The determinants of the Spanish marriage market at the end of the XIX century: A spatial econometric approach

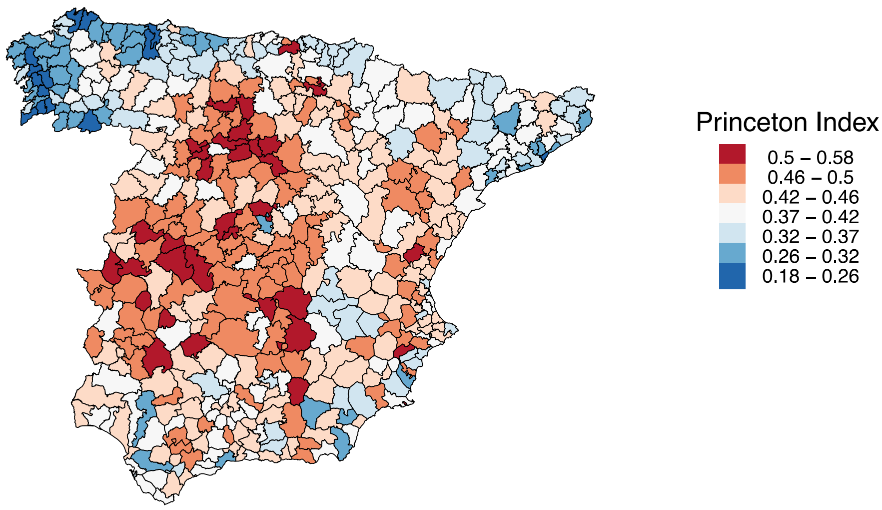
Abstract

The second half of 19th century was a time of great demographic changes in Spain, both in terms of mortality improvements and fertility decrease. However, such changes were far from homogeneous, as the Hibernian peninsula exhibited substantial diversity in demographic characteristics. The literature mostly concentrates on advancements in mortality and on economic determinants that lead to a fertility decline. However little is known on the delicate gender balance at local level, which led to female or male excess in Spain, and that, as a result, deeply impacted nuptiality and childbearing dynamics. The present study aims at providing a view of nuptiality and childbearing dynamics focusing on gender balance in Spain, employing data from the 1887 census for 467 juridical areas (comarcas) of mainland Spain. We employ a spatial-lag regression model to explain variations in fertility and nuptiality, focusing on variables that capture the imbalance in the sex structure, selective migration, celibacy as well as other socio-economic determinants.

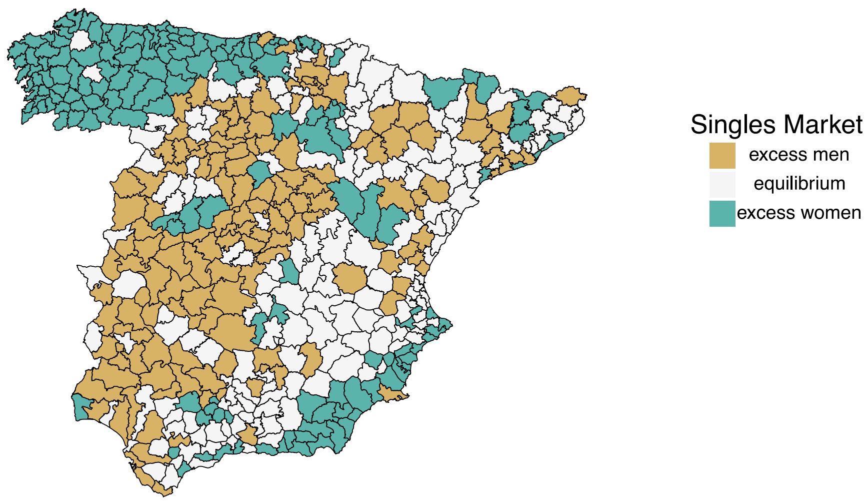
1. Data

The data used in this study come from the 1887 Spanish census, which has a rich collection of data with sub-provincial detail of 467 juridical areas (comarcas). The variables we employ describe the age and sex structure of the population, main demographic indicators (Princeton Indexes If and Ig, mortality), as well as the economic sector of employment, age at marriage, marriage market characteristics, and family structure (nuclear or extended).

Map1. Princeton Index, If, Spain 1887.



Map 2. Market of single men (21-35 years old) and women (16-30 years old)



2. Method and Preliminary Findings

2.1 Spatial autocorrelation and spatial lag model

The first step in spatial analysis is to build neighboring relations between geographical units, the 462 Spanish juridical districts.

Adjacency between regions can be defined in many ways. In this paper, First Order Rook adjacency is used to define neighboring relations between Spanish provinces, so that spatial units are considered neighbors if they share common borders but not vertices.

Once the spatial neighbors 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, establishing a row standardized matrix W_{ij}, where the diagonal of the matrix is by convention set to 0.

$$W_{ij} = \begin{bmatrix} 0 & \cdots & d_{1,52} \\ \vdots & \ddots & \vdots \\ d_{52,1} & \cdots & 0 \end{bmatrix}$$

A first exploratory measure to evaluate the strength of spatial patterns across the considered variables is Moran's I test (Cliff & Ord, 1970; Moran, 1950). In order to measure spatial autocorrelation, Moran's I index is required 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:

$$\mathrm{I} = \frac{n}{\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}} \frac{\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}\,(y_i - \overline{y})(y_j - \overline{y})}{\sum_{i=1}^{n}(\mathrm{y_i} - \overline{\mathrm{y}})^2} \quad (1)$$

where n is the number of spatial units i and j, y_i is the ith spatial unit, $\overline{y}$ 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 between -1 and 1, where – 1 represents strong negative autocorrelation, 0 no spatial autocorrelation and 1, strong positive spatial autocorrelation. Positive and statistically significant values of Moran's I for a given variable evidence spatial autocorrelation.

Moran's I test for spatial autocorrelation is a global measure of spatial autocorrelation, meaning that it is computed from the local relationships between the values observed for the geographical unit and its neighbors. 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 (Messner, Baller, Hawkins, Deane, & Tolnay, 1999). 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:

$$I_i = \frac{(y_i - \bar{y}) \sum_{j=1}^{n} w_{ij} (y_i - \bar{y})}{\frac{\sum_i^n (y_i - \bar{y})}{n}} \quad (2)$$

Where $\bar{y}$, the global mean, is assumed to be an adequate representation of the variable of interest y.

There are two main ways to model areal data. The first is known as Spatial Lag Model, which integrates the spatial dependence explicitly by adding a spatially lagged dependent variable, lag(y), on the right hand side of the regression equation. The main assumption of this model is that spatial neighbors of the dependent variable exercise a direct effect on the value of the independent variable.

The chosen method of analysis is the Spatial Lag Model, where spatial autocorrelation figures in the dependent variable. The Spatial Lag Model is defined as follows:

$$y_i = \beta_0 + \beta_1 x_{1i} + \cdots + \beta_k x_{ki} + \rho \sum_j w_{ij} y_i + \varepsilon_i \quad (3)$$

where y is the vector of error terms spatially weighted using the weight matrix W, ρ is the spatial lag coefficient and ε is a vector of uncorrelated error terms. If there is no spatial autocorrelation, ρ is equal to 0.

The Spatial Lag Model examines spatial autocorrelation between the dependent variable and its adjacent areas and handles spatial autocorrelation as a nuisance. Positive spatial error may reflect a mispecified model with omitted variables or spatial clusters. Ignoring spatial errors in the residuals might lead to biased coefficients and wrong standard errors or p-values.

We first run a geographically weighted linear regression model through means of OLS regression, defined as:

$$y_i = \beta_{0i} + \beta_{1i}x_{1i} + \cdots + \beta_{ki}x_{ki} + \varepsilon_i \ (4)$$

Where the coefficient is calculated through the weight matrix W:

$$\hat{\beta}_i = (X^T W_i X)^{-1} X^T W_i y \ (5)$$

In this phase we consider five model subsets, labeled A to E, where each one uses the set of fertility related indicators (SBFM, SNMB and MAC) plus a labor (general unemployment, female activity and female employment rates) or economic indicators (GDP or disposable household income) as independent variables.

After this step, in order to eliminate spatial autocorrelation we run the spatial error model for each of the five model subsets.

2.2 Preliminary Findings

Maps are a useful tool to visualize spatial patterns. Map.1 shows how the Princeton Index for total fertility, If, has clear spatial patterns of lowest-low fertility in the North-Eastern provinces (Galicia, Asturias and Cantabria), while the Southern and Eastern areas show higher fertility levels.

Map 2, summarizes the substantial diversity of the Spanish marriage market in the XIX century, where internal and international migration produced important imbalances in the sex ratios.

As mentioned earlier, Moran's I measures the global level of spatial correlation. Table 1 shows that all variables are spatially autocorrelated and significative.

Table 1: Moran's I test for selected variables.

Variable	Moran's I*	Variable	Moran's I*
Princeton Index If	0.59	Agriculture (%)	0.33
Princeton Index Ig	0.48	Industry (%)	0.39
Crude Mortality Rate	0.58	Industry women (%)	0.29
SMAM women	0.46	Industry men (%)	0.22
SMAM men	0.76	Servants women (%)	0.21
Celibacy women (%)	0.62	Servants men (%)	0.26
Celibacy men (%)	0.55	Literacy women (%)	0.76
Population increase	0.33	Literacy men (%)	0.82
Migratory balance	0.25	Urban population	0.30
Religious men (%)	0.54	Family size	0.59

*p-value<0.0001

We first investigate the effect of socio-economic determinants on the marriage market of single men and women (dependent variable), defined as the proportion single between 21-35 (men) and 16-30 (women) (therefore if the dependent variable takes a value >1, there is an excess men in the marriage market).

Table 2: Preliminary results of the lag model.

	Model 1	Model 2	Model 3
Intercept	3.35***		3.25***
SMAM	-3.10***		-2.29***
Migr. balance	0.10*		0.08*
Illiteracy men	-0.01		-0.002**
Female Servants	0.01*	-0.07*	0.002
If			0.003***
Agriculture		-0.2***	-0.005**
Industry (women)		-0.07*	-0.01
Clergy (men)		-0.01*	-0.001*
LM test	0.0053	1.12	2.0192
AIC	-511.04	-356.34	-572.25
Moran I statistic	-0.037	-0.079	-0.027

Preliminary results indicate the role of SMAM, singulate mean age at marriage (Hajnal 1953) for women. An increase in the number of single men in a region would substantially decrease the mean age at marriage of women.

This preliminary model, far from being the ultimate result, is an indication that regional diversity in marriage pattern are deeply spatial in nature (as indicated also by the LM test).

We intend to further investigate the social and economic determinants of the marriage market, as well as that of fertility, taking into consideration the family structure (such as the number of components and the gender composition of the family).

3. References

Cliff, A. D., & Ord, K. (1970). Spatial Autocorrelation: A Review of Existing and New Measures with Applications. *Economic Geography*, *46*, 269–292.

Hajnal, J. (1953). Age at Marriage and Proportions Marrying. Population Studies, 7(2), 111–136.

Messner, S. F., Baller, R. D., Hawkins, D. F., Deane, G., & Tolnay, S. E. (1999). The Spatial Petterning of Country Homicide Rates: An Application of Exploratory Spatial Data Analysis. *Journal of Quantitative Criminology*, *15*(4), 423–450.

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