EXTENDED ABSTRACT

Title: ARE THERE SPILLOVERS IN PARENTS SPECIALIZATION IN CHILD CARE ACTIVITIES? EVIDENCE FROM THE SPANISH TIME USE SURVEY 2009/10

Authors and e-mail of all: José-María Montero(a) jose.mlorenzo@uclm.es; María-Ángeles Medina (b) amedina@ucm.es; Gema Fernández-Avilés(a) gema.faviles@uclm.es

Department: Statistics

University: University of Castilla-La Mancha; (b) Complutense University of Madrid

Subject area: 08

Extended Abstract: This research deals with one of the most debated topics in developed countries and especially in Spain: The distribution of the time dedicated by mothers and fathers to child care activities. We selected the child care activities from the list that mirrors the list published in EUROSTAT's 2008 guidelines. This debate is mostly based in opinions rather than in scientific work. In the case of scientific research in the topic, official information is not questioned (which is a key task) and spatial and spatio-temporal dependencies are not taken into account, with results in non-trustable conclusions.

In this study, the data were provided by the Time-Use Survey conducted by the Spanish Statistics Office in 2009-2010 (STUS). The three basic units of observation and analysis that are considered in STUS are (i) the individual members of the household aged 10 and above, (ii) private households residing in main family dwellings, (iii) the days of the week.

We first list the main deficiencies of this survey and show the way it has been depurated to be correctly used. This is crucial information for the users of the STUS because, otherwise, their results and conclusions might lead to misunderstandings. More in detail, the size of the planned sample was around 11,538 dwellings, but after removing the empty dwellings and the dwellings that could not be sampled, the sample was reduced to 9,541. Since the households of interest for childcare research are those made up of at least one heterosexual couple with children, we initially selected households where the reference person was part of a heterosexual couple. However, surprisingly, we could not use the classification used in STUS because of the discrepancy between the type of household and the kinship of household members (this is a serious drawback of the survey). Consequently, we set up our own classification and selected 6,259 households of interest (including a heterosexual couple). Finally, only 1,878 of these households reported having devoted at least ten minutes to childcare activities the day they filled the one-day diary (we excluded Ceuta y Melilla from the database). Therefore, the final database of households with heterosexual parents and children contains 1,878 units (households). It is of note that even though STUS collects information on both main and secondary activities, we only proceed with main activities because of the small number of households reporting that they perform secondary childcare activities (less than 800) and the inconsistency of their responses. This cannot be considered a problem if we do not conflate primary child care activities with the time that parents spend with children.

Second, in order to study parent specialization in child care activities, we use the dissimilarity index, a particular case of the Duncan and Duncan index (Duncan and Duncan 1955a, b) which has been widely used in the literature to study segregation, but could be interpreted as a specialization index. Finally, we check for local spatial autocorrelation using Moran's I and Geary's C statistics. Finding patterns of spatial autocorrelation could help to explain father participation patterns in childcare activities. As far as we know, this is the first article that has attempted to identify spatial patterns in dissimilarity indexes for genders.

Spatial autocorrelation is the two-dimensional equivalent of redundancy. It measures the extent to which the occurrence of an event in an areal unit either constrains the occurrence of an event in a neighboring areal unit or makes it more probable.

MI is based on cross-products of the deviations from the mean and is calculated for n observations on a variable x at locations *i*, *j*, as:

$$MI = \frac{n}{S_0} \frac{\sum_i \sum_j w_{ij}(x_i - \bar{x})(x_j - \bar{x})}{\sum_i (x_i - \bar{x})^2},$$

where $\bar{x}$ is the mean of the x variable, w_{ij} are the elements of the weight matrix (see Getis 2009 for details on the weight matrix) and S_0 is the sum of the elements of the weight matrix: $S_0 = \sum_i \sum_j w_{ij}$. We use the queen criterion, which determines neighboring units as those that have any point in common, including both common boundaries and common corners. *MI* is similar but not equivalent to a correlation coefficient. In the absence of autocorrelation and regardless of the specified weight matrix, the expected value of *MI* is $-1/(n-1)$. It is of note that for a row-standardized spatial weight matrix, the normalizing factor S_0 equals n (since each row sums to 1) and *MI* simplifies to a ratio of a spatial cross product to a variance. An *MI* coefficient greater than $-1/(n-1)$ indicates positive spatial autocorrelation; by contrast, a *MI* value less than $-1/(n-1)$ indicates negative spatial autocorrelation. The variance of *MI* is:

$$V[MI] = \frac{n\left[\left(n^2 - 3n + 3\right)S_1 - nS_2 + 3S_0^2\right] - k\left[n\left(n-1\right)S_1 - 2nS_2 + 6S_0^2\right]}{\left(n-1\right)\left(n-2\right)\left(n-3\right)S_0^2} - \frac{1}{\left(n-1\right)^2}$$

where

$S_1 = \frac{1}{2}\sum\sum_{i \neq j}\left(W_{ij} + W_{ij}\right)^2 = 2S_0$ for symmetric W containing 0's and 1's.

$S_2 = \sum_i \left(W_{i0} + W_{0i}\right)^2$ where $W_{i0} = \sum_j W_{ij}$ and $W_{0i} = \sum_j W_{ji}$

GC coefficient is defined as:

$$GC = \frac{n-1}{2S_0} \frac{\sum_i \sum_j w_{ij}(x_i - x_j)^2}{\sum_i (x_i - \bar{x})^2}.$$

GC is inversely related to *MI*. It ranges from 0 (maximal positive autocorrelation) to a positive value for high negative autocorrelation (usually 2). In the absence of

autocorrelation and regardless of the specified weight matrix, its expected value is 1 (Sokal and Oden 1978). If the value of *GC* is less than 1, it indicates positive spatial autocorrelation. Values between 1 and 2 indicate negative spatial autocorrelation. The variance of *GC* is given by the following expression:

$$V[GC]=\frac{1}{n(n-2)(n-3)S_0^2}\left\{\begin{array}{l} S_0^2\left[\left(n^2-3\right)-k\left(n-1\right)^2\right]+S_1\left(n-2\right)\left[n^2-3n+3-k\left(n-1\right)\right] \\ +\frac{1}{4}S_2\left(n-1\right)\left[k\left(n^2-n+2\right)-\left(n^2+3n-6\right)\right] \end{array}\right\}$$

where S_0, S_1, and S_2 are the same as in *MI*.

MI and *GC* yield similar conclusions. *MI* provides a more global indicator, whereas *GC* is more sensitive to differences in small areas. However, *MI* has usually been preferred since Cliff and Ord (1975, 1981) showed that it is consistently more powerful than *GC*.

Finally, in order to find spatial patterns that could help to explain father participation in childcare activities, we have conducted a model-based cluster procedure implemented in the MCLUST algorithm. This procedure is based on the assumption that data are generated by normal multivariate distributions with different covariance matrices. That is to say, the data generation processes are a mixture of normal distributions (Fraley and Raftery 1999, 2002, 2010)[1]. The covariance matrices are decomposed in terms of volume, shape and orientation —which makes it possible to define a range of models— and their implied type of distributions, considering a complete range of decomposition possibilities.

The choice of both a specific model and a specific number of groups allows for maximum likelihood estimation of the different group matrices (assuming a mixture of normal distributions). Then, observations can be assigned to a group. In order to select both the model class and the number of groups, MCLUST algorithms proceed to maximize a re-parameterization of the Bayesian Information Criterion (BIC) where the maximum is taken over all the models and number of potential groups considered. BIC is the value of the maximized log-likelihood with a penalty for the number of parameters in the model, and allows comparison of models with differing parameterizations and/or differing numbers of clusters. In general, the higher the value

[1] Version 3 of MCLUST for R is available as a contributed package (MCLUST) in the R language. It can be obtained from CRAN at http://cran.r-project.org/web/packages/mclust/index.html.

of BIC, the stronger the evidence for the model and number of clusters (see Fraley and Raftery 2002).

As for results, it can be noted that father participation in childcare activities is far from uniform across Spanish provinces. As for the those derived from the spatial distribution of father participation in childcare activities across provinces, the value of both *MI* and *GC* statistics are found not to be significant when considering all childcare activities together. However, significant results are obtained for specific activities.

Finally, the MCLUST algorithm selects VEV,2 (256.8595), VEV,3 (245.1425) and EEV,2 (229.6044) as the best cluster models[2] (number of groups after the comma and BIC values in parentheses) for grouping the Spanish provinces according to the vector of father participation in the five childcare activities considered. Although the best model is VEV,2 we have selected VEV,3 because the BIC value for the two models is similar and the one with three groups facilitates the interpretation of the results.

Of course, much more can be done because this is a hot topic of special interest for society. Some interesting avenues for future research include the comparing Spanish results to those stemming from the Time-Use Surveys of other countries, analysis of the advances that have taken place in father participation in childcare since the STUS 2002-2003, searching for the latent factors that explain the low level of male participation in childcare activities, analysis of the disparity in the amount of time devoted by mothers and fathers when analyzing the households that dwell in the areas of interest, etc.

References

Cliff, A. D. and J. K. Ord. (1975). The choice of a test for spatial autocorrelation. In Display and Analysis of Spatial Data. In J. C. Davies and M. J. McCullagh eds. John Wiley and Sons, London.

Duncan, O. B. and B. Duncan (1955a). "A methodological analysis of segregation indexes", American Sociological Review, nº 20, p. 210–217.

Duncan, O. B. and B. Duncan (1955b). "Residential distribution and occupational stratification", American Journal of Sociology, nº 60(5), p. 493–503.

[2] VEV: Ellipsoidal distribution with variable volume, equal shape and variable orientation. EEV: Ellipsoidal distribution, equal volume and shape and variable orientation.

Fraley C. and E. Raftery (1999). "MCLUST: Software for model-based cluster analysis", Journal of Classification, nº 16, p. 297–306.

Fraley C. and E. Raftery (2002). "Model-based clustering, discriminant analysis and density estimation", Journal of the American Statistical Association. nº 97, p. 611–631.

Fraley C. and E. Raftery (2010). "MCLUST Version 3 for R: Normal Mixture Modeling and Model-Based Clustering", Technical Report nº 504, Department of Statistics. University of Washington, Washington DC.

Getis, A. (2009). "Spatial Weights Matrices", Geographical Analysi, nº 41(4), p. 414-410.

Sokal, R. R. and N. L. Oden (1978). "Spatial autocorrelation in biology. Methodology". Biological Journal of the Linnean Society, nº 10, p. 199–228.

Keywords: Time use, child care, dissimilitude index, spatial econometrics
JEL codes: C14, C15, C21, J16