An attempt to measure social stratification and changes in terms of social distances *

Sandra Fachelli[1], Pedro López-Roldán[2]
[1]Research Group on Education and Work (GRET)
Department of Sociology, Universitat Autònoma de Barcelona, Barcelona, Spain.
[2]Centre for Sociological Studies on Daily Life and Work (QUIT)
Department of Sociology, Universitat Autònoma de Barcelona, Barcelona, Spain
Email: Sandra.Fachelli@uab.cat, Pedro.Lopez.Roldan@uab.cat

September, 2013

In the context of a research whose purpose was to analyse the changes in the social and economic model of Argentina in the period 1997-2006, we observed that the social structure, measured by a variable that typifies stratification, remains basically unchanged during the years analysed in spite of the profound crisis that occurred at the end of 2001 and along 2002. Within this general framework, the hypothesis that we believe is consistent with the observed behaviour is that the distance among social strata can be measured, in an objective way, taking into account the relative positions of the households with respect to primary goods, both intra temporarily (dispersion within the stratus) as well as inter temporarily (changes of the distance among strata). From the Data of the Permanent Survey of Households of Argentina and using Multiple Correspondence Analysis and Cluster Analysis, we have measured these distances in each moment of time. The results show that the distances among the strata express the changes in time, and that they increase during crisis. These distances, measured in terms of inertia in a multidimensional space and starting from the typologies obtained in each year, allow us to have a measure of the dynamics of the social change.

Keywords: social stratification; typology construction; social dynamics; social distance; multivariate analysis.

* A first version of this paper was presented to the XVII ISA World Congress of Sociology, 11-17 July, 2010, Gothenburg, Sweden. Research Committee 55, Social Indicators.

1. Introduction and Framework of Analysis

In Argentina the early 21st century was marked by the political crisis that broke out in December 2001 with the fall of President De la Rúa and, in the economic sphere, by the default on public debt in late December and the devaluation of the Argentine peso in January 2002. We will not give a detailed description of the events herein, but this context is essential for understanding the social, political and economic background to the phenomenon we discuss.

The aim of this paper is to propose a measure of distance between the social strata of Argentina based on a typology of social stratification. The hypothesis that guides the study is that the distance between social strata can be observed objectively by the relative positions of households in intra-temporal and inter-temporal terms. The study was carried out on the years 1997, 2002, 2003 and 2006.

We took as a basis for analysis the developments presented by Fachelli (2009), who configured a model of social stratification in Argentina using normative criteria. This model offers a multidimensional perspective that transcends the traditional contributions of stratification. In addition to including the classical dimensions such as employment and income, conditions of housing and education were also selected as essential elements for stratifying society. The study used the household as the unit of analysis.

The four years that we selected to carry out the analysis are representative of different situations. The year 1997 was one of growth and of economic and social stability; 2002 showed the effects of the crisis and a great deal of social conflict; 2003 saw the beginning of the recovery from the economic crisis and social conflict; and 2006 saw some consolidation of the economic and social recovery.

The conclusion of this analysis (Fachelli, 2009) was that in the period covered the Argentine social structure was composed of four strata. Furthermore, the distance between social strata can be observed objectively by their intra-temporal positions (with respect to their stratum in the same year) and their inter-temporal positions (over time in comparison with the average household).

We considered how to measure the social distance between strata from a multidimensional perspective so that the intra-temporal distances show the dispersion among households, i.e. whether they are more cohesive or more polarized within each stratum. The inter-temporal distance gives us a measure of inequality between strata in their evolution over time, so it can be expected that in periods of crisis there will be a greater distance between social strata than in periods of stability.

2. Model of Analysis and Methodology

The analysis model used to define the social strata is based on the adoption of John Rawls' principle of "fair equal opportunities" (Rawls, 2002; 1979) to select primary goods: those that any rational person wants in order to develop. This

selection is not exhaustive because it is adapted to the availability of data for carrying out the analysis. The variables that account for the primary goods possessed by households are: employment, education, housing and income (Fachelli, 2009). The hypothesis to be tested is that the distance between social strata may be observed objectively through the intra-temporal and inter-temporal positions of households with respect to primary goods.

The scheme of the model that we used consists in studying social stratification through four dimensions: 1) opportunities for access to the labour market, which are approximated empirically through the indicators of employment, unemployment and inactivity; 2) opportunities for access to education, which are explored through the average number of years of schooling of the household, calculated by adding the years of education of each household member aged 18 and over and dividing the result by the total number of household members aged 18 and over); 3) opportunities for access to housing, which is based on three indicators, overcrowding (more than two people per room excluding kitchen and bathroom), possession and use of a bathroom, and housing tenure; and 4) opportunities for access to household income, which is analyzed on the basis of the decile of family income per capita. See the complete outline in Figure 1.

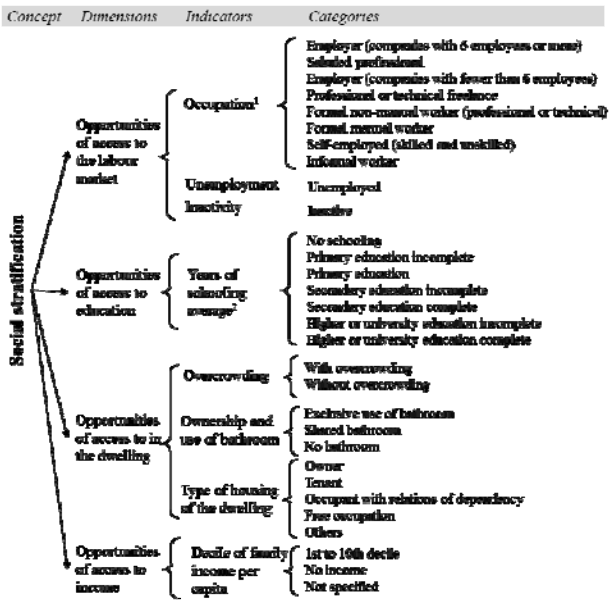


Figure 1.
Operationalization of the concept of social stratification

[1] The occupational status of the household is given by the occupied member of the household (man or woman) with the hierarchically highest job according to the categories drawn up by Portes and Hoffman (2003).
[2] The average number of years of schooling of the household is calculated by adding the years of education of each household member aged 18 and over and dividing the total by the number of household members aged 18 or over.

We selected the indicators of primary goods using data from the Permanent Household Survey (EPH) of Argentina, following a methodical construction of typologies that we call a structural and articulated process (López-Roldán, 1996). We then obtain the model of social stratification from the application of multivariate analysis techniques, which reduce the complexity and diversity of a society to a relatively small and significant number of strata that are homogeneous inside themselves and heterogeneous with each other. These techniques are mainly Multiple Correspondence Analysis (MCA) and Cluster Analysis (CA).

2.1 Data

The source used to prepare this analysis was the EPH, which is a national programme of systematic and ongoing production of social indicators conducted by the Argentine National Institute of Statistics and Censuses (INDEC). Its objective is to determine the demographic and socioeconomic characteristics of the population, and it provides regular official rates of employment, unemployment, underemployment and poverty (INDEC, 2003). In its original form it has been applied in Argentina twice a year (in May and October) since 1973. Following a plan of progressive incorporation, it takes a representative sample of Argentina's urban population from 31 urban agglomerations and an urban-rural area.

At one point in the development of the programme, it was necessary to adapt all the measuring instruments in order to account for the changes taking place in society. For this reason, without altering the initial purpose of the survey, a major reformulation of the EPH was undertaken in order to redevelop the measurement methodology and the operations in line with the current socioeconomic characteristics, the new types of integration in the labour market and the dynamics of change. The reformulation of the EPH covered thematic aspects, modifying the data-gathering instruments and the dimensions of analysis and presentation of results in order to adapt the sample design to the most frequent changes in the phenomena measured. It also covered organizations aspects, adapting the type of work and the computer processes to the established thematic and sample designs. The procedure was set up in the second quarter of 2003 (INDEC, 2005a).

Table 1.
Coverage of the Permanent Household Survey, Argentina

Permanent Household Survey	Edition			
	October 1997	October 2002	2nd half 2003	2nd half 2006
Total households	36,056	28,361	26,548	37,521
Surveyed households	29,360	22,832	26,505	37,521
Expanded households	6,354,293	7.115.643	6.914.843	7.245.436
Surveyed persons	109,302	83,403	93,244	129,410
Expanded persons	22,020,826	24,583,971	23,176,246	24,039,574
Sample error	0.5	0.6	0.6	0.5
Argentina, Census	Year			
	1991	2001	2001	2001
Urban population	28,832,126	32,431,950	32,431,950	32,431,950
% EPH	76.4	75.8	71.5	74.1
Total population	32,615,528	36,260,130	36,260,130	36,260,130
% EPH	67.5	67.8	63.9	66.3

Source: drawn up by the author from micro-data of the EPH and the 1991 and 2001 Censuses

Unlike the specific EPH performed in May and October each year, the new method distributes the sample over the

four quarters of year. This is why it is called the continuous EPH and it provides quarterly, half-yearly and annual figures. Since there have been changes at several levels (in the sample, the variables and the periods of collecting the information, among others) the standardization of the variables was a difficult task. In geographical terms the EPH gathers information on Argentina's urban population living in private households. Significantly, the urban population in Argentina is generally very high, with almost 90% of the population living in urban agglomerations of over 2,000 inhabitants. While the sample of the continuous EPH did not change the geographical area surveyed, it underwent a transformation in the factors of expansion from the first quarter of 2005, resulting in the adjustment of the weights after the completion of the final population projections of the 2001 Census (INDEC, 2005b).

Table 1 presents the sample with which we worked and its level of representativeness based on the expansion factors of the EPH and an additional calculation taking into account the population data from the 1991 and 2001 Censuses. The last database in Argentina that we can reliably use dates from the second half of 2006 due to serious institutional conflicts in the INDEC after that period associated with government intervention. This decision will be maintained at least until a committee of experts' rules on the figures that have been published since then.

2.2 Analysis Techniques

We followed a typological construction method (structural and articulated) involving the combined use of two multivariate analysis techniques:

Multiple Correspondence Analysis. This is a statistical technique applied to studies that use qualitative variables. It establishes the correspondences (correlations) between them based on their categories. In a broad sense, correspondence is understood to be the connection or reciprocal relationship between two equivalent elements or sets of elements (Cornejo, 1988: 95) and it is a generalization of Simple Correspondence Factorial Analysis for cases in which a large number of variables are analyzed simultaneously (Cornejo, 1988: 135). We used this data reduction technique to define the dimensions of social stratification of Argentine households. Cluster analysis comes into operation after the factor variables have been obtained through MCA.

Cluster analysis. This technique allows us to obtain classifications of statistical units arranged in a matrix form. CA is conceived, first, as a direct instrument for the formation of groups that are not necessarily preconceived and, second, as an intermediate data analysis tool that is mainly exploratory. It can be used to build classificatory typologies of units or individuals and therefore facilitates the articulation of hypotheses in data exploration. However, CA can also be used to test hypotheses arising from a previous theoretical work or combined with other data analysis techniques. It therefore provides a strict descriptive exercise to contribute to the conceptualization and explanation of social phenomena, and to the comparison and validation of claims consistent with certain theoretical and methodological models (López-Roldán, 1994: 166-173).

From the point of view of procedure, the purpose of any classification technique is to obtain units that are as homogeneous as possible within classes and as heterogeneous as possible between them. In particular, in a CA process one must select the variables and the measure of proximity, build the distance matrix, choose the classification method, decide the number of classes and validate the results (López-Roldán and Lozares, 2000: 147). In this process, statistical decisions are interspersed with decisions on analysis of consistency, stability, performance and sociological justification (López-Roldán, 1994: 167), as we will see in the following section. As a statistical technique it is subject to a degree of uncertainty, especially in obtaining the final classifications and in delimiting the constituent units of each class.

In our case we used the automatic classification of the mixed algorithm implemented in the SPAD software (version 5.0) which, based on the factor scores of all individuals, applies a triple classification process (Lebart, Morineau & Piron, 2004: 177-184):

1. An initial classification is obtained with the crossing of several basic partitions built around mobile centres.
2. The stable classes derived from this first procedure are then added through a bottom-up hierarchical classification method according to Ward's method based on the criterion of the loss of minimal inertia.
3. Finally, the different partitions of individuals that can be obtained from the cluster tree of Ward's method are optimized or consolidated through a reallocation of the groups created to each partition with a new process of classification by mobile centers that improves the inertia between the groups.

2.3 Criteria for Identifying and Selecting Social Strata

The following are the technical criteria and the interpretation criteria of the researcher, which allowed us to define the types or groups that reflect the social stratification (which we therefore call "social strata"). These criteria are divided into three, according to the research stage. We can thus observe criteria related to determining the factors, determining the groups and validating the global results:

Table 2.
Criteria used to define social strata

Criteria	Technical	Interpretation of researcher
A. Determining the factors	A.1.1 Choice and analysis of the variables and modalities used	A.2.1. Interpretation of the variables and modalities
	A.1.2. Selection of a number of factors that explain at least 70% of the variance	A.2.2. Interpretation of each factor
	A.1.3 Transformation of the eigenvalues	A.2.3. Parsimony in the selection of factors
	A.1.4 Scree test	
B. Determining the groups	B.1.1. Comparison of different classification techniques and modification of parameters in each technique	B.2.1. Substantial comparison of results and interpretation of the typology obtained
	B.1.2. Analysis of indexed hierarchy of the groups obtained	B.2.2. Parsimony in the selection of groups
	B.1.3. Analysis of the proportion of explained variance of each partition	B.2.3. Inter-temporal coincidence of the hierarchical groups
C. Validating the results	C.1.1. Successive processing until stability (some permanence and reiteration) is obtained in the determination of factors and groups	C.2.1. Corroboration of the groups found with external sources
	C.1.2. Multiple correspondence analysis conditional to time	C.2.2. Theoretical and conceptual validation over time

Determining the factors. The crucial point of the process is the selection and analysis of the variables and modalities used. From the technical point of view one must consider several alternative codings of the variables and check that there are relationships between them, particularly observing the contributions and coordinates of each modality (with their test values) in the factors. These assessments enable us to discern the elements that play an important role in the analysis. From an interpretative viewpoint we use substantive criteria of coding and theoretical interrelation between the variables and their modalities, which allow us to assess the choice of variables at this initial stage of the process.

Furthermore, the above analysis is combined with the decision of the number of factors that must be selected to provide an explanation of the phenomenon. The explained variance should not be lower than 70% (Lozares & López-Roldán, 2000: 92). Consequently, there is an effect of reduction and a loss of information, but this is offset by a gain in significance, because the percentage retained expresses the most important information on the phenomenon analyzed.

The procedure used underestimates the explained variance (or inertia), so a transformation of the eigenvalues proposed by Benzécri (1979) is performed in order to correct this (Bisquerra, 1989: 460; Lebart, Morineau & Piron, 2000: 368). With this transformation we evaluate the variance explained among the first factors to determine the main change in the slope of the graph curve relating factors and eigenvalues: the scree test (Lozares & López-Roldán, 2000: 92).

In general, if we wish to give a fuller explanation, a large number of factors can be selected, but it is important to respect the principle of parsimony in order to achieve a balanced analysis, prioritizing the most important elements in the configuration of the factors. This decision will ultimately be guided by theoretical and interpretive criteria. The interpretation of each factor is increasingly the researcher's own competence. This task involves a process of articulation between theoretical elements on which the researcher is asking questions and the observation of empirical information embodied in the factors obtained. The background information, history and other research work are an important input for reaching a comprehensive interpretation of the data being analyzed.

Determining the groups. As we are combining MCA with CA, the variables that act as classification criteria are the factor scores, so we can guarantee the desired technical conditions of the cluster analysis: no correlation between the variables, no redundant information and the same unit of measurement. Under these conditions various methods of classification may be employed. It is interesting to compare results between them and to change the parameters of each one.

The analysis of the indexed hierarchy is used to make the final selection of the groups. The bottom-up aggregation process generates a sequence of partitions ranging from considering each household as a unit to aggregating them all in one group. The process involves the grouping of the most homogeneous units, generating an ultrametric index that allows us to see the distance at which each new partition is formed (López-Roldán, 1994 171).

Each partition implies a significantly greater proportion of explained variance as the number of groups increases. Using the scree test (with the partitions on the abscissa and the percentage of explained variance on the ordinate), we can technically determine a criterion on the number of groups to be considered when the change of curve slope occurs (Lozares & López-Roldán, 2000: 170).

The heuristic nature of the technique requires the support of the theory and knowledge of the phenomenon in the social context in which it occurs. This will give consistency to the decision on the typology finally adopted in terms of number of groups and their defining characteristics. The principle of parsimony is respected, but this time in the selection of the groups.

When the analysis is inter-temporal, as in our case, it is important to compare the number of partitions obtained at any time. If the main feature of the phenomenon analyzed is not high variability in time, it is best to obtain the same number of partitions or strata. If, however, the phenomenon is subject to great changes, we cannot expect to have a homogeneous number of partitions. This fact is extremely important and must be addressed by comparing the theoretical references with the typologies obtained in the research process.

Validating the results. From the technical point of view, the results can also be validated by replicating the analysis for different random subsamples to see whether the factors and the strata contain a level of internal consistency in comparison with the results for the entire sample (López-Roldán & Lozares, 2000: 171). A conditional MCA (Escofier & Pagès, 1990) can also be performed on the variable year (in our case with four categories: 1997, 2002, 2003 and 2006).

Decisions made throughout the above process give the analysis high variability. This variability must be observed over successive processes to find the elements that give coherent stability to the results, in an interactive articulation of theory and empiricism. Depending on the type of research to be carried out, the researcher should initiate a process with one of the following three possibilities: 1) typologies may be generated through a structured induction process, 2) if the research is based on a theoretical framework, the hypotheses may be corroborated through a deductive process, or 3) a combined procedure may be used. Any of these three procedures will allow you to either reformulate the initial theoretical framework or create your own one based on the findings and test them again. Finally, the results of the work must be consistent with scientific developments and contributions made by other researchers in the same field. If these contributions are placed in doubt, the statements made must be thoroughly and forcefully supported.

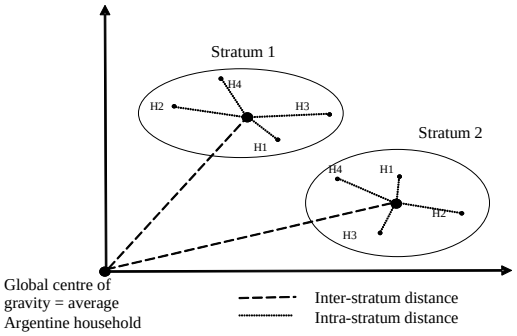
3. Defining Distances

In the context of the research process in which we obtained a typology of stratification in Argentina, we considered the construction of a social distance measure (Fachelli, 2009). For this purpose we used Ward's classification method, which considers the Euclidean distance as a proximity measure.

Households make up a point cloud in a vector space (of three dimensions in our case) with a center of gravity which is defined by the global average for all households. We give the name “total inertia” to the measure of dispersion or variability (expressed through the weighted sum of Euclidean distances) of each household from the global center of gravity of the point cloud. The total inertia can be decomposed, according to Huygens's relation, into the sum of intra-stratum inertias and inter-stratum inertias (Lozares & López-Roldán, 2000: 166-167). The intra-stratum inertia is calculated by summing the squared distances between each household and the center of gravity of the stratum, weighted by the weight of each house-

hold. The inter-stratum inertia is calculated by summing the squared distances of the center of gravity of each group and the global center of gravity, weighting each distance by the number of households in each stratum.

The chart below shows the distances that are used to calculate inter-and intra-stratum inertia.



Total inertia = *inter-stratum inertia* + *intra-stratum inertia*

Figure 2.
Inter-stratum and intra-stratum distance

Ward's method consists in the progressive bottom-up aggregation of the units of analysis so that, in each stage, the households that involve a minimum loss of inertia between the groups are joined, i.e. the nearest or most similar households are joined so that the inertia or variance between groups increases as little as possible. Therefore, the procedure optimizes the variance explained by the joining of households and the minimization of residual variance (within-group variance).

Thus, Ward's method, and particularly the decomposition of the total inertia into inter-stratum inertia and intra-stratum inertia, allows us to measure the distances between Argentine households. To analyze the position of the households in the social stratification, with particular reference to the impact of a social policy or another item of interest, we defined the position of each household in the vector space as the intra-stratum distance based on the following expression, which is presented for each year being analyzed:

$$Inertia_{ik} = \frac{1}{n}\left[\left(F_{1i} - \overline{F_{1k}}\right)^2 + \left(F_{2i} - \overline{F_{2k}}\right)^2 + \left(F_{3i} - \overline{F_{3k}}\right)^2\right]. \quad (1)$$

where:

i: household of stratum k (*i= 1… nk*).

k: group or stratum to which the household belongs.

1/n: inverse of the number of cases in the sample (mass).

F_{1i}, F_{2i}, F_{3i}: coordinates of each household of group k in each factorial axis.

$\overline{F_{1k}}, \overline{F_{2k}}, \overline{F_{3k}}$: coordinates that define the centre of gravity of group k.

Furthermore, we identify the inter-stratum distance in the three-dimensional vector space in order to obtain what we call the "social distance" of each stratum from the average household using the following formula:

$$Social\ Distance_k = \left(\overline{F_{1k}} - \overline{F_1}\right)^2 + \left(\overline{F_{2k}} - \overline{F_2}\right)^2 + \left(\overline{F_{3k}} - \overline{F_3}\right)^2. \quad (2)$$

where:

k: group or stratum to which the household belongs.

$\overline{F_{1k}}, \overline{F_{2k}}, \overline{F_{3k}}$: coordinates that define the centre of gravity of group *k*.

$\overline{F_1}, \overline{F_2}, \overline{F_3}$: coordinates that correspond to the centre of the point cloud (0,0,0) in the space of the three factors.

Consequently:

$$Social\ Distance_k = \overline{F_{1k}}^2 + \overline{F_{2k}}^2 + \overline{F_{3k}}^2. \quad (3)$$

4. Results

4.1 A Brief Reference to the Social Strata

Taking into account the guidance of the regulatory criteria and the use of indicators of employment, education, housing and income, we defined four social strata in each year being analyzed (Fachelli, 2009). The procedure used can be seen in the following table:

Table 3.
Matrices used to obtain the social strata in Argentina

Year	Original matrix X (n×p)	Factorial matrix F (n×m)	Typological matrix T (k×m)
1997	6.354.293 households × 6 variables	6.354.293 households × 3 dimensions	4 strata × 3 dimensions
2002	7.115.643 households × 6 variables	7.115.643 households × 3 dimensions	4 strata × 3 dimensions
2003	6.914.843 households × 6 variables	6.914.843 households × 3 dimensions	4 strata × 3 dimensions
2006	7.245.436 households × 6 variables	7.245.436 households × 3 dimensions	4 strata × 3 dimensions

Where: n=households; p=number of variables; m= number of dimensions (factorial axis) and k= number of strata (clusters).

We label each social stratum as follows: we call the first one the "upper stratum" and the last one the "lower stratum"; the two middle groups have particular characteristics that allow them to be differentiated and we call them the "occupationally active middle stratum" and the "occupationally inactive middle stratum". The frequency distribution of the typology obtained is presented in Table 4.

Table 4.
Social stratification in Argentina (% of households)

Social Strata	Period			
	Stability 1997	Post-crisis 2002	Incipient Recovery 2003	Consolidate Recovery 2006
Upper	15.3	14.0	14.5	16.2
Occupationally active middle	46.5	43.4	42.5	45.8
Occupationally inactive middle	21.2	22.3	21.3	17.9
Lower	17.0	20.2	21.7	20.1
Total	100.0	100.0	100.0	100.0
Expanded households	6,354,293	7,115,643	6,914,843	7,245,436

Source: drawn up by the author from micro-data of the EPH.

Below we describe the most important features of the households of each social group.

The *upper stratum*, mainly households composed of:

- employers or salaried professionals,
- persons who have completed higher or university education,
- home-owners without overcrowding and with exclusive

use of a bathroom,
- persons with a high decile of per capita family income (eighth to tenth)

The *occupationally active middle stratum*, mostly households composed of:
- formal manual workers;
- persons who have or have not completed secondary education;
- home-owners (though a small percentage of households rent) without overcrowding (though a small percentage of households have overcrowding) and with exclusive use of a bathroom;
- persons with a medium decile of per capita family income (fourth to eight).

The *occupationally inactive middle stratum*, mostly households composed of:
- persons not connected to the labour market (over 70% of this stratum, thus giving it its name);
- persons who have or have not completed primary education (some have secondary education);
- home-owners without overcrowding and with exclusive use of a bathroom;
- persons belonging to all deciles of per capita family income but with a lower presence in the fifth to seventh.

The *lower stratum*, mostly households composed of:
- informal workers, skilled or unskilled self-employed workers and some formal workers;
- persons who have completed primary education and some who have started but not finished secondary education;
- home-owners (though this is the stratum with the highest percentage of households that occupy the dwelling free of charge) with overcrowding and with exclusive use of a bathroom (though some households share a bathroom or have no bathroom);
- persons with a low decile of per capita family income (first to third).

Overall, from the point of view of the evolution of these strata, it can be seen that there is a fairly large middle stratum—over 40% of households in all cases—whose members are connected to the labour market. Secondly, the social stratum composed mainly of households not connected to the labour market accounted for about 21% but it was lower in 2006 due to the economic recovery and the increase in employment. Thirdly, the lower stratum of households grew in the 2002 economic crisis and has not recovered. Finally, the smallest group of households is the upper stratum, which represents 14-16% of Argentine households

4.2 Distances towards the Interior of each Stratum and between Strata

Taking into account the guidance of the regulatory criteria and the use of indicators of employment, education, housing and income, we defined four social strata in each year being analyzed (Fachelli, 2009). The procedure used can be seen in the following table.

Since our emphasis is to observe the distances between strata, below we present the intra-stratum distance measure that gives objective references on the dispersion of households within each stratum. This is interesting if we observe it over time to obtain a dimension of the evolution of the degree of cohesion or polarization of the social strata.

We recall that the intra-stratum inertia is a measure that emerges from the composition of the positions of each household on three factorial axes, as reflected in a vector in the three-dimensional space that expresses the degree of dispersion of each stratum. The results are presented below.

Table 5.
Inertias per period according to social strata

Social Strata	Period			
	Stability 1997	Post-crisis 2002	Incipient Recovery 2003	Consolidate Recovery 2006
Upper	0.050	0.041	0.046	0.048
Occupationally active middle	0.093	0.084	0.079	0.094
Occupationally inactive middle	0.088	0.088	0.074	0.048
Lower	0.067	0.073	0.076	0.077
Intra-stratum inertia	0.297	0.286	0.275	0.267
Inter-stratum inertia	0.589	0.582	0.581	0.579
Total Inertia	0.886	0.868	0.856	0.845
Inter-stratum inertia/total inertia	66%	67%	68%	68%

Source: drawn up by the author from micro-data of the EPH.

The most dispersed strata with respect to their centre of gravity are the middle strata. The most homogeneous one is the upper stratum. The lower stratum also has a high level of dispersion but lower than that of the middle strata in 1997 and 2002. Due to the fall in unemployment in 2006 and the move of households towards activity, the occupationally inactive middle stratum was homogenized to such an extent that it reached a similar dispersion to that of the upper stratum.

In 1997 and 2002 the strata that showed the greatest changes in their inertia were the upper stratum, the occupationally active middle stratum and the lower stratum, in that order. This situation shows the importance of the effect of the economic crisis on the strata that are connected to the labour market. In 2003 and 2006, however, the strata most affected were the middle ones. We interpreted this situation above as the effect of the economic recovery of 2006, in which business activity was re-established, the number of unemployed fell and the households that had previously been inactive became occupationally active. Interestingly, the inter-layer inertia divided by the total inertia is an indicator of the explanatory power of the social differences expressed in our model, reaching levels close to 70%.

Table 6.
Social distance per period according to social stratum

Social Strata	Period			
	Stability 1997	Post-crisis 2002	Incipient Recovery 2003	Consolidate Recovery 2006
Upper	1.5770	1.6934	1.4290	1.2862
Occupationally active middle	0.1539	0.1763	0.1783	0.1639
Occupationally inactive middle	0.5969	0.5674	0.6223	0.6987
Lower	0.9980	0.7866	0.8161	0.9112
Global average household	0.0000	0.0000	0.0000	0.0000

Source: drawn up by the author from micro-data of the EPH.

From the inter-stratum distance we measure the distances of each stratum from a global average household that can be taken as the typical Argentine household. We have called this the "social distance". It also allows one to observe the variation in the strata over time (see Table 6).

The table shows the positioning of the social strata. We thus see how the stratum furthest from the average Argentine household is the upper stratum, followed by the lower stratum. On the other hand, the middle strata are closer to the average household. We can conclude that the average social distance between strata functions as an indicator of inequality that, based on the relative position of each stratum, can hierarchize the social inequality per year.

It is interesting to highlight the results for the distance between the upper stratum and the average household, which we consider to be highly significant. 2002 is the year that shows the greatest inequality because its upper stratum is the furthest from the average household, thus confirming the hypothesis that the crisis increases the social distance between strata. The year with least distance between strata is 2006, because of the socio-economic recovery and stabilization of the country.

5. Can a measure of multidimensional stratification be a synthetic indicator?

We considered the problem of obtaining a measure that can synthesize in one-dimensional terms the results of the stratification analysis presented above, which we expressed in three factor dimensions and four social strata. To do this we performed several exercises of construction of the measure, which we present below.

5.1 A Measure Based on the Foster-Greer-Thorbecke (FGT) Index

As a reference for this analysis we applied a measure that is widely used in Latin America and corresponds to the family of poverty indices proposed by Foster, Greer and Thorbecke (1984). The general measure is the following:

$$P_\alpha(y;z) = \frac{1}{n}\sum_{i=1}^{q}\left(\frac{g_i}{z}\right)^\alpha . \qquad (4)$$

where:

y_i are the household income.

z is the poverty line (based on the cost of a basic food basket).

$g_i = z - y_i$.

q is the number of households below the poverty threshold

n is the total number of households.

α is the poverty aversion (α=1 is the poverty ratio).

In order to position the strata in a measure that arranges them in scalar order, in the FGT measure, with α =1, we substitute the cost of the shopping basket that determines the poverty line z with a reference value that is the one occupied by the household positions at the maximum value of each of the three factorial axes. The measure proposed is thus the multidimensional FGT index (FGT_m):

$$FGT_m(F_j;z_j) = \frac{1}{n}\sum_{j=1}^{m}\sum_{i=1}^{n}\frac{z_j - F_{ij}}{z_j} . \qquad (5)$$

where:

F_j are the factor variables ($j=1\ldots m$, with $m=3$).

$z_j = \max(F_j)$

F_{ij} are the coordinates of each household i in the factor variables.

When we apply this measure to our data, which we calculate in each stratum and for each year of study, we obtain the following table:

Table 7.
FGT_m measure for each stratum and year

Social Strata	Period			
	Stability 1997	Post-crisis 2002	Incipient Recovery 2003	Consolidate Recovery 2006
Upper	3.36	3.54	3.32	3.42
Occupationally active middle	2.28	2.20	2.35	2.29
Occupationally inactive middle	3.65	3.44	3.20	3.32
Lower	3.83	3.81	3.89	4.04

Source: drawn up by the author from micro-data of the EPH.

The expected result was the classification of the strata according to the synthetic measure FGT_m, but the combination of the three factorial axes generates a measure that does not arrange them in scalar order. This is because each factor orders the strata differently, and only the first factor (the most important one), reflects the conceptual order that corresponds to the social stratification: i.e. first the upper stratum, then the occupationally active middle stratum, then the occupationally inactive middle stratum and finally the lower stratum. As we obtained few results, we continued the investigation with other measures

5.2 Measure based on the combination of factor variables

The second alternative was to consider a series of one-dimensional measures obtained by combining factor variables with different calculation and weighting criteria. From all the measures developed we chose the one that produced the greatest explained variance according to the stratification variable (Eta-squared statistic). The measure we propose, FL (Fachelli-López index) is the following:

$$FL_t = \sum_{j=1}^{m} w_{jt} \cdot F_{jt} . \qquad (5)$$

where:

F_{jt} are the factor variables ($j=1\ldots m$, with $m=3$) of each year t analyzed (1997, 2002, 2003 and 2006).

w_{jt} are the weights of each factor according to its proportions of inertia in each year t analyzed.

This measure is more effective than the previous one because it manages to order all the households in a continuous rank, with the households in the lower stratum at the bottom, followed by the middle strata and finally the upper stratum. However, it is not a totally effective measure, because the position of the households is not exhaustive, i.e. some households that belong to different strata are superimposed. On evaluating the Eta-squared statistic, we can see that the dimensional measure positions well between 65% and 70% of the households, according to the year analyzed. This situation is shown in the

following chart, which presents the histogram of the *FL* index according to the strata to which the households belong.

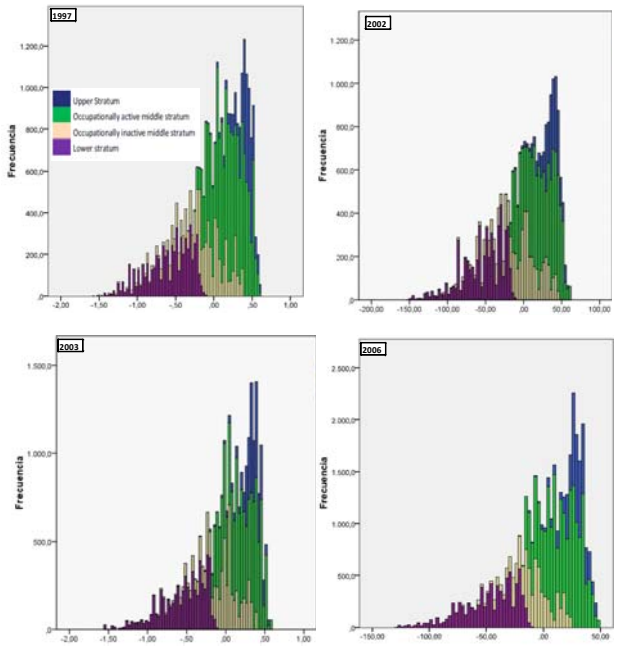


Figure 3.
FL synthetic measure. Ordering of the strata according to the year analyzed

We can conclude that the search for synthetic measures that express the multidimensional results in the four social strata have not yet been implemented successfully.

6. Conclusions

Starting from a differentiation of the strata established previously according to the amount of primary goods (employment, education, housing and income) possessed by each household, we have observed the different positions occupied by households in a multidimensional analysis.

In general terms it can be argued that the social structure we have defined reflects two distinct processes: the socio-economic decline between 1997 and 2002, characterized by an increase in the percentage of households in the lower stratum and a decrease in the percentage of households in the upper stratum; and the recovery process observed from 2003, which was marked by increased economic activity, as evidenced by the higher percentage of the occupationally active middle stratum and the lower percentage of the occupationally active middle stratum. Though the percentage of households belonging to the lower stratum fell by 7% between 2003 and 2006, this change is not as important as the increase in occupational activity.

The positioning of households in different groups allows us to calculate the dispersion of each stratum with respect to its centre and to obtain an objective social distance between strata. These two measures were analyzed in each year.

The intra-stratum distance (which we have called inertia) shows that the strata with most dispersion around their centre of gravity are the middle strata. On the other hand, the most homogeneous stratum is the upper stratum. The lower stratum has a similar dispersion to that of the middle strata, though slightly lower. Therefore, the socioeconomic crisis that began in late 2001 led to an increase in homogenization in the upper stratum and the occupationally active middle stratum, and an increase in dispersion in the lower stratum. The occupationally inactive middle stratum was practically unaffected.

The inter-stratum distance (which we have called the social distance) shows that the strata have different behaviours as they move closer to or further from the average household in each year analyzed. The Argentine crisis increased the distance between the strata. In particular the upper stratum is that which shows the greatest distance from the average household in 2002, though this was inverted in the period of recovery.

Efforts to develop a measure that provides a quantitative synthesis of the strata that express the qualitative characteristics of stratification in Argentina have failed. These efforts were made in two ways: by applying the *FGT* index to the multivariate analysis and by expressing in a single measure the positioning of the households in the three factorial axes. In the first case the measure does not arrange the households in scalar order and in the second it improves the level of measurement but is not exhaustive in the position of the households. We thus conclude that the challenge is still open.

The results achieved so far force us once again to highlight the importance of the typology constructed, which provides a qualitative (and ordinal) synthesis of the social strata identified in Argentina in the years analyzed. This order, which expresses the typology of social stratification, is a structuring of the phenomenon studied. It can therefore be considered as a type of explanation that we can classify as structural because it goes beyond mere description, though it does not meet the conditions of a causal explanation.

REFERENCES

Bisquerra, R. (1989). *Introducción conceptual al análisis multivariable*. Barcelona: Promociones y Publicaciones Universitarias.

Cornejo, J. M. (1988). *Técnicas de investigación social: el análisis de correspondencias (teoría y práctica)*. Barcelona: Promociones y Publicaciones Universitarias.

Escofier, B. (1990). Analyse des Correspondances Multiples Conditionnelle. *Revue de Modulad* , 5, 13-27.

Escofier, B., & Pagès, J. (1990). *Analyses factorielles simples et multiples. Objectifs, méthodes et interprétation*. Paris: Dunod.

Fachelli, S. (2009). *Nuevo modelo de estratificación social y nuevo instrumento para su medición. El caso argentino. Tesis doctoral*. Bellaterra (Barcelona): Universitat Autònoma de Barcelona. URL (last checked 5 January 2013) http://www.tdx.cat/handle/10803/5149

Foster, J., Greer, J. & Thorbecke, E. (1984). A Class of Decomposable Poverty Measures. *Econometrica*, 52(3), 761-766.

Fundació Jaume Bofill (2005). *Estructura social i desigualtats a Catalunya, Volum I, Classes socials, educació, treball i usos del temps a Catalunya*. Barcelona: Editorial Mediterrània.

Greenacre, M. (2008). *La práctica del análisis de correspondencias*. Madrid: Fundación BBVA.

INDEC (2009). *Valorización mensual de la canasta básica alimentaria y de la canasta básica total. Aglomerado Gran Buenos Aires*. Buenos Aires: INDEC.

INDEC (2005a). *Encuesta Permanente de Hogares Continua. Diseño de registro y estructura para las bases preliminares. Hogares y personas*. Buenos Aires: INDEC.

INDEC (2005b). *Encuesta Permanente de Hogares. Cambio de las estimaciones de población en base a las proyecciones definitivas del Censo 2001*. Buenos Aires: INDEC.

INDEC (2003). *La nueva Encuesta Permanente de Hogares de Argentina. Dirección de Encuesta Permanente de Hogares*. Buenos Aires: INDEC.

INDEC (2001). *Censo Nacional de Población, Hogares y Viviendas 2001. Aspectos metodológicos del Censo* 2001. Buenos Aires: IN-

DEC.

Lebart, L., Morineau, A. & Piron, M. (2000). *Statistique exploratoire multidimensionnelle*. Paris: Dunod.

López-Roldán, P. (1996). La construcción de tipologías: metodología de análisis. *Papers. Revista de Sociologia*, 48, 9-29.

López-Roldán, P. (1994). *La construcción de tipologías en sociología: propuesta metodológica de construcción, análisis y validación. Aplicación al estudio de la segmentación del mercado de trabajo en la Región Metropolitana de Barcelona. Tesis Doctoral*. Bellaterra (Barcelona): Universitat Autònoma de Barcelona. URL (last checked 5 January 2013) http://hdl.handle.net/10803/5135.

Lozares, C. & López-Roldán, P. (2000). *Anàlisi multivariable de dades estadístiques*. Bellaterra (Barcelona): Servei de Publicacions.

Portes, A. & Hoffman, K. (2003). *Las estructuras de clases en América Latina, composición y cambio durante la época neoliberal*. Santiago de Chile: CEPAL-Naciones Unidas.

Rawls, J. (2002). *La justicia como equidad. Una reformulación*. Barcelona: Paidós.

Rawls, J. (1979). *Teoría de la justicia*. Madrid: Fondo de Cultura Económica.