Conceptual path diagram for the structural model

We note that this is strong evidence against orthogonal solutions and orthogonality assumptions in exploratory factor analysis with this type of regional development data. In order to explain the found correlations we postulate a structural model in which the economic development is simultaneously determined by structural and demographic factors and

measured by its observed indicators. Furthermore we conjecture that the structural dimension is causally affected by the demographic factor. In terms of model types, this would be a linear recursive (we assume unidirectional causality) structural equation model with latent variables. Putting it all together we arrive at the model shown in Fig. 5.

Note that due to consistency with the notation defined in Table 1 we keep the symbol x_5 for the POPTREND variable and drop x_4 (DISTANCE) from the model. Also, on the basis of the above estimated separate measurement (factor) models and modification indices from preliminary estimation we add the suggested error covariance and complex loadings, but now we put the three measurement models together and add structural relationships among the three latent variables. In matrix notation the endogenous measurement model corresponding to path diagram shown in Fig. 5. is given by

$$\begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \\ y_5 \\ y_6 \end{pmatrix} = \begin{pmatrix} \lambda_{11}^{(y)} & 0 \\ \lambda_{21}^{(y)} & \lambda_{22}^{(y)} \\ \lambda_{31}^{(y)} & 0 \\ \lambda_{41}^{(y)} & \lambda_{42}^{(y)} \\ 0 & \lambda_{52}^{(y)} \\ 0 & \lambda_{62}^{(y)} \end{pmatrix} \begin{pmatrix} \eta_1 \\ \eta_2 \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \varepsilon_4 \\ \varepsilon_5 \\ \varepsilon_6 \end{pmatrix}. \tag{22}$$

Similarly, the exogenous measurement model is given by

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_5 \end{pmatrix} = \begin{pmatrix} \lambda_{11}^{(x)} \\ \lambda_{21}^{(x)} \\ \lambda_{31}^{(x)} \\ \lambda_{41}^{(x)} \end{pmatrix} (\xi_1) + \begin{pmatrix} \delta_1 \\ \delta_2 \\ \delta_3 \\ \delta_4 \end{pmatrix}. \tag{23}$$

Finally, the structural part of the model is specified as follows

$$\begin{pmatrix} \eta_1 \\ \eta_2 \end{pmatrix} = \begin{pmatrix} 0 & \beta_{12} \\ 0 & 0 \end{pmatrix} \begin{pmatrix} \eta_1 \\ \eta_2 \end{pmatrix} + \begin{pmatrix} \gamma_{11} \\ \gamma_{21} \end{pmatrix} (\xi_1) + \begin{pmatrix} \zeta_1 \\ \zeta_2 \end{pmatrix}, \tag{24}$$

with the error covariance matrices specified as

(25) $$\Theta_{\varepsilon} = \begin{pmatrix} \theta_{11}^{(\varepsilon)} & & & & & \\ 0 & \theta_{22}^{(\varepsilon)} & & & & \\ \theta_{31}^{(\varepsilon)} & 0 & \theta_{33}^{(\varepsilon)} & & & \\ 0 & 0 & 0 & \theta_{44}^{(\varepsilon)} & & \\ 0 & 0 & 0 & 0 & \theta_{55}^{(\varepsilon)} & \\ 0 & 0 & 0 & 0 & \theta_{65}^{(\varepsilon)} & \theta_{66}^{(\varepsilon)} \end{pmatrix},$$

(for the y-measurement model) and as

(26) $$\Theta_{\delta} = \begin{pmatrix} \theta_{11}^{(\delta)} & & & \\ 0 & \theta_{22}^{(\delta)} & & \\ 0 & \theta_{32}^{(\delta)} & \theta_{33}^{(\delta)} & \\ 0 & \theta_{42}^{(\delta)} & 0 & \theta_{44}^{(\delta)} \end{pmatrix},$$

for the x-measurement model. The parameter estimates are shown in Table 6 and Table 7. The model chi-square (normal-theory weighted) is 209.41 with 26 degrees of freedom which is appears not well fitting.

Table 6
Maximum likelihood estimates of the coefficient matrices*

$$\mathbf{\Lambda}_y = \begin{pmatrix} .41\,(.04) & 0 \\ .86\,(.06) & .34\,(.14) \\ .23\,(.03) & 0 \\ .23\,(.03) & -.53\,(.05) \\ 0 & .83\,(.04) \\ 0 & .49\,(.04) \end{pmatrix} \qquad \mathbf{\Lambda}_y = \begin{pmatrix} .90\,(.04) \\ -.12\,(.05) \\ -.48\,(.04) \\ -.60\,(.04) \end{pmatrix}$$

$$\mathbf{B} = \begin{pmatrix} 0 & -1.20\,(.14) \\ 0 & 0 \end{pmatrix} \qquad \mathbf{\Gamma} = \begin{pmatrix} .45\,(.07) \\ .53\,(.06) \end{pmatrix}$$

* Standard errors are in the parentheses.

The model modification indices suggested a number of ways to modify the model. A generalisation of the model that would allow correlations between the uniquenesses of the *y* and *x* indicators requires estimation of the $\mathbf{\Theta}_{\delta\epsilon}$ matrix (see in particular Gerbing and Anderson, 1984).

We thus estimated some non-zero elements in this matrix following the specification implied by the model modification indices. The specific form of the matrix is given by

(27)
$$\Theta_{\delta\varepsilon} = \begin{pmatrix} 0 & 0 & 0 & 0 & \theta_{15}^{(\delta\varepsilon)} \\ 0 & 0 & 0 & 0 & 0 \\ 0 & \theta_{32}^{(\delta\varepsilon)} & 0 & \theta_{34}^{(\delta\varepsilon)} & 0 \\ 0 & \theta_{42}^{(\delta\varepsilon)} & 0 & 0 & \theta_{45}^{(\delta\varepsilon)} \end{pmatrix}.$$

Table 7
Maximum likelihood estimates of the error covariances*

$$\Theta_{\varepsilon} = \begin{pmatrix} .59\,(.04) & & & & & \\ 0 & -.22\,(.09) & & & & \\ .29\,(.03) & 0 & .87\,(.05) & & & \\ 0 & 0 & 0 & .20\,(.03) & & \\ 0 & 0 & 0 & 0 & .12\,(.04) & \\ 0 & 0 & 0 & 0 & .01\,(.03) & .70\,(.05) \end{pmatrix}$$

$$\Theta_{\delta} = \begin{pmatrix} .19\,(.05) & & & \\ 0 & .98\,(.06) & & \\ 0 & .15\,(.04) & .77\,(.05) & \\ 0 & .08\,(.04) & 0 & .64\,(.05) \end{pmatrix}$$

* Standard errors are in the parentheses.

The estimation of the modified model (Table 8) produced very similar results to the previous model.

Table 8
Maximum likelihood estimates of the coefficient matrices (modified model)*

$$\mathbf{\Lambda}_y = \begin{pmatrix} .39\,(.04) & 0 \\ .86\,(.06) & .38\,(.15) \\ .20\,(.03) & 0 \\ .21\,(.03) & -.54\,(.05) \\ 0 & .82\,(.04) \\ 0 & .41\,(.04) \end{pmatrix} \qquad \mathbf{\Lambda}_y = \begin{pmatrix} .82\,(.05) \\ -.12\,(.05) \\ -.50\,(.04) \\ -.66\,(.04) \end{pmatrix}$$

$$\mathbf{B} = \begin{pmatrix} 0 & -1.28\,(.15) \\ 0 & 0 \end{pmatrix} \qquad \mathbf{\Gamma} = \begin{pmatrix} .50\,(.08) \\ .55\,(06) \end{pmatrix}$$

* Standard errors are in the parentheses.

The chi-square of the overall fit has dropped to 27.66 with 21 degrees of freedom which is insignificant (p-value = 0.15). Thus we conclude, on statistical grounds, that the estimated model has acceptable fit. Note that now an additional matrix ($\boldsymbol{\Theta}_{\delta\varepsilon}$) is estimated (see Table 9).

Table 9
Maximum likelihood estimates of the error covariances (modified model)*

$$\Theta_{\varepsilon} = \begin{pmatrix} .59\,(.04) & & & & & \\ 0 & -.26\,(.09) & & & & \\ .31\,(.03) & 0 & .86\,(.05) & & & \\ 0 & 0 & 0 & .19\,(.03) & & \\ 0 & 0 & 0 & 0 & .14\,(.04) & \\ 0 & 0 & 0 & 0 & .08\,(.02) & .78\,(.05) \end{pmatrix}$$

$$\Theta_{\delta} = \begin{pmatrix} .32\,(.05) & & & \\ 0 & .99\,(.06) & & \\ 0 & .15\,(.04) & .75\,(.05) & \\ 0 & .08\,(.04) & 0 & .57\,(.05) \end{pmatrix}$$

$$\Theta_{\delta\varepsilon} = \begin{pmatrix} 0 & 0 & 0 & 0 & .30\,(.03) \\ 0 & 0 & 0 & 0 & 0 \\ 0 & -.07\,(.02) & 0 & .11\,(.02) & 0 \\ 0 & .11\,(.02) & 0 & 0 & -.29\,(.04) \end{pmatrix}$$

* Standard errors are in the parentheses.

4. Estimation of latent regional development score

Using the parameters of the estimated LISREL model (Tables 8 and 9) we compute scores for the latent variables following the approach of Jöreskog, 2000. Such methods also allow structural recursive and simultaneous relationships among latent variables. Estimation of factor scores in the pure measurement (factor) models is just a special case of the general procedure (see Lawley and Maxwell, 1971).

We describe a technique capable of computing scores of the latent variables based on the maximum likelihood solution of the Eqs. (1-3) following Jöreskog (2000). Writing Eqs. (2) and (3) in a system

$$\begin{pmatrix} \mathbf{y} \\ \mathbf{x} \end{pmatrix} = \begin{pmatrix} \boldsymbol{\Lambda}_{\mathbf{y}} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Lambda}_{\mathbf{x}} \end{pmatrix} \cdot \begin{pmatrix} \boldsymbol{\eta} \\ \boldsymbol{\xi} \end{pmatrix} + \begin{pmatrix} \boldsymbol{\varepsilon} \\ \boldsymbol{\delta} \end{pmatrix}, \tag{28}$$

and using the following notation

$$\mathbf{\Lambda} \equiv \begin{pmatrix} \mathbf{\Lambda}_y & \mathbf{0} \\ \mathbf{0} & \mathbf{\Lambda}_x \end{pmatrix}, \quad \boldsymbol{\xi}_a \equiv \begin{pmatrix} \boldsymbol{\eta} \\ \boldsymbol{\xi} \end{pmatrix}, \quad \boldsymbol{\delta}_a \equiv \begin{pmatrix} \boldsymbol{\varepsilon} \\ \boldsymbol{\delta} \end{pmatrix}, \quad \mathbf{x}_a \equiv \begin{pmatrix} \mathbf{y} \\ \mathbf{x} \end{pmatrix}, \tag{29}$$

the latent scores for the latent variables in the model can be computed using the formula

$$\boldsymbol{\xi}_a = \mathbf{U}\mathbf{D}^{1/2}\mathbf{V}\mathbf{L}^{-1/2}\mathbf{V}^{\mathrm{T}}\mathbf{D}^{1/2}\mathbf{U}^{\mathrm{T}}\mathbf{\Lambda}^{\mathrm{T}}\mathbf{\Theta}_a^{-1}\mathbf{x}_a\,, \tag{30}$$

where $\mathbf{U}\mathbf{D}\mathbf{U}^{\mathrm{T}}$ is the singular varlue decomposition of $\mathbf{\Phi}_a = \mathrm{E}(\boldsymbol{\xi}_a\boldsymbol{\xi}_a^{\mathrm{T}})$, and $\mathbf{V}\mathbf{L}\mathbf{V}^{\mathrm{T}}$ is the singular value decomposition of the matrix $\mathbf{D}^{1/2}\mathbf{U}^{\mathrm{T}}\mathbf{B}\mathbf{U}\mathbf{D}^{1/2}$, while $\mathbf{\Theta}_a$ is the error covariance matrix of the observed variables. Derivation of the Eq. (30) follows the approach of Jöreskog (2000) and Lawley and Maxwell (1971). The latent scores $\boldsymbol{\xi}_{ai}$ can be computed for each observation x_{ij} in the $(10 \times N)$ sample matrix whose rows are observations on each of our 10 observed variables and $N = 545$, i.e.,

$$\begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1N} \\ x_{21} & x_{22} & \cdots & x_{2N} \\ x_{31} & x_{32} & \cdots & x_{3N} \\ x_{41} & x_{42} & \cdots & x_{4N} \\ \vdots & \vdots & \cdots & \vdots \\ x_{10,1} & x_{10,2} & \cdots & x_{10,N} \end{pmatrix} = (\mathbf{x}_1\ \mathbf{x}_2 \cdots \mathbf{x}_N). \tag{31}$$

Having computed the latent scores for η_1, η_2 and ξ_1 we can use the information from the model to rank and compare development level for all territorial units. The model in Fig. 5 implies an underlying economic development level that is simultaneously determined by the structural and demographic factors and measured by several observed indicators. It is clear that such model, in principle, extracts far more information about the development level then using a single observed indicator of latent economic development such as the GDP per capita (the EU methodology for Structural Funds allocation). With latent scores it is now possible to estimate a simple linear equation (using OLS) with η_1 as endogenous variable. This produces the following result

$$\begin{matrix} \eta_1 = 74.170 - 1.102\xi_1 + 0.087\eta_2 \\ \quad (2.291)\ \ (0.033)\quad (0.006) \end{matrix}\,,$$

with R^2 = 0.825 and the Wald test of 1281.7 which indicates very good fit and well determined coefficients. What the above equation suggests is that η_1 is a linear function of η_2 and ξ_1 and thus the territorial units can be ranked either on the grounds of η_1 or by $-.102\xi_1 + .087\eta_2$ (ignoring the constant). Using 100 least developed municipalities (18%) ranked on the bases of $-.102\xi_1 + .087\eta_2$ and η_1 we can compare how many of them entered the first 100 in each of the two cases. The cross-tabulations in Table 10 shows that 70 municipalities were ranked within 100 least developed by both methods (about 13%). The most likely funding allocation given constraints to the national budget will include 5-10% of least developed municipalities, thus this type of summary statistics shows a clear comparative picture regarding classification and ranking performance of the alternative criteria.

Table 10
Crosstabulation count

Criteria		$-.102\xi_1 + .087\eta_2$	η_1	
	Group	< 100	> 100	Total
$-.102\xi_1 + .087\eta_2$	< 100	70	29	99
η_1	> 100	29	417	446
	Total	99	446	545

Similarly, we can use latent scores from alternative models and subsequently cross-tabulate the results checking how many units enter some predefined cut-off criteria such as least developed group of municipalities.

5. Conclusion

A multivariate statistical approach based on covariance structure analysis for assessment of regional development level was suggested and applied to regional development analysis of 545 Croatian municipalities. The commonly used techniques such as factor and principal component analysis were used only in the preliminary data analysis stage and their initial findings as well as hypothesised data structures were then tested with confirmatory factor analytic models estimated with the maximum likelihood technique. We then estimated a recursive structural equation model making explicit assumptions about causality and simultaneity among the latent variables. The final ranking and estimation of the underlying development level was carried out on the grounds of computed latent scores which allowed fulfilment of the project objectives, namely (i) estimation of the latent development level of the territorial units, (ii) ranking of all units on an interval scale in respect to their latent

development level, and (iii) selection of a given percentage of units for inclusion into a national subsidy programme.

Acknowledgments

We wish to acknowledge the support of the Ministry for Public Works, Reconstruction and Construction of the Republic of Croatia in carrying out the project "Criteria for the Development Level Assessment of the Areas Lagging in Development" which served as the basis for this paper. We wish to thank the University of Zagreb, the Croatian Health Insurance Institute, Ministry of Finance, Ministry of Justice, Administration and Local Government, Ministry of the Interior and the Croatian Employment Office for valuable help in data collection.

References

Anderson, T.W. (1984), "An Introduction to Multivariate Statistical Analysis", New York: John Wiley.

Babakus, E., Ferguson, C.E., Jr., and Joreskog, K.G. (1987), "The sensitivity of confirmatory maximum likelihood factor analysis to violations of measurement scale and distributional assumptions", *Journal of Marketing Research*, **24**, pp. 222-28.

Bai, J. and Ng, S. (2002), "Determining the Number of Factors in Approximate Factor Models", *Econometrica*, **70**, pp. 191-221.

Blasius, J. and Greenacre, M. (1998), *Visualization of Categorical Data*. Academic Press.

Bollen, K.A. (1989), *Structural Equations with Latent Variables*. New York: Wiley.

Bollen, K.A. (2001), "Modeling Strategies: In Search of the Holy Grail", *Structural Equation Modeling*, **7**, pp. 74-81.

Borg, I. And Groenen, P. (1997), *Modern Multidimensional Scaling*. Berlin: Springer.

Bowman, K.O. and Shenton, L.R. (1975), "Omnibus Test Contours for Departures from Normality based on $\sqrt{b_1}$ and b_2", *Biometrika*, **62**, pp. 243-50.

Buntig, B., Saris, W.E. and McCormack, J.A. (1987), "Second-order Factor Analysis of the Reliability and Validity of the 11 plus examination in Northern Ireland", *The Economic and Social Review*, **18**, pp. 137-147

Cudek, R. (1989), "Analysis of Correlation Matrices using Covariance Structure Models", **2**, pp. 317-327.

Curran, P.J., West, S. G., and Finch, J. F. (1996), "The robustness of test statistics to nonnormality and specification error in confirmatory factor analysis", *Psychological Methods*, **1**, pp. 16-29.

Belanger, A. and D'Agostino, R. B. (1990), "A suggestion for using powerful and informative tests of normality", *American Statistician*, **44**, pp. 316-321.

D'Agostino, R.B. (1970), "Transformation to Normality of the Null Distribution of g_1", *Biometrika*, **57**, pp. 679-681.

D'Agostino, R.B. (1971), "An Omnibus Test of Normality for Moderate and Large Sample Size", *Biometrika*, **58**, pp. 341-348.

D'Agostino, R.B. (1986), "Tests for Normal Distribution", in D'Agostino, R.B. and Stephens, M.A. (Eds.), *Goodness-of-fit Techniques*. New York: Marcel Dekker, pp. 367-419.

Doornik, J.A., and Hansen, H. (1994), *A Practical Test for Univariate and Multivariate Normality*, Discussion paper, Nuffield College, Oxford.

Everitt, B.S. (1993), *Cluster Analysis*. New York: John Wiley.

Gerbing, D.W. and Anderson, J.C. (1984), "On the Meaning of Within-factor Correlated Measurement Errors", *Journal of Consumer Research*, **11**, pp. 572-580.

Gerbing, D.W., Hamilton, J.G. and Freeman, E.B. (1994), "A Large-scale Second-order Structural Equation Model of the Influence of Management Participation on Organizational Planning Benefits", *Journal of Management*, **20**, pp. 859-885.

Greenacre, M. (1993), *Correspondence Analysis in Practice*. Academic Press.

Greenacre, M. and Blasius. J., ed. (1994), *Correspondence Analysis in the Social Sciences: Recent Developments and Applications*. Academic Press.

Hayduk, L. (1987), *Structural Equation Modeling with LISREL*. Baltimore: John Hopkins University Press.

Hayduk, L. (1996), *LISREL: Issues, Debates and Strategies*. Baltimore: John Hopkins University Press.

Hendry, D.F. and Doornik, J.A. (1999), *Empirical Econometric Modelling Using PcGive*, Vol. I. West Wickham: Timberlake Consultants.

Jöreskog, K.G. (1973), "A General Method for Estimating a Linear Structural Equation System", in Goldberger, A.S. and Duncan, O.D. (Eds.), *Structural Equation Models in the Social Sciences*. New York: Seminar Press.

Jöreskog, K.G., Sörbom, D. (1996), *PRELIS 2 User's Reference Guide*. Chicago: Scientific Software International.

Jöreskog, K.G. (1999), "Formulas for skewness and kurtosis", Scientific Software International, http://www.ssicentral.com/lisrel.

Jöreskog, K.G. (2000), "Latent variable scores and their uses". Scientific Software International, http://www.ssicentral.com/lisrel.

Jöreskog, K.G., Sörbom, D. du Toit, S. and du Toit, M. (2000), *LISREL 8: New Statistical Features*. Chicago: Scientific Software International.

Jöreskog, K.G. and Sörbom, D. (2001), *LISREL 8: New Statistical Features*. Chicago: Scientific Software International.

Kaplan, D. and Elliott, P.R. (1997a), "A Model-based Approach to Validating Educational Indicators using Multilevel Structural Equation Modeling", *Journal of Educational and Behavioral Statistics*, **22**, pp. 232-348.

Kaplan, D. and Elliott, P.R. (1997b), "A Didactic Example of Multilevel Structural Equation Modeling Applicable to the Study of Organizations", *Structural Equation Modeling*, **4**, pp. 1-24.

Lawley, D.N. and Maxwell, A.E. (1971), *Factor Analysis as a Statistical Method*. 2nd ed. London: Butterworths.

Lipshitz, G. and Raveh, A. (1994), "Application of the Co-plot method in the study of socio-economic differences among cities: a basis for a differential development policy", *Urban Studies*, **31**, pp. 123-135.

Lipshitz, G. and Raveh, A. (1998), "Socio-economic Differences among Localities: A New Method of Multivariate Analysis", *Regional Studies*, **32**, pp. 747-757.

Maleković, S. (2001), *Criteria for the Development Level Assessment of the Areas Lagging in the Development*. Zagreb: IMO.

Mardia, K.V. (1980), "Tests of Univariate and Multivariate Normality", in Krishnaiah, P.R. (Ed.), *Handbook of Statistics*, Vol. I. Amsterdam: North Holland, pp. 279-320.

Mulaik, S.A. (1972), *The Foundation of Factor Analysis*. New York: McGraw-Hill.

Mulaik, S.A. and Quartetti, D.A. (1997), "First Order or Second Order General Factor?", *Structural Equation Modeling*, **4**, pp.193-211.

Shenton, L.R. and Bowman, K.O. (1977), "A Bivariate Model for the Distribution of $\sqrt{b_1}$ and b_2", Journal of the American Statistical Association, **72**, pp. 206-211.

Soares, J.O., Marquês, M.M.L. and Monteiro, C.M.F. (2002), "A multivariate methodology to uncover regional disparities: A contribution to improve European Union and governmental decisions", *European Journal of Operational Research*. Forthcoming.

Sörbom, D. (1989), "Model Modification", *Psychometrika*, **54**, pp. 371-384.

Tacq, J. (1998), *Multivariate Analysis Techniques in Social Science Research*. London: Sage.

Weller, S.C. and Romney, A.K. (1990), *Metric Scaling: Correspondence Analysis*. London: Sage.

West, S. G., Finch, J.F. and Curran, P.J. (1995), "Structural equation models with nonnormal variables: Problems and remedies", in Hoyle R.H. (Ed.), *Structural Equation Modeling: Concepts, Issues, and Applications*. Thousand Oaks, CA: Sage Publications, pp. 56-75.