Available online at www.sciencedirect.com
SCIENCE DIRECT®
European Journal of Operational Research xxx (2005) xxx–xxx
EUROPEAN JOURNAL OF OPERATIONAL RESEARCH
www.elsevier.com/locate/ejor

O.R. Applications

Regional development assessment: A structural equation approach

Dario Cziráky [a,*], Jože Sambt [b], Jože Rovan [b], Jakša Puljiz [c]

[a] Department of Statistics, London School of Economics, Houghton Street, WC2A 2AE London, UK
[b] Faculty of Economics, University of Ljubljana, Kardeljeve Ploščad 17, Ljubljana 1101, Slovenia
[c] Macroeconomics Research Department, IMO 2 Farkaša Vukotinovića, Zagreb 10000, Croatia

Received 14 March 2004; accepted 8 March 2005

Abstract

We propose a multivariate statistical framework for regional development assessment based on structural equation modelling with latent variables and show how such methods can be combined with non-parametric classification methods such as cluster analysis to obtain development grouping of territorial units. This approach is advantageous over the current approaches in the literature in that it takes account of distributional issues such as departures from normality in turn enabling application of more powerful inferential techniques; it enables modelling of structural relationships among latent development dimensions and subsequently formal statistical testing of model specification and testing of various hypothesis on the estimated parameters; it allows for complex structure of the factor loadings in the measurement models for the latent variables which can also be formally tested in the confirmatory framework; and enables computation of latent variable scores that take into account structural or causal relationships among latent variables and complex structure of the factor loadings in the measurement models. We apply these methods to regional development classification of Slovenia and Croatia.

Keywords: Multivariate statistics; Regional development; Structural equation modelling; Cluster analysis

1. Introduction

Assessment of the level of development of territorial units is crucial for regional planning and development policy and is a key criterion for allocation of various structural funds and national subsidies in the

* Corresponding author. Tel.: +44 20 7955 6014
E-mail address: d.ciraki@lse.ac.uk (D. Cziráky).

0377-2217/$ - see front matter © 2005 Elsevier B.V. All rights reserved.
doi:10.1016/j.ejor.2005.03.012

European Union (EU). The EU uses a simple approach based on GDP per capita PPS (purchasing power standards) data to classify European regions into net-receivers and net-payers (NUTS-2 classification). However, there are several major weaknesses associated with this single-criteria approach including too small emphases placed on the socio-economic distinctions and the lack of deeper analysis that takes into account smaller geographical units and a broader spectrum of indicators then merely GDP per capita (Soares et al., 2003; Lipshitz and Raveh, 1998). In addition, the EU regional assessment methodology will be further challenged in 2004 when the Union will expand to include the new members, whose former territorial division was based on classifications different from the EU-NUTS system.

There are several different approaches to regional development level assessment in the literature, most often some form of classification and data reduction is employed (see Cziráky et al., 2003b for a review). Soares et al. (2003) proposed a multivariate methodological framework based on multiple indicators of regional development aimed at improving the EU practice in regional development assessment and allocation of the EU Structural Funds. The statistical framework of Soares et al. (2003) is based on a combination of exploratory factor analysis and cluster analysis, which enabled them to identify a smaller number of development factors and to subsequently classify the Portuguese municipalities according to their regional development level.

While promising in respect to suggesting a more elaborate and efficient approach to regional development assessment from the one currently used in the EU, the statistical framework of Soares et al. (2003) can be extended in several important areas thereby increasing the relevance and implementation potential of the multivariate methods in policy making. The methods suggested by Soares et al. (2003) are exploratory and hence do not allow formal testing of alternative model specifications on the basis of statistical inference (i.e. in terms of goodness-of-fit tests). Without formal testing it would be difficult to chose among alternative models having different specifications or different regional development indicators, which would require structural modelling with inferential estimation procedures.

Subsequently, the results of such analyses tend to be difficult to interpret to the policymakers and can appear as ad hoc. The underlying regional development factors represent unobserved (latent) quantities that are likely to be highly inter-correlated or even structurally related. Such relationships thus cannot be modelled or even accounted for by exploratory factor analysis with orthogonal rotation. In addition, Soares et al. (2003) do not take account of the possible non-normality of the development indicators, hence being limited to less formal non-parametric methods only. Exploratory analysis, however, is useful as a preliminary or descriptive tool, but should be complemented with more powerful estimation methods.

In this paper we propose a methodological framework for regional development assessment based on structural equation modelling with latent variables that treat various development dimensions as latent variables and enable formal modelling of causal recursive and non-recursive relationships among them and, in the same time, provide formal evaluation and fit statistics (Cziráky et al., 2002a,b, 2003a,b).

Such methods can be used in combination with both exploratory factor analysis and cluster analysis additionally providing important advantages such as structural modelling of regional development and statistical testing of the postulated models. Furthermore, an auxiliary non-parametric method such as cluster analysis can be applied to rank the analysed territorial units according to their latent development dimensions. We apply these methods to regional development assessment and ranking of Slovenia and Croatia and compare the specification of the structural model across the two countries thereby enabling cross-sample validation.

The paper is organised as follows. In the second part the data is described and the necessary descriptive statistical analysis is presented. In addition, normality tests are reported for untransformed and transformed variables, where the normal scores technique was used for normalisation. The statistical methodology and estimation methods are described in the third section. Fourth section presents model specification and estimation results for structural equation models, it also describes a technique for computing latent scores from structural equation models. Fifth section presents the results from cluster analysis and the last section concludes.

2. Data and descriptive statistics

In this paper we use municipality data from Slovenia and Croatia, which present the lowest aggregation level available for both countries. The primary source of Slovenian data (see Table 1) was Statistical Office of the Republic of Slovenia (SORS); in some cases the data were published and/or the necessary calculations on data were already done by the Institute of Macroeconomic Analysis and Development (IMAD). We collected Slovenian data on 9 regional development indicators, mostly from the SORS/IMAD sources. The source of the *social aid per capita* (y_3) variable was the Slovenian Ministry of Labour, Family and Social Affairs. The *number of cars per 100 inhabitants* (y_7) was aggregated by Grobler (2002) from micro data provided by the Slovenian Ministry of Interior. The Slovenian census was carried out in 2002 and the final census data were not available at the time of this analysis.

The Croatian data came from the 2001 national census (State Bureau of Statistics). The census data has the advantage of being of higher quality and, as it comes from a single source, it is also less ambiguous. We collected Croatian data on 11 development indicators (Table 1). Moreover, municipalities are the basic territorial units in legal classification of the Croatian territories and are also the basic units used for classification of the Areas of Special State Concern (i.e., national subsidy allocation).

Table 1
Definitions of the variables and notation

Variable description	Symbol[c]
Slovenian data	
Income per capita (in SIT), 2002	y_1
Employment/population ratio, I–IX 2002	y_2
Social aid per capita (in thousands SIT), VI 2002	y_3
Share of agricultural population, VI 2002	y_4
Density (inhabitants per km^2), 30.6.2002	y_5
Students share per 1000 inhabitants (2001–2002)[a]	y_6
Number of cars per 100 inhabitants, 1999	y_7
Age index (65 + /(0 − 14)), 30.6.2002	x_1
Population trend (population 2001/population 1991)	x_2
Croatian data[b]	
Income per capita (in HRK)	$\hat{y}_1$
Population share making income (%)	$\hat{y}_2$
Municipality income per capita (in thousands HRK)	$\hat{y}_3$
Employment/population ratio	$\hat{y}_4$
Social aid per capita (in thousands HRK)	$\hat{y}_5$
Share of agricultural population	$\hat{y}_6$
Education (share of high-school graduates in total population)	$\hat{y}_7$
Age index (65 + /(0 − 20))	$\hat{y}_8$
Population trend (population 2001/population 1991)	$\hat{x}_1$
Density (inhabitants per km^2)	$\hat{x}_2$
Vitality index (live births over number of deceased)	$\hat{x}_3$

[a] Undergraduate students enrolled in the higher education institutions.
[b] Croatian data is from the 2001 census. The population figure for 1991 used to compute $\hat{x}_1$ came from the 1991 census.
[c] The symbols with the "hat" are used to denote Croatian variables to keep the x–y notation.

Table 2 reports results of the normality tests for all variables (D'Agostino, 1986; Doornik and Hansen, 1994; Mardia, 1980). It can be easily seen that most variables are not distributed normally, as the reported normality chi-square (χ^2) tests strongly reject the null hypothesis. The exceptions are *income per capita* (y_1) and *employment* (y_2) for Slovenia, which appear to be normally distributed, thus needing no additional transformation. Because we wish to use Gaussian maximum likelihood techniques in further analysis, it is necessary to have variables that are approximately normally distributed. Therefore, 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 (Babakus et al., 1987; Curran et al., 1996; West et al., 1995).

For this purpose we apply the *normal scores* (NS) technique (Jöreskog et al., 2000; Jöreskog, 1999). Similar transformation of regional development data were applied in Cziráky et al. (2002a) and Cziráky et al. (2002b). We note that the NS technique is widely applicable with other types of data (see Cziráky and Čumpek (2002) for a macro-economic application and Cziráky et al. (2002c) for an application in environmental sciences). Given a sample of N observations on the jth variable, $\mathbf{x}_j = (x_{j1}, x_{j2}, x_{jN})$, the normal scores transformation is computed in the following way. First define a vector of k distinct sample values, $\mathbf{x}_j^k = (x'_{j1}, x'_{j2}, , x'_{jk})$ where $k \leqslant N$ thus $\mathbf{x}^k \subseteq \mathbf{x}$. Let f_i be the frequency of occurrence of the value x_{ji} in x_j so that $f_{ji} \geqslant 1$. Then normal scores x_{ji}^{NS} are computed as $x_{ji}^{\mathrm{NS}} = (N/f_{ji})[\phi(\alpha_{j,i-1}) - \phi\alpha_{ji})]$ where ϕ is the standard Gaussian density function and α is defined as

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

Table 2
Normality tests (raw data)[a]

Variable	Skewness		Kurtosis		Skewness and Kurtosis	
	z-Score	*p*-Value	*z*-Score	*p*-Value	χ^2	*p*-Value
Slovenian data						
$\hat{y}_1$	0.572	0.567	0.282	0.778	0.407	0.816
$\hat{y}_2$	−1.298	0.194	−0.522	0.602	1.957	0.376
$\hat{y}_3$	7.099	0.000	5.468	0.000	80.298	0.000
$\hat{y}_4$	7.088	0.000	4.199	0.000	67.862	0.000
$\hat{y}_5$	10.765	0.000	7.762	0.000	176.126	0.000
$\hat{y}_6$	5.469	0.000	6.581	0.000	73.228	0.000
$\hat{y}_7$	4.035	0.000	2.362	0.018	21.856	0.000
$\hat{x}_1$	10.545	0.000	8.023	0.000	175.557	0.000
$\hat{x}_2$	3.425	0.001	4.542	0.000	32.361	0.000
Croatian data						
$\hat{y}_1$	2.869	0.004	−3.765	0.000	22.408	0.000
$\hat{y}_2$	−2.590	0.010	−2.165	0.030	11.397	0.003
$\hat{y}_3$	16.112	0.000	10.876	0.000	377.894	0.000
$\hat{y}_4$	4.237	0.000	3.414	0.001	29.611	0.000
$\hat{y}_5$	18.271	0.000	12.902	0.000	500.317	0.000
$\hat{y}_6$	11.233	0.000	6.070	0.000	163.022	0.000
$\hat{y}_7$	2.853	0.004	−2.101	0.036	12.553	0.002
$\hat{y}_8$	31.629	0.000	17.529	0.000	1307.683	0.000
$\hat{x}_1$	−2.826	0.005	6.781	0.000	53.967	0.000
$\hat{x}_2$	25.330	0.000	15.886	0.000	893.997	0.000
$\hat{x}_3$	10.209	0.000	6.970	0.000	152.794	0.000

[a] Normality tests were computed with the PRELIS 2 computer programme (Jöreskog and Sörbom, 1996).

Table 3
Normality tests (normalised data)[a]

Variable	Skewness		Kurtosis		Skewness and Kurtosis	
	z-Score	*p*-Value	*z*-Score	*p*-Value	χ^2	*p*-Value
Slovenian data						
$\hat{y}_1$	0.572	0.567	0.282	0.778	0.407	0.816
$\hat{y}_2$	1.298	0.194	0.522	0.602	1.957	0.376
$\hat{y}_3$	0.000	1.000	0.100	0.920	0.010	0.995
$\hat{y}_4$	0.000	1.000	0.100	0.920	0.010	0.995
$\hat{y}_5$	0.000	1.000	0.100	0.920	0.010	0.995
$\hat{y}_6$	0.000	1.000	0.101	0.920	0.010	0.995
$\hat{y}_7$	0.000	1.000	0.100	0.920	0.010	0.995
$\hat{x}_1$	0.005	0.996	0.107	0.914	0.012	0.994
$\hat{x}_2$	0.001	1.000	0.100	0.920	0.010	0.995
Croatian data						
$\hat{y}_1$	0.000	1.000	0.065	0.948	0.004	0.998
$\hat{y}_2$	0.000	1.000	0.065	0.948	0.004	0.998
$\hat{y}_3$	0.000	1.000	0.065	0.948	0.004	0.998
$\hat{y}_4$	0.000	1.000	0.065	0.948	0.004	0.998
$\hat{y}_5$	0.000	1.000	0.065	0.948	0.004	0.998
$\hat{y}_6$	0.000	1.000	0.065	0.948	0.004	0.998
$\hat{y}_7$	0.001	0.999	0.064	0.949	0.004	0.998
$\hat{y}_8$	0.000	1.000	0.065	0.948	0.004	0.998
$\hat{x}_1$	0.000	1.000	0.065	0.948	0.004	0.998
$\hat{x}_2$	0.000	1.000	0.065	0.948	0.004	0.998
$\hat{x}_3$	0.001	1.000	0.064	0.949	0.004	0.998

[a] Normality tests were computed with the PRELIS 2 computer programme (Jöreskog and Sörbom, 1996).

where 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.

Table 3 shows the results of the normality tests computed for the normalised variables (note that the two originally normally distributed variables were not transformed). It is apparent that normalisation procedure successfully removed departures from normality.

3. Structural equation methodology

The proposed econometric methodology aims to model regional development using structural equations methodology based on latent variable "LISREL" models (Jöreskog et al., 2000; Cziráky, 2004). The model is specified as a special case of the general LISREL model as follows. Denoting the latent endogenous variables by $\boldsymbol{\eta}$ and latent exogenous variables by $\boldsymbol{\xi}$, and their respective observed indicators by y and x, the structural part of the model is given by

$$\boldsymbol{\eta} = \mathbf{B}\boldsymbol{\eta} + \boldsymbol{\Gamma}\boldsymbol{\xi} + \boldsymbol{\zeta}, \tag{2}$$

where $\boldsymbol{\eta}$ is the vector of latent endogenous variables, $\boldsymbol{\xi}$ is the vector of latent exogenous variables, $\boldsymbol{\zeta}$ is the vector of latent errors and $\mathbf{B}$ and $\boldsymbol{\Gamma}$are coefficient matrices. The measurement models are given in typical factor analytic form as

$$\mathbf{y} = \boldsymbol{\Lambda}_y\boldsymbol{\eta} + \boldsymbol{\varepsilon}, \tag{3}$$

for latent endogenous, and

$$\mathbf{x} = \boldsymbol{\Lambda}_x\boldsymbol{\xi} + \boldsymbol{\delta}, \tag{4}$$

for latent exogenous variables, where $\mathbf{y}$ $(q \times 1)$ and $\mathbf{x}$ $(p \times 1)$ are the vectors of observable variables; $\mathbf{\Lambda}_y$ and $\mathbf{\Lambda}_x$ are coefficient matrices; and $\boldsymbol{\varepsilon}$ and $\boldsymbol{\delta}$ are vectors of latent errors. Using Jöreskog's LISREL notation we also define the following second-moment matrices: $E[\boldsymbol{\xi}\boldsymbol{\xi}'] \equiv \mathbf{\Phi}$, $E[\boldsymbol{\zeta}\boldsymbol{\zeta}'] \equiv \mathbf{\Psi}$, $E[\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}'] \equiv \mathbf{\Theta}_\varepsilon$, $E[\boldsymbol{\delta}\boldsymbol{\delta}'] \equiv \mathbf{\Theta}_\delta$, and $E[\boldsymbol{\varepsilon}\boldsymbol{\delta}'] \equiv \mathbf{\Theta}_{\varepsilon\delta}$. The covariance matrix implied by the model is comprised of three separate covariance matrices: the covariance matrix of the observed indicators of the latent endogenous variables, the covariances between the indicators of latent endogenous variables and indicators of latent exogenous variables, and the covariance matrix of the indicators of the latent exogenous variables. Arranging these three matrices together we get the joint covariance matrix implied by the model, which is given by

$$\mathbf{\Sigma} = \begin{pmatrix} \mathbf{\Sigma}_{yy} & \mathbf{\Sigma}_{yx} \\ \mathbf{\Sigma}_{xy} & \mathbf{\Sigma}_{xx} \end{pmatrix}. \tag{5}$$

Using (2)–(4), the matrix (5) can be written in terms of the model parameters as

$$\mathbf{\Sigma} = \begin{pmatrix} \mathbf{\Lambda}_y(\mathbf{I}-\mathbf{B})^{-1}(\mathbf{\Gamma\Phi\Gamma}' + \mathbf{\Psi})[(\mathbf{I}-\mathbf{B})^{-1}]'\mathbf{\Lambda}_y' + \mathbf{\Theta}_\varepsilon & \mathbf{\Lambda}_y(\mathbf{I}-\mathbf{B})^{-1}\mathbf{\Gamma\Phi\Lambda}_x' + \mathbf{\Theta}_{\delta\varepsilon}' \\ \mathbf{\Lambda}_x\mathbf{\Phi\Gamma}'[(\mathbf{I}-\mathbf{B})^{-1}]'\mathbf{\Lambda}_y' + \mathbf{\Theta}_{\varepsilon\delta} & \mathbf{\Lambda}_x\mathbf{\Phi\Lambda}_x' + \mathbf{\Theta}_\delta \end{pmatrix}. \tag{6}$$

The maximum likelihood estimates of the model parameters, given the model is identified, are obtained by minimisation of the multivariate Gaussian (discrepancy) log-likelihood function

$$F = \ln|\mathbf{\Sigma}| + \mathrm{tr}(\mathbf{S\Sigma}^{-1}) - \ln|\mathbf{S}| - (p+q), \tag{7}$$

where p and q are the numbers of the observed indicators of latent endogenous and latent exogenous variables, respectively (for more details see e.g. Kaplan, 2000). Estimation procedures for structural equation models might be modified when some of all observed variables are categorical (see Cziráky et al. (in press) for an application, and Bartholomew and Knott (1999) for a detailed overview).

Using the parameter matrices from (2)–(4), the scores for latent variables can be computed following the approach of Lawley and Maxwell (1971) and Jöreskog (2000), which can be summarised as follows. Writing (3) and (4) in a system as

$$\begin{pmatrix} \mathbf{y} \\ \mathbf{x} \end{pmatrix} = \begin{pmatrix} \mathbf{\Lambda}_\mathrm{y} & \mathbf{0} \\ \mathbf{0} & \mathbf{\Lambda}_\mathrm{x} \end{pmatrix} \begin{pmatrix} \boldsymbol{\eta} \\ \boldsymbol{\xi} \end{pmatrix} + \begin{pmatrix} \boldsymbol{\varepsilon} \\ \boldsymbol{\delta} \end{pmatrix} \tag{8}$$

and using the following notation

$$\mathbf{\Lambda} \equiv \begin{pmatrix} \mathbf{\Lambda}_\mathrm{y} & \mathbf{0} \\ \mathbf{0} & \mathbf{\Lambda}_\mathrm{x} \end{pmatrix}, \quad \boldsymbol{\xi}_\mathrm{a} \equiv \begin{pmatrix} \boldsymbol{\eta} \\ \boldsymbol{\xi} \end{pmatrix}, \quad \boldsymbol{\delta}_\mathrm{a} \equiv \begin{pmatrix} \boldsymbol{\varepsilon} \\ \boldsymbol{\delta} \end{pmatrix}, \quad \mathbf{x}_\mathrm{a} \equiv \begin{pmatrix} \mathbf{y} \\ \mathbf{x} \end{pmatrix}, \tag{9}$$

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

$$\boldsymbol{\xi}_\mathrm{a} = \mathbf{UD}^{1/2}\mathbf{VL}^{-1/2}\mathbf{V}'\mathbf{D}^{1/2}\mathbf{U}'\mathbf{\Lambda}'\mathbf{\Theta}_\mathrm{a}^{-1}\mathbf{x}_\mathrm{a}, \tag{10}$$

where $\mathbf{UDU}'$ is the singular value decomposition of $\mathbf{\Phi}_\mathrm{a} \equiv E[\boldsymbol{\xi}_\mathrm{a}\boldsymbol{\xi}_\mathrm{a}']$, and $\mathbf{VLV}'$ is the singular value decomposition of the matrix $\mathbf{D}^{1/2}\mathbf{UTBUD}^{1/2}$, while $\mathbf{\Theta}_\mathrm{a}$ is the error covariance matrix of the observed variables. Derivation of (10) follows the approach of Jöreskog (2000) and Lawley and Maxwell (1971) is described in more detail in Cziráky et al. (2002c). The latent scores $\xi_{\mathrm{a}i}$ can be computed for each observation x_{ij} in the $(p+q) \times n$ data matrix whose rows are observations on each of our $p+q$ observed indicators, where $n = q + p$ is the sample size.

4. Model specification and estimation results

A preliminary exploratory factor analysis indicated 4 latent development dimensions (i.e. factors) for both Slovenia and Croatia (see Table 4). However, the factor solution did not display a 'simple structure',

Table 4
Latent variables

Latent development dimensions		Observed indicators	
Description	Symbol	Slovenia	Croatia
Economic	η_1	y_1, y_3, y_5, y_7	$\hat{y}_1, \hat{y}_2, \hat{y}_3, \hat{y}_4, \hat{y}_7, \hat{y}_8$
Structural	η_2	y_2, y_3, y_4	$\hat{y}_2, \hat{y}_4, \hat{y}_5, \hat{y}_6$
Social	η_3	y_4, y_5, y_6	$\hat{y}_6, \hat{y}_7, \hat{y}_8$
Demographic	ξ_1	x_1, x_2	$\hat{x}_1, \hat{x}_2, \hat{x}_3$

rather it indicated complex factor loadings and high correlation among factors, thus suggesting that factors might be structurally (or causally) related. Therefore, treating latent development dimensions as orthogonal and measured by non-overlapping sets of indicators would be misleading in this particular application, and this is likely to hold in general in regional development research. It follows that the use of exploratory factor analysis for extracting assumingly simple structure (possibly accompanying with an orthogonal rotation that assumes orthogonally of factors) for more then preliminary descriptive analysis (see e.g. Soares et al., 2003) might be inappropriate in this context.

The structure of factor loadings is likely to be complex with ambiguous loadings and structural relationships among factors are possible which requires confirmatory modelling accompanied by formal testing of the model's fit and hence its postulated specification and structure.

In our application we specify a simple non-recursive structural model for both Slovenia and Croatia of the form

$$\begin{pmatrix} \eta_1 \\ \eta_2 \\ \eta_3 \end{pmatrix} = \begin{pmatrix} 0 & \beta_{12} & \beta_{13} \\ 0 & 0 & \beta_{23} \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} \eta_1 \\ \eta_2 \\ \eta_3 \end{pmatrix} + \begin{pmatrix} \gamma_{11} \\ \gamma_{21} \\ \gamma_{31} \end{pmatrix} \xi_1 + \begin{pmatrix} \zeta_1 \\ \zeta_2 \\ \zeta_3 \end{pmatrix}. \tag{11}$$

The model (11) postulates four (partly overlapping) development dimensions each measured by a factor-analytic measurement model. As the measurement models generally have complex structure, separate factors are occasionally allowed to load on the same indicators. The latent variables and their indicators (using notation from Table 1) for Slovenia and Croatia are given in Table 4. Substantively, these four latent variables aim to approximately capture economic, structural, social, and demographic development dimensions, although slightly different substantive interpretation might be given to these factors. As our focus is on methodology, we do not pursue any further substantive interpretation of the factors.

While we specify the structural part of the model for both countries equally, data availability issues render certain differences in the measurement models expectable (Table 4). Specifically, the endogenous measurement model for Slovenia is formally specified by the following matrix equation

$$\begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \\ y_5 \\ y_6 \\ y_7 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ \lambda_{31}^{(y)} & \lambda_{32}^{(y)} & 0 \\ 0 & \lambda_{42}^{(y)} & \lambda_{43}^{(y)} \\ \lambda_{51}^{(y)} & 0 & \lambda_{53}^{(y)} \\ 0 & 0 & 1 \\ \lambda_{71}^{(y)} & 0 & 0 \end{pmatrix} \begin{pmatrix} \eta_1 \\ \eta_2 \\ \eta_3 \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \varepsilon_4 \\ \varepsilon_5 \\ \varepsilon_6 \\ \varepsilon_7 \end{pmatrix}, \tag{12}$$

while the exogenous measurement model is formulated as

$$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} \lambda_{11}^{(x)} \\ \lambda_{21}^{(x)} \end{pmatrix} \xi_1 + \begin{pmatrix} \delta_1 \\ \delta_2 \end{pmatrix}. \tag{13}$$

In the estimation, the covariance matrix of the latent errors in the endogenous measurement model was firstly set to a diagonal matrix; however, preliminary analysis and modification indices (see Sörbom, 1989) suggested that relaxing the zero restriction on $\theta_{42}^{(\varepsilon)}$ improves the fit of the model, thus we specified the $\mathbf{\Theta}_\varepsilon$ matrix as

$$\mathbf{\Theta}_\varepsilon = \begin{pmatrix} \theta_{11}^{(\varepsilon)} & & & & & & \\ 0 & \theta_{22}^{(\varepsilon)} & & & & & \\ 0 & 0 & \theta_{33}^{(\varepsilon)} & & & & \\ 0 & \theta_{42}^{(\varepsilon)} & 0 & \theta_{44}^{(\varepsilon)} & & & \\ 0 & 0 & 0 & 0 & \theta_{55}^{(\varepsilon)} & & \\ 0 & 0 & 0 & 0 & 0 & \theta_{66}^{(\varepsilon)} & \\ 0 & 0 & 0 & 0 & 0 & 0 & \theta_{77}^{(\varepsilon)} \end{pmatrix}, \tag{14}$$

noting that $\theta_{42}^{(\varepsilon)}$ is residual correlation between the share of agricultural population and employment share indicators. Estimation of $\theta_{42}^{(\varepsilon)}$ parameter resulted in significant decrease in the χ^2 statistic from 108 (d.f. = 18) to 67.22 (d.f. = 17). Finally, the covariance matrix for the latent errors in the exogenous measurement model is specified as diagonal matrix of the form

$$\mathbf{\Theta}_\delta = \begin{pmatrix} \theta_{11}^{(\delta)} & \\ 0 & \theta_{22}^{(\delta)} \end{pmatrix}. \tag{15}$$

As noted above, estimation of the system (12)–(15), with the structural part from (11), with Slovenian data (Table 5) produced an overall-fit χ^2 statistic of 67.22 (d.f. = 17) with the goodness of fit index (GFI) = 0.927 and standardised root mean square residual (SRMR) = 0.057, which shows relatively good fit to the data. We note that the estimated model had no significant modification indices and no remaining residual correlation was left un-modelled. The full maximum likelihood estimates of all model parameters are given in Table 6.

For Croatia, the endogenous measurement model is specified as

$$\begin{pmatrix} \hat{y}_1 \\ \hat{y}_2 \\ \hat{y}_3 \\ \hat{y}_4 \\ \hat{y}_5 \\ \hat{y}_6 \\ \hat{y}_7 \\ \hat{y}_8 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ \lambda_{21}^{(y)} & \lambda_{22}^{(y)} & 0 \\ \lambda_{31}^{(y)} & 0 & 0 \\ 0 & 1 & 0 \\ \lambda_{51}^{(y)} & \lambda_{52}^{(y)} & 0 \\ 0 & \lambda_{62}^{(y)} & \lambda_{63}^{(y)} \\ 0 & 0 & 1 \\ \lambda_{81}^{(y)} & 0 & \lambda_{83}^{(y)} \end{pmatrix} \begin{pmatrix} \eta_1 \\ \eta_2 \\ \eta_3 \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \varepsilon_4 \\ \varepsilon_5 \\ \varepsilon_6 \\ \varepsilon_7 \\ \varepsilon_8 \end{pmatrix}, \tag{16}$$

while the exogenous measurement model is given by

$$\begin{pmatrix} \hat{x}_1 \\ \hat{x}_2 \\ \hat{x}_3 \end{pmatrix} = \begin{pmatrix} \lambda_{11}^{(x)} \\ 1 \\ \lambda_{31}^{(x)} \end{pmatrix} \xi_1 + \begin{pmatrix} \delta_1 \\ \delta_2 \\ \delta_3 \end{pmatrix}. \tag{17}$$

Table 5
Correlation matrices (normalised data)

	y_1	y_2	y_3	y_4	y_5	y_6	y_7	x_1	x_2		
Slovenian data											
y_1	1.000										
y_2	0.582	1.000									
y_3	−0.681	−0.591	1.000								
y_4	−0.779	−0.196	0.458	1.000							
y_5	0.329	−0.017	0.063	−0.436	1.000						
y_6	0.647	0.338	−0.296	−0.504	0.393	1.000					
y_7	0.868	0.515	−0.632	−0.643	0.227	0.526	1.000				
x_1	−0.133	−0.259	0.241	0.114	−0.328	−0.233	−0.063	1.000			
x_2	0.547	0.420	−0.596	−0.434	0.633	0.296	0.327	0.005	1.000		
Croatian data											
	$\hat{y}_1$	$\hat{y}_2$	$\hat{y}_3$	$\hat{y}_4$	$\hat{y}_5$	$\hat{y}_6$	$\hat{y}_7$	$\hat{y}_8$	$\hat{x}_1$	$\hat{x}_2$	$\hat{x}_3$
$\hat{y}_1$	1.000										
$\hat{y}_2$	0.478	1.000									
$\hat{y}_3$	0.716	0.371	1.000								
$\hat{y}_4$	0.034	0.715	0.026	1.000							
$\hat{y}_5$	−0.396	−0.577	−0.384	−0.520	1.000						
$\hat{y}_6$	−0.646	0.080	−0.453	0.528	0.014	1.000					
$\hat{y}_7$	−0.013	−0.281	0.047	−0.001	0.117	0.136	1.000				
$\hat{y}_8$	0.235	0.157	0.278	0.226	0.443	−0.165	−0.539	1.000			
$\hat{x}_1$	0.789	−0.283	0.641	−0.072	−0.367	−0.684	−0.208	0.440	1.000		
$\hat{x}_2$	0.046	0.338	0.006	−0.134	−0.032	−0.214	−0.802	0.417	0.250	1.000	
$\hat{x}_3$	0.259	0.137	0.151	0.193	−0.304	−0.251	−0.629	0.589	0.492	0.508	1.000

The error-covariance matrix of the endogenous measurement model $\mathbf{\Theta}_\delta$ is diagonal. The exogenous error-covariance matrix $\mathbf{\Theta}_\varepsilon$ matrix was initially specified as

$$\mathbf{\Theta}_\delta = \begin{pmatrix} \theta_{11}^{(\delta)} & & \\ 0 & \theta_{22}^{(\delta)} & \\ 0 & 0 & \theta_{33}^{(\delta)} \end{pmatrix} \tag{18}$$

and after initial estimation we relaxed the zero restriction on $\theta_{31}^{(\delta)}$ and re-estimated the model with $\mathbf{\Theta}_\delta$ matrix specified as

$$\mathbf{\Theta}_\delta = \begin{pmatrix} \theta_{11}^{(\delta)} & & \\ 0 & \theta_{22}^{(\delta)} & \\ \theta_{31}^{(\delta)} & 0 & \theta_{33}^{(\delta)} \end{pmatrix}, \tag{19}$$

which resulted in a significant decrease in the c2 from 88.65 (d.f. = 34) to 75.57 (d.f. = 33). Using Croatian data, estimation of the system (16)–(19), with specification (11) for the structural equations, produced a χ^2 of 75.57 (d.f. = 33), GFI = 0.98, and SRMR = 0.04, which indicate an approximately good fit to the data. The full parameter estimates are shown in Table 6 (note that particular symbols do not necessarily indicate comparable coefficients because the dimensions of the measurement models as well as the observed indicators themselves differed between the two countries; the structural parameters are, however, identical for both countries and can be thus directly compared).

By comparing the structural equations part of the model (11) between the two countries (see Table 6) we can note that the effect of social factor on economic dimension is positive, strong, highly significant, and of

Table 6
Maximum likelihood estimates

Parameter	Slovenian model		Croatian model	
	Estimate	(S.E.)	Estimate	(S.E.)
$\lambda_{21}^{(y)}$	–	–	0.474	(0.081)
$\lambda_{22}^{(y)}$	–	–	0.789	(0.089)
$\lambda_{31}^{(y)}$	0.391	(0.263)	0.788	(0.088)
$\lambda_{32}^{(y)}$	−1.859	(0.435)	–	–
$\lambda_{42}^{(y)}$	−0.082	(0.159)	–	–
$\lambda_{43}^{(y)}$	−1.109	(0.162)	–	–
$\lambda_{51}^{(y)}$	−2.612	(1.162)	−0.466	(0.080)
$\lambda_{52}^{(y)}$	–	–	−0.552	(0.091)
$\lambda_{53}^{(y)}$	4.619	(1.801)	–	–
$\lambda_{62}^{(y)}$	–	–	0.562	(0.094)
$\lambda_{63}^{(y)}$	–	–	−0.788	(0.092)
$\lambda_{64}^{(y)}$	–	–	3.170	(0.967)
$\lambda_{71}^{(y)}$	0.836	(0.038)	−2.559	(0.932)
$\lambda_{11}^{(x)}$	0.572	(0.074)	−0.943	(0.149)
$\lambda_{12}^{(x)}$	−1.106	(0.079)	–	–
$\lambda_{31}^{(x)}$	–	–	0.742	(0.136)
β_{12}	0.557	(0.178)	−0.014	(0.030)
β_{13}	1.100	(0.170)	1.165	(0.131)
β_{23}	0.586	(0.128)	−0.052	(0.106)
γ_{11}	0.082	(0.030)	−0.391	(0.156)
γ_{21}	−0.199	(0.060)	0.222	(0.140)
γ_{31}	−0.347	(0.056)	0.570	(0.129)
$\mathrm{Var}(\zeta_1)$	0.031	(0.012)	0.029	(0.032)
$\mathrm{Var}(\zeta_2)$	0.164	(0.052)	0.963	(0.168)
$\mathrm{Var}(\zeta_3)$	0.334	(0.064)	0.616	(0.103)
$\theta_{11}^{(\varepsilon)}$	−0.036	(0.017)	1.092	(0.105)
$\theta_{22}^{(\varepsilon)}$	0.559	(0.061)	1.298	(0.113)
$\theta_{33}^{(\varepsilon)}$	0.164	(0.076)	1.435	(0.105)
$\theta_{44}^{(\varepsilon)}$	0.369	(0.039)	1.014	(0.147)
$\theta_{55}^{(\varepsilon)}$	0.121	(0.221)	1.513	(0.109)
$\theta_{66}^{(\varepsilon)}$	0.545	(0.056)	1.218	(0.110)
$\theta_{77}^{(\varepsilon)}$	0.276	(0.030)	1.199	(0.096)
$\theta_{88}^{(\varepsilon)}$	–	–	1.081	(0.270)
$\theta_{42}^{(\varepsilon)}$	0.201	(0.037)	–	–
$\theta_{11}^{(\delta)}$	0.672	(0.079)	1.494	(0.125)
$\theta_{22}^{(\delta)}$	−0.223	(0.146)	1.431	(0.125)
$\theta_{33}^{(\delta)}$	–	–	1.687	(0.122)
$\theta_{32}^{(\delta)}$	–	–	−0.404	(0.095)
χ^2	67.224		75.565	
d.f.	17		33	
GFI[a]	0.927	16	0.975	
SRMR[b]	0.057		0.042	

similar magnitude for both countries. The effect of structural factor on the economic one is positive and significant in Slovenia, while in Croatia it is of much smaller magnitude and negative. Another difference can be seen in the effect of demographic factor on structural and social dimensions. Namely, demographic factor seems to affect structural dimension negatively in Slovenia and positively in Croatia while its effect on the economic dimension is significant and positive in Slovenia and insignificant in Croatia.

This, and also an observed difference in the effect of demographic on social dimension which is negative in Slovenia and positive in Croatia, is actually a consequence of normalisation as in the Croatian case the demographic measurement model was normalised in respect to density, which was not the case in the Slovenian model, thus the opposite signs of the estimated coefficients were expected.

An important difference in endogenous measurement models between the two countries can be observed in the relationship between the share of agricultural population and employment, which is positive in Croatia, where large unemployment trend affects primarily urban areas, and negative in Slovenia, where agricultural areas appear to suffer from lower employment (see Tables 5 and 6). This difference is the most likely cause of different signs of the effect of structural on economic factors between the two countries.

So far we have estimated an econometric model for regional development using Slovenian and Croatian data with fully parametric inferential procedures (maximum likelihood), tested specific model formulation and assessed model fit.

5. Clustering territorial units

Having computed the latent variable scores ξ_{ai} using (10), we perform cluster analysis with the purpose of grouping (clustering) municipalities into several groups with similar characteristics (for more details see Everitt, 2001; also see Soares et al., 2003 for an application using scores from an exploratory factor model).

In the first step, we used the Ward hierarchical procedure to define the number of clusters and the group centroids. The graphical presentation of results with dendrogram (Fig. 1) suggests three clusters for both countries. In the second step we used the *K*-means method by taking the centroids from the Ward (hierarchical) method as initial seed-points for the *K*-means method (see e.g. Ferligoj, 1989; Rovan and Sambt, 2003).

While clustering of the original variables has the advantage of producing results in terms of directly observable quantities (e.g. collected by governmental agencies), the clustering of the latent variables provided similar picture while being more clear in interpretation since a lower number of centroids had to be analysed, which indicates an advantage of clustering on the basis of latent variables. On the other hand, a procedure based on factor scores obtained from exploratory factor analysis (e.g. Soares et al., 2003) assumes that factors are orthogonal in population (as an orthogonal rotation has to be used), and ignores a possibly complex structure, specially cases with ambiguous (compound) loadings and causal relationships.

The 3-cluster solution converged in 10 iterations with Slovenian and 8 with Croatian data. Table 7 gives the ANOVA results, which indicate highly significant discriminatory power of each latent variable. We note, however that ANOVA results in this context present merely a descriptive tool and are not adjusted

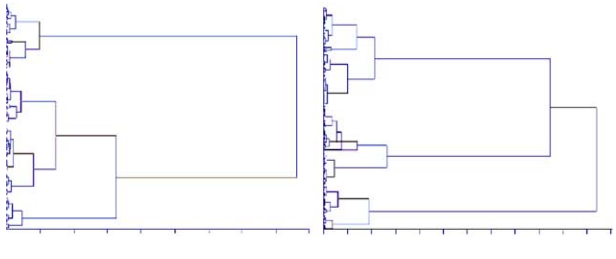

Fig. 1. Dendrogram for Slovenia (left) and Croatia (right).

Table 7
Analysis of variance (ANOVA) table (d.f. = 189)

	Mean square	Mean square error	F-test	p-Value
Slovenian data				
η_1	73.724	0.283	260.368	0.000
η_2	28.567	0.165	172.704	0.000
η_3	28.611	0.143	200.091	0.000
ξ_1	43.491	0.550	79.031	0.000
Croatian data				
η_1	48.623	0.128	381.302	0.000
η_2	98.439	0.291	338.142	0.000
η_3	51.701	0.143	361.301	0.000
ξ_1	2.848	0.153	18.617	0.000

Table 8
Final cluster centers

	Cluster 1 $N = 55$	Cluster 2 $N = 89$	Cluster 3 $N = 48$
Slovenian data			
η_1	2.18	1.04	−0.22
η_2	1.30	0.45	−0.18
η_3	1.16	0.46	−0.33
ξ_1	−0.90	0.05	0.93
Croatian data			
	$N = 204$	$N = 115$	$N = 227$
η_1	1.59	0.36	0.57
η_2	0.18	1.32	−0.82
η_3	1.71	0.38	0.69
ξ_1	0.03	−0.26	0.10

either for the fact that variables were clustered or for the fact that the criterion variables for clustering were linear combinations of the observable development indicators. Table 8 shows centroids for each cluster (expressed in standardised units).

For Slovenia, cluster 1 consist of Ljubljana, and the municipalities from its larger metropolitan area, some municipalities from the western part of Slovenia and some other municipalities, which are mostly regional centers (Fig. 2, left). Those are the most developed municipalities with the highest scores on the latent economic dimension (η_1), structural dimension (η_2) and social dimension (η_3). This cluster, at the same time, has the most favorable average values on all indicators. The picture for cluster 3 is just the opposite. Municipalities in this third cluster are mainly concentrated in the eastern part of Slovenia; these are rural municipalities and most of them lie near the border. They face severe socio-economic situation with low income, low employment and high social aid. Population density is low and population trend is negative. In short, these are the least developed municipalities, for which all latent scores and all indicators are most unfavorable. Cluster 2 represents the group of medium-developed municipalities, located in eastern, northwestern and southern part of Slovenia. It is clear that the given groups can be ranked with regard to the socio-economic development level, which can be also seen from distances between cluster centers (see Table 9).

For Croatia, it can be easily inferred that cluster 1 includes the most economically developed municipalities (with higher per capita and municipality incomes, higher population share making income and lower

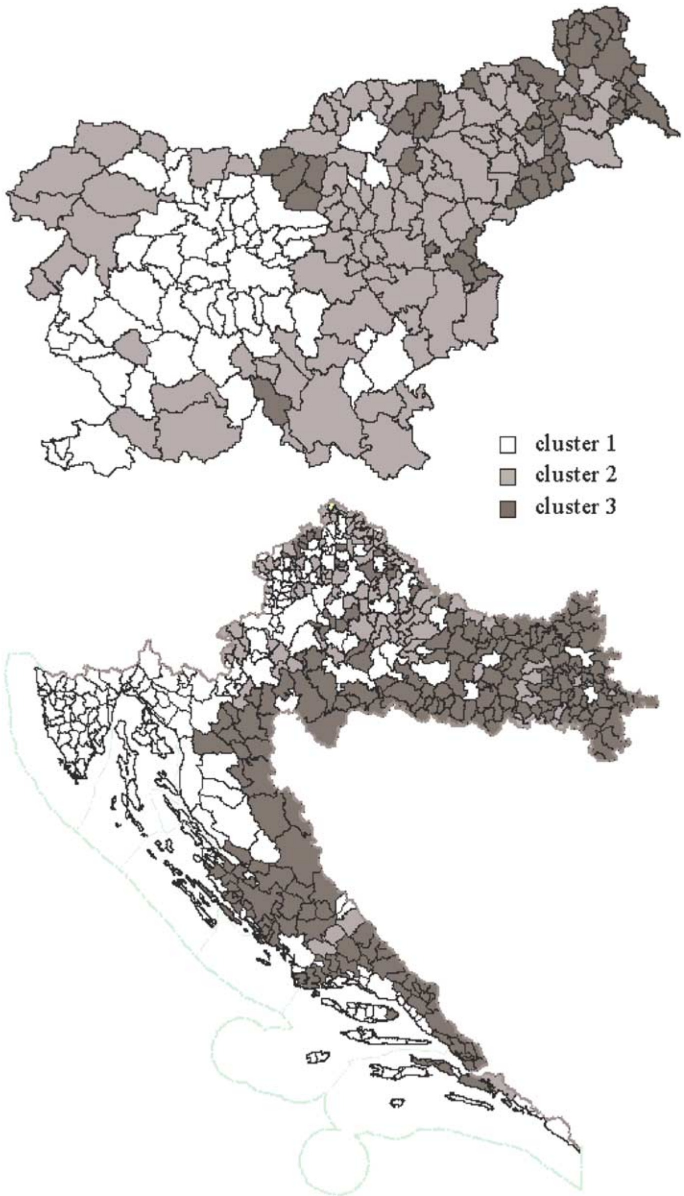

Fig. 2. Development-level map.

social aid per capita). These are northern Adriatic region municipalities, including most of Istria and part of western continental Croatia (see Fig. 2, right). Cluster 1 is also characterised by higher score on the latent social dimension (η_3), which basically indicates lower share of agricultural population, higher educational level, although the age index is higher than in other two groups.

Clusters 2 and 3 are less developed than first one. The situation/problems however differ for those two groups. Cluster 3 has higher values of economic and social dimension, but it has low value of structural dimension (lower employment and population share making income).

Table 9
Distances between final cluster centers

Cluster	1	2
Slovenian data		
1		
2	1.849	
3	3.680	1.842
Croatian data		
1		
2	2.153	
3	1.754	2.195

Finally, cluster 2 includes medium-developed municipalities. The high share of agricultural population indicates that those are the rural territories. Distances between pairs of centroids (in standardised scale) are shown in Table 9.

In addition, we note that the computed latent scores on each development dimension can be used to rank-order municipalities within each cluster, which might provide valuable information for inclusion in regional subsidy funds.

6. Conclusion

In this paper we proposed a multivariate statistical framework for regional development assessment based on structural equation modelling with latent variables and showed how such methods can be combined with cluster analysis to obtain development grouping of territorial units. These methods have several important advantages over the previously taken approaches in the literature in that they (i) take account of distributional issues in turn enabling application of more powerful inferential techniques, (ii) enable modelling of structural relationships among latent development dimensions and subsequently allow formal statistical testing of model specification and testing of various hypothesis on the estimated parameters, (iii) allow for complex structure of the factor loadings in the measurement models for the latent variables which can also be formally tested in the confirmatory framework, and (iv) enable computation of latent variable scores that take into account structural (possibly simultaneous) inter-relationships among latent variables and complex structure of the factor loadings in the measurement models. In addition, these methods can be straightforwardly combined with non-parametric classification techniques such as cluster analysis therefore enabling development level classification of territorial units.

In this respect, combining formal structural equation methods and non-parametric classification methods, such as cluster analysis, gives a broader, more encompassing methodological approach, however structural equation methods can be used for assessing regional development without cluster analysis (Cziráky et al., 2002a). The addition of cluster analysis primarily aids the simplicity of interpretation and also provides a link to the exiting methodological approaches (Soares et al., 2003). The use of exploratory factor analysis, on the other hand, is useful only for the preliminary analysis when there are numerous potential regional development indicators and no established models, hence these methods are useful primarily for initial exploration of the data. Once potential indicators are identified, more formal statistical procedures, such as structural equation methods, capable of testing statistical suitability of alternative indicators as well of other model specification issues are needed. The main issue is that without such formal testing classification might be done on the basis of poor indicator or using models that assume independence of development dimensions (factors). The structural equation methods, similarly to the classical confirmatory maximum likelihood factor analysis take into consideration distributional issues and provide basis for formal statistical inference, however, unlike the confirmatory factor analysis, structural equation methods allow for structural (e.g. causal) relationships among factors (i.e. latent variables). This is particularly important when attempting to model regional development dimensions that are not mutually independent and possibly causally inter-related.

Using Slovenian and Croatian municipality data, we estimated a structural equation model with four-dimensional measurement models which displayed relatively good fit for both countries. Furthermore, we performed cluster analysis on the estimated latent variable scores and found three clusters of municipalities in both countries, grouped on the basis of their latent development characteristics.

Consideration more sophisticated multivariate methods for regional development assessment and classification has high policy relevance especially for the EU countries where they might provide a better alternative to the currently used simple GDP/PPS rule. Consideration of such methods might have even higher relevance for the Accession countries where territorial classifications were not traditionally based on the EU NUTS system and where the GDP/PPS rule might be inefficient in distinguishing between regions in greater need for structural subsidies from the regions with smaller need for such support.

References

Babakus, E., Ferguson, C.E., Jöreskog, K.G., 1987. The sensitivity of confirmatory maximum likelihood factor analysis to violations of measurement scale and distributional assumptions. Journal of Marketing Research 24, 222–228.

Bartholomew, D.J., Knott, M., 1999. Latent variable models and factor analysis. Kendall's Library of Statistics, second ed. Arnold, London.

Curran, P.J., West, S.G., Finch, J.F., 1996. The robustness of test statistics to non-normality and specification error in confirmatory factor analysis. Psychological Methods 1, 16–29.

Cziráky, D., 2004. LISREL 8.54: A programme for structural equation modelling with latent variables. Journal of Applied Econometrics 19, 135–141.

Cziráky, D., Čumpek, T., 2002. Multivariate analysis of the European economic and defence structure. Croatian International Relations Review 8 (24), 79–91.

Cziráky, D., Puljiz, J., Jurlin, K., Maleković, S., Polić, M., 2002a. Covariance structure analysis of regional development data: An application to municipality development assessment. In: Bade, V., Heinrich, H., Brauser, H.L. (Eds.), From Industry to Advanced Services: Perspectives of European Metropolitan Region. Proceedings from the 2002 ERSA Conference. European Regional Science Association, Dortmund.

Cziráky, D., Puljiz, J., Jurlin, K., Maleković, S., Polić, M., 2002b. A multivariate methodology for modelling regional development in Croatia. Croatian International Relations Review 8 (25), 35–52.

Cziráky, D., Pisarović, A., Filipan, T., 2002c. Multivariate modelling in experimental designs. In: Ferligoj, A., Mrvar, A. (Eds.), Developments in Applied Statistics. Metodoloski zvezki, Ljubljana, FDV, pp. 137–155.

Cziráky, D., Puljiz, J., Jurlin, K., Maleković, S., Polić, M., 2003a. An econometric model for development level assessment with an application to municipality development classification. In: Proceedings of the 5th International Conference Enterprise in Transition. University of Split, Split.

Cziráky, D., Sambt, J., Rovan, J., Puljiz, P., Maleković, S., Polić, M., 2003b. Regional development assessment using parametric and non-parametric methods: A comparative analysis of Slovenia and Croatia. In: Proceedings from the 43rd European Regional Science Association (ERSA) Congress, Jyväskyklä: ERSA (R-session papers).

Cziráky, D., Tišma, S., Pisarović, A., in press. Determinants of the low SME loan approval rate in Croatia: A latent variable structural equation approach. Journal of Small Business Economics.

D'Agostino, R.B., 1986. Tests for normal distribution. In: D'Agostino, R.B., Stephens, M.A. (Eds.), Goodness-of-fit Techniques. Marcel Dekker, New York, pp. 367–419.

Doornik, J.A., Hansen, H., 1994. A practical test for univariate and multivariate normality. Discussion paper, Nuffield College, Oxford.

Everitt, B.S., 2001. Cluster Analysis, fourth ed. Oxford University Press, Arnold and New York.

Ferligoj, A., 1989. Razvrščanje v skupine. Metodoloski zvezki 4.

Grobler, P., 2002. Proucevanje družbeno-ekonomske razvitosti slovenskih obcin s pomocjo metod multivariatne analize. Unpublished BA thesis. University of Ljubljana, Faculty of Economics.

Jöreskog, K.G., 1999. Formulas for skewness and kurtosis. Scientific Software International. Available from: <http://www.ssicentral.com/lisrel>.

Jöreskog, K.G., 2000. Latent variable scores and their uses. Scientific Software International. Available from: <http://www.ssicentral.com/lisrel>.

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

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

Kaplan, D., 2000. Structural Equation Modeling: Foundations and Extensions. Sage Publications, Thousands Oaks.

Lawley, D.N., Maxwell, A.E., 1971. Factor Analysis as a Statistical Method, second ed. Butterworths, London.

Lipshitz, G., Raveh, A., 1998. Socio-economic differences among localities: A new method of multivariate analysis. Regional Studies 32, 747–757.

Mardia, K.V., 1980. Tests of univariate and multivariate normality. In: Krishnaiah, P.R. (Ed.), Handbook of Statistics, vol. I. North Holland, Amsterdam, pp. 279–320.

Rovan, J., Sambt, J., 2003. Socio-economic differences among Slovenian municipalities: A cluster analysis approach. Metodoloki zvezki 19, 265–278.

Soares, J.O., Marquŝ, M.M.L., Monteiro, C.M.F., 2003. A multivariate methodology to uncover regional disparities: A contribution to improve European Union and governmental decisions. European Journal of Operational Research 145, 121–135.

Sörbom, D., 1989. Model modification. Psychometrika 54, 371–384.

West, S.G., Finch, J.F., Curran, P.J., 1995. Structural equation models with non-normal variables: Problems and remedies. In: Hoyle, R.H. (Ed.), Structural Equation Modeling: Concepts, Issues, and Applications. Sage, Thousand Oaks, pp. 56–75.