Maza, Adolfo; Villaverde, José

Conference Paper

Migratory flows in Spain: a nonparametric and semiparametric approach


Provided in Cooperation with:
European Regional Science Association (ERSA)

Suggested Citation: Maza, Adolfo; Villaverde, José (2004) : Migratory flows in Spain: a nonparametric and semiparametric approach, 44th Congress of the European Regional Science Association: "Regions and Fiscal Federalism", 25th - 29th August 2004, Porto, Portugal, European Regional Science Association (ERSA), Louvain-la-Neuve

This Version is available at:
http://hdl.handle.net/10419/116955

Terms of use:
Documents in EconStor may be saved and copied for your personal and scholarly purposes.
You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.
If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.
MIGRATORY FLOWS IN SPAIN: A NONPARAMETRIC AND SEMIPARAMETRIC APPROACH

Adolfo Maza
José Villaverde

University of Cantabria (Spain)
E-mail: mazaaj@unican.es; villavej@unican.es

Abstract
Labour market is characterized in Spain by a high persistence in unemployment rates. One of the main reasons of this persistence is the lack of labour mobility. The present paper addresses this issue empirically and analyses the determinants of migration in Spain from a regional standpoint. We used a panel data set that includes annual bilateral migratory flows between the 17 Spanish regions from 1995 through 2000. For this purpose, after a descriptive analysis, we develop a nonparametric approach to show the factors that influence in the magnitude of migratory flows. Later on, semiparametric estimation techniques are applied to provide more econometric evidence regarding migratory flows.

Main conclusions are as follows: first, a high inertia in the migratory flows exists, that is to say, migratory movements are very persistent; second, migratory flows mainly respond, though weakly, to the differentials of wages, unemployment rates and house prices between regions; third, migratory flows are also affected, to a great extent, by non economic factors.

Keywords: migratory flows, regions, unemployment, wages.
JEL classification: J61, R23, C14
1. INTRODUCTION

One of the most worrying aspects of the Spanish economy in recent decades has been, and indeed continues to be, the deficient functioning of its labour market. In an economic context like the present, with Spain fully integrated in the European Monetary Union, income levels converging slowly but steadily towards the European average and a low inflation rate, the labour market is still a very interesting research topic. Although it is true that the situation has improved somewhat, it is still far from what would be desirable. The deficiencies in this market are many and various, although the persistence of high unemployment is without doubt one of its most worrying features.

In this article we analyse one of the reasons normally given to explain this persistence: the low level of interregional migration that exists in Spain. Different studies have already examined this question, either directly or indirectly (Ahn, Jimeno and Garcia, 2002; Bover and Arellano, 2002; Bover and Velilla, 2002; Antolín and Bover, 1997; Bentolila, 1997a).

This current paper is framed within the same line of analysis, its main contribution being its use of novel techniques for the study of migration. In particular we employ nonparametric and semiparametric estimation methods. In order to ensure homogeneity in the data series under analysis, our data (provided by FUNCAS, INE, IVIE-BANCAJA and the Development Ministry (Ministerio de Fomento)) only cover the period 1995-2000. Given the reduced timescale, the conclusions we come to must be treated with some caution, and only an extension of the series looked at would permit these conclusions to be confirmed or qualified.

---

1 An analysis of the situation of the labour market in Spain is carried out in López-Bazo, Barrio and Artís (2002) and Villaverde and Maza (2002). The persistence of the effects of a shock in the Spanish regions is addressed in Maza (2002) and Maza and Sánchez-Robles (2004).

INE: “Survey of residential variations”; IVIE and BANCAJA: “Human capital and Economic Activity”;
Development Ministry: “Statistical Bulletin”.
The remainder of this article is divided into four sections. In Section 2, we carry out a descriptive analysis of the current situation of internal migration in Spain. In Section 3, we identify and analyse the factors affecting interregional migratory flows. Extending the previous analysis, Section 4 proposes and estimates various regression equations that allow us to precisely identify the joint influence of these factors. As is customary, in the final section we outline our most significant conclusions.

2. INTERREGIONAL MIGRATION IN SPAIN: RECENT DEVELOPMENTS

It is well known that in the 1960’s and first half of the 1970’s migratory movements in Spain grew in strength; internal migration was very strong, contributing significantly to reducing regional inequalities in income levels and unemployment rates. In that period, the flows were generally in one direction – from poor to rich regions – consequently the net flows were very high.

For a decade following the mid 1970’s internal migratory flows slowed somewhat. However, subsequently interregional migration started to grow again, until in the 1990’s migration approached the levels last seen in the early 1960’s. Nevertheless, the pattern of these new migratory flows was totally different from that of earlier decades, and net migration was very low this time. As well as the traditional flows there were now flows from rich to poor regions and from regions of low unemployment to regions of very high unemployment. These migratory movements, in flagrant contradiction of economic theory, have become known as inverse migration.

In view of the above, it is instructive to look at developments over the past few years. A simple description of migratory flows during the period under analysis is shown in Figure 1, which presents, for each year, interregional migration rates. In the figure it is noticeable that the aforementioned rate falls in the first year, but from then on recovers, reaching 7.8 per 1000 in the year 2000.

\[ \text{Interregional Migration Rate} = \frac{\sum_{\text{total population}_{t-1}} \text{entries} - \sum_{\text{departures}}}{*1000} \]
Similarly, the new migration pattern is clearly shown in Figure 2. As can be seen, the internal migration is very balanced: most regions are close to the diagonal, which indicates that their net migration is close to zero. It is the objective of the rest of this article to more fully understand the factors provoking this type of migratory flows.

3. INTERREGIONAL MIGRATION: A NONPARAMETRIC ANALYSIS

Simplifying, we could say that the critical factor determining migration is the search for a higher quality of life. However, in this section we shall try to look more deeply into this question, carrying out an exhaustive analysis of the factors that lead to migratory flows according to economic theory. Among them we might mention the national unemployment rate, unemployment rate differentials between the regions, differences between the per capita GDPs, the cost of housing and educational levels. However, it is also worth remembering that there are other factors of significance -the cost of emigration, population density, climate, public policies, etc.- although their influence on migratory flows is difficult to measure.

One of the elements that significantly affects the size of interregional migratory flows according to economic theory is the level of national unemployment (Bentolila, 1997a). Indeed, a high unemployment rate in the country as a whole discourages the movement of people by diminishing the potential benefits of migration, since it makes it less likely they will find employment in their proposed destination region.

In order to determine the influence of this factor we carried out a nonparametric analysis revealing the sensitivity of the net interregional migration rate to the national unemployment rate. Specifically, we calculated the bi-dimensional nonparametric density function between both variables, computed using a Gaussian kernel with optimal bandwidth – following Silverman’s rule of thumb. The results obtained are reported on

\[ \text{Net Interregional Migration Rate} = \left( \frac{\text{Immigration}_i - \text{Emigration}_i}{\text{Population}_{i-1}} \right) \times 1000 \]
the left of Figure 3, which shows the net migration rate on the X-axis, the national unemployment rate for the previous period on the Y-axis, and the probability density for each point (X,Y) in the Z-axis. Likewise, on the right of Figure 3 is shown the contour plot, obtained by taking a cut parallel to the (X,Y) plane in the three dimensional graph and representing the distribution of the unemployment rate determined by a fixed rate of net migration. According to this, and given that the kernel (the probability mass) sits on the vertical, we can conclude that the net interregional migration rate seems to be independent of the national unemployment rate. This result could be explained by the fact that many of the workers who move between regions emigrate with a work contract, or that their main objective for moving is not to find employment, as Bentolila (1997b) points out.

Another factor behind the development of migration that is directly related with the labour market is the different unemployment rates between the regions (see, for example, Dickie and Gerking, 1998). Analysing this factor, it is important to distinguish between relative and absolute differences. Specifically, in terms of relative differences Figure 4, which is to be interpreted in the same way as Figure 3, shows that their influence on migratory movements is very limited. The same result is found for absolute differences. Likewise, it is clear that regional differences in unemployment rates are very marked, as can be seen from the values on the Y-axis of the contour plot. Finally, we also find a local (or second order) maximum that indicates that regions with a considerably higher unemployment rate than the national average show a negative net migration rate; in other words, it seems that only when the unemployment rate differentials are very significant does the migration follow the direction predicted by economic theory.

Obviously, migration depends not only on unemployment. Migratory movements can also be affected by, for example, the per capita GDPs of the different regions. Migratory flows occur, in principle, from regions with low income levels to regions with higher income levels. However, in practice there is empirical evidence of migratory flows in the reverse direction (Bentolila, 1997b). Thus, again applying nonparametric estimation techniques, Figure 5 shows that differing per capita GDP levels have little effect on
internal migration in Spain. In short, according to this analysis we cannot say that people change their regions of residence seeking higher income levels. However, this result will not be absolutely confirmed in the next section.

Another important factor for its effect on migratory flows is the cost of housing. In fact, one of the causes that may be behind the direction of internal migration in Spain is the poor functioning of the housing market, reflected in the high cost of housing compared to incomes. However, the empirical evidence is not conclusive on this factor either, as can be seen in the contour plot (Figure 6).

In addition to the variables mentioned above, the composition of the population by educational levels may also affect the size and direction of migratory flows, since it is accepted that it is generally the most educated individuals who are the most mobile (Bover and Arellano, 2002). Similarly, Mauro and Spilimbergo (1999) find that qualified people respond to drops in the demand for labour in their region by emigrating to other regions, while less qualified people either abandon the labour market or remain unemployed.

In this case, the results obtained in our estimation of a stochastic kernel between the net migration rate and the level of human capital (proportion of the population of working age with secondary or higher studies) do not agree closely with what we have just noted (Figure 7). Thus, initially it seems that human capital has a negligible effect on interregional migratory movement in Spain. We shall attempt to confirm this result, along with all of the others, in the next section.

In short, the analysis carried out in the previous paragraphs has considered some of the main causes behind migratory flows according to economic theory. However, to conclude this section we feel that we cannot ignore the influence of the net migration rate from the previous period; this variable combines, at least in part, the influence of non-economic factors. Thus, we need to carry out an analysis to determine the level of inertia in migratory flows.
Figure 8 shows without room for doubt that this variable has a significant effect on internal migration in Spain. Specifically, in the contour plot we can see that the lines are distributed along the positive diagonal, a clear sign that the persistence of migratory flows is very strong. This result makes it even more doubtful that the above-mentioned factors affect migration, since it seems that changes in the economic situation of a region do not affect its pattern of migration to a great extent; quite the opposite, this appears to persist in time.

4. INTERREGIONAL MIGRATION: A SEMIPARAMETRIC ANALYSIS

The results obtained in the previous section do not appear to provide much support for the conclusions founded on theory. With a view to studying this question in more detail we now analyse determinants of net migratory flows once more, but this time employing a different approach. We analyse the joint behaviour of the flows and some of the explanatory variables considered previously. The reason for changing our approach is that it may be that the various factors exert more influence combined than in isolation – indeed it appears that this is what happens in our case.

Parametric estimation techniques are traditionally employed to carry out this type of analysis. The main characteristic of this approach is that it considers that there is a known functional form (generally linear) between the explanatory variables and the dependent variable. However, there is often no apparent reason (either economic or otherwise) to assume that the relation is in fact of this type; quite the opposite: in many cases one can guess that the relation is nonlinear, or at least that the functional form linking the endogenous variable with the exogenous variables is unknown, as is the case here. Then it becomes necessary to use more flexible estimation techniques than the parametric method.

In view of this, the main innovation of the current study lies precisely in the technique of analysis it employs, which is a semiparametric estimation with panel data. This implies
the estimation of an equation in which no strong restrictions are imposed on the functional form of some of its components; it is simply assumed that it is a smooth function – i.e., continuous and with a certain degree of differentiability – whose form is unknown.

As its name implies, the semiparametric estimation consists of two elements: the first is estimated nonparametrically, while in the second a group of parameters is estimated. The general form of a model of this type is as follows:

\[ Y = \beta^T X + m(T) + \varepsilon \]

where \( X \) is the vector of explanatory variables that has a linear influence on the endogenous variable; \( \beta \) is the vector of parameters associated with those variables; \( m(T) \) is an unknown function of the vector \( T \), which represents the group of explanatory variables whose influence is – or might be – nonlinear; and \( \varepsilon \) is the error term, with \( E(\varepsilon / X, T) = 0 \) and \( V(\varepsilon / X, T) = \sigma^2 \).

The process of estimation carried out in this paper is based on that of Li and Stengos (1996), in which they combine semiparametric estimation techniques with the use of panel data. A detailed description of this process can be found in the Appendix.

Taking into account these considerations, and following the guidelines of Pissarides and McMaster (1990), we estimated various regression equations, introducing in all of them the variables mentioned in the previous section. Prior to carrying out this estimation, we built an origin-destination migration matrix; by working with the net interregional flows of each of the regions vis-à-vis the others we sought to gain in informational content and precision, following the example of Raymond and García (1996).

Thus, the equation which in principle seems to best reflect the situation of migration in Spain is the following:
\[
mr_{ij,t} = \alpha_i + m\left(\frac{u}{u_j}\right)_{t-1} + \beta_1\left(\frac{Y}{Y_j}\right)_{t-1} + \beta_2\left(\frac{H}{H_j}\right)_{t-1} + \beta_3 K_i + \epsilon_{ij,t}
\]  \hspace{1cm} (1)

where \(mr\) denotes the net migration rate; \(u\) the unemployment rate; \(Y\) the per capita GDP; \(H\) the cost of housing; \(K\) the stock of human capital; and the subindices \(i, j, t\) refer to region “i”, region “j” and time period “t”, respectively. It should be noted, as the equation specifies, that the nonparametric variable represents the differences between unemployment rates.

However, and given that the variable considered to be nonparametric in Equation (1) seems, according to the results obtained and which will be presented shortly, to be linearly related to the dependent variable, we opted to estimate this equation again but with an important difference: we associated a coefficient to the unemployment differentials variable, and we allowed the influence on each region’s net migration rate of the GDP differentials variable (which is, in this case, the nonparametric variable) to be nonlinear. In this way, the second equation is estimated as follows:

\[
mr_{ij,t} = \alpha_i + m\left(\frac{Y_i}{Y_j}\right)_{t-1} + \beta_1\left(\frac{u_i}{u_j}\right)_{t-1} + \beta_2\left(\frac{H_i}{H_j}\right)_{t-1} + \beta_3 K_i + \epsilon_{ij,t}
\]  \hspace{1cm} (2)

The results obtained in both equations are shown in Table 1; figures 9 and 10 present the variable considered nonparametric in each case. The most relevant conclusions from this analysis are as follows:

1. Unemployment rates do not appear to play an important role in determining migration (Figure 9). When this variable is estimated parametrically (Equation 2) the results confirm this impression and indicate that although unemployment rate differentials between the regions do exert a negative effect on net migration rates, as predicted by economic theory, this effect is weaker than expected (coefficient of \(-0.23\)). Thus, it appears that a high level of unemployment in the destination
region does discourage – although only moderately – migratory movements, since it diminishes the likelihood of finding work.

2. In contrast to what we noted in the previous section, differences in income levels do exert a certain influence on internal migration in Spain; this effect appears to be stronger than that of unemployment (Equation 1). Looking at this point more closely, the nonparametric analysis (Figure 10) provides new information and indicates that the effect is especially intense when the differences in GDP are very great (more than 50%). Only then does a higher per capita GDP act as a magnet for immigrants.

3. Another of the factors that appears to be behind net interregional migration in Spain is housing cost differentials; the coefficient associated with this variable is statistically significant in both Equation 1 and Equation 2. A high cost of housing in the destination region discourages migratory flows to it.

4. The level of human capital does not appear to exert an effect on net migratory flows. Although in Equation 1 the coefficient is significant, its value is very low, while in Equation 2 it does not differ statistically from zero.

5. Although not shown in the table for reasons of simplicity, the fixed effects of each region, which represent all those factors that differentiate them from other regions and which scarcely change over time, are in some cases statistically significant. This indicates that apart from the explanatory variables we have looked at, there are other determining factors of migratory movements, as we suspected in the previous section. The study of some of these factors is in our research agenda for the near future.

5. CONCLUSIONS

Starting from a descriptive analysis of interregional migration in Spain, which shows that net flows have been very low between 1995 and 2000, this paper has analysed the determinants of migration using both nonparametric as well as semiparametric techniques. The first of these points to the existence of a marked inertia in interregional migration as its most significant conclusion. Moreover, it shows that numerous factors
that according to theory should affect net flows actually do so much less than expected. This appears to indicate that along with the traditional economic factors there are other determining factors of migration that are non-economic in nature and whose influence is difficult to quantify.

Subsequently we estimated various regression equations using semiparametric techniques. In this case the results showed that the variable that affects migration most is the one representing differentials in income levels between the regions. Likewise, we found that differentials in unemployment and housing costs also appear to explain net migration rates, although with less power.

In view of the above, and as we suggested at the beginning of this study, we might ask if migratory flows can contribute to resolving the problems of the labour market in Spain, and particularly to reducing the persistently high unemployment rates. The results do not allow us to be very optimistic on this point, since they show that the influence of unemployment is very limited and that income levels only appear to be of relevance when the differences are very great. As we have said, alongside the traditional migratory movements there is substantial inverse migration in Spain, which points to the loss of importance of those factors that theory signals as determinants of migration. Only if the migratory flows were very high and only if they followed patterns predicted in economic theory would the movement of people help to improve the situation of the labour market in this country.

**APPENDIX: SEMIPARAMETRIC ESTIMATION PROCESS**

In this appendix we explain in general terms the method of estimation. In the first part of this paper we carried out a nonparametric estimation. This type of estimation came about from the conviction that traditional estimation methods tend to be badly specified. According to many experts the parametric approach is very restrictive, since it only allows freedom in the vector of parameters, which can distort the results. In contrast, the nonparametric models are aimed at obtaining much more flexible and robust forms, and
this is their main advantage. However, parametric methods permit a much simpler and
direct interpretation of the results than nonparametric ones. For this reason, in the
second part of this study we carried out a semiparametric estimation, a technique that
combines the best of the nonparametric and parametric methods: on the one hand it is
more flexible than parametric methods; and on the other, interpretation of its results is
simple and direct.

In the estimation process carried out we start from the following original model:
\[ Y = \beta^T X + m(T) + \varepsilon \]
Next, we take the conditional expectation to \( T = t \) and we obtain:
\[ E(Y/T = t) = \beta^T E(X/T = t) + m(T) \]
Subtracting this expression from the original model we get:
\[ Y - E(Y/T = t) = \beta^T (X - E(X/T = t)) + \varepsilon \]
or equivalently:
\[ \tilde{Y} = \beta^T \tilde{X} + \varepsilon \]
Finally, and with regards the nonparametric component, this can be expressed as
follows:
\[ m(T) = E(Y - \beta^T X / T = t) \]
In accordance with the above expressions, the stages we followed in practice in the
estimation process were as follows:

1. Estimate \( E(Y/T = t) \) and \( E(X/T = t) \) – for the \( p \) explanatory variables included in
the parametric part – with a nonparametric estimation method.
\[ E(Y/T = t) = h(T) \]
\[ E(X/T = t) = g(T) \]
2. With nonparametric estimations the following variables are generated:
\[ \tilde{X} = X - \hat{E}(X/T = t) \]
\[ \tilde{Y} = Y - \hat{E}(Y/T = t) \]
3. With these new variables the regression function \( \tilde{Y} = \beta' \tilde{X} + \epsilon \) is formed. Now it is possible to estimate the vector of parameters by ordinary least squares:

\[
\hat{\beta} = \left( \tilde{X}' \tilde{X} \right)^{-1} \tilde{X}' \tilde{Y}
\]

4. Having estimated the parameter \( \beta \), the following variable can be generated:

\[
\tilde{Y} = Y - \hat{\beta}' X
\]

5. Finally the equation \( \tilde{Y} = m(T) \) is considered, and \( m(T) \) is estimated using a nonparametric regression of \( \tilde{Y} \) on \( T \); the nonparametric estimator of the function \( m(T) \) is:

\[
\hat{m}(T) = \frac{1}{nh} \sum_{i=1}^{T} K \left( \frac{T - T_i}{h} \right) \hat{Y}_i
\]

where

\[
\hat{P}(T) = \frac{1}{nh} \sum_{i=1}^{T} K \left( \frac{T - T_i}{h} \right)
\]
REFERENCES


Figure 1
INTERREGIONAL MIGRATION RATE

Figure 2

Legend: and=Andalusia; ara=Aragón; ast=Asturias; bal=Balearic Islands; can=The Canary Islands; cant=Cantabria; cl=Castile-León; cm=Castile-La Mancha; cat=Catalonia; cv=Valencian C.; ext=Extremadura; gal=Galicia; mu=Murcia; nav=Navarre; pv=Basque Country; rio=La Rioja
Figure 3
STOCHASTIC KERNEL BETWEEN NET INTERREGIONAL MIGRATION RATE AND NATIONAL UNEMPLOYMENT RATE (t-1)

Figure 4
STOCHASTIC KERNEL BETWEEN NET INTERREGIONAL MIGRATION RATE AND RELATIVE DIFFERENCES IN UNEMPLOYMENT RATE (t-1)
Figure 5
STOCHASTIC KERNEL BETWEEN NET INTERREGIONAL MIGRATION RATE AND RELATIVE DIFFERENCES IN PER CAPITA GDP (t-1)

Figure 6
STOCHASTIC KERNEL BETWEEN NET INTERREGIONAL MIGRATION RATE AND RELATIVE DIFFERENCES IN HOUSE PRICES (t-1)
Figure 7
STOCHASTIC KERNEL BETWEEN NET INTERREGIONAL MIGRATION RATE AND THE LEVEL OF HUMAN CAPITAL (t-1)

Figure 8
STOCHASTIC KERNEL BETWEEN NET INTERREGIONAL MIGRATION RATE AND INTERREGIONAL MIGRATION RATE (t-1)
Table 1

<table>
<thead>
<tr>
<th>Dependent variable: $m_{ij,t}$</th>
<th>Equation 1</th>
<th></th>
<th>Equation 2</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>$\left(\frac{u_i}{u_j}\right)_{t-1}$</td>
<td>n.p.v.</td>
<td></td>
<td>-0.229*</td>
<td>-6.79</td>
</tr>
<tr>
<td>$\left(\frac{Y_i}{Y_j}\right)_{t-1}$</td>
<td>0.548*</td>
<td>7,62</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\left(\frac{H_i}{H_j}\right)_{t-1}$</td>
<td>-0.395*</td>
<td>-7.51</td>
<td>-0.188*</td>
<td>-4.88</td>
</tr>
<tr>
<td>$K_{i,t-1}$</td>
<td>0.011**</td>
<td>2.31</td>
<td>0.009</td>
<td>1.66</td>
</tr>
</tbody>
</table>

Notes:
- (*) Significant 99%; (**) Significant 95%.
- “n.p.v” denotes the nonparametric variable in each case.

Sources: INE, FUNCAS, IVIE, Ministerio de Fomento and own elaboration.
Figure 9

NONPARAMETRIC VARIABLE EQUATION-1

\[ m \left( \frac{u_i}{u_{i-1}} \right) \] -

Figure 10

NONPARAMETRIC VARIABLE EQUATION-2

\[ m \left( \frac{Y_i}{Y_{i-1}} \right) \] -