An Asset Index Proposal for Households in Mexico Applying the Mixed Principal Components Analysis Methodology

Lorena DelaTorre-Díaz(✉) and Román Rodriguez-Aguilar

Facultad de Ciencias Económicas y Empresariales, Universidad Panamericana, Augusto Rodin 498, 03920 Mexico City, Mexico
lotorre@up.edu.mx

Abstract. The development of assets indices has grown as an alternative to measure wealth from different generations in the evaluation of social mobility. A proposal of the development of an asset index is presented using the GSVD-based mixed principal components analysis (PCAMix package in R). The contribution rests in the combination of both numerical and categorical data and the integration of the simultaneous effect of these variables in the index. It was used in profiling the Mexican households according to the information from the 2018 National Household Income and Expenditure and the determination of the Gini coefficient to evaluate the inequality of distribution at the state level. Results show a high level of disparity in the distribution of assets with only 0.01% of the households possessing 40% or more of the assets included in the index, being the southern region where greatest challenges for ascending social mobility.

Keywords: Asset index · Mixed principal components · Social mobility · Mexico

1 Introduction

Social mobility refers to the changes experienced by individuals in their socioeconomic condition, reflected in a variation in their relative position according to an educational, employment or income indicator [1, 2]. Its analysis makes it possible to determine whether aspects such as effort and talent determine the achievement of objectives and the change in their living conditions, regardless of the individual's physical and personal characteristics or the socioeconomic position of their parents [3].

A society that favors social mobility allows individuals to improve their living status on their own merits and are not predetermined by their conditions of origin. Social mobility is therefore one of the aspects of the study of inequality of opportunities in a society.

The World Economic Forum [4] recently developed the Global Social Mobility Index (GSMI) that offers a tool to identify areas for improvement in this indicator, evaluating ten pillars: health, education access, education quality and equity, lifelong

J. A. Marmolejo-Saucedo et al. (Eds.): COMPSE 2021, LNICST 393, pp. 88–106, 2021.
https://doi.org/10.1007/978-3-030-87495-7_7

learning, social protection, technology access, work opportunities, fair wages, working conditions, efficient and inclusive institutions. In the first edition, 82 countries were compared with 51 indicators. In the results report, it is estimated that an increase of 10 points in the index could translate into an additional growth in GDP of 4.41% by 2030. Hence the importance for countries to identify and invest in the right mix of factors determinants of social mobility. Mexico was ranked 58th out of the 82 countries evaluated, with the Nordic economies showing the best levels of social mobility and therefore greater equality of opportunities for their population.

Countries with greater inequality experience less mobility between generations, as a greater fraction of the economic advantages and disadvantages are passed from parents to children. This relationship is represented by the so-called Great Gatsby Curve, which is constructed using income inequality measured by the Gini Coefficient on the horizontal axis, and a measure of intergenerational economic mobility on the vertical axis [5].

This curve highlights that inequality of opportunities is the missing link between income inequality and social mobility: if greater inequality makes intergenerational mobility more difficult, it is due to greater inequality in the distribution of economic growth opportunities for the new generations.

The measurement of social mobility from an economic perspective has used as the main variable the level of income from wages and salaries or the level of family income that includes other elements such as transfers or financial assets [6]. However, there are studies that propose an alternative estimate based on the wealth of families under the assumption that the accumulation of assets constitutes a better approximation to household wealth [7].

In the economic research literature, different studies can be found that use asset indices as the indicator of economic mobility, in substitution of income level: [6–9] are some examples. These indices can be a valid predictor of the manifestation of poverty, as well as become an approximation of long-term wealth with a lower degree of error than the measurement of expenditures [10].

Particularly in developing countries, the use of asset indices has increased in studies related to poverty and inequality since these indicators present fewer measurement problems or resistance from interviewees to provide the information. Additionally, when comparing asset possessions, it is possible to establish differences in living conditions between households in the same country, between countries or even in periods over time [10, 11]. Most countries periodically prepare national representation surveys that provide this type of data, which favors the creation of these indices and their use in intergenerational analysis.

In Mexico, studies that have developed asset indices have used data from the available specialized social mobility surveys (Encuesta de Movilidad Social [EMOVI by its acronym in Spanish]) from 2006, 2011 and 2017. However, the National Institute of Statistics and Geography (Instituto Nacional de Estadística y Geografía [INEGI by its acronym in Spanish]) provides robust information on the profile and living conditions of households in Mexico obtained through the National Household Income and Expenditure Surveys (Encuesta Nacional de Ingreso y Gasto de los Hogares [ENIGH by its acronym in Spanish]) and this information has not been used for the development of assets indices before. This survey includes not only dichotomous variables, but categorical multi-level variables, as well as qualitative variables derived from a broader objective of offering an overview of the behavior of household income and expenditure, the sociodemographic characteristics of its members and the household infrastructure characteristics.

Moreover, the calculated indices have used methods focused exclusively on numerical or categorical variables, but no asset index has been developed by combining both.

Derived from the above, the present study aims to develop an asset index representing the households in Mexico with the information from ENIGH 2018. The selection of assets that would be part of the index takes as a reference the pillars conforming the Global Social Mobility Index [4], and the mixed principal components method is applied. Additionally, the asset index will be used to rank households in Mexico and compare it with the Gini Coefficient.

The structure of the study is as follows: in Sect. 2 the framework of social mobility is addressed, including a background on the use of asset indices. Section 3 describes the methodology used in the development of the index and the variables selected for the analysis. Section 4 contains the results of the analysis as well as the conclusions obtained and further analysis possibilities.

2 Social Mobility Framework

Social mobility refers to changes in the socioeconomic condition of individuals, and can be defined in educational, employment or income terms [2]. Sociologists study social mobility as changes in class and job configuration, while economists evaluate mobility in terms of an income vector or some other measure of well-being [12].

Social mobility can also be analyzed in its intragenerational dimension, that is, the mobility that the same individual experiences during his or her life, or intergenerational that refers to the mobility of individuals with respect to their parents [13]. Furthermore, mobility can be absolute when it refers to the rise or fall in an absolute income scale, or relative when the change is measured in relation to the position occupied in the reference period [2].

The absence of upward social mobility in a society gives rise to situations such as the loss of potential talents that remain hidden, decrease in productivity levels, loss of investment opportunities, the hoarding of educational, economic or financial opportunities on the part of the higher socioeconomic classes, waste in the allocation of human resources, and finally a breakdown of the social cohesion when citizens perceive barriers that prevent them from accessing better conditions [14, 15].

2.1 Assets Indices for Measuring Social Mobility

The variable that is traditionally used to estimate social mobility in economic terms is income. This variable can include not only income from wages and salaries but also other factors such as financial assets and public and private transfers [6].

Nonetheless, this methodology is limited by the need to have information on the income of different generations, which is not always available in all countries or regions. Therefore, a growing trend in the literature is the development of asset indices that allow estimating household wealth and, based on them, assessing its mobility. These indices can be considered approximations of permanent household income [6, 7, 9].

Although a combination of human, physical, social and financial assets is required to improve the socioeconomic situation of people, the accumulation of physical assets can generate greater wealth, and become a possible indicator of the capacity of that condition improvement [16].

The models of Filmer-Pritchett [17] and Sahn-Stifel [10] are the most cited. The former developed an index of household assets and characteristics based on the principal component analysis methodology (PCA) to assess the impact of wealth on the educational level of households in Brazil, India and Kenya, and later to assess the relationship between wealth and school enrollment in India [18]. The value of the first principal component is the latent variable that represents the possession of household assets.

On the other hand, in [10] and [19] a factor analysis was used to estimate a single common factor that explains the variances in the possession of a set of assets, and this factor is considered as the metric of economic status or well-being.

Other studies followed the PCA methodology for the elaboration of asset indices that are used to measure social mobility in Mexico. One measures educational, occupational, and economic intergenerational social mobility [7]; another uses an asset index to measure intergenerational social mobility in Mexico between 1950 and 1980 [6].

The information from the 2011 Social Mobility Survey (EMOVI) was used in an study where PCA was used to develop the asset index based on three types of assets: consumer durable goods, household features and financial assets including the possession of a bank account, credit card, vacuum cleaner, toaster, domestic service, telephone, savings account, water heater, washing machine, refrigerator, automobile, inside toilet, stove, electricity service, tubing water, own house and the household crowding index [7].

Similarly, the 2006, 2011 and 2017 EMOVI information was used [6] and created an asset index including the ownership of personal computer, cellular phone, landline phone, internet access, cable TV, shop or business, land or farm, second residence, animals, agricultural equipment, stove, washing machine, refrigerator, inside toilet, electricity, domestic service, savings account, checking account, credit cards and cars. In a previous study, [20] it was also included the parents' and respondents' occupation status.

There are examples of studies using PCA in other regions, such as Bangladesh where a wealth index was used in the evaluation of intergenerational mobility [21], in Colombia [22], or in Pakistan [23].

A variation from the previous studies is where it was decided to use the multiple correspondence analysis (MCA) given that most of the variables included in the index were categoric (mostly binaries) [8]. They included additional variables such as vacation home, apartment for rent and investment in shares.

As noted above, most studies have relied on methodologies focused only on numerical variables. However, there are scarce studies reporting the use of a mixed component analysis method, being an example of the use of a multiple-factor-analysis method that handles a combination of quantitative and qualitative variables a study used in the assessment of poverty alleviation programs in China [24].

It is worth to mention that no studies have been identified using mixed principal components analysis in the evaluation of social mobility.

3 Method

3.1 PCAMix Method

The method used for the construction of the asset index in this study is the mixed principal components analysis applying the generalized singular value decomposition methodology (GSVD), given the different nature of the variables selected. The PCA Mix (also called PCA with metrics) is a generalization of standard PCA using the GSVD to decompose the matrix Z obtained after processing the original information in order to have a particular case of PCA for the numerical variables and MCA for the categorical ones [25, 26].

The Z matrix is the real matrix $Z = [Z_1, Z_2]$ of dimension $n \times p$ where Z_1 is the standard version of the $n\,x\,p$ quantitative matrix and Z_2 is the centered version of the $n \times m$ indicator matrix G of the $n \times p$ qualitative matrix (n being the number of observations and p the number of variables).

The standard PCA aims to reduce the number of dimensions under analysis, maintaining the maximum representation of the original information, and even creating a latent measure that takes the form of an index. PCA analysis is useful when the differences or distances between continuous variables can be captured, however its interpretation is less clear when categorical variables are included.

The alternative method of multiple correspondence analysis is constructed using categorical variables and its purpose is also to reduce dimensionality, using the relative frequencies of each category as a substitute for distances [8].

However, when it is desired to generate a latent variable from a combination of quantitative and categorical variables, a mixed method is used [27], maintaining the same objective of reduction of dimensions. The information provided by surveys such as the national income and expenses surveys include different types of variables, and when many of those variables are considered relevant in the evaluation of the household socioeconomic condition, a methodology that uses mixed date is preferred.

In the construction of an asset index using a mixed methodology, the impact of the variable is not only limited to the possession or not of the asset -as it is in the traditional PCA and MCA methods -, but the amount of money the household spends on it.

PCA Mix splits the original dataset into a numerical matrix and a categorical matrix, and then uses PCA on the quantitative variables and MCA on the categorical variables to obtain a linear combination of the observed variables that accounts for the largest inertia (o variance) [24]. Equation 1 shows the decomposition of the matrix Z.

$$\mathrm{Z} = \mathrm{U}\Lambda\mathrm{V}^{\mathrm{T}} \tag{1}$$

Where:

Z is the real matrix of dimension $n \times p$.

N and M are the diagonal matrixes of the weights of the n rows and p columns.

Λ is the diagonal $r \times r$ matrix $(\sqrt{\lambda 1}, \ldots, \sqrt{\lambda r})$ of the singular values of ZMZ^TN and Z^TNZM, where r denotes the rank of Z.
U is the $n \times r$ matrix of the first r eigenvectors of ZMZ^TN where $U^TNU = I_r$ with I_r the identity matrix of size r
V^T is the $p \times r$ matrix of the first r eigenvectors of Z^TNZM such that $V^TMV = I_r$

After the GSVD method, PCA Mix produces a matrix of dimension $(p_1 + m) \times r$ of the factor coordinates of the p_1 quantitative variables and the m levels of the p_2 categorical variables [28].

The decision of the number of dimensions to maintain is based on the proportion of the total inertia for each dimension.

3.2 Measure of Concentration

The Gini coefficient is a common measure of inequality, evaluating the degree of distribution of income or wealth among individuals or households from a perfectly distributed economy. Its values range from cero (a perfectly equitable distribution) to 1, where a single individual or household concentrates wealth [29] (Eq. 2).

$$G = 1 - \frac{2}{N-1}\left(N - \frac{\sum_{i=1}^{N} i x_i}{\sum_{i=1}^{N} x_i}\right) \tag{2}$$

Where N is the population size and x_i is the variable under evaluation of the *ith* individual or household. Although it is commonly used to evaluate the inequality of the distribution of the income, this coefficient can be used to measure the degree of concentration of any variable. In the present study, the Gini coefficient is calculated using the asset index as the proxy for household wealth.

The Gini coefficient from the *reldist* package in R is used. This model was developed by Handcock as one especial case of the models described in Handcock & Morris [30].

3.3 Segmentation

The clustering method of K-mean is used to segment the Mexican states in groups presenting maximum intra-group homogeneity and inter-group heterogeneity [24]. This clustering technique is a form of unsupervised classification, where there is no external criterion used for the grouping of the cases. On the contrary, the groups are formed after evaluating the intrinsic similarities and dissimilarities among the different cases [31].

The k-means clustering from R was used.

3.4 Data and Variables

The information was taken from the ENIGH carried out in 2018. The size of this national coverage sample was 87,826 households representing 125 million inhabitants from Mexico and the dataset is distributed in 11 tables containing normalized data and an additional table offering a household-level summary. The units of analysis are dwelling, household and the members of the household.

The selection of the variables was based on seven of the ten pillars used by the World Economic Forum's Global Social Mobility Index (GSMI) including a group of qualitative variables representing the ownership of different types of assets and services for each household, and a group of quantitative variables that represent the average monthly expenses (or income) destined by (obtained from) each household to certain income/expense activities.

Additionally, other financial-inclusion variables were included, such as possession of credit card, life insurance policy, mortgage, or similar housing loan, the financial and capital monthly perception and the monthly deposits on savings accounts.

To complete the household profile according to their assets ownership, the following indicators were considered: the possession of vehicles, radio, toaster, microwave, refrigerator, stove, washing machine, sewing machine, vacuum cleaner, domestic service, water availability, toilet, electricity, home ownership, water heater (gas and solar), and the monthly spending on household goods, vehicles acquisitions and home property (Table 1).

3.5 Index Estimation

The information related to income and expenses obtained from ENIGH was deflated in preparation for the analysis. Some new categorical variables were created to present the ownership of assets such as school loan, medical expenses insurance or life insurance. Some variables were converted from multi-level to binary to account only for the possession of the asset regardless of the number of items. A total of 49 variables were selected and gathered in a master dataset using the household folio number as the key in merging the different tables.

The package PCAmixdata [28] from R was applied to perform the mixed principal components analysis, and the dimensions and eigenvalues were obtained. The proportion of the total inertia explained by each dimension is used to determine the number of dimensions to keep.

The factor coordinates values of the dimensions selected are then used in the creation of the asset index, weighted by its own percentage of inertia explained (Eq. 3).

$$Y_i = \frac{x_{1i}w_1 + x_{2i}w_2 + \ldots + x_{ki}w_k}{w_1 + w_2 + \ldots + w_k} \tag{3}$$

Where Y_i is the value of the index for household i, $x_{\mathrm{k}i}$ is the value of the k factor coordinate selected, and w_i is the proportion of the total inertia explained by that coordinates. Finally, the value of Y_i is adjusted to a range between [0,100] to facilitate its interpretation. The asset index density distribution is estimated.

The index is used to compare Mexican households according to two attributes: their geographic location based on the state and its rural or urban condition. The mean and median of the value of the asset index per state is calculated, and this value is contrasted with the results of the Gini coefficient calculated on the same index per state.

The clustering k-mean method is used to identify the regions where Mexican households present low, medium, and higher levels of assets accumulation. The flowchart summarizing the process followed in the creation of the index is shown in Fig. 1.

Table 1. Variables selected for the analysis

Global Social Mobility Index Pillar	Variables selected from national survey of household income and expenditure	Type of variable
Health	Monthly spending in health Possession of medical expense insurance	Numerical Categorical (1 yes, 0 no)
Education access	Educational level of the head of household School loan	Categorical (from 1 to 11) Categorical (1 yes, 0 no)
Education quality and equity	Possession of scholarship Monthly scholarship received Monthly spending in education	Categorical (1 yes, 0 no) Numerical Numerical
Social protection	Retirement fund Access to "Seguro Popular" Affiliation for health care	Categorical (1 yes, 0 no) Categorical (1 yes, 0 no) Categorical (1 yes, 0 no)
Technology access	Telephone Cellular phone Pay TV Computer Printer Internet access Analog or Digital TV DVD or VCR Videogames Monthly spending in communications	Categorical (1 yes, 0 no) Categorical (1 yes, 0 no) Categorical (1 yes, 0 no) Categorical (1 yes, 0 no) Categorical (1 yes, 0 no) Categorical (1 yes, 0 no) Categorical (1 yes, 0 no) Categorical (1 yes, 0 no) Categorical (1 yes, 0 no) Numerical
Work opportunities	Own business Number of household members working and receive a salary	Categorical (1 yes, 0 no) Numerical
Working conditions	Written contract	Categorical (1 yes, 0 no)

Source: authors

4 Results

4.1 Asset Index Estimates

Based on the proportion of total inertia explained by each dimension, a total of 39 dimensions were selected accumulating 70.62% of proportion (Fig. 2). It is worth mentioning that discriminating some dimensions imply the loss of 29.38% of the variance of the information contained in the total dataset. However, this limitation is lessened by the fact that each of the dimensions maintained contains information on the total 49 variables selected.

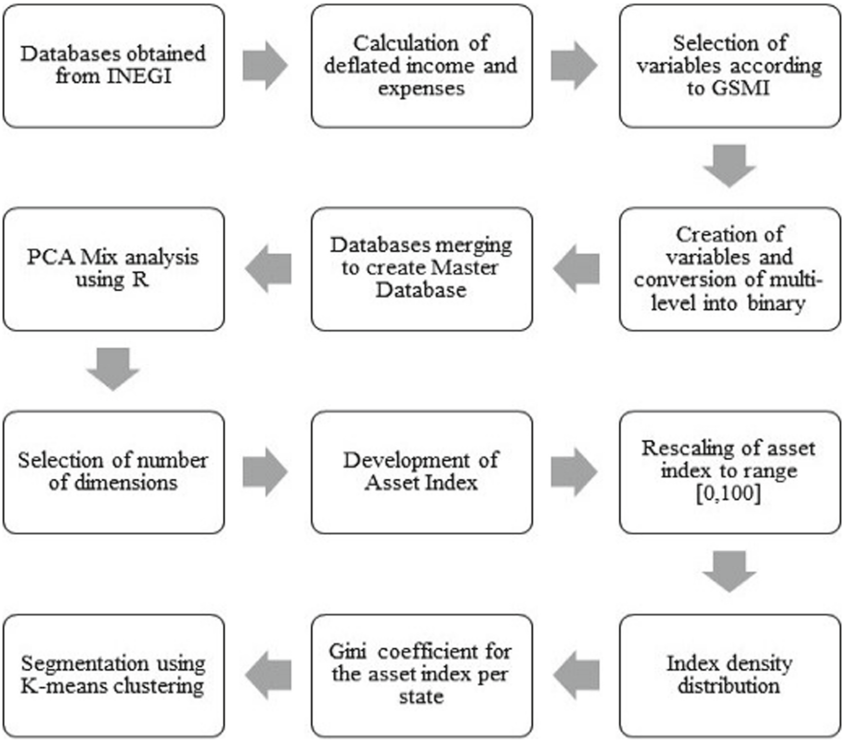

Fig. 1. Flowchart of the creation of the asset index

The coordinates corresponding to these first 39 dimensions were weighted by its own proportion of inertia and rescaled to create an asset index limited to values from 0 to 100.

The density distribution of the asset index calculated for the total households comprehended in the ENIGH survey shows a high right skewness, indicating high degree of inequality in the possession of assets among households (Fig. 3). Only 0.01% of the households possess 40% or more of the assets included in the index.

The main benefit obtained from the construction of the asset index comes from the additional disaggregated analysis that can be conducted. In the case of the Mexican households, the first analysis confirms that when separating rural and urban households, the distribution maintains the right skewness (Fig. 4).

Additionally, the index density was determined for each state. Even though the distribution shown in all states present the same skewed shape, the concentration of assets is higher in some states compared to others (Fig. 5).

The mean of the asset index for the entire sample of households contained in the survey is 6.93, which makes the skewness in the distribution even more evident, considering that the value ranges from 0 to 100. This number indicates that the average household in Mexico owns 6.93% of the total assets included in the index.

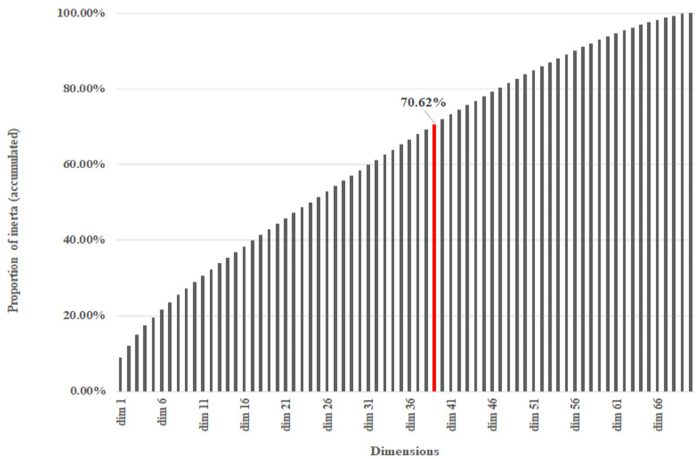

Fig. 2. Accumulated proportion of total inertia explained by each dimension

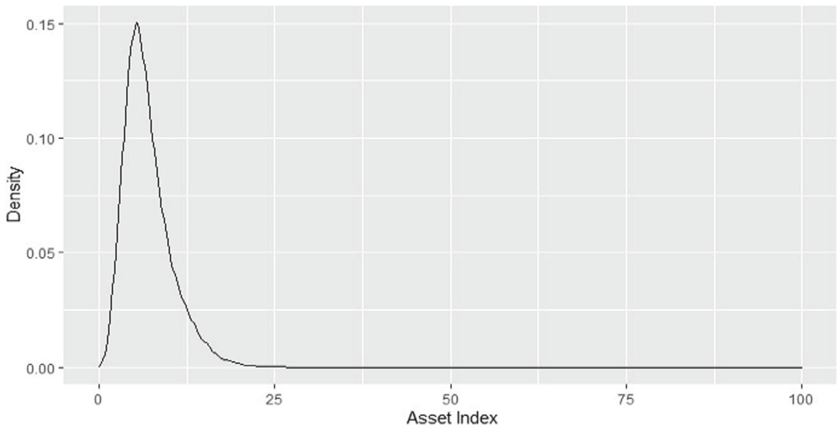

Fig. 3. Density distribution on the Asset Index of Mexican household

Segmenting households according to their rural or urban condition, it can be observed that urban households own a greater level of assets, but the disparity in the distribution is high in both conditions. In urban households, it is possible to find families possessing 100% of the assets included in the index, but at the same time families with indices as low as 0.08%. The median among urban households is 7.07, indicating that the asset possession level is low for most families (Table 2).

Contribution of the Assets to the Index (Loadings)

The contribution of each variable to each component (dimension) is called the loading. It can be observed that the first dimension is mostly influenced by assets related to technology (computer and internet access) and by the level of education of the household head. The second dimension presents a higher contribution from variables related to house

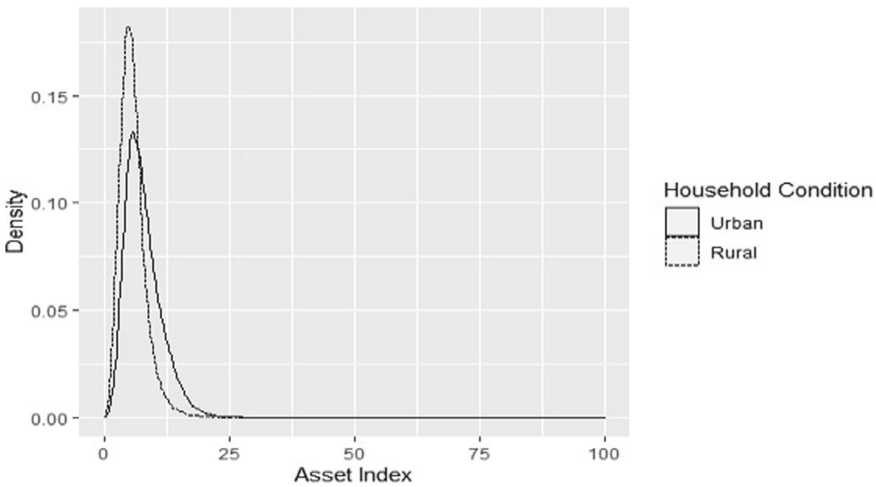

Fig. 4. Asset Index density distribution for urban condition of Mexican households

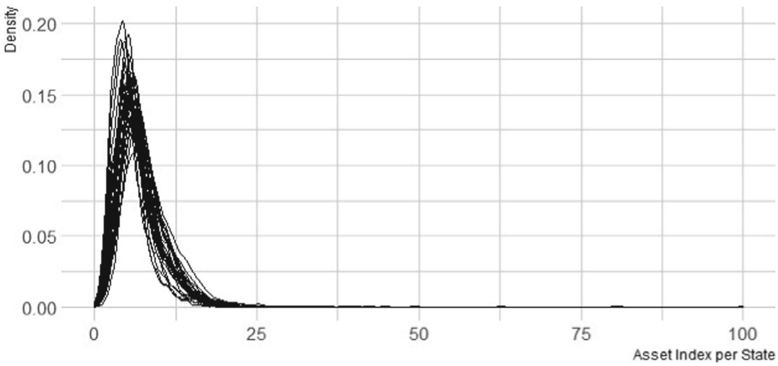

Fig. 5. Density distribution on the Asset index per Mexican states

Table 2. Descriptive statistics on the Asset Index according to Rural and Urban condition

	Mean	95% CI		Median	SD	Min	Max
		LL	UL				
Rural	5.63	5.6	5.66	5.26	2.68	0	80.58
Urban	7.73	7.7	7.77	7.07	3.68	0.08	100

Source: authors

ownership, and the sixth dimension shows a higher contribution form the possession of television (Table 3).

Here are presented only the most relevant variables contributing in the first ten dimensions, however, the asset index developed takes into consideration the contribution of all variables in the 39 dimensions kept. This is a variation compared to other indices created using standard PCA, where only the first component is taken. The methodology presented in this study creates the index as a weighted average of the contribution of the first 39 dimensions, and each dimension is at the same time computing different weights of contributions of the variables.

Table 3. Contribution of selected variables to the components (first ten dimensions)

Variable	dim 1	dim 2	dim 4	dim 5	dim 6	dim 7	dim 8	dim 10
Num. Members receive salary			0.418					
Monthly education expenses							**0.328**	
Scholarship							0.429	
Analog TV					0.664			
Digital TV					0.461			
Microwave	0.348							
Computer	0.430							
Monthly communic. expenses						**0.445**		
Internet access	0.467							
Education of household head	0.398							
Water	0.305							0.400
Electricity								
House ownership		0.700						
House loan		0.701		0.367				
Water heater (gas)	0.301							

Source: authors

An important aspect to highlight is that the variables that present the most important contributions to the components or dimensions include both categorical and numerical variables, reflecting the benefit of the method used. Had the standard PCA or MCA methods been used, the impact of the monthly expenses in education and communications would not have been considered because of its numerical nature -not categorical as most of the other variables are-.

The variables that are common among the households whose asset index is superior to 40 are the possession of cellular phone, water availability, electricity, computer, and particularly higher levels of education of the household head (graduate studies).

Index Concentration Degree by State

The asset index and the Gini Coefficient were contrasted for all 32 states. The asset index evaluates the ownership of assets among households, and therefore families will rank higher on the index if they own more assets. The Gini coefficient, on the other hand, measures the inequality in the distribution of assets among households. States with lower levels of Gini coefficient are those in which the distribution of assets is more equitable.

It can be observed the behavior of the Gini coefficient compared to the median of the asset index in every state, and it is mostly inverse. The decision of using the median in this comparison intends to reduce the impact of the extreme values of few households with higher asset indices (Fig. 6).

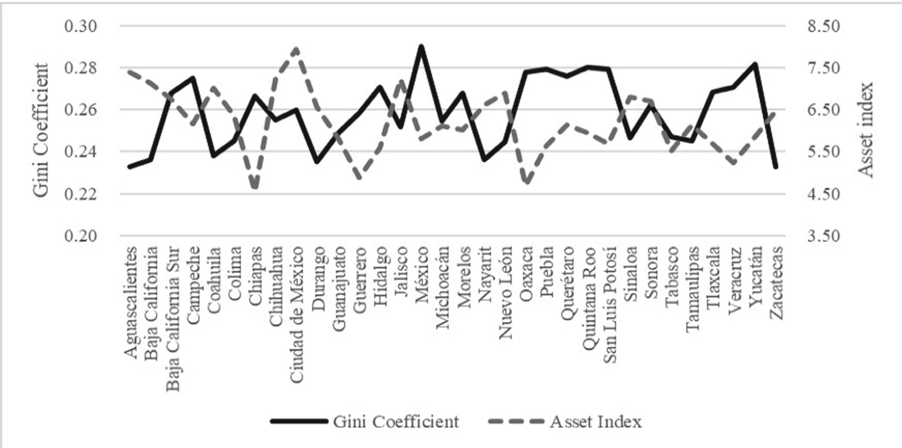

Fig. 6. Median of asset index compared to Gini coefficient per state

Ciudad de México is the state with the highest median of asset index, while Chiapas has the lowest. México State is the region where the distribution of assets is more unequal, while Zacatecas is the state with the most equitable distribution. However, this lower level of Gini coefficient is not always related to a higher ownership of assets; it could reflect regions inhabited by families with similar low or medium level of asset index.

Special attention must be taken to those states with the lowest level of asset possession that at the same time show high levels of concentration, because this inequality in the distribution of the assets may represent an obstacle for social mobility.

The comparison presented in Appendix 1 ranks the 32 states in ascending order according to their Gini coefficient and in descending order according to their asset index value using median and mean. Ciudad de México, the city capital of the country, ranks number one on the asset index but it ranks 17 according to the inequality of distribution. It is worth highlighting Aguascalientes that presents the second highest value of the median in the asset index, and it is also ranked second according to the values of the Gini coefficient, which indicates a state where the distribution of assets is more equitable, and households are able to own a higher number of these assets.

The Gini coefficient for rural households is 0.24 and for urban households is 0.25. Urban households own more assets than rural households, but at the same time they are more concentrated, that is, a less equitable distribution.

Disaggregation of the Index per State

The partition of groups of Mexican states using the k-mean clustering method was made based on the average asset index value of the households located in every state. Three clustering calculations were performed with 10, 5 and 3 clusters, and there was a drop in the intra-group sum of squares after every reduction in the number of clusters. The proportion of intra-group sum of squares of the total sum of squares dropped from 98.6% in a K(10) to 84.3% in K(3). Therefore, the analysis of the partitions of Mexican states is based on 3 groups (Fig. 7).

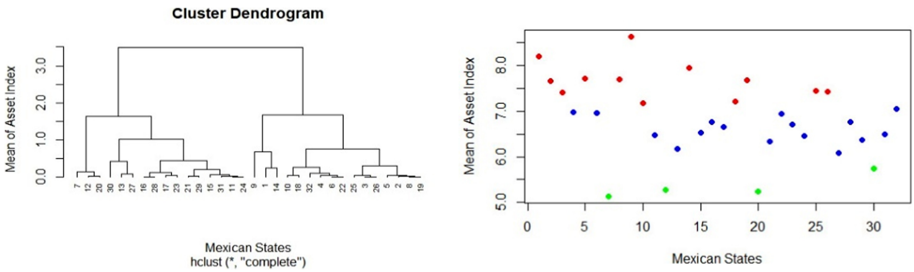

Fig. 7. Clustering of Mexican states according to the asset-index level

Four states belong to the group with the lowest level of asset accumulation, 16 states are in the medium-range group, and 12 states belong to the highest-level group. Cluster number one corresponds to the states with a low mean of asset index, and it groups Veracruz, Oaxaca, Guerrero and Chiapas. The states that form the Cluster number 2 are Baja California, Baja California Sur, Sonora, Chihuahua, Coahuila, Nuevo León, Sinaloa, Durango, Nayarit, Jalisco, Aguascalientes and Ciudad de México, and this group shows a high level of average asset index. Finally, Tamaulipas, San Luis Potosí, Zacatecas, Guanajuato, Querétaro, Colima. Michoacán, Estado de México, Hidalgo, Puebla, Morelos, Tlaxcala Tabasco, Campeche, Quintana Roo and Yucatán are the states grouped in the Cluster number 3, with medium level of asset index (Fig. 8).

When representing these clusters in the map, it is clear the difference in the three regions: the northern and western states as well as Ciudad de México belong to the group with the highest asset accumulation levels. The states located at the center and at the Yucatán Peninsula form the middle-range group, and the southern states of Oaxaca, Guerrero, Chiapas and Veracruz are the region where households own less assets.

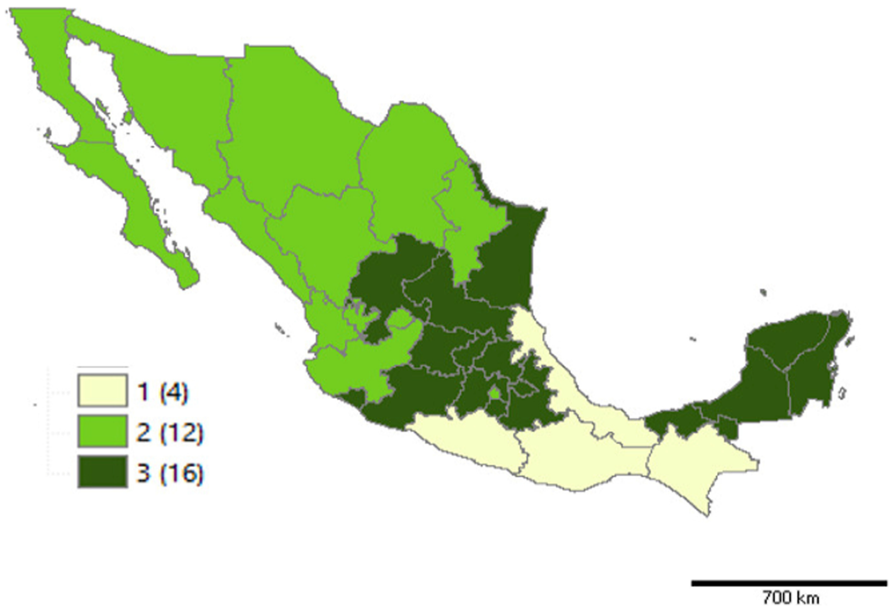

Fig. 8. Clustering of Mexican states according to Asset Index

5 Conclusions

A number of recent studies are using asset indices as alternative wealth measures when the information of other metrics such as income is not available, especially if the objective of the analysis is the evolution of the socio-economic status of households in different countries and different generations.

The index built in this study was estimated using the mixed principal components analysis, and presents some important differences compared to other indices aiming to evaluate socio-economic conditions:

1. Mixed principal components analysis allows the use of different types of variables: binary, categorical (multi-level) and numerical. This variety broadens the range of information that can be included in the index coming from surveys such as the national surveys of income and expenses, where not all variables are categorical.
2. An asset index that uses only categorical variables, particularly binary, would not include the simultaneous impact of the possession of a given asset (such as education) and the magnitude of the investment made by the household on that asset (measured by the amount of monthly expenses). The methodology used allows the inclusion of this simultaneous effect.

   Although these amounts are registered as expenses in the survey where the data is obtained from, they are considered in this study as assets since families benefit as they increase the monetary resources used in these activities.

3. Even though one of the main purposes of the method is the reduction of dimensions, each one of the dimensions selected for the integration of the index contain information of all the variables, lessening the negative effect of the loss of dimensions.
4. The weights assigned to every dimension and to every variable are not arbitrary. The dimensions are weighted by the proportion of the total inertia each dimension explains when they are averaged to compute the index. In addition, every dimension contains information of all the variables, showing a different proportion of contribution in every case. These proportions are derived from the correlations between the variable and the dimension.

The index was used in the evaluation of the how asset accumulation is distributed throughout the country and the degree of inequality in its distribution. The outcome show a clear segmentation of the Mexican states: northern states, western states and Ciudad de Mexico show the highest mean of asset index; center states and the Yucatan Peninsula region rank in the medium level; and southern states are those where the asset accumulation is the lowest.

The Gini Coefficient was useful to prove that regions where asset possession is lower, tend to present greater levels of inequality in its distribution. Southern states, thus, are the states where challenges for ascending social mobility are higher due to less availability of assets and greater levels of inequality.

These results are in line with other studies using wealth indices to measure inequality in the different Mexican regions [32], which shows the adequacy of the Mixed principal components analysis-based index in profiling the socioeconomic condition. Identifying the regions in which families have accumulated less assets allows the development of focalized policies intended to improve the access to different types of assets that are relevant for social mobility.

The methodology used in the present study is not restricted to the Mexican region. It can be useful in the development of asset indices in other countries where information of asset ownership, income and expenses at the household level is available.

Further analysis may use the asset index to measure intergenerational social mobility, overcoming the lack of sufficient income information from different generations. Other applications of this index may use it in the identification of the assets with highest impacts on social mobility and the probabilities of households to grow in their living conditions by possessing those assets.

Appendix 1. Comparison of Gini Coefficient and Mean and Median of the Asset Index of Mexican States

State	*Gini Coef*	*State*	*Median of Index*	*State*	*Mean of Index*
Zacatecas	0.23	Ciudad de México	7.93	Ciudad de México	8.63
Aguascalientes	0.23	Aguascalientes	7.39	Aguascalientes	8.19

(*continued*)

(*continued*)

State	*Gini Coef*	*State*	*Median of Index*	*State*	*Mean of Index*
Durango	0.24	Chihuahua	7.22	Jalisco	7.94
Nayarit	0.24	Jalisco	7.21	Coahuila	7.71
Baja California	0.24	Baja California	7.13	Chihuahua	7.69
Coahuila	0.24	Coahuila	7.02	Nuevo León	7.67
Nuevo León	0.24	Nuevo León	6.89	Baja California	7.66
Tamaulipas	0.24	Sinaloa	6.80	Sinaloa	7.45
Colima	0.24	Baja California Sur	6.76	Sonora	7.42
Sinaloa	0.25	Sonora	6.70	Baja California Sur	7.41
Tabasco	0.25	Nayarit	6.62	Nayarit	7.22
Guanajuato	0.25	Durango	6.56	Durango	7.18
Jalisco	0.25	Zacatecas	6.47	Zacatecas	7.05
Michoacán	0.25	Colima	6.35	Campeche	6.99
Chihuahua	0.25	Campeche	6.17	Colima	6.96
Guerrero	0.26	Querétaro	6.13	Querétaro	6.95
Ciudad de México	0.26	Tamaulipas	6.13	Michoacán	6.77
Sonora	0.26	Michoacán	6.10	Tamaulipas	6.76
Chiapas	0.27	Morelos	6.01	Quintana Roo	6.70
Baja California Sur	0.27	Quintana Roo	5.95	Morelos	6.65
Morelos	0.27	Yucatán	5.85	México	6.53
Tlaxcala	0.27	México	5.81	Yucatán	6.49
Hidalgo	0.27	Guanajuato	5.80	Guanajuato	6.48
Veracruz	0.27	San Luis Potosí	5.69	San Luis Potosí	6.46
Campeche	0.27	Tlaxcala	5.69	Tlaxcala	6.36
Querétaro	0.28	Puebla	5.64	Puebla	6.33
Oaxaca	0.28	Hidalgo	5.59	Hidalgo	6.18
San Luis Potosí	0.28	Tabasco	5.52	Tabasco	6.09
Puebla	0.28	Veracruz	5.23	Veracruz	5.74
Quintana Roo	0.28	Guerrero	4.87	Guerrero	5.28
Yucatán	0.28	Oaxaca	4.68	Oaxaca	5.25
México	0.29	Chiapas	4.55	Chiapas	5.14
Average	0.26	Average	6.20	Average	6.85

Source: authors with information from ENIGH 2018

Note. Gini index ordered in ascending order; mean and median ordered in descending order

References

1. Delajara, M., De la Torre, R., Díaz-Infante, E.: El México del 2018. Movilidad social para el bienestar. Centro de Estudios Espinosa Yglesias (2018)
2. Vélez-Grajales, R., Monroy-Gómez, L.A.: Movilidad social en México: hallazgos y pendientes. Revista de Economía Mexicana (2), 97–142 (2017)
3. Vélez-Grajales, R., Campos-Vázquez, R.M., Huerta-Wong, J.E.: Informe de movilidad social en México 2013: imagina tu futuro. CEEY Centro de Estudios Espinosa Yglesias, Ciudad de México (2013)
4. World Economic Forum: The global social mobility report 2020, Switzerland (2020)
5. Corak, M.: Income inequality, equality of opportunity, and intergenerational mobility. J. Econ. Perspect. **27**(3), 79–102 (2013)
6. Torche, F.: Changes in intergenerational mobility in Mexico: a cohort analysis. Centro de Estudios Espinosa Yglesias Documento de trabajo 03 (2020).
7. Behrman, J., Vélez-Grajales, V.: Patrones de movilidad intergeneracional para escolaridad, ocupación y riqueza en el hogar: el caso de México. In: Vélez-Grajales, H.W. (ed.) México, ¿el motor inmóvil?, pp. 299–346. CEEY Centro de Estudios Espinosa Yglesias, A.C., Ciudad de México (2015)
8. Vélez-Grajales, R., Vélez-Grajales, V., Stabridis, O.: Construcción de un índice de riqueza intergeneracional a partir de la encuesta ESRU de movilidad social en México (EMOVI). Centro de Estudios Espinosa Yglesias, Documento de trabajo 02/2015 (2015)
9. Vélez-Grajales, R., Stabridis Arana, O., Minor Campa, E.: Still looking for the land of opportunity: regional differences in social mobility in Mexico. Sobre México, Temas de Economía **1**, 54–69 (2018)
10. Sahn, D.E., Stifel, D.C.: Poverty comparisons over time and across countries in Africa. World Dev. **28**(12), 2123–2155 (2000)
11. Moser, C., Felton, A.: The construction of an asset index measuring asset accumulation in Ecuador. CPRC Vol. Working Paper 87 Washington DC: The Brookings Institution (2007)
12. Ferreira, F.H., Messina, J., Rigolini, J., López-Calva, L.-F., Lugo, M., Vakis, R.: La movilidad económica y el crecimiento de la clase media en América Latina. Banco Mundial, Washington, D.C (2013)
13. Inter-American Development Bank. La realidad social. Módulo 1 - Pobreza, Desigualdad y Movilidad Social (2020)
14. Vélez-Grajales, R., Campos-Vázquez, R., Fonseca, C.E.: El concepto de movilidad social: dimensiones, medidas y estudios en México. Centro de Estudios Espinosa Yglesias, Documento de trabajo (2015)
15. OCDE: A Broken Social Elevator? How to Promote Social Mobility. OECD Publishing, Paris (2018)
16. Cotler, P., Rodríguez-Oreggia, E.: Microfinanzas y la tenencia de activos no financieros en México. Investigación económica **69**(274), 63–86 (2010)
17. Filmer, D., Pritchett, L.: The effect of household wealth on educational attainment around the world: Demographic and health survey evidence. World Bank (1998)
18. Filmer, D., Pritchett, L.H.: Estimating wealth effects without expenditure data—or tears: an application to educational enrollments in states of India. Demography **38**, 115–132 (2001)
19. Sahn, D.E., Stifel, D.: Exploring alternative measures of welfare in the absence of expenditure data. Rev. Income Wealth **49**(4), 463–489 (2003)
20. Torche, F., Spilerman, S.: Influencias intergeneracionales de la riqueza en México. In: Serrano Espinosa, J., Torche, F. (eds.) Movilidad Social en México, población, desarrollo y crecimiento, pp. 229–274. Centro de Estudios Espinosa Yglesias, A.C., México (2010)

21. Asadullah, N.: Intergenerational economic mobility in rural Bangladesh. In: Royal Economic Society (RES) Annual Conference, Nottingham (2006)
22. Fajardo-Gonzalez, J.: Inequality of opportunity in adult health in Colombia. J. Econ. Inequal. **14**(4), 395–416 (2016). https://doi.org/10.1007/s10888-016-9338-2
23. Muhammad, M., Jamil, M.: Intergenerational mobility in educational attainments. Pak. Dev. Rev. **59**(2), 179–198 (2020)
24. Zeng, Z., Zhu, M.: Poverty groups identification and assessment of poverty alleviation programs in rural China. In: 6th International Conference on Humanities and Social Science Research, pp. 174–189. Atlantis Press (2020)
25. Rodriguez-Aguilar, R.: Main metric components in the generation of mixed indicators: an application of SGVD methodology. In: Vasant, P., Litvinchev, I., Marmolejo-Saucedo, J.A., Rodriguez-Aguilar, R., Martinez-Rios, F. (eds.) Data Analysis and Optimization for Engineering and Computing Problems. EICC, pp. 195–206. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-48149-0_14
26. Beaton, D., Fatt, C.R.C., Abdi, H.: An ExPosition of multivariate analysis with the singular value decomposition in R. Comput. Stat. Data Anal. **72**, 176–189 (2014)
27. Kalantan, Z.I., Alqahtani, N.A.: A study of principal components analysis for mixed data. Int. J. Adv. Appl. Sci. **6**(12), 99–104 (2019)
28. Chavent, M., Kuentz-Simonet, V., Labenne, A., Saracco, J.: Multivariate analysis of mixed data: the R Package PCAmixdata. arXiv:1411.4911 (2017)
29. Flores, J.A.F., Gutiérrez, H.A., Zea, J.F.: Estimación por muestreo del índice de Gini para las localidades de Bogotá usando funciones en R. Comunicaciones en Estadística **8**(1), 59–79 (2015)
30. Handcock, M.S., Morris, M.: Relative Distribution Methods in the Social Sciences. Springer, Heidelberg (1999)
31. Morissette, L., Chartier, S.: The k-means clustering technique: general considerations and implementation in Mathematica. Tutor. Quant. Methods Psychol. **9**(1), 15–24 (2013)
32. Plassot, T., Rubio, G., Soloaga, I.: Movilidad social intergeneracional y desigualdad de oportunidades en México. Educación y activos: un enfoque territorial (2019)