

RESEARCH

Open Access



A longitudinal study of social lag: regional inequalities of growth in Mexico 2000 to 2015

Delfino Vargas-Chanes*  and Servando Valdés-Cruz

* Correspondence: dvchanes@gmail.com

Universidad Nacional Autónoma de México, Ciudad de México, México

Abstract

Social lag is an indicator that measures social development in Mexico. The institutions in charge of measuring poverty require studies for measuring the extent to which social programs are efficient to combat poverty. In the past, SL has been measured using principal component analysis but this approach cannot be used for longitudinal studies. We present an alternative analysis for measuring social lag to overcome this difficulty. For this study, we use the census data for the years 2000, 2005, 2010, and 2015, which include 2446 municipalities. The results are summarized in three steps: (1) The number of SL indicators is reduced from 11 to 6 using confirmatory factor analysis techniques and the adjustment is found to be satisfactory; (2) mixture latent growth curve models were used to estimate growth trajectories of the municipalities from 2000 to 2015; (3) we used LISA Maps and the Moran index to identify regions of potential growth. In conclusion, we observed an unequal development of the municipalities in Mexico during 15 years of application of the social policy to reduce poverty and inequality. The present work contributes to providing evidence for elaborating public policies for targeting communities in need.

Keywords: Social lag, Inequality, Latent growth curve model, Spatial analysis

Introduction

Social lag in Mexico

The social lag index (SLI) in Mexico measures the household living conditions that include several aspects: education, access to health, basic infrastructure in housing, quality services, and space in housing and household assets. The SLI is a measure based on a methodology proposed by the National Council for the Evaluation of Social Development Policy (CONEVAL) in Mexico since 2007, using data from the Census of Population and Housing. The purpose of the SLI is to classify the municipalities of the country according to the levels of their social indicators, as well as to provide information of the locations that contain very similar characteristics (CONEVAL 2011).

In Mexico, several official publications indicate the need for measuring poverty, marginalization, and social lag indices, which are useful for assessing the degree of development of the communities. However, inequalities have prevailed over time and have lasted for years. In this article, we propose an alternative social lag index using a longitudinal database that consists of 2540 municipalities in Mexico, measured from 2000

to 2015. It allows us to describe the longitudinal trajectory of social lag of these municipalities.

The General Law of Social Development was approved by the Mexican Senate in 2004, after its promulgation became mandatory to incorporate institutional mechanisms for evaluating and monitoring social development policies. As a result, CONEVAL became a public entity with technical and autonomous administration. Since 2010 CONEVAL measures multidimensional poverty of Mexico every 2 years with inferences at the state level and every 5 years at the municipal level. Also, CONEVAL is tasked to regulate and coordinate evaluation of social development policies and programs as well as to establish the guidelines and criteria for defining, identifying, and measuring poverty and degree of social cohesion (CONEVAL 2010).

Besides poverty, there is a constant problem of inequality in the country and causes some regions to experience lag or stagnation generated by the slow development in some urban, agricultural, industrial, and touristic areas (Rubalcava and Ordaz 2016).

CONEVAL proposed the SLI in 2015, using indicators of education, access to health services, basic housing infrastructure, quality of housing spaces, and household assets. Although it is not a measure of poverty (among other reasons because it does not incorporate the dimensions of income, social security, and food), it generates useful information at the state, municipal, local, and AGEB (acronym in Spanish, Geo-statistical geographic areas) levels to guide social development efforts of state and municipal governments. On May 23, 2013, an agreement was published in the Official Gazette of the Federation, announcing the incorporation of the SLI into the National Catalog of Indicators. The estimation of this index uses census data, as well as inter-census surveys.

High social lag, marginalization, poverty, and inequality constitute a cocktail of undesirable factors associated with lack of growth and development of México. Some of these political, social, and geographic factors might be associated with inequality and poverty that affect the less favored population living in isolated regions. Public policies must aim at overcoming inequality and poverty in areas with high disadvantage and providing equalizing conditions to the population in need. For example, the indigenous population is vulnerable, since they are concentrated in some of the sub-regions that lack opportunities and high levels of inequality (Calderón Villareal and Herreros, 2016).

SL is associated with deteriorated living conditions that individuals and families face in everyday life. We include in this analysis several variables measured by the percentage of population without access to education, healthcare, basic housing infrastructure, quality housing infrastructure, housing space, and household assets.

The SLI measures poverty-related indicators from a territorial perspective (although is not a measure of poverty, as previously annotated)—that is, the units of analysis are regions, states, municipalities, and localities. The objective of the index is to classify areas based on whether there is an advancement or a stagnation of the living conditions of the population within a given geographic space. There are 11 explanatory variables for the social lag, and they are divided into three thematic axes: (a) characteristics of the home, (b) characteristics of household members, and (c) multidimensional poverty.

Unlike poverty, the SLI considers several characteristics distinctive from the monetary income of families or individuals to explain the stagnation or development of geographical areas, which allows for comparison within and between regions. Also, the

SLI takes into account aspects that are related to development. Therefore, this can be useful to allocate resource to where it is needed.

SLI helps policy-making target the population in the areas with the greatest lag in the country and thereby promotes the advancement for the disadvantaged population. The objective of the social policy is the search for homogeneous development of the population, and the SLI helps monitor areas within the national territory that require specific attention for equal development.

Data and methods

Our data is obtained from Census of Population and Housing published by CONEVAL for 2450 municipalities in México for years 2000, 2005, 2010, and 2015¹. The social lag index-CONEVAL (SLI-C) contains 11 variables related to social lag. They are divided into three thematic axes: Characteristics of household members, characteristics of the home, and household assets. Respectively, each variable indicates the percentage of the population (1) that are illiterate ($x1$), (2) with children of school age of 6–14 years who do not attend school ($x2$), (3) with truncated or incomplete basic education ($x3$), (4) not entitled to any health service ($x4$), and (5) that live in houses without the following basic infrastructure: drainage ($x5$), ground floor ($x6$), toilet ($x7$), piped water ($x8$), electricity ($x9$), washing machine ($x10$), or refrigerator ($x11$).

The analysis of this article begins with the construction of an alternative index of social lag from a longitudinal perspective. Three phases are proposed: (1) the first consists of reviewing the 11 variables of social lag using confirmatory factor analysis (CFA) and determine the psychometric properties of a social lag index-revised (SLI-R). In this way, we compare our SLI-R with the one obtained by CONEVAL (SLI-C) and verify if the alternate index has acceptable psychometric properties. Once these properties are met, this index can be used for longitudinal purposes. (2) The second phase consists of fitting a latent growth curve and mixture models for determining the number of social lag trajectory groups of the 2450 municipalities measured in the years 2000, 2005, 2010, and 2015. (3) Finally, we use spatial models to identify the geographical locations of the latent trajectory classes of municipalities previously identified in a map.

Principal components analysis

The SLI-C was calculated using 11 variables via principal component analysis (PCA). However, this approach presents several challenges that prevent us from exploring the social lag of the municipalities over time. PCA standardizes the variables for each year, that is, each original variable is subtracted from its mean and divided by its standard deviation. The implication of the standardization is that the scores estimated via PCA cannot be compared over time since the scores produced are relevant only for the year when they are calculated, so we cannot observe an increase or decrease of social lag in years of study (Cortés and Vargas, 2017).

PCA aims to maximize the variance for each component with a linear combination of the original variables. When using the principal components technique, a substantial reduction of the dimensions is sought, and each component contains as much information as possible for all variables. In other words, when dimensions are reduced, loss of variance is minimized. When the first component is calculated, the variables are

standardized and generate a linear combination, named c_1 , whose weights allow us to account for the maximum variance. But it is necessary to impose a restriction on the vector of components, $c_1^t c_1 = 1$; otherwise, this variance could be infinitely large, that is, the vector has a unitary norm ².

The first component accounts for the maximum variance and has associated the eigenvalue λ_1 with the corresponding linear combinations c_1 . The second component corresponds to the vector c_2 , which has a maximum variance subjected to the normalization condition $c_2^t c_2 = 1$ and to the restriction that the first and second components must be orthogonal (independent), which is equivalent to establishing that $c_1^t c_2 = 0$; the second component has the eigenvalue associated with it and fulfils $\lambda_1 \geq \lambda_2$.

This same procedure is repeated as many times as there are variables, thus obtaining p -eigen values. By construction, the principal components are not correlated to each other, and each eigenvalue is equal to the variance of its corresponding factor, so that the percentage of variance explained by each component is equal to $100 \times \lambda_i / \sum_{i=1}^p \lambda_i$ and fulfills the property that $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p \geq 0$.

In our case, the number of variables p for the SLI is 11, and the number of municipalities n is approximately 2450. The SLI scores of each municipality can be obtained using the solution provided by the PCA through the linear combinations (C).

$$\begin{aligned} C_1 &= c_{11}x_1 + c_{12}x_2 + \dots + c_{1p}x_p \\ C_2 &= c_{21}x_1 + c_{22}x_2 + \dots + c_{2p}x_p \\ &\vdots \\ C_p &= c_{p1}x_1 + c_{p2}x_2 + \dots + c_{pp}x_p \end{aligned}$$

where x_j are the observed values of each variable, c_{ij} are the loads estimated by the PCA. This analysis transforms the original observations into a new coordinate system, in which the coordinates are the C (components), instead of the x (the variables). However, both systems are of the same dimension (p) to reduce the number of coordinates (or dimensions) very often the criterion is used which consists in retaining those components that have an eigenvalue greater than one (Kaiser 1960). It must be reminded, as has been noted at the beginning, that the matrix X contains standardized variables that have an average of 0 and a variance of 1. The criterion for retaining components is equivalent to considering only those with a greater variance than that of the variables. Also, it must be remembered that standardization prevents the comparison over time of the marginal index values as well as their degrees.

Confirmatory factorial analysis

To solve the problem of generating an alternative SLI and compare scores of social lag longitudinally, we use factorial analysis. This alternative index could be sensible to the evolution of social lag over time by using a first- or second-order confirmatory factor analysis (CFA) and generate the scores, accordingly.

The factorial analysis is a method used to explore the variation and covariation of a set of variables x_j ($j = 1, \dots, p$) also called manifest variables, as a function of factors F_k ($k = 1, \dots, m$), $m < p$, also called latent variables.

In our case, the factorial model considers the p variables of social lag x_1, x_2, \dots, x_p and the covariance matrix. The factorial model can be represented as follows:

$$\begin{aligned}
 F_1 &= \alpha_1 + \lambda_{11}x_1 + \lambda_{12}x_2 + \dots + \lambda_{1p}x_p + \varepsilon_1 \\
 F_2 &= \alpha_2 + \lambda_{21}x_1 + \lambda_{22}x_2 + \dots + \lambda_{2p}x_p + \varepsilon_2 \\
 &\vdots \\
 F_m &= \alpha_p + \lambda_{m1}x_1 + \lambda_{m2}x_2 + \dots + \lambda_{mp}x_p + \varepsilon_m
 \end{aligned}$$

where $m < p$, α_i are the means, λ_{jk} represent the factor loadings, F_m are the common factors, and ε_{ij} are the residual (specific factors) with mean zero and are independent of the factors, this means that they are not related with them.

Mixture latent growth curve model

We use the CFA scores produced in the previous step to generate the alternate SLI-R. The second phase consists of fitting a latent growth curves (LGC) model for estimating the intersections and random slopes that describe the initial value and the rate of change of SLI-R. The LGC model can be written as follows

$$\begin{aligned}
 Y_{it} &= \pi_{0i} + \pi_{1i}\lambda_t + \varepsilon_{it} \\
 \pi_{0i} &= \beta_0 + r_{0i} \\
 \pi_{1i} &= \beta_1 + r_{1i}
 \end{aligned}$$

where y_{it} is the SLI-R scores measured for the i -th municipality at time t ; the effect of time is λ_t , with the corresponding ε_{it} error terms, and π_{0i} and π_{1i} denote the intercept and slope, respectively, for each municipality; the terms r_{0i} and r_{1i} denote the random errors of the corresponding intercepts and slopes for each social lag trajectory for each municipality i -th.

Within this second phase, we use mixture models to assess latent classes of growth. That is, to further understand the trajectories, we need to find the appropriate number of classes of trajectories over time. These trajectory classes are helpful for identifying the number of groups that characterize the evolution of social lag trajectories for all municipalities. This step is relevant for generating social policies. To select the number of classes, we use five criteria based on the LGC model.

- i. The Bayes Information Criterion (BIC). The smaller the value, the better the model fit (Sclove 1987; Schwartz 1978).
- ii. The entropy value. The closer to the unit, the better fitness (Jones et al., 2001).
- iii. All classes must have at least 5% of the total observations. As Collins and Lanza (2010) suggest, if any of the groups has less observations than that, it indicates an incorrect grouping (when incorporating municipalities with outliers).
- iv. The probability of belonging to the municipalities in each group must be greater than or equal to 0.70 (Jones et al., 2001).
- v. The Lo-Mendell-Rubin test, which test null hypothesis $H_0: p$ against alternative hypothesis $H_1: p-1$ (Lo, Mendell, and Rubin 2001).

Spatial autocorrelation model

The last step is to identify the geographical locations of the latent trajectory classes of municipalities previously identified in a map. This article emphasizes that social lag is a part of a series of indicators that measure the development and inequality of a country like Mexico. We underline the importance of linking our SLI-R to the geographical location of the municipalities for studying the regional aspect of social lag. In this way,

we shift the focus of targeted policy and social programs from municipalities to regions.

Spatial models are helpful for the geolocation of high-lagged areas and regions.³ It will be interesting to observe whether there is a spatial distribution pattern of high-lagged municipalities. The use of spatial models is helpful for producing maps and uses geolocation of municipalities to identify high-lagged regions.

Spatial tools have gained great relevance in the field of social sciences in recent years. For example, excellent publications by Urry (1987) and Massey (1991) both attributed space a fundamental role in social sciences for it “[allows] us [to examine] theoretically and empirically the interaction between social structures and space” (Massey 1991, p 271).

Before presenting the spatial-analytical tool used for this article, we define spatial autocorrelation as the degree of association of a variable of interest between regions that are usually adjacent. Another way of interpreting the spatial autocorrelation is the extent to which a specific municipality is related to other adjacent municipalities nearby (Anselin 1988).

The Moran index allows us to verify the existence of spatial autocorrelation in such a way that the relationship between the municipalities of the study can be identified. This index seeks to contrast the null hypothesis that spatial autocorrelation is absent (that is, there is no relationship regarding the index of social lag in adjacent municipalities) versus the alternative hypothesis that spatial autocorrelation exists (as either a positive or a negative value).

There are several indexes for autocorrelation, such as the C test (Geary 1954) or the G test (Getis and Ord 1992). However, the most frequently used is the I-Moran index (Anselin 1995). In this article, we use the I-Moran index at the local level, since the global level does not allow us to observe whether the spatial correlation detected over the entire territory as a whole is maintained at the municipal level. This index is given by the following expression:

$$I_i = \frac{(x_i - \bar{x})}{m_2} \sum_{j=1}^n W_{ij} (x_j - \bar{x}) \quad \text{for } i \neq j$$

where $m_2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n}$; W_{ij} is the weight spatial matrix for i -th municipality and j -th adjacent municipality; x_i is the SLI-R for the i -th municipality; x_j is the SLI-R for the j -th adjacent municipality, and \bar{x} is the mean of the SLI-R.

To graphically observe the spatial autocorrelation, we use the LISA (Local Indicators of Spatial Association) map, which allows us to locate geographically the spatial clusters and then to measure the spatial autocorrelation. A positive autocorrelation means that there is a “contagion” process, suggesting that a specific municipality attracts others with similar SLI-R. In other words, it would be easy to detect agglomerations of municipalities that have similar characteristics. Meanwhile, a negative autocorrelation tells us another story, namely, that of “absorbent” states. This means those municipalities with a low level of social lag are surrounded by municipalities with a high social lag, suggesting that some municipalities with low social lag can form regions with high potential for development, most likely because they are urban municipalities with some industrial development or tourism centers.

Results

The SLI-R includes descriptive statistics of the social lag variables that are shown in Table 1 for the years 2000, 2005, 2010, and 2015. The information shows clear differences between the means of each variable. We observe a reduction of the percentages over time, which could indicate an improvement on the variables that measure social lag. However, the standard deviation and the coefficient of variation indicate the opposite, particularly for the year 2015, so it is not appropriate to conclude that the social lag has decreased during this period. In the same way, this table provides very little information about whether the social lag indicators have improved over the years of study. For this reason, we need an SLI to summarize this information.

Principal components analysis

CONEVAL has proposed SLI-C using principal components analysis (PCA). The results of this fitting are shown in Table 2, for the years 2000, 2005, 2010, and 2015. Using Kaiser (1960) criteria,⁴ we observe that the first two components accumulate more than 60% of the variance. Note that the first component for the year 2000 explains 59.5% of the variance, but has decreased to 47.2% for the year 2015. This decrease suggests that two components are needed for the years 2005, 2010, and 2015 to satisfy the Kaiser criteria.

We foresee two problems with using this approach to construct the SLI-C. First, as we have annotated in previous paragraphs, PCA produces standardized scores that are not sensitive to changes over time. Second, the first component explains a small percentage for year 2000 and the amount of variance for the first component is smaller as time progresses.

Second-order confirmatory factor analysis

For the construction of the alternative index, SLI-R, we conducted two CFA phases.⁵ First, we defined a second-order CFA and selected the most relevant variables that are eligible for the index among the 11 originally proposed. Secondly, we used a first order CFA to generate the SLI-R scores that will be used for the fitting the longitudinal trajectories of social lag for each municipality. The second order CFA model is shown in Fig. 1; the first order factors are *F1: education*, *F2: Household characteristics*, and *F3: Household assets*.

Table 1 Descriptive statistics of the social lag variables

Description	2000			2005			2010			2015		
	Mean	Std. Dev.	Coef. Variation	Mean	Std. Dev.	Coef. Variation	Mean	Std. Dev.	Coef. Variation	Mean	Std. Dev.	Coef. Variation
x1 15+ illiterate	18.40	11.99	65.2	16.69	11.04	66.1	13.98	9.63	68.9	11.63	8.51	73.1
x2 6-14 w/school	10.07	5.19	51.6	6.42	3.61	56.2	5.41	3.17	58.6	3.86	2.60	67.5
x3 15+ w/elementary school	72.30	14.42	19.9	65.33	14.61	22.4	59.37	14.65	24.7	52.63	14.53	27.6
x4 w/health service	76.24	18.54	24.3	68.56	22.36	32.6	37.88	17.46	46.1	14.60	7.31	50.1
x5 ground floor	31.45	25.20	80.1	24.48	22.04	90.0	12.88	11.82	91.8	8.45	8.95	106.0
x6 w/toilet	23.79	16.44	69.1	14.09	12.87	91.3	8.82	9.55	108.3	5.62	7.19	127.9
x7 w/pipe water	27.04	22.13	81.9	20.73	20.81	100.4	20.08	19.95	99.3	8.83	11.34	128.4
x8 w/drainage	49.52	29.36	59.3	31.47	27.20	86.4	24.80	24.33	98.1	18.70	22.47	120.2
x9 w/electricity	10.95	12.54	114.5	7.72	8.07	104.6	4.53	6.20	136.9	2.51	3.44	137.1
x10 w/washing machine	72.29	23.58	32.6	61.83	25.62	41.4	53.80	24.93	46.3	49.30	24.90	50.5
x11 w/refrigerator	57.76	26.28	45.5	44.70	26.39	59.0	36.26	24.05	66.3	30.98	23.04	74.4

Source: Own elaboration bureau census data 2000-2015

Table 2 Results for the first two principal components for SLI-C. Mexico 2000, 2010, and 2015^a

Principal component	2000		2005		2010		2015	
	Eigen value	Accumulated proportion	Eigen value	Accumulated proportion	Eigen value	Accumulated proportion	Eigen value	Accumulated proportion
1	6.544	59.5%	5.911	53.7%	5.379	48.9%	5.191	47.2%
2	0.962	68.2%	1.170	64.4%	1.236	60.1%	1.232	58.4%
3	0.862	76.1%	1.036	73.8%	1.043	69.6%	1.104	68.4%

Source: Own elaboration based on Census Bureau data 2000-2015

^aNote that we only include only the solution of first three components for saving space, actually there are nine components

Table 3 shows the loads corresponding to the second-order CFA. For the alternative index (SLI-R), we should take into account the proposed dimensions, for the selection of variables that will be part of the index. A good index must show the composite reliability coefficient (Omega) with a value greater than 0.7 (McDonald, 1999); additionally, Comparative Fit Index (CFI) values must be greater than 0.9 and Root Mean Square Error Approximation (RMSEA) less than 0.05.⁶ However, some variables do not meet these criteria.

The second-order CFA loadings in this table show that some variables do not meet the condition of having all factorial loads greater than 0.70.⁷ For example, variables *x2*, *x4*, *x6*, *x7*, and *x9* all have very low factorial loads, suggesting a poor fit for the SLI. The rest of the variables fulfill two properties: (1) the loadings are greater than 0.70 and (2) they meet the factorial invariance property, which means that the factorial loads for all the years of study are similar. For example, *x1* has loadings of 0.857, 0.891, 0.926, and 0.940 for the years 2000, 2005, 2010, and 2015, respectively. These loads are practically the same and guarantee the comparison of the results of the factorial analysis over time, so that we used the assumption of longitudinal invariance. If loads of the variables of SLI are invariant, that is a guarantee that the index has the same metric and can be used longitudinally. Then, it is possible to argue whether the SLI of a given municipality improved or worsened over the years of the study.⁸

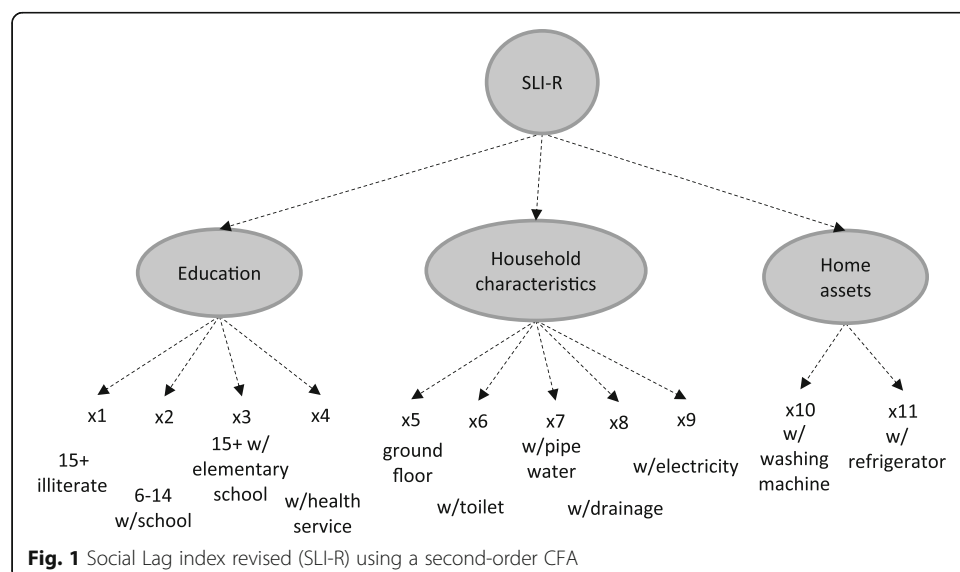


Table 3 Factor loadings corresponding to the second-order CFA

		2000	2005	2010	2015
F1 (education)					
x1	15+ Illiterate	0.857	0.891	0.926	0.94
x2	6–14 Without school	0.541	0.513	0.495	0.373
x3	15+ Without elementary school	0.870	0.845	0.836	0.824
x4	without health service	0.718	0.629	0.327	-0.098
	Omega	0.84	0.819	0.761	0.645
F2 (household characteristics)					
x5	Ground floor	0.881	0.91	0.735	0.849
x6	Without toilet	0.506	0.252	0.325	0.273
x7	Without pipe water	0.63	0.621	0.56	0.519
x8	Without drainage	0.841	0.787	0.797	0.764
x9	Without electricity	0.654	0.576	0.605	0.617
	Omega	0.835	0.781	0.75	0.755
F3 (household assets)					
x10	Without washing machine	0.942	0.953	0.961	0.952
x11	Without refrigerator	0.986	0.966	0.951	0.943
	Omega	0.964	0.959	0.955	0.946
G (second-order)					
F1	Education	0.932	0.892	0.875	0.866
F2	Household characteristics	0.992	0.958	0.923	0.923
F3	Household assets	0.909	0.928	0.945	0.946
	CFI	0.874	0.885	0.887	0.913
	RMSEA	0.168	0.149	0.136	0.115
	SRMR	0.054	0.061	0.059	0.055

Source: Own elaboration based on Census Bureau data 2000-2015

Note: factor loadings, the bolded numbers indicate the most relevant variables to preserve for the first-order CFA solution

It is interesting to see that the factorial loads decrease for some variables. For example, the scores for variable x_4 decreases over time (0.718, 0.629, 0.327, -0.098), which can be interpreted as an improvement in the indicator during these years, possibly related to the implementation of some social programs such as popular insurance.⁹ However, this variable does not fulfill the property of factorial invariance, and the reliability decreases over time.

The next step consists of selecting the variables from Table 3 that have factor loadings greater than 0.70 and calculate new loadings for the first-order CFA; the selected variables are those that have a significant load contribution to the SLI-R. The results are shown in Table 4, where all variables fulfill the property of factorial invariance. That is, the factor loads are similar between 2000 and 2015 and greater than 0.70 (see Table 4). In this table, the Omega reliability coefficients for all years are higher than 0.90, suggesting that the first-order CFA with a reduced number of SLI-R indicators, from 11 to 6, shows better-fit statistics. We observe that the factorial loads for this solution are very similar for all years, and fulfill the condition of factorial invariance, and have satisfactory adjustment fit statistics (CFI = 0.921–0.925, SRMR = 0.044, Omega 0.92–0.94). Besides, the parsimony principle suggests that six variables are preferred for the construction of the SLI-R that explains the same information as the original index with 11 variables.

Table 4 Factor loadings corresponding to the first-order CFA

Variable	Description	2000	2005	2010	2015
x1	15+ illiterate	0.787	0.795	0.800	0.807
x3	15+ w/ elementary school	0.747	0.707	0.708	0.705
x5	Ground floor	0.837	0.846	0.701	0.783
x8	W/ drainage	0.804	0.739	0.714	0.700
x10	W/ washing machine	0.941	0.945	0.954	0.941
x11	W/ refrigerator	0.970	0.957	0.948	0.936
	CFI	0.925	0.923	0.921	0.921
	RMSEA	0.224	0.218	0.209	0.209
	SRMR	0.044	0.044	0.044	0.044
	Cronbach-alpha	0.927	0.918	0.896	0.889
	Omega	0.940	0.933	0.919	0.923

Source: Own elaboration based on Census Bureau data 2000-2015

The alternative index is obtained by calculating the scores. It is compared with the one constructed by CONEVAL, obtaining for each year a Spearman correlation higher than 0.93 (the correlation matrix is not shown). Both indices measure the same concept, while the alternative index is more parsimonious and can be used for longitudinal purposes.

Mixture latent growth curve model

We use the alternative index of social lag, SLI-R, constructed from the reduction of the 11 original variables, to construct social lag trajectories. Table 5 shows that the mean of SLI-R is decreasing with standard deviations almost constant. However, the coefficient of variation indicates a systematic increase during the period of study and suggests a heterogeneous increase among the municipalities.

In order to analyze the SLI-R, we fit a longitudinal latent growth curve model, as described in the previous section. The overall fit is given by $SLI_{ij} = 68.81 - 7.33 * time$, which means that social lag decreases at a rate of 7.33 points every 5 years. This implies that after ten more years, the SLI-R score will be 24.63 (see Fig. 2). However, we observe in this figure a great deal of variability. One model fitted to all trajectories cannot tell us much about the evolution of the municipalities. As the SLI-R score of some municipalities could decrease at a slower pace than that of others, this model does not allow us to propose regional public policies. Therefore, we need to classify the trajectories into classes.

Table 5 Descriptive statistics of social lag factorial scores. Mexico 2000–2015

Variable	Mean	Std. Dev.	C.V.	Min	Max
fs_2000	70.13	26.96	0.38	7.30	119.01
fs_2005	60.01	27.42	0.46	6.80	125.75
fs_2010	53.07	25.15	0.47	7.92	112.91
fs_2015	48.06	25.31	0.53	5.80	120.56

Source: Own elaboration Census Bureau data 2000-2015

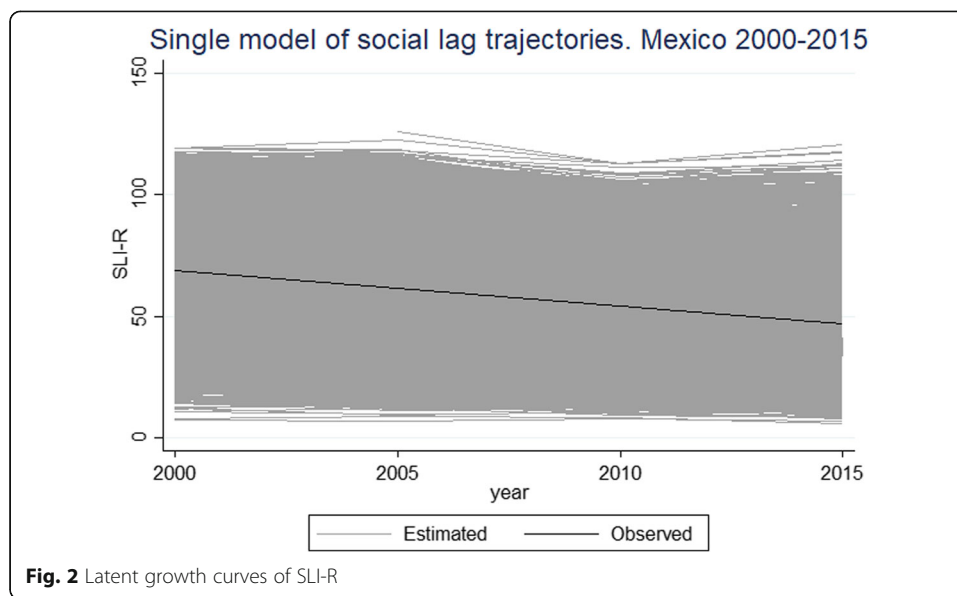


Fig. 2 Latent growth curves of SLI-R

The mixture latent growth curve models can help us to identify trajectory groups of SLI. First of all, we need to determine the number of trajectory groups, using the five criteria previously described. The results are shown in Table 6. Column (1) shows the value of the BIC, which decreases as more groups are identified. The lower the BIC, the better the fit. Column (2) indicates the entropy, for which the value closest to the unit indicates an acceptable adjustment, and in this case, any solution between three and four groups is satisfactory. Column (3) shows the minimum percentage of belonging to a group. In this case, the acceptable solutions, which have a percentage higher than 5%, are three groups and four groups. Column (4) indicates the probability of membership of the municipalities to each group, which must be greater than 0.7. All solutions satisfy this condition. Finally, in column (5), the Lo-Mendel-Rubin test indicates that solutions of 2 to 5 groups are acceptable. In conclusion, using the criteria (1) to (5), four trajectory classes are selected.¹⁰

With the adjustment of the mixture latent growth curves with the scores of the alternative index (SLI-R), we generate four curves, one for each class and another model that shows the general trend. In Table 7, the “complete model” column shows the slope (time) that indicate the social lag index decreases at a rate of 7.33 points every 5 years, and the intercept at 68.81, which is the score of year 2000.

An interesting result is the negative relationship between the intercepts and the slopes $-.39$, $-.87$, and $-.64$ for the years 2000, 2005, and 2010, respectively. This suggest there in a heterogeneous growth.¹¹ For example, for the full model, the covariance of the intercept

Table 6 Determination of the number of trajectory classes

Classes	BIC (1)	Entropy (2)	Pct Min–Max (3)	Prob. clasif (4)	LMR-adjusted (5)
2	66,903.09	0.908	16.7–83.2%	0.94–0.98	$p < 0.0001$
3	66,415.15	0.811	17.1–43.3%	0.89–0.92	$p < 0.0001$
4	66,115.20	0.8	12.6–35.9%	0.83–0.92	$p < 0.0001$
5	66,004.55	0.805	5.4–36.7%	0.83–0.91	$p = 0.0001$

Note: Using the criteria (1) to (5) explained in previous sections we select four classes

Source: Own elaboration Census Bureau data 2000-2015

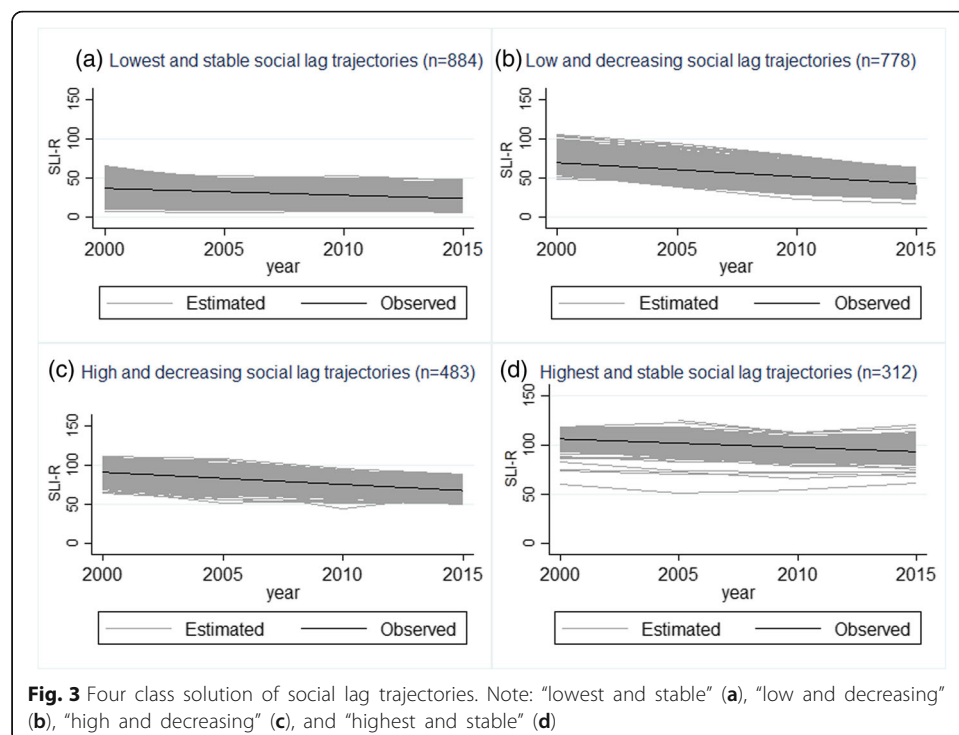
Table 7 Latent growth curves models of SLI-R. Mexico 2000–2015

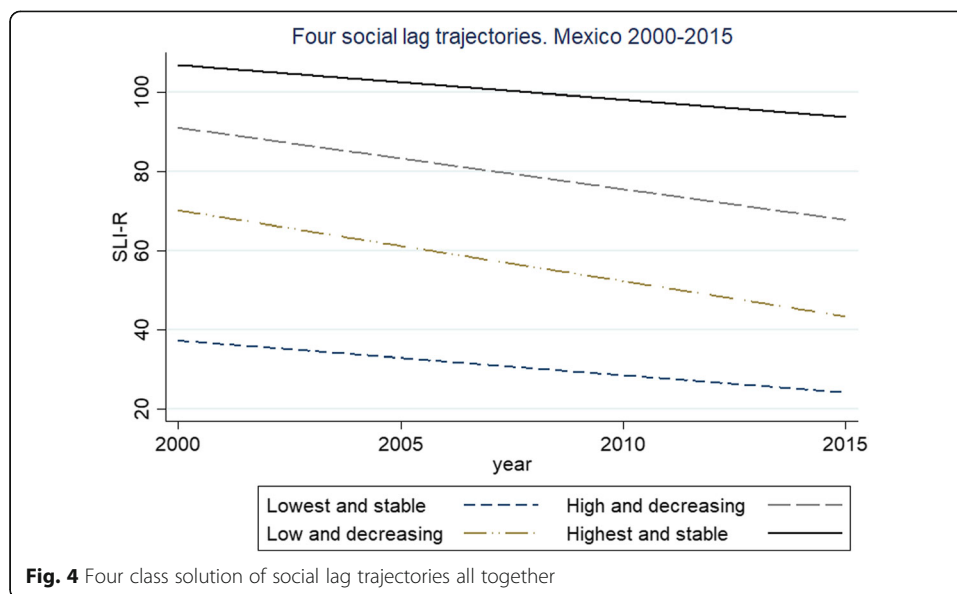
Fixed effects	Full model	Lowest and stable (<i>n</i> = 884)	Low and decreasing (<i>n</i> = 778)	High and decreasing (<i>n</i> = 483)	Highest and stable (<i>n</i> = 312)
Intercept	68.81 (0.06)	38.43 (0.39)	73.17 (0.39)	92.56 (0.44)	107.24 (0.41)
time	− 7.33 (0.55)	− 5.17 (0.07)	− 10.26 (0.08)	− 8.28 (0.08)	− 4.65 (0.09)
Random effects					
S.D. (time)	2.71	1.73	1.27	1.09	0.69
S.D. (intercept)	27.17	11.4	10.04	9.26	6.8
Cov (time, intercept)	− 0.39	− 0.87	− 0.64	− 0.53	0.49
S.D. (residual)	3.39	3.01	4.00	3.05	3.21

Note: Estimates and standard errors in parenthesis
Source: Own elaboration Census Bureau data 2000-2015

and slope is -0.39 , that is, the lower the intercept, the higher the slopes. In other words, if a municipality had a high score in 2000, its slope will most likely will be negative, suggesting the expectation of some improvement in social lag in the following years. However, when we explore the covariance for all classes, the coefficient increases from -0.87 to -0.53 from 2000 to 2010 and becomes positive by the year 2015.¹² It is possible to confirm what is observed in Fig. 3 that represents the social lag trajectories, where a downward trend is clearly seen on average (panels a to c), although the “highest and stable” class shows the trajectories in the opposite direction to that observed at the national level (Fig. 4).

The following models in this table identify four classes of municipality trajectories “lowest and stable” ($n = 884$, or 35.9%), “low and decreasing” ($n = 778$ or 31.7%), “high





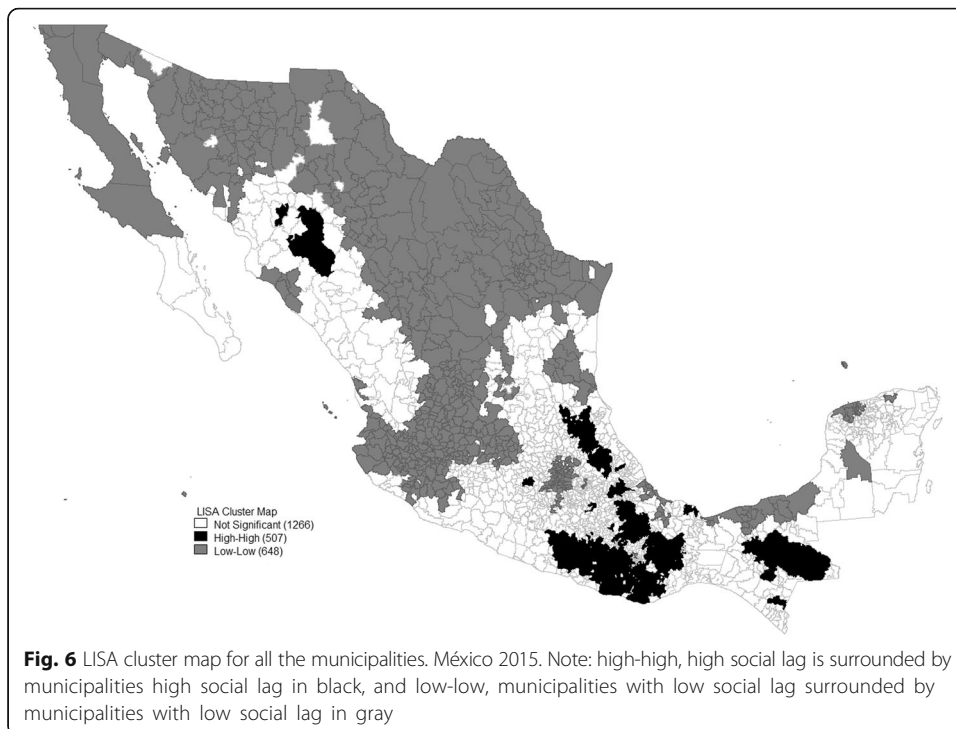
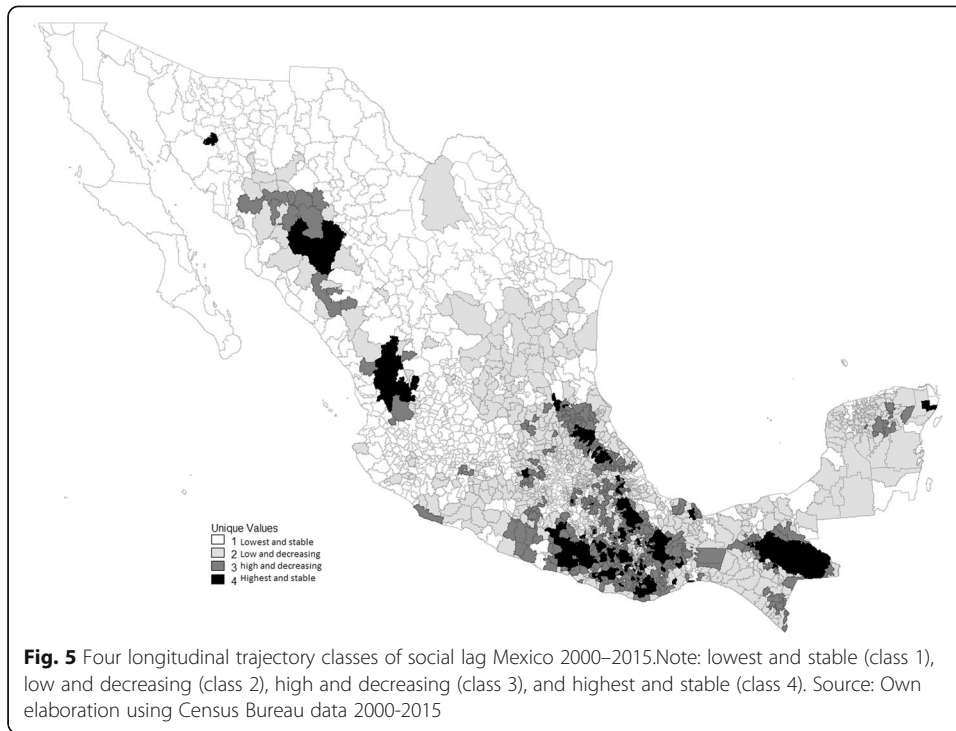
and decreasing" ($n = 483$ or 19.7%), and "highest and stable" ($n = 312$ or 12.7%). The observed and estimated curves are shown in Fig. 3. Additionally, we can observe in Fig. 4 all four class models in a single plot.

Moran index and LISA maps

Now, we proceed with the analysis by using geo-referenced information to fit spatial models. First, when we use the four classes of social lag trajectories observed in Fig. 4, the territorial division between the north, center, and southeast regions of the country becomes evident. This division is visualized in the map of Fig. 5. The municipalities with "low and stable" social lag, as well as those with "low and decreasing" trajectories are located throughout the northern part and the central region of the country, where large metropolitan and industrial centers are located. On the other hand, the municipalities of "high and decreasing" social lag, as well as those of "highest and stable" trajectories, are located mainly in the southeast region (in the states of Oaxaca, Guerrero, Chiapas, Puebla, Veracruz, and Chihuahua). Most of them are in mountainous regions where various indigenous populations live and where access is inadequate and services unsatisfactory or nonexistent.

This map is a useful tool for locating municipalities with high or highest social lag but have had minimal or no improvement in their living conditions over the past 15 years. However, an additional analytic tool is needed for identifying with more detail the areas where living conditions are stagnated for the purpose of public policy application. The LISA map will help us to identify the conglomerates of municipalities with high and low social lag more specifically.

The local Moran's $I = 0.7867$, indicating a high positive correlation between municipalities with the same degree of social lag. This means we can define clusters of high and low social lag. Figure 6 shows the LISA map for 2015, where again we see a story similar to Fig. 5—there is a clear division between the center, north, and southeast regions of the country.

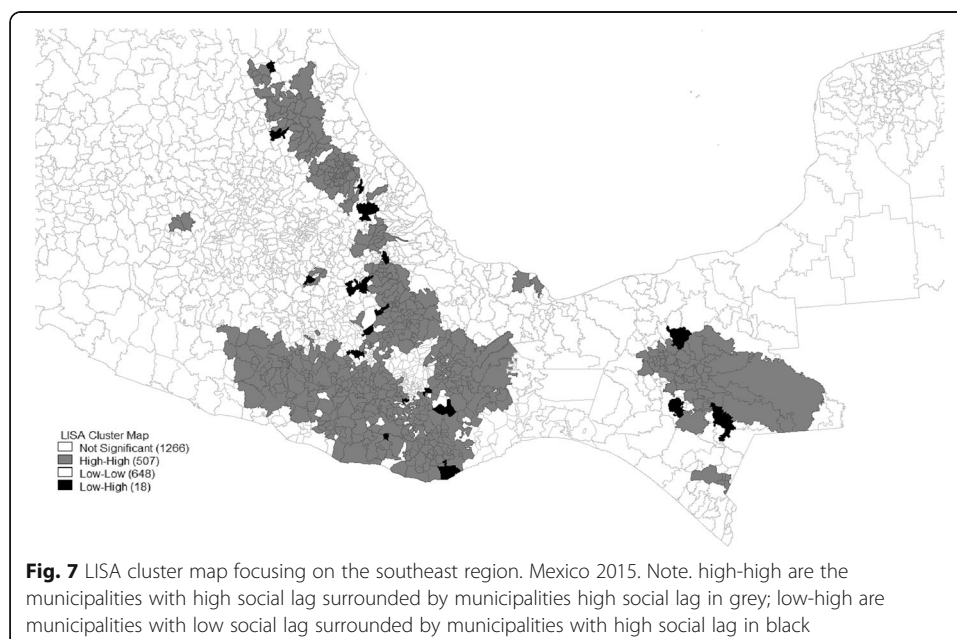


There are two categories relevant for this study: the low-low and the high-high. The low-low shows regional homogeneity: the municipalities of low social lag are surrounded by municipalities of low social lag. One of these groups is located in the northern region, near the border with the USA; other groups include the Bajío region, the highlands of Jalisco, and the region of the Valley of Mexico; these three regions are mainly characterized by their industrial development, foreign investment, and rapid growth of the cities,¹³ as well as the region of the oil exploration zone of Campeche.

On the other hand, the high-high conglomerates tell us another story. These groups are located in the Lacandon jungle, highlands in Chiapas; in Oaxaca, we identify the northern, southern highlands, the Mixteca, and “La Cañada;” in Guerrero, we identified regions of “Tierra Caliente,” mountains and the coast; in Puebla, we found the northern highlands, and “La Huasteca” region in San Luis Potosí and Veracruz (Fig. 7). Finally, in the northern region, we found the Tarahumara highlands in Chihuahua.

A total of 507 municipalities belong to these regions. Most of them lack basic services such as medical services and schools. These municipalities are located mainly in the country’s mountains, which makes it difficult to get to the municipal centers for food, medical services and the like. The population in these regions are of indigenous origin with diverse ethnicities. In some municipalities, residents do not speak Spanish, but their indigenous language. The population is mainly peasants with a self-consumption economy, in which they raise livestock and fish, make handcrafts, and sell their products in local markets.

Within the regions of high social lag, there are some atypical and interesting cases. Municipalities of low social lag surrounded by municipalities with high social lag (the low-high cases) are centers of greater development that provide medical services, education, and work. Examples of these type of municipalities are San Pedro Juchatengo and Santa María Huatulco in Oaxaca (southeast region of Mexico), both of which are touristic centers. Other medium-sized cities such as Comitán de Domínguez, Teziutlán, and Perote provide services to the smaller communities nearby. Others are



manufacturing centers, such as Tehuacán in Puebla (a production center of egg and poultry), as well as Teotitlán and Tlacolula in the central valleys of Oaxaca. It is important to emphasize that these municipalities, among many others, are service and employment providers of the regions of high social lag. For this reason, it is necessary to focus public policies for generating development to improve the living conditions of the population of the communities.

Conclusions

The elaboration of indexes has been a recurrent theme when it comes to measuring the effectiveness of public policies for social development in México. The SLI is considered as another indicator, similar to the marginalization index for measuring poverty. This article starts with the SLI proposed by CONEVAL and has constructed a revised version using confirmatory factor analysis. This more parsimonious alternative reduces the number of variables from 11 to 6. Also, it has the following properties: (i) factorial invariance in the years of study, (ii) high-reliability Omega and acceptable adjustment indices, (iii) correlations greater than 0.93 with the original SLI, and (iv) that it can be used for longitudinal measurement purposes.

The advantage of using the proposed alternative SLI-R is that it enables tracking of the evolution of social lag in municipalities over time. We found four typologies of growth—“lowest and stable,” “lowest and decreasing,” “high and decreasing,” and “highest and stable.” When these municipalities are located in a map, we defined a regional division of growth. Mexico is divided into three regions: north, center, and southeast. Cities located in the northern and central part of the country have trajectories with less SLI, while those with high SLI are located mainly in the southeastern part. In particular, we focus our attention on the “highest and stable” social lag category, which represents 12.7% of the municipalities located mainly in the mountains, where small communities live.

Adding the information of the mixture latent growth curve model to the LISA map gives us a more precise picture of social lag. These results are useful for generating social policies addressed to regions instead of municipalities. The proposed regions are based in the so-called “absorbent” states. That means that those municipalities with low social lag are surrounded by municipalities with high social lag. With this information, we located regions with high potential of growth, creating a similar contagious process. LISA maps are a good complement to the longitudinal analysis and provide better tools for social policy designed to improve the quality of life of citizens in Mexico.

Endnotes

¹Data obtained from CONEVAL (2010)

²A vector c has a unitary norm when $c^t c = 1$, that means, the variance of the components must be equal to 1 since the original variables are standardized.

³It is known that zones with high-social lag are located in the mountains (where the indigenous population lives) and in states where typically there is a greater incidence of poverty and marginalization. Likewise, it is observed that low-lagged zones are grouped into urban areas, as well as industrial regions with better living conditions.

⁴These criteria establish: (1) that the meaningful number of components to interpret must accumulate 60% or more percentage of variance (2) or the number of components with Eigen values higher or equal to 1.

⁵Previous to this phase, we performed an Exploratory Factor Analysis to observe the natural formation of the factors without constraints; however, the solution was not helpful since the factors were not satisfactorily identified. Instead, we proceed with the confirmatory phase to obtain a better solution.

⁶The Root Mean Error Approximation (RMSEA) is useful for calculating the degree to which the proposed model fits the population reasonably well (Steiger and Lind 1980; Browne and Cudeck 1993). Values less than 0.05 are desirable, but values between 0.06 and 0.08 are acceptable. As values close to 0.09 or greater are undesirable (Ridgdon, 1996). The Standardized Root Square Residual (SRMR) values range between 0 and 1.0, values less than 0.08 indicates a good fit (Hu and Bentler 1999). The CFI belongs to a category of incremental adjustment measures that compare the proposed model against the null model, called the Comparative Fit Index (CFI; Bentler 1990). This index must exceed the recommended level of 0.90 in providing additional evidence to accept the proposed model.

⁷If the factor loading is 0.7, the residual variance is $1 - 0.70^2 = 0.51$, the higher the factor loading the lower the residual variance. The SLI has a better fit when the residual variance decreases to values less than 0.5

⁸The factorial invariance property states that if the probability of observing a given score does not depend on the group to which it belongs, then the factor scores are invariant and measure the same variable over time (Millsap 2007: 463; Meredith 1993: 528; Millsap and Meredith 2007).

⁹The so-called “Popular insurance” is a program that gives financial support to public health care institutions that give access to the population in need. This program started in 2002 so that the coverage increased in the following years.

¹⁰Possibly the solution of three groups is eligible; however, for public policy purposes, we chose 4.

¹¹Some municipalities having a high intercept and a small slope, this means by 2000 they have a high score in SLI and decrease in the following years; some others having a small intercept and a high slope, this means that by 2000 they have a small SLI but increase over time.

¹²A positive covariance suggests a fan pattern of growth that is the higher the intercept, the higher the slope and the other way too, the lower the intercept, the lower the growth.

¹³In the regions, we observe a rapid growth of new urban cities surrounded by population with poor living conditions. Some scholars name this phenomenon as the new urban cities, for example Queretaro and the Bajío region.

Abbreviations

AGEB: Geo-statistical geographic areas; CFA: Confirmatory factor analysis; CFI: Comparative Fit Index; CONEVAL: National Council for the Evaluation of Social Development Policy; LISA: Local Indicators of Spatial Association; PCA: Principal component analysis; RMSEA: Root Mean Square Error Approximation; SLI: Social lag index; SLI-C: Social lag index-CONEVAL; SLI-R: Social lag index-Revised; SRMR: Standardized Root Mean Square Residual

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Acknowledgements

We acknowledge Dr. Lukasz Czarnecki for inviting us to submit this contribution.

Authors' contributions

DV, conducted part of the analysis and shaping the idea of the paper. SV equally contributed for another part of the analysis and complement the manuscript. Both authors read and approved the final manuscript.

Funding

The authors declare no founding support other than our job position at the National Autonomous University of México.

Availability of data and materials

We used Census Bureau data 2000-2015 accessed at https://www.coneval.org.mx/Medicion/IRS/Paginas/Indice_Rezago_Social_2015.aspx.

Competing interests

The authors declare that they have no competing interests.

Received: 12 December 2018 Accepted: 7 June 2019

Published online: 12 July 2019

References

- Anselin, L. 1988. *Spatial Econometrics: Methods and models*. The Netherlands: Kluwer Academic Publishers.
- Anselin, L. 1995. Local indicators of spatial association-LISA. *Geographical Analysis* 27 (2): 93–115.
- Bentler, P.M. 1990. Comparative fit indexes in structural models. *Psychological Bulletin* 107: 238–246.
- Browne, M.W., and R. Cudeck. 1993. Alternative ways of assessing model fit. In *Testing structural equation models*, ed. Kenneth A. Bollen and J. Scott Long, 136–162. Newbury Park: CA: Sage.
- Calderón Villareal, C., and Peláez Herreros. 2016. Condiciones de vida en áreas de alto rezago social y factores sociodemográficos de la pobreza multidimensional en Baja California. *Revista de Ciencias Sociales y Humanidades* 27: 78–104.
- Collins, L.M., and S.T. Lanza. 2010. *Latent class and latent transition analysis: with applications in the social, behavioral, and health sciences*. New York: John Wiley & Sons.
- CONEVAL. 2010. Metodología para la medición multidimensional de la pobreza en México. Ciudad de México, México
- CONEVAL. 2011. Metodología para la medición multidimensional de la pobreza en México. *Realidad, datos y espacios* 2 (1): 36–63.
- Cortés, F., and D. Vargas. 2017. *Origen es destino: un análisis de la marginación municipal. México 1990-2015*. Ciudad de México: Siglo XXI-UNAM.
- Geary, R. 1954. The contiguity ratio and statistical mapping. *The Incorporated Statistician* 5: 67–80.
- Getis, A.Y., and J. Ord. 1992. The analysis of spatial association by use of distance statistics. *Geographical Analysis* 24: 189–206.
- Hu, L., and P.M. Bentler. 1999. Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. *Structural Equation Modeling* 6: 1–55.
- Jones, B.L., D.S. Nagin, and K. Roeder. 2001. A SAS procedure based on mixture models for estimating developmental trajectories. *Sociological Methods & Research* 29 (3): 374–393.
- Kaiser, H.F. 1960. The application of electronic computers to factor analysis. *Educational and Psychological Measurement* 20: 141–151.
- Lo, Y., N. Mendell, and D.B. Rubin. 2001. Testing the number of components in a normal mixture. *Biometrika* 88: 767–778.
- Massey, D. 1991. The political place of locality studies. *Environment and Planning A* 23: 267–281.
- McDonald, R.P. 1999. *Test theory: a unified treatment*. Mahwah: Lawrence Erlbaum Associates, Inc.
- Meredith, W. 1993. Measurement invariance, factor analysis models of factorial invariance: a multifaceted approach. *Psychometrika* 58: 525–543.
- Millsap, R.E. 2007. Invariance in measurement and prediction revisited. *Psychometrika* 72 (4): 461–473.
- Ridgdon, Edward E. 1996. CFI versus RMSEA: a comparison of two fit indexes for structural equation modeling. *Structural Equation Modeling* 3 (4): 369–379.
- Rubalcava, R., and J. Ordaz. 2016. Desigualdad intrarregional: las propensiones como indicio de desigualdad social crónica. In *La heterogeneidad de las políticas sociales en México: instituciones, derechos sociales y territorios Volumen II*, 195–221. Ciudad de México, México: Universidad Iberoamericana.
- Schwartz, G. 1978. Estimating the dimension of a model. *The Annals of Statistics* 6: 461–464.
- Sclove, L.S. 1987. Application of model selection criteria to some problems in multivariate analysis. *Psychometrika* 52: 333–343.
- Steiger, J.H., and J.C. Lind. 1980. Statistically based test for the number of common factors. In *Paper read at Annual Meeting of the Psychometric Society, at Iowa City, IA*.
- Urry, J. 1987. Survey 12: Society, space and locality. *Environment and Planning D: Society and Space* 5: 435–444.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.