Regional development assessment using parametric and non-parametric ranking methods: A comparative analysis of Slovenia and Croatia

Dario Cziráky[a], Jože Sambt[b], Jože Rovan[b], Jakša Puljiz[c], Mario Polić[c], and Sanja Maleković[c],

[a]*Resource Economics Department, IMO, Zagreb, Croatia*
[b]*Faculty of Economics, University of Ljubljana, Ljubljana, Slovenia*
[c]*Macroeconomics Research Department, IMO, Zagreb, Croatia*

Abstract

This paper develops a multivariate methodological framework for development level assessment of territorial units by merging two different methodological traditions based on parametric and non-parametric techniques. We consider a parametric, inferential approach based on maximum likelihood estimation of a structural equation model with latent variables for metric-scale development ranking, and subsequently combine it with a non-parametric approach based on cluster analysis for development grouping. Both methodological frameworks are applied to data on Slovenian and Croatian municipalities with an aim of assessing their regional development level. Within the parametric approach, a simultaneous equation econometric model is estimated and latent scores are computed for each underlying latent development variable, where four latent constructs are postulated corresponding to economic, structural, social, and demographic development dimensions. In the non-parametric approach, a combination of Ward's hierarchical method and *K*-means clustering procedure is applied to classify the territorial units. The advantages of the combined parametric/non-parametric approach are shown in respect to applying each approach individually, and a methodological framework capable of estimating the development level of territorial units or regions on a metric scale, while in the same time preserving the robustness of the non-parametric techniques is presented.

JEL Classification*:* R1, C3, C1
Keywords: Regional development; linear structural equation modelling with latent variables; cluster analysis; multivariate development ranking

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. Within the European Union, a simple approach based on GDP per capita PPS (purchasing power standards) data is used to classify European regions into net-receivers and net-payers (NUTS-2 classification).[1] However, there are several major weaknesses associated with this single-criteria approach. Primary problems with the NUTS-2 classification concern too small emphases placed on the socio-economic distinctions (Lipshitz and Raveh, 1998) 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).

While the issue of using GDP as the key regional development indicator[2] is questionable even within the EU, where such data generally exist on the level of basic territorial units (NUTS-2 regions), in many countries outside the EU (in particular those not using NUTS system) the appropriate GDP data on the level of basic territorial units does not exits, and alternative development indicators play a central role in regional development assessment.

Slovenia and Croatia are examples of countries with territorial division based on micro-units (municipalities) for which no GDP data exists. While both countries are aspirants to full EU membership, their present territorial division precludes the application of NUTS-2 criteria to the existing territorial units, which either calls for adoption of a variant of NUTS-5 or for a redesign of territorial division. Therefore, in addition to the general weaknesses of NUTS-2 criteria there is the problem of their inapplicability. Similar situation exists in most other EU-accession countries, which together with the above mentioned problems with NUTS-2 classification calls for serious consideration of alternative development indicators and more sophisticated regional development assessment methods.

There are several different approaches to regional development level assessment in the literature—most often some form of classification and data reduction is employed. Soares, *et al.* (2003) suggest a combination of factor and cluster analysis and provide an example of a regional classification for Portugal. Rovan and Sambt (2002) and Bregar, *et al.* (2002) used a combination of hierarchical and non-hierarchical cluster analysis methods to classify Slovenian municipalities into several clusters of differing development level. 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. They are based on non-parametric and rather informal methods from the statistical-inference point of view. While not imposing any distributional assumptions on the regional development data, these methods have two general weaknesses. Firstly, they have rather limited potential for formal econometric modelling of regional development as they do not provide any model fit and diagnostic statistics. Secondly, they, at best, provide broad territorial groupings while at the same time failing to assign development ranks (ordinal or interval) to the analysed territorial units. In addition to these two problems, there is also a known but less problematic issue of subjective interpretation of territorial groups or clusters in terms of their true development level.

An alternative parametric approach based on inferential multivariate methods was proposed by Cziráky, *et al.* (2002a;b), who used multiple regional development indicators to estimate the underlying development level of the territorial units.[3] The approach taken by Cziráky, *et al.* (2002a;b) is to formally model regional development by treating several development dimensions as latent variables imperfectly measured by various (available) regional development indicators. This approach is based on structural equation modelling with latent variables (LISREL) and it has two major advantages over the above mentioned (non-parametric) approaches. Firstly, it allows formal statistical testing of the estimated model and consequently specific formulation of causal and simultaneous relationships among latent development dimensions as well as their respective measurement structures. Secondly, it enables computation of interval-level latent scores (e.g. values of latent development variables) for each individual territorial unit, thereby allowing interval ranking and mutual comparisons across territories in respect to various development dimensions. This later aspect is particularly important when regional assessment is used for policy purposes such as subsidy allocation or inclusion/exclusion in structural funds. This is because interval-ranking enables straightforward selection of any share of territorial units while the same is not possible if only group or cluster membership information is available.

Nevertheless, non-parametric grouping methods such as cluster analysis do offer some advantages which are best seen in the ability to identify groups of territorial units with similar development level but without any within-cluster interval information on the relative development differences among identically clustered units. This is the point where inferential techniques can be of highest utility and a unified framework based on a combination of formal inferential econometric modelling with model fit assessment and non-parametric grouping methods can provide a powerful tool for regional development modelling and classification.

In this paper we develop an integrated framework that combines formal parametric structural equation latent variable modelling with non-parametric cluster analysis and subsequently apply it to regional development modelling and development level assessment of Slovenia and Croatia. We estimate several latent development dimensions and then perform cluster analysis on the computed scores of the latent variables. Such an approach has two main advantages. Firstly, explicit modelling of the underlying relationships among development indicators takes into account substantive causal relationships. Secondly, using smaller number of latent variables in cluster analysis allows clearer interpretation of the clusters as well as rank-ordering of municipalities within each cluster on the bases of estimated (latent) development dimensions.

Our approach starts from specifying and estimating a general structural equation model with latent variables and proceeds with computation of latent variable scores, which are finally used in cluster analysis. The application to Slovenia and Croatia suggests the same structural development model with smaller differences in the latent measurement models, mainly due to data differences between the two countries. Cluster analysis using latent scores gave clear and well-interpretative results for both countries finding smaller number of clusters with different development level. Preserving the advantage of structural equation modelling, the clustering (aside of providing cluster membership information) also retained the interval-level information on the latent variable scores—on each latent development dimension, for all clustered territorial units. Therefore, the final results gave us territorial groupings and interval-level values for territorial units within each group, thus allowing additional development ranking within clusters.

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 econometric methodology and estimation methods are described in the third section. Fourth section presents model specification and estimation results for structural equation econometric models for Slovenia and Croatia, while fifth section describes a technique for computing latent scores from structural equation models. Sixth section presents the results from hierarchical and *K*-means cluster analysis including numerical and graphical representation of the identified clusters and the last section concludes.

2. Data and descriptive analysis

The collected data are on municipality level and presents 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; see Maleković, 2001).

Table 1
Definitions of the variables and notation

Slovenian data	
Variable description	Symbol
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)[4]	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**	
Variable description	Symbol**
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)	$\hat{x}_3$

*All Croatian data come from the 2001 census. The population figure for 1991 used to compute $\hat{x}_1$ came from 1991 census.
** The symbols with the "heat" are used to denote Croatian variables to keep the *x*-*y* notation.

Table 2 reports results of the normality tests for all variables (see 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 (X^2) tests strongly reject the null hypothesis. The exceptions are Income per capita (y_1) and Employment (y_2) for Slovenia, which seem 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 (see e.g. Babakus, *et al.*, 1987; Curran, *et al.*, 1996; West, *et al.*, 1995).

Table 2

Normality tests (raw data)*

Slovenian data						
	Skewness		Kurtosis		Skewness and Kurtosis	
Variable	z-score	p-value	z-score	p-value	X^2	p-value
y_1	0.572	0.567	0.282	0.778	0.407	0.816
y_2	−1.298	0.194	−0.522	0.602	1.957	0.376
y_3	7.099	0.000	5.468	0.000	80.298	0.000
y_4	7.088	0.000	4.199	0.000	67.862	0.000
y_5	10.765	0.000	7.762	0.000	176.126	0.000
y_6	5.469	0.000	6.581	0.000	73.228	0.000
y_7	4.035	0.000	2.362	0.018	21.856	0.000
x_1	10.545	0.000	8.023	0.000	175.557	0.000
x_2	3.425	0.001	4.542	0.000	32.361	0.000

Croatian data						
	Skewness		Kurtosis		Skewness and Kurtosis	
Variable	z-score	p-value	z-score	p-value	X^2	p-value
$\hat{y}_1$	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

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

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;b). We note that the NS technique is widely applicable with other types of data (see Cziraky and Čumpek, 2002 for a macro-economic application and Cziráky, *et al*. 2002c;d for an application in environmental sciences). 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, $\mathbf{x}_j^{\text{k}} = \{x_{j1}', x_{j2}', \ldots, x_{jk}'\}$ where $k \leq N$ thus $\mathbf{x}^{\text{k}} \subseteq \mathbf{x}$. Let f_i be the frequency of occurrence of the value x_{ji} in $\mathbf{x}_j$ so that $f_{ji} \geq 1$. 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

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

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.

Table 3

Normality tests (normalised data)*

Slovenian data

	Skewness		Kurtosis		Skewness and Kurtosis	
Variable	z-score	p-value	z-score	p-value	X^2	p-value
y_1	0.572	0.567	0.282	0.778	0.407	0.816
y_2	1.298	0.194	0.522	0.602	1.957	0.376
y_3	0.000	1.000	0.100	0.920	0.010	0.995
y_4	0.000	1.000	0.100	0.920	0.010	0.995
y_5	0.000	1.000	0.100	0.920	0.010	0.995
y_6	0.000	1.000	0.101	0.920	0.010	0.995
y_7	0.000	1.000	0.100	0.920	0.010	0.995
x_1	0.005	0.996	0.107	0.914	0.012	0.994
x_2	0.001	1.000	0.100	0.920	0.010	0.995

Croatian data

	Skewness		Kurtosis		Skewness and Kurtosis	
Variable	z-score	p-value	z-score	p-value	X^2	p-value
$\hat{y}_1$	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

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

3. Econometric methodology

The proposed econometric methodology first aims to model regional development using structural equations models with latent variables (LISREL) and then subsequently to use the computed latent scores in secondary cluster analysis (for LISREL references see Jöreskog, 1973; Hayduk, 1987, 1996; Bollen, 1989; Jöreskog, *et al.* 2000). In Cziráky, *et al.* (2002a;b) a structural equation model including latent variables corresponding to economic, structural/social, and demographic development dimensions was estimated, however, no clustering was attempted and territorial units were assigned (metric) latent scores on each of the tree dimensions. In the present paper we aim to re-estimate this model using newly available Croatian census data, and to estimate a similar model for Slovenia.

The econometric model is specified as a special case of the general structural equation model with latent variables (Jöreskog, *et al.*, 2000). Denoting the latent endogenous variables by $\boldsymbol{\eta}$ and latent exogenous variables by ξ, and their respective observed indicators by $\mathbf{y}$ and $\mathbf{x}$, the structural part of the model is given by

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

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

$$\mathbf{y} = \boldsymbol{\Lambda}_{\mathrm{y}}\boldsymbol{\eta} + \boldsymbol{\epsilon}, \tag{3}$$

for latent endogenous, and

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

for latent exogenous variables, where, $\mathbf{y} \in \mathbf{R}^{q}$, $\mathbf{x} \in \mathbf{R}^{p}$; $\boldsymbol{\Lambda}_{\mathrm{y}}$ and $\boldsymbol{\Lambda}_{\mathrm{x}}$ are matrices with factor loadings; and $\boldsymbol{\epsilon}$ and $\boldsymbol{\delta}$ are vectors of latent errors. Using Jöreskog's LISREL notation we also define the following second-moment matrices: $E(\xi\xi^{\mathrm{T}}) = \boldsymbol{\Phi}$, $E(\zeta\zeta^{\mathrm{T}}) = \boldsymbol{\Psi}$, $E(\boldsymbol{\epsilon}\boldsymbol{\epsilon}^{\mathrm{T}}) = \boldsymbol{\Theta}_{\epsilon}$, $E(\boldsymbol{\delta}\boldsymbol{\delta}^{\mathrm{T}}) = \boldsymbol{\Theta}_{\delta}$, and $E(\boldsymbol{\epsilon}\boldsymbol{\delta}^{\mathrm{T}}) = \boldsymbol{\Theta}_{\delta\epsilon}$. 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

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

Using Eq. (2)-(5) the implied covariance matrix can be written in terms of model parameters as

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

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

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

where p and q are the numbers of the observed indicators of latent endogenous and latent exogenous variables, respectively (for more details see Kaplan, 2000).

4. Model specification and estimation results

In previous regional development research, Cziráky, *et al.* (2002a;b) performed principal component analysis on a set of development indicators similar to those used in this paper, using 1999 and 2000 Croatian non-census data. Having identified 3-4 latent development dimensions (i.e. factors) they separately tested measurement models for each dimension using maximum likelihood confirmatory factor analysis technique and subsequently estimated a general LISREL model that included both structural (causal) relationships and factor analytic measurement models. Following that work, Cziráky, *et al.* (2003) further refined the structural part of the model using a non-recursive structural equation model of the form (see Table 4 for notation)

$$\begin{pmatrix} \eta_1 \\ \eta_2 \\ \eta_3 \end{pmatrix} = \begin{pmatrix} 0 & \beta_{12} & \beta_{13} \\ 0 & 0 & \beta_{23} \\ 0 & 0 & 0 \end{pmatrix} \cdot \begin{pmatrix} \eta_1 \\ \eta_2 \\ \eta_3 \end{pmatrix} + \begin{pmatrix} \gamma_{11} \\ \gamma_{21} \\ \gamma_{31} \end{pmatrix} \cdot \xi_1 + \begin{pmatrix} \zeta_1 \\ \zeta_2 \\ \zeta_3 \end{pmatrix}. \tag{10}$$

This specification proved to be robust to data alterations and cross-sample validation using Croatian data and subsequently demonstrated the best fit in respect to alternative specifications using Slovenian data. Therefore, a rather interesting feature of this structural model is cross-sample validity using samples for two countries. The model postulates four (partly overlapping) development dimensions each measured by a factor-analytic measurement model. The measurement models generally have complex structure, thus 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. These four dimensions resulted from both substantive reasoning and from empirical findings. Substantively, it was aimed to capture separate economic, structural/social, and demographic development dimensions.[5]

Table 4
Latent variables

Latent development dimensions		Observed Indicators	
Description	LISREL notation	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	$\hat{x}_1, \hat{x}_2, \hat{x}_3$

Empirical tests and examination of alternative specifications strongly suggested a refinement in the structural/social dimension, namely we found a significant improvement in the overall model fit as a consequence if splitting this dimension into separate structural and social latent variables. Clearly, these are related concepts both theoretically and in respect to their indicators (some of which load on both dimensions), which can be easily incorporated in a confirmatory framework of a general LISREL model.

While the structural part of the model has the same latent variables and the same specification for both countries, data availability issues render certain differences in the measurement models expectable.

In the case of Slovenia, the *Economic dimension* (η_1) is measured by Income per capita (y_1); Social aid per capita (y_3); Density (y_5); and the Number of cars per 100 inhabitants (y_7), while in the case of Croatia the observed indicators are Income per capita ($\hat{y}_1$); Population share making income ($\hat{y}_2$); Municipality income per capita ($\hat{y}_3$); Employment ($\hat{y}_4$); Social aid per capita ($\hat{y}_5$); Share of agricultural population ($\hat{y}_6$); Education ($\hat{y}_7$); and Age index ($\hat{y}_8$). Difference between the measurement models for the two countries is, in part, a consequence of data (un)availability, as the number of cars indicator was available only for Slovenia and municipality income data was collected only for Croatia. On the other hand, population share making income was available for Croatia in addition to the employment indicator and used here as an additional economic indicator partly due to its availability and partly due to specific Croatian situation where employment/unemployment figures do not capture sufficiently well economic potential of the municipal population.[6]

Structural dimension (η_2) for Slovenia is measured by Employment, i.e., employment/population ratio, (y_2); Social aid per capita (y_3); and Share of agricultural population (y_4), while for Croatia the observed indicators are Population share making income ($\hat{y}_2$); Employment ($\hat{y}_4$), Social aid per capita ($\hat{y}_5$); and Share of agricultural population ($\hat{y}_6$), thus the only difference is in the inclusion of the above mentioned $\hat{y}_2$ indicator which was not collected for Slovenia. *Social dimension* (η_3), a concept closely related to the previous one, is measured in the Slovenian model by Share of agricultural population (y_4); Density (y_5); and Students share per 1000 inhabitants (y_6), while in the Croatian model this dimension is measured by the Share of agricultural population ($\hat{y}_6$); Share of high-school graduates in total population ($\hat{y}_7$); and Age index ($\hat{y}_8$). Therefore, differences in this measurement model between the two countries are in the construction of the education indicator, namely in Slovenia, the students-share variable is used as opposite to high school graduates variable used in Croatia and in inclusion of age index versus population density in Croatia and Slovenia, respectively.

Finally, the *Demographic dimension* (ξ_1) is measured by Age index (x_1) and Population trend (x_2) in Slovenia and by Population trend ($\hat{x}_1$); Density ($\hat{x}_2$); and Vitality index ($\hat{x}_3$) in Croatia, where the vitality indicator was available. Population density was empirically problematic in Slovenia as its inclusion in the demographic dimension caused non-convergence of the optimisation algorithm used to estimate the model parameters; on the other hand density fitted well in the structural/social dimensions in Slovenia, which might indicate a country-specific characteristic.[7]

The endogenous measurement model for Slovenia is formally specified in the following matrix equation

$$\begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \\ y_5 \\ y_6 \\ y_7 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ \lambda_{31}^{(y)} & \lambda_{32}^{(y)} & 0 \\ 0 & \lambda_{42}^{(y)} & \lambda_{43}^{(y)} \\ \lambda_{51}^{(y)} & 0 & \lambda_{53}^{(y)} \\ 0 & 0 & 1 \\ \lambda_{71}^{(y)} & 0 & 0 \end{pmatrix} \cdot \begin{pmatrix} \eta_1 \\ \eta_2 \\ \eta_3 \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \varepsilon_4 \\ \varepsilon_5 \\ \varepsilon_6 \\ \varepsilon_7 \end{pmatrix}, \tag{11}$$

while the exogenous measurement model is given by

$$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} \lambda_{11}^{(x)} \\ \lambda_{21}^{(x)} \end{pmatrix} \cdot \xi_1 + \begin{pmatrix} \delta_1 \\ \delta_2 \end{pmatrix}. \tag{12}$$

The structural equation model, repeating Eq. (10), is of the same form for both countries, namely

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

where the latent errors are assumed to be uncorrelated across equations with a diagonal covariance matrix

$$\mathbf{\Psi} = \begin{pmatrix} \mathrm{var}(\zeta_1) & 0 & 0 \\ 0 & \mathrm{var}(\zeta_2) & 0 \\ 0 & 0 & \mathrm{var}(\zeta_3) \end{pmatrix}. \tag{14}$$

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

$$\mathbf{\Theta}_\varepsilon = \begin{pmatrix} \theta_{11}^{(\varepsilon)} & & & & & & \\ 0 & \theta_{22}^{(\varepsilon)} & & & & & \\ 0 & 0 & \theta_{33}^{(\varepsilon)} & & & & \\ 0 & \theta_{42}^{(\varepsilon)} & 0 & \theta_{44}^{(\varepsilon)} & & & \\ 0 & 0 & 0 & 0 & \theta_{55}^{(\varepsilon)} & & \\ 0 & 0 & 0 & 0 & 0 & \theta_{66}^{(\varepsilon)} & \\ 0 & 0 & 0 & 0 & 0 & 0 & \theta_{77}^{(\varepsilon)} \end{pmatrix}, \tag{15}$$

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 chi-square statistic from 108 (d.f. = 18) to 67.2 (d.f. = 17). Finally, the covariance matrix for the latent errors in the exogenous measurement model is specified as diagonal matrix of the form

$$\mathbf{\Theta}_{\delta} = \begin{pmatrix} \theta_{11}^{(\delta)} & \\ 0 & \theta_{22}^{(\delta)} \end{pmatrix}. \tag{16}$$

Estimation of the Eq. (11)−(16) with Slovenian data (Table 5) produced an overall fit chi-square statistic of 67.224 (d.f. = 17) with GFI = 0.927 and SRMR = 0.057, which shows approximately good fit to the data. We note that the estimated model resulted in 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

$$\begin{pmatrix} \hat{y}_1 \\ \hat{y}_2 \\ \hat{y}_3 \\ \hat{y}_4 \\ \hat{y}_5 \\ \hat{y}_6 \\ \hat{y}_7 \\ \hat{y}_8 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ \lambda_{21}^{(y)} & \lambda_{22}^{(y)} & 0 \\ \lambda_{31}^{(y)} & 0 & 0 \\ 0 & 1 & 0 \\ \lambda_{51}^{(y)} & \lambda_{52}^{(y)} & 0 \\ 0 & \lambda_{62}^{(y)} & \lambda_{63}^{(y)} \\ 0 & 0 & 1 \\ \lambda_{81}^{(y)} & 0 & \lambda_{83}^{(y)} \end{pmatrix} \cdot \begin{pmatrix} \eta_1 \\ \eta_2 \\ \eta_3 \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \varepsilon_4 \\ \varepsilon_5 \\ \varepsilon_6 \\ \varepsilon_7 \\ \varepsilon_8 \end{pmatrix}, \tag{17}$$

while the exogenous measurement model is given by

$$\begin{pmatrix} \hat{x}_1 \\ \hat{x}_2 \\ \hat{x}_3 \end{pmatrix} = \begin{pmatrix} \lambda_{11}^{(x)} \\ 1 \\ \lambda_{31}^{(x)} \end{pmatrix} \cdot \xi_1 + \begin{pmatrix} \delta_1 \\ \delta_2 \\ \delta_3 \end{pmatrix}. \tag{18}$$

The structural part of the model is specified as a recursive system of equations of the form

$$\begin{pmatrix} \eta_1 \\ \eta_2 \\ \eta_3 \end{pmatrix} = \begin{pmatrix} 0 & \beta_{12} & \beta_{13} \\ 0 & 0 & \beta_{23} \\ 0 & 0 & 0 \end{pmatrix} \cdot \begin{pmatrix} \eta_1 \\ \eta_2 \\ \eta_3 \end{pmatrix} + \begin{pmatrix} \gamma_{11} \\ \gamma_{21} \\ \gamma_{31} \end{pmatrix} \cdot \xi_1 + \begin{pmatrix} \zeta_1 \\ \zeta_2 \\ \zeta_3 \end{pmatrix}, \tag{19}$$

with the latent error covariance matrix given by

$$\mathbf{\Psi} = \begin{pmatrix} \text{var}(\zeta_1) & 0 & 0 \\ 0 & \text{var}(\zeta_2) & 0 \\ 0 & 0 & \text{var}(\zeta_3) \end{pmatrix}. \tag{20}$$

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

$$\mathbf{\Theta}_\delta = \begin{pmatrix} \theta_{11}^{(\delta)} & & \\ 0 & \theta_{22}^{(\delta)} & \\ 0 & 0 & \theta_{33}^{(\delta)} \end{pmatrix}, \tag{21}$$

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

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

which resulted in a significant decrease in the chi-square statistic for the overall model fit from 88.65 (d.f. = 34) to 75.57 (d.f. = 33).

Estimation of the Eq. (17)−(22) with maximum likelihood technique produced an overall-fit chi-square statistic of 75.57 (d.f. = 33), GFI = 0.98, and SRMR = 0.04, which jointly 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 in Table 6.

By comparing the structural equations part of the model (Eq. (10)) 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 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.

Table 5

Correlation matrices (normalised data)

Slovenian data

	y_1	y_2	y_3	y_4	y_5	y_6	y_7	x_1	x_2
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

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 Table 5 and Table 6). This difference is the most likely cause of different signs of the effect of structural on economic factors between the two countries.

5. Clustering territorial units

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. At this point we still do not have regional development information on the level of territorial micro units, i.e., municipalities, which is necessary for development classification and ranking. This problem could be solved if we could obtain values on each of the modelled latent variables, for each territorial unit. While the latent variables are, by definition, "unobserved" it can be shown that interval-level values (scores) for latent variables can be straightforwardly computed with the following technique.

Using the parameters of the estimated LISREL model from Eqs. (11)−(22), i.e., estimates shown in Table 6, we compute the scores for the latent variables following the approach of Lawley and Maxwell (1971) and Jöreskog (2000)[viii], which is based on the maximum likelihood solution of structural equation models such as the models estimated above for Slovenian and Croatia. Writing Eqs. (3) and (4) in a system as

$$\begin{pmatrix} \mathbf{y} \\ \mathbf{x} \end{pmatrix} = \begin{pmatrix} \mathbf{\Lambda}_y & \mathbf{0} \\ \mathbf{0} & \mathbf{\Lambda}_x \end{pmatrix} \cdot \begin{pmatrix} \boldsymbol{\eta} \\ \boldsymbol{\xi} \end{pmatrix} + \begin{pmatrix} \boldsymbol{\varepsilon} \\ \boldsymbol{\delta} \end{pmatrix}, \tag{23}$$

Table 6
Maximum likelihood estimates

	Slovenian model		Croatian model	
Parameter	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)
$\text{var}(\zeta_1)$	0.031	(0.012)	0.029	(0.032)
$\text{var}(\zeta_2)$	0.164	(0.052)	0.963	(0.168)
$\text{var}(\zeta_3)$	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)
X^2	67.224		75.565	
d.f.	17		33	
GFI*	0.927		0.975	
SRMR**	0.057		0.042	

* Goodnes of fit index; ** Standardised root-mean-square residual.

and using the following notation

$$\mathbf{\Lambda} \equiv \begin{pmatrix} \mathbf{\Lambda}_y & \mathbf{0} \\ \mathbf{0} & \mathbf{\Lambda}_x \end{pmatrix}, \ \boldsymbol{\xi}_a \equiv \begin{pmatrix} \boldsymbol{\eta} \\ \boldsymbol{\xi} \end{pmatrix}, \quad \boldsymbol{\delta}_a \equiv \begin{pmatrix} \boldsymbol{\varepsilon} \\ \boldsymbol{\delta} \end{pmatrix}, \quad \mathbf{x}_a \equiv \begin{pmatrix} \mathbf{y} \\ \mathbf{x} \end{pmatrix}, \tag{24}$$

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

$$\boldsymbol{\xi}_a = \mathbf{U}\mathbf{D}^{1/2}\mathbf{V}\mathbf{L}^{-1/2}\mathbf{V}^{\mathrm{T}}\mathbf{D}^{1/2}\mathbf{U}^{\mathrm{T}}\mathbf{\Lambda}^{\mathrm{T}}\mathbf{\Theta}_a^{-1}\mathbf{x}_a, \tag{25}$$

where $\mathbf{UDU}^T$ is the singular value decomposition of $\mathbf{\Phi}_a = E(\xi_a \xi_a^T)$, and $\mathbf{VLV}^T$ is the singular value decomposition of the matrix $\mathbf{D}^{1/2}\mathbf{U}^{\mathrm{T}}\mathbf{BUD}^{1/2}$, while $\mathbf{\Theta}_a$ is the error covariance matrix of the observed variables. Derivation of Eq. (25) follows the approach of Jöreskog (2000) and Lawley and Maxwell (1971) and is described in more detail in Cziráky, *et al*. (2002c). The latent scores ξ_{ai} can be computed for each observation x_{ij} in the $(p + q) \times n$ sample matrix whose rows are observations on each of our $p + q$ observed indicators, where $n = q + p$ is the sample size.

It is straightforward now to use the computed micro-level values of each of the estimated latent variables to assign ordinal or interval ranks to each territorial unit. In Cziráky, *et al*. (2002c) such an approach was used to directly rank territorial units in respect to each individual latent dimension. The advantage of that approach is in the ability to rank territorial units on an interval scale, which is in principle suited for policy purposes such as inclusion/exclusion in structural subsidy funds. However, that potentially leads to ambiguities when there is more then one latent development dimension. This happens because territorial ranking is straightforward only for each latent dimension separately; attempting to deduce some general development criteria (as opposite to separate ranking on the basis of different development dimensions)[9] in this methodological context would require a higher (second) order factor that can explain all other factors, which is empirically highly unlikely to exist. What would be needed is a secondary technique capable of classifying territorial units on the basis of the estimated latent variable scores. A non-parametric classification technique such as cluster analysis would be particularly suitable for this purpose. This way it would be possible to use parametrically computed latent scores as an input for non-parametric cluster analysis and thus potentially obtain more interpretable clustering results as a consequence of using a smaller number of statistically computed (and assessed) latent dimensions (Cziráky, *et al*. 2003).

Having computed the latent variable scores (ξ_{ai}), we perform cluster analysis with the purpose of grouping (clustering) municipalities into several groups with similar characteristics (for more details see Everitt, 1993).

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 (Figure 1, left) shows a fairly clear picture − three clusters are suggested for both countries. Croatia appears to have either 3 or 5 clusters, according to dendrogram (Figure 1, right). In the second step we used the *K*-means method. Namely, the main deficiency of the Ward's method (and also of all other hierarchical methods) is that the allocation of units is final, with no possibility of reassignment to another (more appropriate) group during the procedure. *K*-means method, on the other hand, is sensitive to the initial value setting and if we are unfortunate we can get trapped into a local optimum which can be far from the global one. Empirical evidence suggests that there is a high probability of achieving the global optimum if we take centroids from the hierarchical methods as initial seed-points for the *K*-means method (Ferligoj, 1989). In our case centroids from the Ward's method have been used for that purpose. The *K*-means procedure assigns cases to clusters based on their distance from the centroids and updates the locations of centroids based on the mean values of the cases in each cluster. These steps are repeated until any reassignment of cases would not make the clusters more internally cohesive (homogeneous) and more clearly separated from each other. This way, the *K*-means method was used to improve the results of the Ward's method.

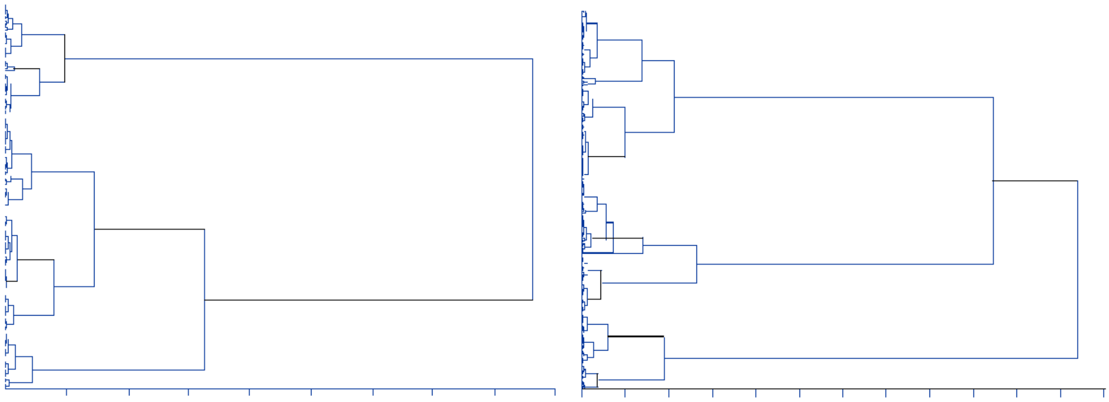

Figure 1. Dendrogram for Slovenia (left) and Croatia (right)

On the basis of the dendrogram obtained from preliminary hierarchical clustering (Figure 1) of the estimated latent variables, for both countries a 3-cluster solution is suggested.

We note that clustering of the original development indicators, while providing to large degree similar picture, lacked clarity in interpretation due to a need to analyse higher number of centroids, which indicates a major advantage of clustering on the basis of latent variables.[10]

In Slovenian case the results with latent variable differed considerably from those with original development indicators (results are omitted). The method presented in this paper seems to give very reasonable results for the Slovenian case. 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 either for the fact that the variables were clustered or for that the criteria variables for clustering were actually linear combinations of observable development indicators.

The principal advantage of clustering the smaller number of latent variables instead of the original variables (i.e. observed development indicators) is in easier interpretation and more meaningful cluster centres. Table 8 shows centroids for each cluster (expressed in standardised units).

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

For Slovenia the picture is very clear. 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 capitals of the regions (see Figure 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 favourable 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; they 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 unfavourable. Cluster 2 represents the group of medium-developed municipalities, located in eastern, north-western and southern part of Slovenia. It is clear that the given groups can be ranked with regard to the socio-economic development level. The study reinforces a well-known fact in Slovenia: more developed western part and less developed eastern part of Slovenia.

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 social aid per capita). These are northern Adriatic region municipalities, including most of Istria and part of western continental Croatia (see Figure 2, right). Cluster 1 is also characterised by higher score on the latent *social dimension* (η_3), which basically indicates more positive population trend, higher vitality index and younger population (i.e. lower age index).

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

Table 8

Final cluster centers

	Slovenian data		
	Cluster 1	Cluster 2	Cluster 3
	$N = 55$	$N = 89$	$N = 48$
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		
	Cluster 1	Cluster 2	Cluster 3
	$N = 204$	$N = 227$	$N = 115$
η_1	1.59	0.57	0.36
η_2	0.18	−0.82	1.32
η_3	1.71	0.69	0.38
ξ_1	0.03	0.10	−0.26

Finally, cluster 3 includes least developed municipalities. The high share of agricultural population indicates that those are the rural territories. Although these municipalities have favourable *structural dimension* – higher population share making income and employment − the situation of other three dimensions is less favourable, so we define them as "least developed municipalities". Distances between pairs of centroids (in standardised scale) are shown in Table 9.

Table 9
Distances between final cluster centers

Slovenian data			
Cluster	1	2	3
1		1.849	3.680
2	1.849		1.842
3	3.680	1.842	

Croatian data			
Cluster	1	2	3
1		1.754	2.153
2	1.754		2.195
3	2.153	2.195	

In the Slovenian case, largest distance is between clusters 1 and 3, while distance between clusters 1 and 2 is similar to distance between 2 and 3. In Croatia, it can be seen that cluster 3 is more distant from cluster 1 (which represents most developed municipalities) than from cluster 2. But the distance between cluster 2 and 3 is also high, indicating that the characteristics and problems of those 2 groups differ.

Additionally, the computed latent scores on each development dimension can be used to rank-order municipalities within each cluster.

6. Conclusion

In this paper we combined structural equation econometric modelling with more descriptive cluster analysis techniques in order to obtain a development grouping of Slovenian and Croatian municipalities. This approach allowed us to model regional development by postulating various latent development dimensions and specifying recursive causal relationships among them. While the use of more powerful inferential techniques, such as maximum likelihood estimation of structural equation models, offers significant potential for modelling regional development data, such data often fail to satisfy necessary distributional assumptions required by these techniques.

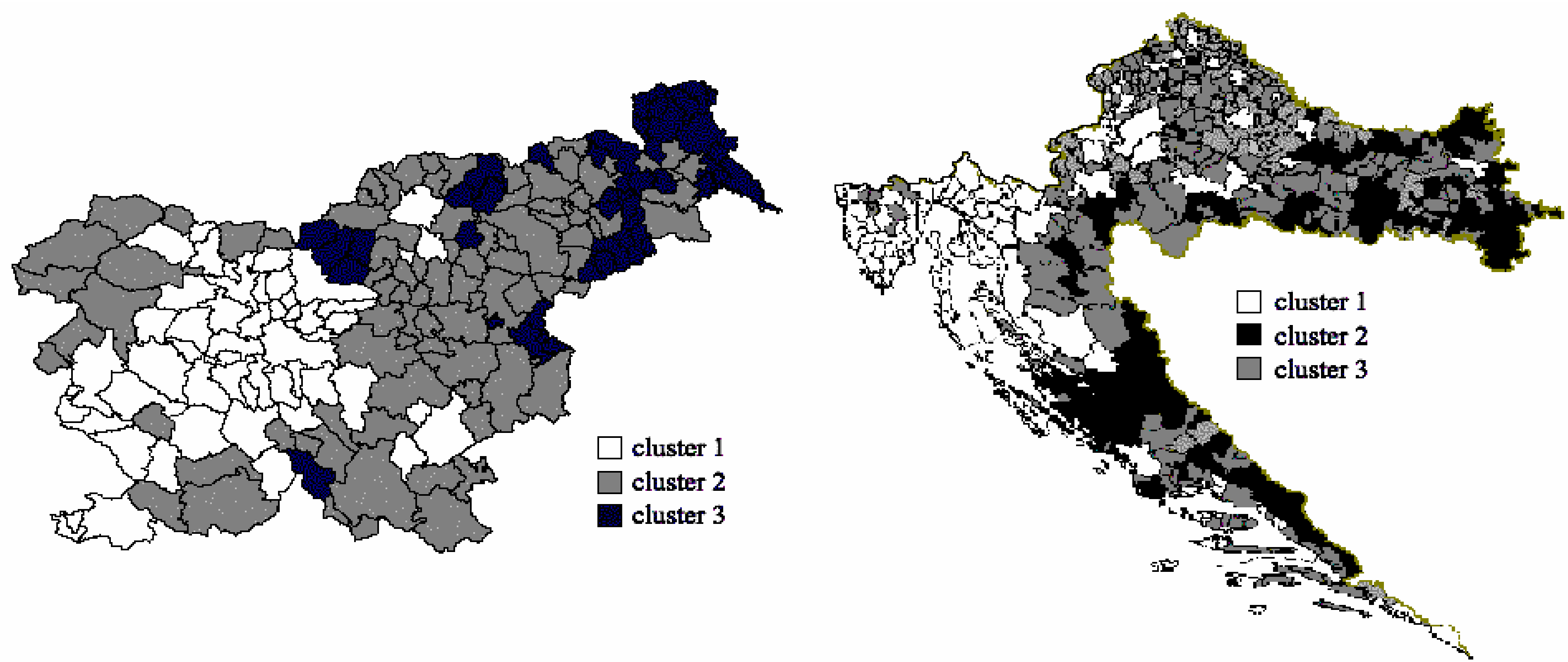

Figure 2. Development-level map of Slovenia (left) and Croatia (right)

The solution taken in this analysis was to normalise the variables using the normal scores technique, and this resulted in satisfactory distribution of all considered variables. This formally allowed application of maximum likelihood estimation methods based on the assumption of normal (Gaussian) distribution of the modelled variables.

Using Slovenian municipality data and the newly available Croatian census data, we estimated the same structural equation model for both countries, having four (to some degree different) measurement models with 9 and 11 observed indicators, for Slovenian and Croatia, respectively. It was found that a four-dimensional model fits the data relatively well in both countries, which suggests a similar but complex structure of regional development in both analysed countries.

Secondly, we performed cluster analysis on the estimated latent variable scores, thus clustering latent development dimensions instead of raw variables. While similar to the results previously obtained with the raw data, latent variable clustering offered a clearer picture and much easier interpretability. In effect, we were able to distinguish three clusters of municipalities in both Slovenian and Croatia, grouped on the basis of their latent development characteristics.

Finally, we emphasise that estimation of latent variables, i.e., underlying development dimensions also resulted in metric-type development scores for each municipality, thereby allowing for a possibility of subsequent ranking of municipalities within each cluster which can be used for policy purposes such as subsidy and structural funds allocation or regional development planning.

Notes

[1] The European Union's criteria for Structural Funds allocation is the "objective I", which allocates Funds' subsidies to all regions with GDP/PPS per capita under 75% of the EU average (European Council, 1999).

[2] GDP per capita is considered by the EC as "the standard measure of the size and performance of a regional economy" (European Commission, 1999).

[3] This methodological framework was used in designing legislation for regional development assessment and national subsidy classification of Croatian territorial units through 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 between 2000 and 2002. 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 were receiving state support under the "Law on Areas of Specific Governmental Concern". Subsequently, the World Bank's Global Development Network (GDN) featured this work as a case study of a successful research-policy link.

[4] Undergraduate students enrolled in the higher education institutions.

[5] For example, the Croatian territorial-division legislation defined these categories as general development dimensions.

[6] Note that population share making income indicator is also used as indicator of the *Structural dimension*, which was additionally confirmed by empirical testing.

[7] Slovenia has a very diverse configuration of the territory on the one hand and historically based formation of the municipalities on the other. For that reason the expected relationship between population density and the level of socio-economic development is likely to be specific. Therefore, population density might not be taken as necessarily demographic development indicator, and in fact, this view was strongly supported by the empirical findings.

[viii] See also Cziráky, *et al.* (2002c;d).

[9] For example, the 2002 Croatian Law on Areas of Special State Concern specifies three (latent) development dimensions (economic, structural/social, and demographic) and explicitly assigns quotas to each dimension; thus municipalities can receive subsidy funding on the basis of multiple criteria, not a single "overall" development level. A methodological problem occurs when such quotas need to be assessed empirically, from data, in case political decision is not made a priori.

[10] We note that a similar procedure can be employed on the basis of factor scores obtained from a varimax-rotated principal components solution (see e.g., Soares, *et al*. 2003), however doing so assumes that factors are orthogonal in the population (which is clearly not acceptable in this case) and furthermore it ignores a likely more complex structure, specially cases with ambiguous (compound) loadings and causal relationships.

References

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.

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

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

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

Bregar, L., Bavdaž Kveder, M., Lavrač, I., Pavlin, B., Pečar, J., Repar, B., Sambt, J., Stanovnik, T., Špiranec, K., and Ograjenšek, I. (2002), *Needs assessment report − Subnational Statistical Capacity Building.* Final version. Ljubljana: Statistical Institute of the Faculty of Economics, University of Ljubljana.

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.

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

Cziráky, D. and Čumpek, T. (2002), "Multivariate Analysis of the European Economic and Defence Structure", *Croatian International Relations Review*, 24, pp. 79-91.

Cziráky, D., Filipan, T. and Pisarović, A. (2002c), "Multivariate modelling in experimental design", *International Conference on Methodology and Statistics*, Ljubljana, Slovenia 15-18, 2002.

Cziráky, D., Pisarović, A., Tišma, S. (2002d) "Experimental Design with Latent Variables: An application to Spruce (*Picea abies* K.)" *Ecological Modelling* (under review).

Cziráky, D., Puljiz, J., Jurlin, K., Maleković, S. and 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 (CD). Dortmund: European Regional Science Association.

Cziráky, D., Puljiz, J., Jurlin, K., Maleković, S. and Polić, M. (2002b), "A Multivariate Methodology for Modelling Regional Development in Croatia", *Croatian International Relations Review*, Vol. VIII, **25**.

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.

European Commission (1999), *Sixth periodic report on the social and economic situation and development of the regions of the European Union*.

European Council (1999), Regulation (EC) No. 1260/1999 of the European Council, *Official Journal of the European Communities*, 12.7.1999.

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

Ferligoj, A. (1989), "Razvrščanje v skupine", *Metodološki zvezki*, **4**.

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.

Grobler, P. (2002), *Proučevanje družbeno-ekonomske razvitosti slovenskih občin s pomočjo metod multivariatne analize*. Unpublished BA thesis. University of Ljubljana, Faculty of Economics.

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.

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. (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. (1996), *PRELIS 2 User's Reference Guide*. Chicago: Scientific Software International.

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

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

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.

Rovan, J. and Sambt J. (2002), "Development of Slovenian municipalities", *International Conference on Methodology and Statistics*, Ljubljana, Slovenia 15-18, 2002.

Soares, J.O., Marquês, M.M.L. and 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**, pp. 371-384.

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 non-normal variables: Problems and remedies”, in Hoyle R.H. (Ed.), *Structural Equation Modeling: Concepts, Issues, and Applications*. Thousand Oaks, CA: Sage Publications, pp. 56-75.