An Econometric model for development level assessment with an application to municipality development classification

Dario Cziráky\textsuperscript{a†}, Jakša Puljiz\textsuperscript{b}, Krešimir Jurlin\textsuperscript{b}, Sanja Maleković\textsuperscript{b}, Mario Polić\textsuperscript{b}
\textsuperscript{aDepartment for Resource Economics, IMO}
\textsuperscript{bDepartment for International Economic Relations, IMO}
2 Farkaša-Vukotinovića, 10000 Zagreb, Croatia

February 2003

Abstract

This paper develops a structural equation econometric model with latent variables for modelling regional development in Croatia. Using newly available census data on 546 Croatian municipalities two latent variable structural equation models were estimated assuming three and four latent development dimensions, respectively. Scores for the latent variables from the best fitting econometric model were computed and used in secondary $K$-means cluster analysis for the purpose of development grouping of Croatian territorial units. The best fitting model indicated four latent development dimensions, generally corresponding to economic, structural, demographic, and population development dimensions. Cluster analysis resulted in identification of four clusters of municipalities with differing development levels and characteristics, two of which were noticeably less developed on all accounts. The use of latent variable scores for clustering greatly facilitated substantive interpretation thus providing valuable information about each cluster.

\textit{JEL Classification:} R1, C3, C1

\textit{Keywords:} Regional development; Structural equation modelling; Latent variables, Cluster analysis

\textsuperscript{†}Corresponding author; e-mail: dario@irmo.hr; fax: +385 1 48 28 361.
1. Introduction

Assessment of the regional development level is generally a difficult task though its policy relevance is exceptionally high. Information about the level of development is a crucial input in regional planning and development policy and a key criterion for allocation of various structural funds and national subsidies. The European Commission 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. The 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 is questionable even within European Union where such data generally exist on the level of basic territorial units, in many countries outside of the EU the appropriate GDP data on the level of basic territorial units does not exits, and alternative development indicators play a crucial role in regional development assessment. In Croatia, the basic territorial units are municipalities and GDP data is not available on the municipality level. An alternative multivariate approach was proposed in Cziráky, et al. (2002a;b), where multiple (available) regional development indicators were used to estimate the underlying development level of the territorial units. In this paper we extend the previous analysis with classification methods similar to those used by Soares, et al. (2003), and by using the most recently available census data.

There are different possible approaches to regional development level assessment; most often some form of classification and data reduction is employed. Soares, et al. (2003) suggest a combination of factor and cluster analysis (Everitt, 1993) and provide an example of a regional classification for Portugal. Lipshitz and Raveh (1998; 1994) proposed the use of a co-plot technique for the study of regional disparities. Multidimensional scaling techniques

---

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 The analysis reported in this paper originates from 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 (Maleković, 2001). The purpose of the project was to provide an analytical base for evaluation of the development level of the Croatian territorial units (municipalities) with an aim of widening the span of territorial units which were receiving state support under the “Law on Areas of Specific Governmental Concern”. 
(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.

In this paper we develop a general structural equation econometric model for regional development modelling and use it as an input for grouping of 546 Croatian municipalities with non-parametric cluster analysis methods. The main approach taken in this paper is based on methods that combine the results of an econometric model with non-parametric cluster analysis techniques, namely, 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.

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 were reported for untransformed and transformed variables, where the normal scores technique was used for normalisation. The econometric methodology and estimation method was described in the third section. Fourth section presents model specification and estimation results for structural equation econometric models, while fifth section describes a technique for computing latent scores from structural equation models. Sixth section presents the results from the $K$-means cluster analysis including numerical and graphical representation of the identified clusters and seventh section concludes.

### 2. Data and descriptive analysis

In previous regional development analysis of Croatia (Cziráky, et al. 2002a;b) data from several different Croatian sources was used. Such data, though only available at that time, suffered from likely low quality and inconsistency across different data sources. The present analysis, on the other hand, uses most recent national census data, which is certainly of much higher quality and it comes from a single source (State Bureau of Statistics). In addition to
added quality and non-ambiguity, the census data included measurements on previously unavailable indicators, namely education (share of high-school graduates in total population) and share of agricultural population. We collected data on 12 development indicators, which are shown in Table 1 along with brief descriptive definitions. Note that we include symbolic notation that will be used later in order to simplify exposition.

### Table 1

<table>
<thead>
<tr>
<th>Code</th>
<th>Description</th>
<th>Symbol</th>
</tr>
</thead>
<tbody>
<tr>
<td>INCOME</td>
<td>Income per capita (in thousands HRK)</td>
<td>y₁</td>
</tr>
<tr>
<td>SHARE</td>
<td>Population share making income (%)</td>
<td>y₂</td>
</tr>
<tr>
<td>DIRECT</td>
<td>Municipality income per capita (in thousands HRK)</td>
<td>y₃</td>
</tr>
<tr>
<td>EMPLOY</td>
<td>Employment</td>
<td>y₄</td>
</tr>
<tr>
<td>UNEMP</td>
<td>Unemployment</td>
<td>y₅</td>
</tr>
<tr>
<td>SOCAID</td>
<td>Social aid per capita (in thousands HRK)</td>
<td>y₆</td>
</tr>
<tr>
<td>AGRICULT</td>
<td>Share of agricultural population</td>
<td>y₇</td>
</tr>
<tr>
<td>DENSITY</td>
<td>Density (inhabitants per km²)</td>
<td>y₈</td>
</tr>
<tr>
<td>EDUCA</td>
<td>Education (share of high-school graduates in total population)</td>
<td>x₁</td>
</tr>
<tr>
<td>AGEINDEX</td>
<td>Age index (number of people older than 65 divided by the number of people younger than 20)</td>
<td>x₂</td>
</tr>
<tr>
<td>POPTREND</td>
<td>Population trend (population 2001 divided by population 1991)</td>
<td>x₃</td>
</tr>
<tr>
<td>VITIINDE</td>
<td>Vitality index (live births over number of deceased)</td>
<td>x₄</td>
</tr>
</tbody>
</table>

The collected data is on the municipality level and presents lowest aggregation level available for Croatia. 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 (see Maleković, 2001).

Table 2 reports the results of the univariate normality tests for all variables (see D’Agostino, 1986; Doornik and Hansen, 1994; Mardia, 1980). It can be easily seen that neither variable is distributed normally, as the reported chi-square tests strongly reject. However, 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
Tests of univariate normality (raw data)*

<table>
<thead>
<tr>
<th>Variable</th>
<th>Skewness</th>
<th>P-Value</th>
<th>Kurtosis</th>
<th>Z-Score</th>
<th>P-Value</th>
<th>Skewness and Kurtosis</th>
<th>X²</th>
<th>P-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>INCOME</td>
<td>2.869</td>
<td>0.004</td>
<td>-3.765</td>
<td>0.000</td>
<td>22.408</td>
<td>0.000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SHARE</td>
<td>-2.590</td>
<td>0.010</td>
<td>-2.165</td>
<td>0.030</td>
<td>11.397</td>
<td>0.003</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DIRECT</td>
<td>16.112</td>
<td>0.000</td>
<td>10.876</td>
<td>0.000</td>
<td>377.894</td>
<td>0.000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>EMPLOY</td>
<td>4.237</td>
<td>0.000</td>
<td>3.414</td>
<td>0.001</td>
<td>29.611</td>
<td>0.000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>UNEMP</td>
<td>12.123</td>
<td>0.000</td>
<td>9.420</td>
<td>0.000</td>
<td>235.710</td>
<td>0.000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SOCAID</td>
<td>18.271</td>
<td>0.000</td>
<td>12.902</td>
<td>0.000</td>
<td>500.317</td>
<td>0.000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AGRICULT</td>
<td>11.233</td>
<td>0.000</td>
<td>6.070</td>
<td>0.000</td>
<td>163.022</td>
<td>0.000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AGEINDEX</td>
<td>31.629</td>
<td>0.000</td>
<td>17.529</td>
<td>0.000</td>
<td>1307.683</td>
<td>0.000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>POPTREND</td>
<td>-2.826</td>
<td>0.005</td>
<td>6.781</td>
<td>0.000</td>
<td>53.967</td>
<td>0.000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DENSITY</td>
<td>25.330</td>
<td>0.000</td>
<td>15.886</td>
<td>0.000</td>
<td>893.997</td>
<td>0.000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>EDUCA</td>
<td>2.853</td>
<td>0.004</td>
<td>-2.101</td>
<td>0.036</td>
<td>12.553</td>
<td>0.002</td>
<td></td>
<td></td>
</tr>
<tr>
<td>VITINDEX</td>
<td>10.209</td>
<td>0.000</td>
<td>6.970</td>
<td>0.000</td>
<td>152.794</td>
<td>0.000</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

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

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

and $\Phi^{-1}$ 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. It is clear that normalisation successfully removed departures from normality resulting in variables that are, individually, normally distributed.
### Table 3
Tests of univariate normality (normalised data)*

<table>
<thead>
<tr>
<th>Variable</th>
<th>Skewness Z-Score</th>
<th>Skewness P-Value</th>
<th>Kurtosis Z-Score</th>
<th>Kurtosis P-Value</th>
<th>Skewness and Kurtosis X²</th>
<th>P-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>INCOME</td>
<td>0.000</td>
<td>1.000</td>
<td>0.065</td>
<td>0.948</td>
<td>0.004</td>
<td>0.998</td>
</tr>
<tr>
<td>SHARE</td>
<td>0.000</td>
<td>1.000</td>
<td>0.065</td>
<td>0.948</td>
<td>0.004</td>
<td>0.998</td>
</tr>
<tr>
<td>DIRECT</td>
<td>0.000</td>
<td>1.000</td>
<td>0.065</td>
<td>0.948</td>
<td>0.004</td>
<td>0.998</td>
</tr>
<tr>
<td>EMPLOY</td>
<td>0.000</td>
<td>1.000</td>
<td>0.065</td>
<td>0.948</td>
<td>0.004</td>
<td>0.998</td>
</tr>
<tr>
<td>UNEMP</td>
<td>0.000</td>
<td>1.000</td>
<td>0.065</td>
<td>0.948</td>
<td>0.004</td>
<td>0.998</td>
</tr>
<tr>
<td>SOCAID</td>
<td>0.000</td>
<td>1.000</td>
<td>0.065</td>
<td>0.948</td>
<td>0.004</td>
<td>0.998</td>
</tr>
<tr>
<td>AGRICULT</td>
<td>0.000</td>
<td>1.000</td>
<td>0.065</td>
<td>0.948</td>
<td>0.004</td>
<td>0.998</td>
</tr>
<tr>
<td>AGEINDEX</td>
<td>0.001</td>
<td>0.999</td>
<td>0.064</td>
<td>0.949</td>
<td>0.004</td>
<td>0.998</td>
</tr>
<tr>
<td>POPGTREND</td>
<td>0.000</td>
<td>1.000</td>
<td>0.065</td>
<td>0.948</td>
<td>0.004</td>
<td>0.998</td>
</tr>
<tr>
<td>DENSITY</td>
<td>0.000</td>
<td>1.000</td>
<td>0.065</td>
<td>0.948</td>
<td>0.004</td>
<td>0.998</td>
</tr>
<tr>
<td>EDUCA</td>
<td>0.000</td>
<td>1.000</td>
<td>0.065</td>
<td>0.948</td>
<td>0.004</td>
<td>0.998</td>
</tr>
<tr>
<td>VITINDEX</td>
<td>0.001</td>
<td>1.000</td>
<td>0.064</td>
<td>0.949</td>
<td>0.004</td>
<td>0.998</td>
</tr>
</tbody>
</table>

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

### 3. Econometric methodology

The aim of the proposed econometric methodology is to first model regional development using structural equations models with latent variables (LISREL) and then subsequently 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 three latent variables corresponding to economic, structural 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. Our purpose in this paper is to, firstly, re-estimate a similar structural model assuming three latent development dimensions using newly available data and, secondly, to fine-tune the model and finally use its results as an input in clustering of territorial units.

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 $\eta$ and latent exogenous variables by $\xi$ and their respective observed indicators by $y$ and $x$, the structural part of the model is given by

$$\eta = B\eta + \Gamma\xi + \zeta$$

(2)

The measurement models are given in the form of typical factor analytic models as

$$y = \Lambda y + \epsilon,$$

(3)
for latent endogenous, and
\[ x = \Lambda \xi + \delta, \]  
(4)
for latent exogenous variables. Using Jöreskog’s LISREL notation we also define the following second moment matrices: \( E(\xi \xi^T) = \Phi, E(\xi \delta^T) = \Psi, E(\delta \delta^T) = \Theta_\delta, \) and \( E(\epsilon \epsilon^T) = \Theta_\epsilon. \) 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 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
\[
\Sigma = \begin{pmatrix} \Sigma_{xx} & \Sigma_{xy} \\ \Sigma_{yx} & \Sigma_{yy} \end{pmatrix}.
\]
(5)
Using Eq. (2)-(5) the implied covariance matrix can be written in terms of model parameters as
\[
\Sigma = \begin{pmatrix} \Lambda_x (I - B)^{-1} (\Gamma \Phi \Gamma^T + \Psi)[(I - B)^{-1}]^T \Lambda_y + \Theta_\epsilon & \Lambda_x (I - B)^{-1} \Gamma \Phi \Lambda_y + \Theta_\delta \\ \Lambda_x \Phi \Gamma^T [(I - B)^{-1}]^T \Lambda_y + \Theta_\epsilon \delta & \Lambda_x \Phi \Lambda_y + \Theta_\delta \end{pmatrix}.
\]
(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
\[
F = \ln |\Sigma| + \text{tr} \left( \Sigma^{-1} S \right) - \ln |S| - (p + q),
\]
(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. Estimation of the structural equation models

Estimation of the structural equation model uses the covariance matrix of the standardised normalised variables (Table 4), which is in effect a correlation matrix. The use of standardised variables, or equivalently a correlation matrix is necessary because the variables are measured on different scales.
We first estimated a model similar to the one in Cziráky, et al. (2002a;b) using, however, newly available data, and two additional variables (EDUCA and AGRICULT). The model we initially considered consisted of the endogenous measurement model given by

\[
\begin{align*}
Y_1 &= \mathbf{1} \text{.} \\
Y_2 &= \lambda_{21}^{(y)} \eta_1 + \varepsilon_1 \\
Y_3 &= \lambda_{31}^{(y)} \eta_1 + \varepsilon_3 \\
Y_4 &= \lambda_{41}^{(y)} \eta_1, \\
Y_5 &= \lambda_{51}^{(y)} \eta_2 + \varepsilon_5 \\
Y_6 &= \lambda_{61}^{(y)} \eta_2 + \varepsilon_6 \\
Y_7 &= \lambda_{71}^{(y)} \eta_2 + \varepsilon_7 \\
Y_8 &= \lambda_{81}^{(y)} \eta_2 + \varepsilon_8
\end{align*}
\]

and the exogenous measurement models, which is specified as

\[
\begin{align*}
x_1 &= \lambda_{11}^{(x)} \delta_1 + \xi_1 \\
x_2 &= \lambda_{21}^{(x)} \delta_2 \\
x_3 &= 1 \delta_3 \\
x_4 &= \lambda_{41}^{(x)} \delta_4
\end{align*}
\]

Note that the above specification implies two endogenous latent variables, \(\eta_1\) and \(\eta_2\), measured by various structural and economic indicators (see Table 1). Namely, \(\eta_1\) presents an economic development dimension and is measured by income per capita \((y_1)\), municipality income \((y_3)\), social aid per capita \((y_6)\), population density \((y_8)\), share of agricultural population \((y_7)\) and share of population earning income \((y_2)\). We descriptively term this latent variable “economic development dimension”. The second endogenous latent variable, \(\eta_2\), is assumed to denote “structural development dimension” and is measured by employment \((y_4)\),
unemployment \((y_3)\), share of population earning income \((y_2)\) and share of agricultural population \((y_7)\). Note that \(y_2, y_6,\) and \(y_7\) are compound indicators of both \(\eta_1\) and \(\eta_2\). Finally, a “demographic development dimension”, denoted by \(\xi_1\), is measured by education \((x_1)\), age index \((x_2)\), population trend \((x_3)\) and vitality index \((x_4)\).

The structural part of the model, i.e., relationships among latent variables, is given by

\[
\begin{bmatrix}
\eta_1 \\
\eta_2
\end{bmatrix} = \begin{bmatrix} 0 & \beta_{12} \\ 0 & 0 \end{bmatrix} \begin{bmatrix}
\eta_1 \\
\eta_2
\end{bmatrix} + \begin{bmatrix} y_{11} \\
y_{21}
\end{bmatrix} + \begin{bmatrix} \xi_1 \\
\xi_2
\end{bmatrix},
\]

(12)

where the latent error terms have a covariance matrix

\[
\Psi = \begin{bmatrix} \text{var}(\xi_1) & 0 \\ 0 & \text{var}(\xi_2) \end{bmatrix}.
\]

(13)

The path diagram for the Model I together with the estimated coefficients is shown in Figure 1.

**Figure 1.** Model I path diagram with standardised coefficient estimates (drawn by LISREL 8.53)

The model was estimated using two different residual covariance matrix specifications with the maximum likelihood technique. Both versions of the model (M1A & M1B) used the same specification for the \(\Theta_e\) matrix, namely
\[
\Theta_\epsilon = \begin{bmatrix}
\theta_{11}^{(e)} & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & \theta_{22}^{(e)} & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & \theta_{33}^{(e)} & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & \theta_{44}^{(e)} & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & \theta_{55}^{(e)} & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & \theta_{66}^{(e)} & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & \theta_{77}^{(e)} & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & \theta_{88}^{(e)}
\end{bmatrix},
\]

(14)

while the \( \Theta_\delta \) matrix was firstly estimated by setting all off-diagonal elements to zero. This produced an overall chi-square statistic of 316.4 (d.f. = 48) thus indicating poor fit (see Table 5 for estimates). Inspection of standardised residuals and modification indices (Sörbom, 1989) suggested relaxing zero restrictions on several off-diagonal elements of \( \Theta_\delta \), and the specification (M1B)

\[
\Theta_\delta = \begin{bmatrix}
\theta_{11}^{(\delta)} & 0 & 0 & 0 \\
0 & \theta_{22}^{(\delta)} & 0 & 0 \\
0 & \theta_{32}^{(\delta)} & \theta_{33}^{(\delta)} & 0 \\
0 & \theta_{42}^{(\delta)} & \theta_{43}^{(\delta)} & \theta_{44}^{(\delta)}
\end{bmatrix}
\]

(15)

produced a chi-square statistic of 163.5 (d.f. = 45), which is still indicative of relatively poor fit, even if we assume that the model holds in population only approximately.

It is interesting to note that the above estimated structural equation model fits less well then a similar model estimated in Cziráky, et al. 2002a;b using non-census data, which might be a consequence of different data source and the introduction of two new development indicators, namely education and share of agricultural population. A more detailed inspection of the estimation results, and especially diagnostic statistics, indicated that the problem might be in an omitted, fourth, development dimension. Fitted standardised residuals particularly indicated unsuitably modelled variables. Therefore, we considered an alternative specification that introduced another latent variable accounting for “population development dimension”, which is rather broad term that might be, to some degree, differently elaborated.
Table 5
Maximum likelihood estimates for Model 1

<table>
<thead>
<tr>
<th>Parameter</th>
<th>M1A</th>
<th>(S.E.)</th>
<th>M1B</th>
<th>(S.E.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\lambda_{21}^{(v)}$</td>
<td>0.475 (0.082)</td>
<td></td>
<td>0.468 (0.079)</td>
<td></td>
</tr>
<tr>
<td>$\lambda_{31}^{(v)}$</td>
<td>0.723 (0.081)</td>
<td></td>
<td>0.727 (0.082)</td>
<td></td>
</tr>
<tr>
<td>$\lambda_{22}^{(v)}$</td>
<td>0.809 (0.095)</td>
<td></td>
<td>0.806 (0.091)</td>
<td></td>
</tr>
<tr>
<td>$\lambda_{52}^{(v)}$</td>
<td>-1.011 (0.096)</td>
<td></td>
<td>-1.012 (0.096)</td>
<td></td>
</tr>
<tr>
<td>$\lambda_{61}^{(v)}$</td>
<td>-0.526 (0.085)</td>
<td></td>
<td>-0.507 (0.081)</td>
<td></td>
</tr>
<tr>
<td>$\lambda_{62}^{(v)}$</td>
<td>-0.539 (0.078)</td>
<td></td>
<td>-0.541 (0.078)</td>
<td></td>
</tr>
<tr>
<td>$\lambda_{71}^{(v)}$</td>
<td>-0.828 (0.097)</td>
<td></td>
<td>-0.800 (0.092)</td>
<td></td>
</tr>
<tr>
<td>$\lambda_{72}^{(v)}$</td>
<td>0.622 (0.083)</td>
<td></td>
<td>0.617 (0.082)</td>
<td></td>
</tr>
<tr>
<td>$\lambda_{81}^{(v)}$</td>
<td>0.536 (0.088)</td>
<td></td>
<td>0.468 (0.083)</td>
<td></td>
</tr>
<tr>
<td>$\lambda_{11}^{(x)}$</td>
<td>2.069 (0.347)</td>
<td></td>
<td>2.642 (0.562)</td>
<td></td>
</tr>
<tr>
<td>$\lambda_{21}^{(x)}$</td>
<td>-0.676 (0.176)</td>
<td></td>
<td>-0.486 (0.175)</td>
<td></td>
</tr>
<tr>
<td>$\lambda_{41}^{(x)}$</td>
<td>0.650 (0.174)</td>
<td></td>
<td>0.511 (0.182)</td>
<td></td>
</tr>
<tr>
<td>$\beta_{12}$</td>
<td>0.123 (0.062)</td>
<td></td>
<td>0.129 (0.062)</td>
<td></td>
</tr>
<tr>
<td>$\gamma_{11}$</td>
<td>1.523 (0.268)</td>
<td></td>
<td>1.831 (0.367)</td>
<td></td>
</tr>
<tr>
<td>$\gamma_{12}$</td>
<td>-0.105 (0.138)</td>
<td></td>
<td>-0.156 (0.160)</td>
<td></td>
</tr>
<tr>
<td>$\text{var}(\zeta_1)$</td>
<td>0.200 (0.089)</td>
<td></td>
<td>0.239 (0.119)</td>
<td></td>
</tr>
<tr>
<td>$\text{var}(\zeta_2)$</td>
<td>0.921 (0.127)</td>
<td></td>
<td>0.918 (0.127)</td>
<td></td>
</tr>
<tr>
<td>$\theta_{11}^{(e)}$</td>
<td>1.235 (0.101)</td>
<td></td>
<td>1.193 (0.099)</td>
<td></td>
</tr>
<tr>
<td>$\theta_{22}^{(e)}$</td>
<td>1.293 (0.099)</td>
<td></td>
<td>1.287 (0.099)</td>
<td></td>
</tr>
<tr>
<td>$\theta_{33}^{(e)}$</td>
<td>1.500 (0.105)</td>
<td></td>
<td>1.476 (0.104)</td>
<td></td>
</tr>
<tr>
<td>$\theta_{44}^{(e)}$</td>
<td>1.076 (0.101)</td>
<td></td>
<td>1.078 (0.101)</td>
<td></td>
</tr>
<tr>
<td>$\theta_{55}^{(e)}$</td>
<td>1.057 (0.101)</td>
<td></td>
<td>1.055 (0.101)</td>
<td></td>
</tr>
<tr>
<td>$\theta_{66}^{(e)}$</td>
<td>1.478 (0.103)</td>
<td></td>
<td>1.484 (0.103)</td>
<td></td>
</tr>
<tr>
<td>$\theta_{77}^{(e)}$</td>
<td>1.195 (0.102)</td>
<td></td>
<td>1.203 (0.101)</td>
<td></td>
</tr>
<tr>
<td>$\theta_{88}^{(e)}$</td>
<td>1.781 (0.113)</td>
<td></td>
<td>1.823 (0.115)</td>
<td></td>
</tr>
<tr>
<td>$\theta_{11}^{(d)}$</td>
<td>0.977 (0.147)</td>
<td></td>
<td>0.824 (0.228)</td>
<td></td>
</tr>
<tr>
<td>$\theta_{22}^{(d)}$</td>
<td>1.891 (0.118)</td>
<td></td>
<td>1.960 (0.120)</td>
<td></td>
</tr>
<tr>
<td>$\theta_{33}^{(d)}$</td>
<td>1.761 (0.114)</td>
<td></td>
<td>1.831 (0.116)</td>
<td></td>
</tr>
<tr>
<td>$\theta_{44}^{(d)}$</td>
<td>1.899 (0.118)</td>
<td></td>
<td>1.956 (0.120)</td>
<td></td>
</tr>
<tr>
<td>$\theta_{32}^{(d)}$</td>
<td>-</td>
<td></td>
<td>-0.457 (0.086)</td>
<td></td>
</tr>
<tr>
<td>$\theta_{42}^{(d)}$</td>
<td>-</td>
<td></td>
<td>-0.760 (0.091)</td>
<td></td>
</tr>
<tr>
<td>$\theta_{43}^{(d)}$</td>
<td>-</td>
<td></td>
<td>0.331 (0.085)</td>
<td></td>
</tr>
<tr>
<td>$X^2$</td>
<td>48</td>
<td></td>
<td>45</td>
<td></td>
</tr>
<tr>
<td>d.f.</td>
<td>316.413</td>
<td></td>
<td>163.495</td>
<td></td>
</tr>
<tr>
<td>GFI*</td>
<td>0.912</td>
<td></td>
<td>0.952</td>
<td></td>
</tr>
<tr>
<td>SRMR**</td>
<td>0.079</td>
<td></td>
<td>0.064</td>
<td></td>
</tr>
</tbody>
</table>

* Goodness of fit index; ** Standardised root-mean-square residual.
We postulated that the fourth dimension, accounted for by an additional endogenous latent variable $\eta_3$, measured by education ($x_1$), population density ($y_8$), and share of agricultural population ($y_7$) where the last one is also a compound indicator of the structural dimension. Otherwise, this model (M2) is similar to the model M1 estimated above with the difference that we now dropped the unemployment variable which was very highly (negatively) correlated to the employment indicator. The endogenous measurement model is now specified as

\[
\begin{bmatrix}
y_1 \\
y_2 \\
y_3 \\
y_4 \\
y_5 \\
y_6 \\
y_7 \\
y_8 \\
\end{bmatrix} =
\begin{bmatrix}
1 & 0 & 0 \\
\lambda_{21}^{(y)} & \lambda_{22}^{(y)} & 0 \\
\lambda_{31}^{(y)} & 0 & 0 \\
0 & 0 & 1 \\
\lambda_{51}^{(y)} & \lambda_{52}^{(y)} & 0 \\
0 & \lambda_{61}^{(y)} & \lambda_{63}^{(y)} \\
0 & 0 & 1 \\
\lambda_{81}^{(y)} & 0 & \lambda_{83}^{(y)} \\
\end{bmatrix}
\begin{bmatrix}
\eta_1 \\
\eta_2 \\
\eta_3 \\
\eta_4 \\
\eta_5 \\
\eta_6 \\
\eta_7 \\
\eta_8 \\
\end{bmatrix} +
\begin{bmatrix}
\varepsilon_1 \\
\varepsilon_2 \\
\varepsilon_3 \\
\varepsilon_4 \\
\varepsilon_5 \\
\varepsilon_6 \\
\varepsilon_7 \\
\varepsilon_8 \\
\end{bmatrix},
\]  

(16)

while its exogenous measurement model is given by

\[
\begin{bmatrix}
x_2 \\
x_3 \\
x_4 \\
\end{bmatrix} =
\begin{bmatrix}
\lambda_{11}^{(x)} \\
1 \\
\lambda_{31}^{(x)} \\
\end{bmatrix}
\begin{bmatrix}
\xi_1 \\
\xi_2 \\
\xi_3 \\
\end{bmatrix} +
\begin{bmatrix}
\delta_1 \\
\delta_2 \\
\delta_3 \\
\end{bmatrix}.
\]  

(17)

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

\[
\begin{bmatrix}
\eta_1 \\
\eta_2 \\
\eta_3 \\
\end{bmatrix} =
\begin{bmatrix}
0 & \beta_{12} & \beta_{13} \\
0 & 0 & \beta_{23} \\
0 & 0 & 0 \\
\end{bmatrix}
\begin{bmatrix}
\eta_1 \\
\eta_2 \\
\eta_3 \\
\end{bmatrix} +
\begin{bmatrix}
\gamma_{11} \\
\gamma_{21} \\
\gamma_{31} \\
\end{bmatrix}
\begin{bmatrix}
\xi_1 \\
\xi_2 \\
\xi_3 \\
\end{bmatrix} +
\begin{bmatrix}
\zeta_1 \\
\zeta_2 \\
\zeta_3 \\
\end{bmatrix},
\]  

(18)

with the latent error covariance matrix given by

\[
\Psi =
\begin{bmatrix}
\text{var}(\xi_1) & 0 & 0 \\
0 & \text{var}(\xi_2) & 0 \\
0 & 0 & \text{var}(\xi_3) \\
\end{bmatrix}.
\]  

(19)

The path diagram for model M2 is shown in Figure 2. The shown coefficient estimates come from the final, best fitting model (M2B, see below).
We first estimate model $M_2$ with diagonal residual covariance matrices ($M_{1A}$), where the $\Theta_\epsilon$ matrix is specified as in Eq. (14), and $\Theta_\delta$ is a diagonal matrix of the form

$$
\Theta_\delta = \begin{bmatrix}
\theta_{11}^{(\delta)} & 0 & 0 \\
0 & \theta_{22}^{(\delta)} & 0 \\
0 & 0 & \theta_{33}^{(\delta)} \\
\end{bmatrix}.
$$

(20)

The maximum likelihood estimates are given in Table 6. The overall fit of the model is much better than before with chi-square of 88.5 (d.f. = 35) and goodness of fit index (GFI) of 0.97. This strongly supports the inclusion of the fourth development dimension showing evidently superior fit to model $M_1$. Finding one error-covariance coefficient ($\theta_{31}^{(\delta)}$) with a relatively large modification index we re-estimated the model with $\Theta_\delta$ matrix specified as

$$
\Theta_\delta = \begin{bmatrix}
\theta_{11}^{(\delta)} & 0 & 0 \\
0 & \theta_{22}^{(\delta)} & 0 \\
\theta_{31}^{(\delta)} & 0 & \theta_{33}^{(\delta)} \\
\end{bmatrix}.
$$

(21)

We name this model $M_{2b}$. As the results in Table 6 indicate, the chi-square dropped to 77.3 (d.f. = 34) with GFI of 0.98 and SRMR of 0.04, which jointly indicates an approximately good fit to the data. Consequently, we will use the specification of model $M_{2b}$ in further secondary analysis.
Table 6
Maximum likelihood estimates for Model 2

<table>
<thead>
<tr>
<th>Parameter</th>
<th>M2A Estimate (S.E.)</th>
<th>M2B Estimate (S.E.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\lambda_{21}^{(v)}$</td>
<td>0.468 (0.080)</td>
<td>0.467 (0.080)</td>
</tr>
<tr>
<td>$\lambda_{22}^{(v)}$</td>
<td>0.707 (0.106)</td>
<td>0.706 (0.105)</td>
</tr>
<tr>
<td>$\lambda_{31}^{(v)}$</td>
<td>0.788 (0.088)</td>
<td>0.789 (0.089)</td>
</tr>
<tr>
<td>$\lambda_{51}^{(v)}$</td>
<td>-0.461 (0.080)</td>
<td>-0.461 (0.080)</td>
</tr>
<tr>
<td>$\lambda_{52}^{(v)}$</td>
<td>-0.538 (0.090)</td>
<td>-0.535 (0.090)</td>
</tr>
<tr>
<td>$\lambda_{62}^{(v)}$</td>
<td>0.557 (0.094)</td>
<td>0.562 (0.094)</td>
</tr>
<tr>
<td>$\lambda_{63}^{(v)}$</td>
<td>-0.793 (0.093)</td>
<td>-0.791 (0.092)</td>
</tr>
<tr>
<td>$\lambda_{81}^{(v)}$</td>
<td>-2.694 (1.001)</td>
<td>-2.733 (1.042)</td>
</tr>
<tr>
<td>$\lambda_{83}^{(v)}$</td>
<td>3.334 (1.053)</td>
<td>3.368 (1.085)</td>
</tr>
<tr>
<td>$\lambda_{11}^{(x)}$</td>
<td>-1.470 (0.212)</td>
<td>-0.983 (0.159)</td>
</tr>
<tr>
<td>$\lambda_{31}^{(x)}$</td>
<td>1.276 (0.187)</td>
<td>0.814 (0.147)</td>
</tr>
<tr>
<td>$\beta_{12}$</td>
<td>-0.045 (0.028)</td>
<td>-0.038 (0.027)</td>
</tr>
<tr>
<td>$\beta_{13}$</td>
<td>1.103 (0.111)</td>
<td>1.159 (0.129)</td>
</tr>
<tr>
<td>$\beta_{23}$</td>
<td>0.045 (0.083)</td>
<td>0.054 (0.082)</td>
</tr>
<tr>
<td>$\gamma_{11}$</td>
<td>-0.355 (0.130)</td>
<td>-0.374 (0.152)</td>
</tr>
<tr>
<td>$\gamma_{13}$</td>
<td>0.491 (0.124)</td>
<td>0.569 (0.131)</td>
</tr>
<tr>
<td>var($\zeta_1$)</td>
<td>0.047 (0.029)</td>
<td>0.030 (0.029)</td>
</tr>
<tr>
<td>var($\zeta_2$)</td>
<td>1.004 (0.173)</td>
<td>1.001 (0.173)</td>
</tr>
<tr>
<td>var($\zeta_3$)</td>
<td>0.699 (0.108)</td>
<td>0.623 (0.108)</td>
</tr>
<tr>
<td>$\theta_{11}^{(e)}$</td>
<td>1.089 (0.105)</td>
<td>1.092 (0.105)</td>
</tr>
<tr>
<td>$\theta_{22}^{(e)}$</td>
<td>1.305 (0.113)</td>
<td>1.303 (0.113)</td>
</tr>
<tr>
<td>$\theta_{33}^{(e)}$</td>
<td>1.434 (0.105)</td>
<td>1.434 (0.105)</td>
</tr>
<tr>
<td>$\theta_{44}^{(e)}$</td>
<td>0.994 (0.150)</td>
<td>0.997 (0.150)</td>
</tr>
<tr>
<td>$\theta_{55}^{(e)}$</td>
<td>1.521 (0.109)</td>
<td>1.520 (0.109)</td>
</tr>
<tr>
<td>$\theta_{66}^{(e)}$</td>
<td>1.220 (0.110)</td>
<td>1.222 (0.110)</td>
</tr>
<tr>
<td>$\theta_{77}^{(e)}$</td>
<td>1.206 (0.097)</td>
<td>1.204 (0.097)</td>
</tr>
<tr>
<td>$\theta_{88}^{(e)}$</td>
<td>1.043 (0.288)</td>
<td>1.039 (0.292)</td>
</tr>
<tr>
<td>$\theta_{11}^{(i)}$</td>
<td>1.152 (0.123)</td>
<td>1.482 (0.127)</td>
</tr>
<tr>
<td>$\theta_{22}^{(i)}$</td>
<td>1.608 (0.113)</td>
<td>1.464 (0.123)</td>
</tr>
<tr>
<td>$\theta_{33}^{(i)}$</td>
<td>1.361 (0.115)</td>
<td>1.645 (0.124)</td>
</tr>
<tr>
<td>$\theta_{31}^{(i)}$</td>
<td>- -</td>
<td>-0.373 (0.098)</td>
</tr>
<tr>
<td>$X^2$</td>
<td>35</td>
<td>34</td>
</tr>
<tr>
<td>d.f.</td>
<td>88.482</td>
<td>77.285</td>
</tr>
<tr>
<td>GFI*</td>
<td>0.971</td>
<td>0.975</td>
</tr>
<tr>
<td>SRMR**</td>
<td>0.047</td>
<td>0.042</td>
</tr>
</tbody>
</table>

* Goodness of fit index; ** Standardised root-mean-square residual.
5. Latent regional development scores

Using the parameters of the estimated LISREL model from Eqs. (14), (16)-(19), and (21), 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)\(^4\). This approach is based on the maximum likelihood solution of structural equation models such as M\(_1\) and M\(_2\) estimated above. Writing Eqs. (3) and (4) in a system as

\[
\begin{pmatrix}
y' \\
x'
\end{pmatrix} = \begin{pmatrix}
\Lambda_x & 0 \\
0 & \Lambda_x
\end{pmatrix} \begin{pmatrix}
\eta' \\
\xi'
\end{pmatrix} + \begin{pmatrix}
\epsilon' \\
\delta'
\end{pmatrix},
\]  

(22)

and using the following notation

\[
\Lambda \equiv \begin{pmatrix}
\Lambda_x & 0 \\
0 & \Lambda_x
\end{pmatrix}, \quad \xi' \equiv \begin{pmatrix}
\eta' \\
\xi'
\end{pmatrix}, \quad \delta' \equiv \begin{pmatrix}
\epsilon' \\
\delta'
\end{pmatrix}, \quad x' \equiv \begin{pmatrix}
y' \\
x'
\end{pmatrix},
\]  

(23)

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

\[
\xi_a = UD^{1/2}VL^{-1/2}V^T D^{1/2}U^T \Lambda_a^{1/2} \Theta_a^{-1} x_a,
\]  

(24)

where \(UD^T\) is the singular value decomposition of \(\Phi_a = E(\xi_a \xi_a^T)\), and \(VL^T\) is the singular value decomposition of the matrix \(D^{1/2}U^T B U D^{1/2}\), while \(\Theta_a\) is the error covariance matrix of the observed variables. Derivation of Eq. (24) 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 \(\xi_{ai}\) can be computed for each observation \(x_{ij}\) in the \((11 \times N)\) sample matrix whose rows are observations on each of our 11 observed indicators and \(N = 546\), i.e.,

\[
\begin{pmatrix}
x_{11} & x_{12} & \cdots & x_{1N} \\
x_{21} & x_{22} & \cdots & x_{2N} \\
x_{31} & x_{32} & \cdots & x_{3N} \\
x_{41} & x_{42} & \cdots & x_{4N} \\
\vdots & \vdots & \cdots & \vdots \\
x_{11,1} & x_{11,2} & \cdots & x_{11,N}
\end{pmatrix} = \begin{pmatrix}
x_1 \\
x_2 \\
x_3 \\
x_4 \\
\vdots \\
x_{11}
\end{pmatrix}.
\]  

(25)

\(^4\)See also Cziráky, et al. (2002c;d).
6. Cluster analysis

Having computed the latent variable scores, we perform $K$-means cluster analysis with the purpose of grouping (clustering) municipalities into several groups with similar characteristics (for more details on the cluster analysis procedure see Everitt, 1993). In this context we use cluster analysis as purely exploratory technique and perform clustering of the 546 municipalities on the basis of the four latent variables. The $K$-means procedure was applied due to relatively large number of cases (546) and the construction of initial cluster centres was based on $k$ well spaced observations, selected by the algorithm (Euclidian distances).\footnote{Alternatively, initial cluster centres can be specified a priori, for example, they can result from a hierarchical procedure (e.g., Ward's method) or from prior (or subjective) information (see e.g., Rovan and Sambt, 2002, for a similar application to clustering of Slovenian municipalities).} The $K$-means technique requires \textit{a priori} specified number of clusters and thus the final solution largely depends on the substantive interpretability of various solutions, i.e., for different $k$, (see e.g. Cziráky and Čumpek, 2002). On the basis of the specified $k$ and the initially obtained cluster centres, the $K$-means procedure assigns cases to clusters based on their distance from the cluster centres and updates the locations of cluster centres based on the mean values of the cases in each cluster. These steps are repeated until any reassignment of cases would make the clusters more internally variable or externally similar.

On the basis of the dendrogram obtained from preliminary hierarchical clustering (Figure 3) and after investigation of a number of alternative solutions (varying $k$ from 2 to 5) and substantive interpretability, we accepted a 5-cluster solution obtained on the basis of four latent variables.

![Dendrogram](image)

\textbf{Figure 3. Dendrogram}
We note that clustering of the original development indicators (results are omitted), while providing to large degree similar picture, lacked clarity in interpretation mainly due to a need to analyse 11 instead of only 4 cluster centres, which is a major advantage of clustering on the basis of latent variables. The 5-cluster solution converged in 23 iterations, noting that the fifth cluster from the start of iterations included a single municipality only (Table 7).

<table>
<thead>
<tr>
<th>Iteration</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1.034</td>
<td>1.692</td>
<td>1.390</td>
<td>1.523</td>
<td>0.000</td>
</tr>
<tr>
<td>2</td>
<td>0.589</td>
<td>0.388</td>
<td>0.171</td>
<td>0.187</td>
<td>0.000</td>
</tr>
<tr>
<td>3</td>
<td>0.428</td>
<td>0.169</td>
<td>0.092</td>
<td>0.078</td>
<td>0.000</td>
</tr>
<tr>
<td>4</td>
<td>0.274</td>
<td>0.078</td>
<td>0.034</td>
<td>0.022</td>
<td>0.000</td>
</tr>
<tr>
<td>5</td>
<td>0.096</td>
<td>0.027</td>
<td>0.006</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>6</td>
<td>0.075</td>
<td>0.014</td>
<td>0.007</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>7</td>
<td>0.057</td>
<td>0.005</td>
<td>0.012</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>8</td>
<td>0.099</td>
<td>0.000</td>
<td>0.030</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>9</td>
<td>0.078</td>
<td>0.022</td>
<td>0.045</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>10</td>
<td>0.040</td>
<td>0.024</td>
<td>0.039</td>
<td>0.019</td>
<td>0.000</td>
</tr>
<tr>
<td>11</td>
<td>0.072</td>
<td>0.000</td>
<td>0.031</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>12</td>
<td>0.029</td>
<td>0.005</td>
<td>0.018</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>13</td>
<td>0.063</td>
<td>0.000</td>
<td>0.034</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>14</td>
<td>0.032</td>
<td>0.005</td>
<td>0.024</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>15</td>
<td>0.032</td>
<td>0.009</td>
<td>0.028</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>16</td>
<td>0.033</td>
<td>0.014</td>
<td>0.034</td>
<td>0.009</td>
<td>0.000</td>
</tr>
<tr>
<td>17</td>
<td>0.029</td>
<td>0.023</td>
<td>0.041</td>
<td>0.010</td>
<td>0.000</td>
</tr>
<tr>
<td>18</td>
<td>0.033</td>
<td>0.015</td>
<td>0.037</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>19</td>
<td>0.019</td>
<td>0.036</td>
<td>0.039</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>20</td>
<td>0.012</td>
<td>0.030</td>
<td>0.028</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>21</td>
<td>0.007</td>
<td>0.000</td>
<td>0.009</td>
<td>0.010</td>
<td>0.000</td>
</tr>
<tr>
<td>22</td>
<td>0.005</td>
<td>0.000</td>
<td>0.004</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>23</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
</tr>
</tbody>
</table>

This situation persisted across all reasonably looking solutions and a closer investigation of the fifth cluster revealed that each time it included municipality Civljani. As it is actually the case, Civljani have extremely high age index value due to significant preponderance of elderly people (5200 or over 22 in z-score value), which even after standardisation pertained to cause outlier problems. Therefore, we conclude there are no more then four true clusters and treat this outlier as a special case. Clustering municipalities after dropping this outlier

---

6 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.
produces equivalent results for other four clusters, thus there is no need to separately present these results.

Table 8 gives the ANOVA results, which indicate highly significant discriminatory power of each latent variable. The newly introduced, fourth, development dimension, while still highly significant, has notably lower value of the F test, thus showing somewhat less discriminatory potential than the other two latent variables. 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 nor for that the criteria variables for clustering were actually linear combinations of observable development indicators.

<table>
<thead>
<tr>
<th>Cluster</th>
<th>Mean square</th>
<th>d.f.</th>
<th>Mean square error</th>
<th>d.f.</th>
<th>F-test</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>ECON</td>
<td>48.623</td>
<td>4</td>
<td>0.128</td>
<td>541</td>
<td>381.302</td>
<td>0.000</td>
</tr>
<tr>
<td>STRUC</td>
<td>98.439</td>
<td>4</td>
<td>0.291</td>
<td>541</td>
<td>338.142</td>
<td>0.000</td>
</tr>
<tr>
<td>POPUL</td>
<td>51.701</td>
<td>4</td>
<td>0.143</td>
<td>541</td>
<td>361.301</td>
<td>0.000</td>
</tr>
<tr>
<td>DEMO</td>
<td>2.848</td>
<td>4</td>
<td>0.153</td>
<td>541</td>
<td>18.617</td>
<td>0.000</td>
</tr>
</tbody>
</table>

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 9 shows cluster centres for each cluster (expressed in standardised units). It can be easily inferred that cluster 2 includes the most economically developed municipalities (with higher per capita and municipality incomes, higher share of income earning population and lower social aid per capita). These are northern Adriatic regions including most of Istria and part of western continental Croatia (see Figure 4). Cluster 2 is also characterised by higher score on the latent population dimension (POPUL), which basically indicates more positive population trend, higher vitality index and younger population (i.e. lower age index). Cluster 3, on the other hand, includes areas relatively high in per capita and municipality income, but lower in employment (structural dimension). This cluster corresponds primarily to continental areas of northern and middle Croatia.

<table>
<thead>
<tr>
<th>Cluster</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>N = 131</td>
<td>0.40</td>
<td>1.77</td>
<td>0.93</td>
<td>0.33</td>
</tr>
<tr>
<td>N = 136</td>
<td>-1.10</td>
<td>0.26</td>
<td>-0.17</td>
<td>1.43</td>
</tr>
<tr>
<td>N = 176</td>
<td>0.51</td>
<td>1.91</td>
<td>1.02</td>
<td>0.35</td>
</tr>
<tr>
<td>N = 102</td>
<td>0.14</td>
<td>0.07</td>
<td>-0.01</td>
<td>-0.26</td>
</tr>
</tbody>
</table>
Finally, clusters 4 and 1 are both low in income indicators with the difference that cluster 1 is apparently least developed as it has not only low income score but also low structural, mainly employment characteristics, while cluster 4 includes higher-employment, more agricultural areas, which are nevertheless low on earnings parameters. On the other hand, cluster 4 is characterised by a poor demographic dimension, namely bad population trend and vitality index with higher share of elderly population.

Distances between pairs of cluster centres (in standardised scale) are shown in Table 10. It can be seen that the largest overall gap exists between clusters 1 and 2, and similarly noticeable is the difference between cluster 1 and 4.

<table>
<thead>
<tr>
<th>Cluster</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2.569</td>
<td>1.854</td>
<td>2.441</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>2.569</td>
<td>1.196</td>
<td>2.390</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>1.854</td>
<td>1.196</td>
<td>1.299</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>2.441</td>
<td>2.390</td>
<td>1.299</td>
<td></td>
</tr>
</tbody>
</table>

* Fifth single-member cluster is omitted

Figure 4. Development-level map of Croatia
On the other hand, clusters 1 and 4 appear to be closer together thus jointly comprising generally underdeveloped part of Croatia. Similarly, clusters 2 and 3 appear closer together then to the other clusters, and these two clusters actually comprise the more developed Croatian territories. Therefore, we are most likely looking into a major development split where each development category is further split into two groups of territorial units each having certain development specifics. Additionally, the computed latent scores on each development dimension can be used to rank-order municipalities within each cluster.

7. Conclusion

In this paper we combined structural equation econometric modelling with more descriptive cluster analysis techniques in order to obtain a development grouping of Croatian municipalities. This approach allowed us to model regional development by postulating various latent development dimension 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 fails to satisfy necessary distributional assumptions required by these techniques. 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 newly available Croatian census data, we estimated two structural equation models: with three and four latent development variables (dimensions), respectively. It was found that a four-dimensional model fits the data significantly better then the three-dimensional model, which, after including few more newly available development indicators, suggested a more complex structure of Croatian regional development then previously anticipated (Cziráky, et al. 2002a;b).

Secondly, we performed cluster analysis on the estimated latent variable scores, thus in effect clustering latent development dimensions instead of raw variables. While similar to the results we obtained with the raw data, latent variable clustering offered a clearer picture and much easier interpretability. In effect, we were able to distinguish four clusters of municipalities, grouped on the basis of their latent development characteristics.
Finally, we emphasise that estimation of latent variables, i.e., underlying development dimensions, resulted also 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.

**Acknowledgments**

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

**References**


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


