Multilevel factor models: identification of three-level model parameters for the study of regional development in Argentina

María del Pilar Díaz, José M. Vargas, Margarita Díaz

Ponencia presentada en 28th International Workshop on Statistical Modelling realizada en 2013 en Universitá di Palermo. Palermo, Italia

Multilevel factor models: Identification of Three-level Model Parameters for the Study of Regional Development in Argentina

Maria del Pilar Díaz[1], José M. Vargas[23], Margarita Díaz[3]

[1] Biostatistics Unit, Medical Sciences Department, UNC, Córdoba, Argentina
[2] ICBA, UNVM, Villa María, Argentina
[3] Institute of Mathematics and Statistics, UNC, Córdoba, Argentina

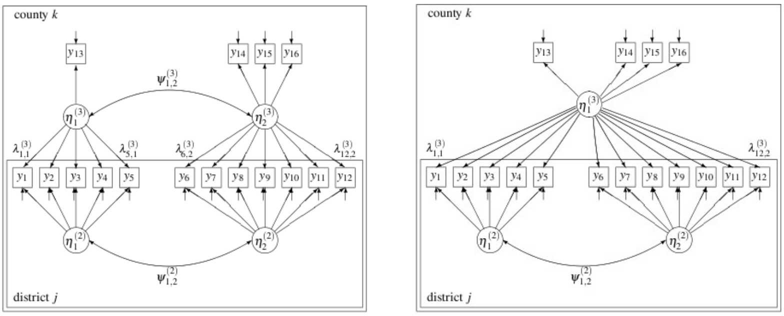
E-mail for correspondence: pdiaz@fcm.unc.edu.ar

Abstract: Based on a theoretical-social model which states that Communication and Information Technologies (CITs) infuence the human development, due to the impact they have on economic growth, this work explores the relationship between the social, economic and technological dimensions or constructs, in a province, Córdoba, of Argentina. Confirmatory common framework (GLLAMMs, Skrondal & Rabe-Hesketh, 2004) was performed considering three level modeling. This approach allowed us to identify some indicators that measure constructs and help characterize the level of development of our region using hierarchical information. Since the estimation method was based on full information maximum likelihood, we devoted special attention to the identifiability of the parameters. Two constructs to describe the socioeconomic and technological (SET) development at the district level were obtained. Due to correlation between the two latent variables at department level was near one, a new common factor model containing only one dimension was appropriate.

Keywords: GLLAMM;latent variable; CFA; socio-technological dimensions; identifiability

1 Motivation and Modeling

The measurement of access and use of ICTs, as well as its dynamics, is indispensable to understand the development of today's information societies and to support adequate design of policies. The present work contributes to this measurement, since it identifies the factors associated with SET development and it studies its regional distribution in Córdoba, Argentina. It examines the dimensions of SET constructs in the nested political divisions of Córdoba, and analyzes the relationship between them. We used confirmatory factor analysis (CFA) to explore the dimensionality of constructs. In the multidimensional case, an important example of a restricted model is that of a complexity one model or independent clusters model, where Λ



(a) First model. (b) Second model.

FIGURE 1. Models considered. Short arrows pointing to boxes y_i correspond to residuals ϵ_i; long arrows from circles $\eta_s^{(l)}$ pointing to boxes y_i correspond to loadings $\lambda_{i,s}^{(l)}$, where l stands for levels 2 or 3.

has enough elements set to zero so that each indicator measures one and only one factor. Such a configuration makes sense if one set of indicators is designed to measure one factor and another one is designed to measure another factor. A single-level factor model would not be appropriate in the present study because our information was hierarchical (municipal districts arranged into departments). Thus, we used a three-level factor model, considering districts as level 2 and departments as level 3, with two latent variables in each one. Twelve indicators measured at the district level were used to define two latent variables at each level. A graphical representation of the model is given in figure 1(a). Latent variables are represented by circles, observed variables by rectangles and arrows connecting circles; rectangles also represent regressions residuals being the short arrows pointing at circles or rectangles. Curved double-headed arrows connecting two variables indicate that they are correlated. It is typically assumed that common and unique factors have multivariate normal distributions. Four additional indicators were included at the department level in order to generate the department constructs, giving a total number of indicators of $I = 16$. The three-level model chosen for our study has the following matrix formulation:

$$\mathbf{y} = \beta + \Lambda^{(2)}\eta^{(2)} + \Lambda^{(3)}\eta^{(3)} + \epsilon,$$

where β is the expectation of the observed variables $\mathbf{y}$ (indicators), $\Lambda^{(2)}$, $\Lambda^{(3)}$ denote the 16×2 factor loadings matrices at second and third level respectively, $\eta^{(2)}$, $\eta^{(3)}$ denote the 2×1 latent factors at second and third level respectively, and ϵ is the error term that is assumed independent of latent factors. Additionally we introduce the following matrices of parameters

$$\Psi^{(l)} = \mathbb{V}[\eta^{(l)}] = \mathbb{E}[\eta^{(l)}\eta^{(l)\prime}],\ l = 2, 3,$$
$$\Theta = \mathbb{V}[\epsilon] = \mathbb{E}[\epsilon\epsilon'],$$

with the usual convention that $\mathbb{E}[\eta^{(l)}] = 0$ and $\mathbb{E}[\epsilon] = 0$. Then the covariance matrix of the observed variables is

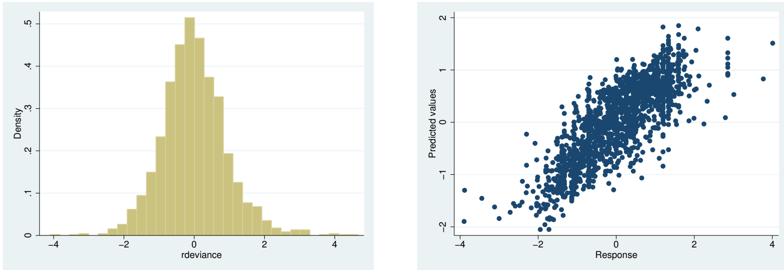
$$\Sigma = \mathbb{V}[\mathbf{y}] = \Lambda^{(2)}\Psi^{(2)}\Lambda^{(2)'} + \Lambda^{(3)}\Psi^{(3)}\Lambda^{(3)'} + \Theta. \tag{1}$$

2 Identification of Three-level Model Parameters

The identification of parameters becomes an issue given the relative large number of them and we understand it as described in Skrondal(2004), p. 135-138, and O'Brien(1994). Bollen(1989) summarizes and extends several rules that establish the identifiability of models; O'Brien(1994) extends those rules even further. These rules apply only to models with a factor complexity of one; that is, models in which each indicator loads only on a single latent variable. Our model being of complexity two, does not satisfies these conditions requiring a new proof specially tailored for the case at hand. Nevertheless, the identification process of parameters is essentially a hierarchical one and this will prove to be enough. We show that parameters of our model can be determined uniquely from information of measured variables. Through algebraic manipulation, we show that if $\Lambda^{(l)}$, $\Psi^{(l)}$ and Θ exist such that the relationship above for Σ holds, then $\Lambda^{(l)}$, $\Psi^{(l)}$ and Θ must be unique. First of all, we notice for further convenience that levels two, three, and latent factors, induce a block partition on the above equation for Σ, $\Lambda^{(2)}\Psi^{(2)}\Lambda^{(2)'}$, $\Lambda^{(3)}\Psi^{(3)}\Lambda^{(3)}$ and Θ have a block partition of shape 3 by 3. We label those blocks $A, D', E', D, B, F', E, F, C'$, from left to right and top to bottom respectively, where A is 5×5, B is 7×7, C is 4×4, D is 7×5, E is 4×5 and F is 4×7. We will refer to those blocks in any of the matrices involved to locate entries that will be of interest to us. We proceed in steps: first we prove that third level parameters can be identified by blocks E, F and C of Σ, that is, all of $\Lambda^{(3)}\Psi^{(3)}\Lambda^{(3)'}$ and block C of Θ can be identified, then we show that all remaining parameters in $\Lambda^{(2)}\Psi^{(2)}\Lambda^{(2)'}$ and blocks A and B of Θ can be identified. Proceeding in this way, we show that the identification of parameters in the first model proposed with four latent factors, $\eta_1^{(2)}$, $\eta_2^{(2)}$, $\eta_1^{(3)}$ and $\eta_2^{(3)}$ is guaranteed; the second model proposed with one factor in the third level, can be shown to be identified by similar algebraic manipulations.

3 Results

The performance of model was suitable. Figure 2 illustrates the behavior of deviance residuals and predicted values versus indicators. For the first one we also observed that percentiles 5, 50 and 95 were respectively -1.4113, -0.0492 and 1.5953, which confirm appropriate representation of our model. Our results indicated that two constructs were significant to



(a) Deviance residuals. (b) Predicted values vs indicators.

FIGURE 2. Behavior of deviance residuals and predicted values versus indicators

describe the socio-economic and technological development at the district level ($\hat{\rho}^{(2)} = 0.55$. All the indicators included in the common factor model were significant ($p < 0.05$) at this lower level. However, at the department level, the correlation between the two latent variables was near one ($\hat{\rho}^{(3)} = 0.88$), suggesting a new common factor model containing only one dimension at this level (figure 1(b)). The percentage of households with telephone was the most important socio-economic indicator at the district level, while the percentage of households with health insurance coverage the least important. Also at this level, the most important technological indicator was the percentage of the population with secondary-level education or beyond. When the new common factor model, with only a latent variable at the department level, was fitted similar factor loading estimates were obtained at both levels. We have chosen this model even though a negligible change was obtained for AIC statistic. All the socio-economic items were significant, except for percentage of homeowner at a higher level, whereas percentage of municipalities with web-site was the only non significant technological item. In addition, a correlation equal to 0.58 was observed at the district level, confirming two constructs for it.

References

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

O'Brien, R. M. (1994). *Identification of Simple Measurement Models with Multiple Latent Variables and Correlated Errors*. In: *Sociological Methodology*, Vol. 24.

Skrondal, A. and Rabe-Hesketh, S. (2004). *Generalized Latent Variable Modeling: multilevel, longitudinal, and structural equation models.*. Chapman & Hall/CRC.