The relationship between space, time, fertility, and employment: a fixed effectsspatial panel approach.

Alessandra Carioli[1]

Daniel Devolder[2]

Joaquin Recaño[2]

L.J.G. van Wissen[3]

[1] University of Groningen and Netherlands Interdisciplinary Demographic Institute
[2] Centre d'Estudis Demografics, Universitat Autonoma de Barcelona
[3] Netherlands Interdisciplinary Demographic Institute

Abstract

Heterogeneity and time dependency are a challenging topic in demography. In this study, we aim to describe the spatial effect of socio-economic indicators over the evolution of fertility among 910 Spanish municipalities measured cross-sectionally at macro level in 1986, 1991, 2001 and 2011. We apply a fixed effects spatial error model with GMM estimation that extends the cross-sectional model of Kelejian and Prucha (2010). Preliminary analysis finds strong spatial autocorrelation for all the variables and years taken into account in the panel and show significant clusters substantially transforming over time. Indeed, during these four decades major transformations took place in the family formation process, generating significant changes in the effect of female participation into the labour force and employment as shown by (Adsera, 2004). Moreover, such evolution displays important heterogeneity across areas revealing underlying leaders and laggers in fertility behaviour changes.

Data and Method

1. Data

The data used for the analysis come from microcensus data for years 1986, 1991, 2001 and 2011.

Data on the number of births consist of raw numbers of births by mothers' age group (5 years age groups from 15 to 49) by single year and by birth order (1 to 3+). For years 2001 and 2011, data on births contain also information on mothers and fathers' nationality.

Data on female population exposures consist of population numbers by five-years age groups measured on the 1st of March of each year. From 1998 to 2012 single year population estimates are available. For the previous period, 1981-1997, three calendar years measurements are available: 1981, 1986 and 1991, which were used to obtain inter year estimates. From 1998 onwards data contain information on mothers and fathers' nationality.

All data are grouped by 910 comarcas, which are administrative agglometarions of municipalities, each geographical unit having at least 20,000 inhabitants.

2. Measures

In this study several fertility measures are computed to help investigating spatial patterns of fertility. These measures are constructed using the data described in section 1. Some considerations are necessary on the assumptions of data and methods applied to obtain the indicators.

The dependent variable used in this analysis is the Princeton Index for total birth order and first birth order:

$$I_f = \frac{B_x}{\sum_x p_x F_x}$$

where B_x is the number of births by women in age group x, $\mathbf{F}_x$ represents the Hutterites fertility for age group x and p_x the female population in age group x. The choice for I_f instead of Total Fertility Rate relies on its stability, being less affected by small changes in the number of births by age and by small and scarcely populated rural areas births variability, as it is expressed as a proportion of Hutterites fertility.

The explanatory variables considered are:

1. Mean Age at Childbearing, MAC for total birth order and first birth order:

$$MAC = \frac{\sum_{x_{\min}}^{x_{\max}} \bar{x} \cdot f(x)}{\sum_{x_{\min}}^{x_{\max}} f(x)}$$

2. Socio-economic variables
    - 2.1. Foreigners presence as % of the total;
    - 2.2. Share of population with Primary Education by sex;
    - 2.3. Share of population with University Education by sex;
    - 2.4. Share of Unemployed by sex;
    - 2.5. Share of Active by sex;
    - 2.6. Share of population with permanent contract (with social security benefits) by sex;

2.7. Share of population with temporary contract by sex;
2.8. Share of self-employed by sex;
2.9. Share of self-employed with employees by sex;
2.10. Share of self-employed without employees by sex;
2.11. Share of employed by sector (Primary, Secondary, Tertiary and Constructions) by sex;

3. Rural-Urban indicator (number of inhabitants per square km);
4. Distance from largest nearest city;
5. Easting and Northing;

3. Method

Spatial panels have recently attracted much attention, especially with the increasing availability of micro and macro-level panel data, especially in the field of economics (see Baltagi:2007vz and Driscoll & Kraay, 1995). Although spatial econometrics techniques are not new in demography, (Bocquet-Appel & Jakobi, 1996; 1998; Galloway, Hammel, & Lee, 1994; among others: Klusener, Szoltysek, & Goldstein, 2012; Szołtysek, Gruber, Klüsener, & Goldstein, 2010), spatial panel techniques are seldom applied to analyse demographic phenomena, especially fertility. The spatial panel approach assumes that the structure of cross-sectional correlation is related to location and distance among areas and it is defined according to a pre-specified metric given by a connection or spatial matrix that characterizes the pattern of spatial dependence according to pre-specified rules. Cross-sectional correlation is represented by means of a spatial processes, which explicitly relate each area to its neighbors via a neighborhood matrix (Anselin, 1988; Anselin, Florax, & Rey, 2004; Cliff & Ord, 1981; Moran, 1950).

In this study we define the neighborhood relationship to follow a First Order Rook, defined as a set of boundary points b of unit i, which share a positive proportion of their boundary, thus having length l_{ij} >0:

$$w_{ij} = \begin{cases} 1 & l_{ij} > 0 \\ 0 & l_{ij} = 0 \end{cases} \quad (1)$$

Once the spatial neighbor list has been defined, in spatial analysis it is necessary to set the weight matrix for each relationship. The spatial weight matrix has been constructed so that the weights for each areal item sum up to unity.

$$W_{ij} = \begin{bmatrix} 0 & d_{1,2} & \dots & d_{1,n} \\ d_{2,1} & 0 & \dots & d_{2,1} \\ \vdots & \vdots & \ddots & \vdots \\ d_{m,1} & d_{m,2} & \cdots & 0 \end{bmatrix} \quad (2)$$

To establish the presence of significative and non-random spatial autocorrelation we apply first Moran't I test for spatial autocorrelation, to evaluate the strength of spatial patterns across the considered variables (Moran, 1950) and is computed on the model's residuals.

Moran's I is the index obtained through the product of the variable considered, let's call it y, and its spatial lag, with the cross product of y and adjusted for the spatial weights considered:

$$I = \frac{n}{\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}} \cdot \frac{\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}(y_i - \bar{y})(y_j - \bar{y})}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \quad (3)$$

where n is the number of spatial units i and j, y_i is the i^{th} spatial unit, is the mean of y, and w_{ij} is the spatial weight matrix, where j represents the regions adjacent to i. Moran's I can take on values [-1,+1], where – 1 represents strong negative autocorrelation, 0 no spatial autocorrelation and 1, strong positive spatial autocorrelation.

It is possible to break down this measure its components in order to identify clusters and hotspots. Clusters are defined as observations with similar neighbors, while hotspots are observations with very different neighbors (Anselin, 1995). The procedure is knows as Local Indicators of Spatial Association or LISA, where the Local Moran's I decomposes Moran's I into its contributions for each location. These indicators detect clusters of either similar or dissimilar values around a given observation. The relationship between global and local indicators is quite simple, as the sum of LISAs for all observations is proportional to Moran's I. Therefore, LISAs can be interpreted both as indicators of local spatial clusters or as pinpointing outliers in global spatial patterns.

The measure for LISAs is defined as:

$$I_i = \frac{(y_i - \bar{y})\sum_{j=1}^{n} w_{ij}(y_i - \bar{y})}{\frac{\sum_{i=1}^{n}(y_i - \bar{y})}{n}} \quad (4)$$

The aim of this preliminary analysis is to establish the presence of spatial autocorrelation and of spatial clusters. Once this step is done, we define our spatial panel model as a fixed effects spatial error model, thus allowing for presence of endogeneity and correlation in the residuals:

$$y_{ti} = X'_{ti}\beta + u_{ti} \quad \text{(i=1,…,N; t=1,…,T)} \quad (5)$$

where y_{ti} is the observation on the i^{th} comarca at time t, X'_{ti} is the k × 1 vector of observations and u_{ti} the regression error.

In order to allow for endogeneity in the model, we choose a fixed effects approach using a spatial error model, thus:

$$u_{ti} = \mu + \varepsilon_t \quad (6)$$

with:

$$\varepsilon_t = \lambda W_N \varepsilon_t + v_t \quad (7)$$

where $\mu' = (\mu_1, ..., \mu_N)$ represents the random effects for the selected geographical units and are assumed to be $\text{IIN}(0, \sigma^2)$. λ is the scalar spatial auroregressive parameter and it's $|\lambda|<1$, W_N is the N × N spatial weights matrix with 0s diagonal elements and $I_n - \lambda W_N$ is non-singular. $v'_t = (v_{t1}, ..., v_{tN})$ with $v_{ti} \sim IIN(0, \sigma^2)$ and indipendent of μ_i.

In this paper we will not apply MLE as the number of cross-sectional units is large (>400), thus we will use a generalized method of moments, as suggested by Kapoor et al.(2007). This GM procedurehas the advantage of being less demanding with respect to LME is specified for T fixed and $N \rightarrow \infty$, where the error specification in (7) follows a first-order spatial autoregressive process:

$$u = \lambda(I_T \otimes W_N)u + \varepsilon \quad (8)$$

with:

$$\varepsilon = (\iota_T \otimes I_N)\mu + v \quad (9)$$

which allows also each geographical unit to be spatially correlated.
To carry out the analysis we use R package *splm* spgm estimation method (G. Millo, Piras, & Millo, 2013), which estimates λ through GM and the model coefficients by a feasible GLS estimator.

4. Preliminary Results

1. Moran's I is statistically significant and non-random for all variables considered under Monte-Carlo bootstrap permutation test, where the observations are randomly assigned, run for 999 simulations.

Table 1: Moran's I Monte Carlo bootstrap permutation test for selected variables and years. P-value <0.001.

Variable	Year	Moran's I MC
Lower Edu Men	1991	0.523
	2001	0.6102
Lower Edu Women	1991	0.55
	2001	0.6307
University Edu Men	1991	0.3386
	2001	0.4135
University Edu Women	1991	0.297
	2001	0.4344
Activity Rate Men	1991	0.6029
	2001	0.6265
Activity Rate Women	1991	0.6052
	2001	0.6749
Unemployment Men	1991	0.7819
	2001	0.7524
Unempl. Women	1991	0.7258
	2001	0.7954
Permanent Contract	1991	0.6281
	2001	0.7218

2. Clusters are significant and are present for all four time frames for all considered variables. Cluster analysis shows important clusters, which can vary across censuses changing mostly from North (Low-Low)-South(High-High) to East(Low-Low)-West(High-High) clusters, centered around the capital, Madrid. Indeed, from empirical analysis we can deduce that in those years 1986-2011, two important transformations took place:

a. From high fertility in rural areas to high fertility in urban areas, mostly due to the concentration of migrants in metropolitan areas after the mid 1990s;

b. massive female entrance into the labor force transformed the dynamics of childcare, thus in a first phase, female activity had a negative impact, while in a second phase this relationship reversed.

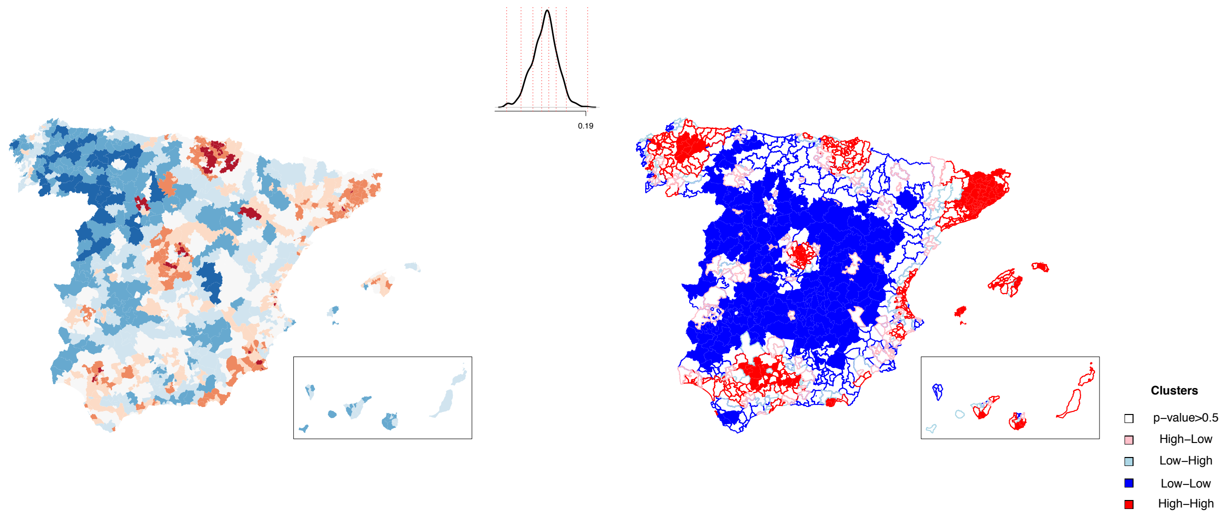
3. Preliminary analysis of cross sectional data underline the important role of labour indicators in explaining fertility changes. For instance female activity rate has a negative effect on fertility until the late 1990s, when women started entering the labor force in significative numbers. This effect becomes positive in the lastest period, in areas with considerable female participation to the labor force (North-East area), while for areas with higher unemployment and more traditional family roles, South, the effect stays negative. This multifaceted effect of activity rate is mirrored by men unemployment rates, which also negatively impact fertility in reacher areas with lower unemployment rates, but not in areas with high unemployment (South).

Table 2. Preliminary exploratory SAR model estimation for selected variables.

YEAR	VARIABLES	Model 1	Model 2	Model 3	Model 4	Model 5	YEAR	VARIABLES	Model 1	Model 2	Model 3	Model 4	Model 5
1986	**Intercept**	0.46***	0.142***	0.141***	0.12***	0.404***	**2001**	**Intercept**	0.284***	0.087***	0.11***	0.191***	0.218***
	Unempl. Men	-0.007***						**Unempl. Men**	-0.004***				
	Self Empl. Women		-0.001***					**Self Empl. Women**		0.002***			
	MAC				0.001			**MAC**				-0.003***	
	MAC First					-0.011***		**MAC First**					-0.004***
	Adj.r2	**0.41**	0.01	0.03	0	**0.14**		**Adj.r2**	**0.26**	**0.12**	**0.26**	0.02	0.06
1991	**Intercept**	0.463***	0.117***	0.123***	0.18***	0.381***	**2011**	**Intercept**	0.135***	0.097***	0.124***	0.123***	0.131***
	Unempl. Men	-0.008***						**Northing**	-0.001				
	Self Empl. Women		0					**Easting**		0.003***			
	MAC				-0.002*			**MAC**				0	
	MAC First					-0.01***		**MAC First**					-0.001
	Adj.r2	**0.53**	0	0.05	0	**0.14**		**Adj.r2**	0	**0.12**	**0.25**	0	0

*** p-value<0.0001

Map 1: Princeton Index and Female Activity Rate LISA cluster map (right), year 2001.



References

Adsera, A. (2004). Changing Fertility Rates in Developed Countries. The Impact of Labor Market Institutions. *Journal of Population Economics*, *17*(1), 17–43. doi:10.1007/s00148-003-0166-x

Anselin, L. (1988). *Spatial Econometrics: Methods and Models*. Kluwer Academic Publishers.

Anselin, L. (1995). Local Indicators of spatial association-LISA. *Geographical Analysis*, *27*, 93–115.

Anselin, L., Florax, R. J. G. M., & Rey, S. J. (2004). *Advances in Spatial Econometrics: Methodology, Tools and Applications*. Berlin.

Bocquet-Appel, J.-P., & Jakobi, L. (1996). Barriers to the spatial diffusion for the demographic transition in Western Europe. *Spatial Analysis of Biodemographic Data*, 117–129.

Bocquet-Appel, J.-P., & Jakobi, L. (1998). Evidence for a spatial diffusion of contraception at the onset ofthe fertility transition in victorian Britain. *Population*, (1), 181–204.

Cliff, A. D., & Ord, J. K. (1981). *Spatial Processes: Models and Applications.* London, England: Pion Ltd.

Driscoll, J., & Kraay, A. (1995). Spatial Correlations in Panel Data.

Galloway, P. R., Hammel, E. A., & Lee, R. D. (1994). FertilityDecline in Prussia,1875-1910: A Pooled Cross-SectionTime SeriesAnalysis. *Population Studies*, *48*(1), 135–158.

Kapoor, M., Kelejian, H. H., & Prucha, I. R. (2007). Panel data models with spatially correlated error components. *Journal of Econometrics*, *140*(1), 97–130. doi:10.1016/j.jeconom.2006.09.004

Klusener, S., Szoltysek, M., & Goldstein, J. R. (2012). Towards an Integrated Understanding of Demographic Change and its Spatio-Temporal Dimensions:Concepts, Data Needs and Case Studies. *Die Erde*, *143*(1-2), 75–104.

Millo, G., Piras, G., & Millo, M. G. (2013). Package "splm," 1–24.

Moran, P. A. P. (1950). Notes on continuous stochastic phenomena. *Biometrika*, *37*, 17–23.

Szołtysek, M., Gruber, S., Klüsener, S., & Goldstein, J. R. (2010). Spatial Variation in Household Structure in 19 th-Century Germany.