Abstract. This paper presents a novel method for modelling internal migration dynamics using non-stationary causative matrices. Its application to annual migration flows in Spain reveals changes in location preferences of migrants between 1986 and 2003. Our approach also indicates some instability in Spain’s migratory system prior to the late 1990s, as opposed to a negligible instability after that.

JEL classification: C61, O18, R23

Key words: Internal migration, non-stationary transition probabilities, non-stationary causative matrices, stability, Spain

1 Introduction

Research on internal migration in Spain has demonstrated important structural shifts in migration patterns since the early 1980s. For example, there has been a rapid increase in internal migration as the Spanish population widely dispersed itself in cities (that is, significant intra-regional movements), and the so-called ‘inverse migration’ pattern suggests that particular location-specific amenities affect migrant decisions (Antolin and Bover 1997; Bentolila 1997; Maza 2006).

Significant research to date has focused on the determining factors of migration (for example, see Antolin and Bover 1997; Maza and Villaverde 2004; Pons et al. 2007), but considerably less attention has been paid to the implications of migration (Raymond and García 1996; Basile and De Benedictis 2008) and the dynamics of migration patterns. The present work aims to contribute to our understanding of internal migration by addressing the dynamics of migration patterns. To this end, we present a flexible model of migration dynamics based on non-stationary causative matrices.

The migration data used in this paper is taken from Statistics of Residential Variations, as published by the Spanish National Institute of Statistics (INE). This archive consists of officially

* The author is grateful to the anonymous reviewers for their valuable comments. I would also like to thank my colleague Adolfo Maza for his review of the current paper and for many helpful and stimulating discussions on this topic.
registered migration data, specifically changes in the municipality of residence as submitted each year. The data consists of gross internal migration flows for the 17 Spanish regions between the years 1986 and 2003.

The paper is organised as follows. Section 2 provides a short overview of internal migration trends in Spain during the last few decades. Section 3 then presents a brief discussion on the theoretical Markov chain model of migration, paying special attention to the main stationary and non-stationary models explored previously in the literature. Section 4 then proposes a new non-stationary Markov chain model, and section 5 illustrates the results of applying this model to migration patterns in Spain between 1986 and 2003. Finally, section 6 presents our conclusions.

2 Internal migration in Spain: An overview

Internal migration patterns in Spain have changed drastically over the last few decades. In the early 1980s, the industrial restructuring associated with the economic recession after the first oil price shock was accompanied by a great deal of return migrations to poor regions in the south and southwest of Spain. The subsequent expansion of the welfare state in the late 1980s and the upward trend in housing costs from the early 1990s onwards exerted crucial influences on the evolution of new migration patterns in Spain.

Several characteristics should be noted for these trends. First, there has been a steady increase in internal migration. As can be seen from Table 1, the gross internal migration rate almost doubled, rising from 15.25‰ in 1986–1991 to 25.67‰ in 1995–2003 (both being periods of economic expansion).

Second, intra-regional movements have intensified in all regions since the 1980s (Bover and Arellano 2002). Although this migration has not received much attention to date, it nonetheless has been the main determinant of the rapid increase in internal migrations (see Table 1). In a manner similar to the experiences of other European countries such as Italy and Portugal (see for instance, Faini et al. 1997, Fachin 2007), the metropolitan expansion in many regions and rapidly increasing housing prices in large cities have led to an unforeseen rise in intra-regional migration to smaller towns. Additionally, Antolin and Bover (1997) suggest that the rise in employment opportunities in the service sector within large cities has helped to stop migrants from moving out of their respective regions.

Table 1. Gross internal migration rate in Spain (annual averages)

<table>
<thead>
<tr>
<th>Period</th>
<th>Total internal migration</th>
<th>Intra-regional migration</th>
<th>Inter-regional migration</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Number</td>
<td>Rate ‰</td>
<td>Number</td>
</tr>
<tr>
<td>1995–2003</td>
<td>1,038,479</td>
<td>25.67</td>
<td>718,400</td>
</tr>
</tbody>
</table>

1 This data source computes migration movements, not migrant persons. This fact may bias upwards the net regional migration rate, since some people (due to temporary contracts) may migrate throughout the year more than once. Notwithstanding that, it is likely that this fact does not involve a chronic situation in data due to the lack of incentives among most migrants to register all movements in the Civil Register.

2 Gender or age-disaggregated data in migration flows (migration data by region of origin and destination) would have been desirable in order to examine migration dynamics of heterogeneous subgroups of the population. Unfortunately, these data were not available for Spanish regions.

3 This rate is measured as gross migration flows as a proportion of total population.
Finally, the increasingly important role played by some location-specific amenities in affecting migrant decisions has led to some low-income regions becoming attractive destinations (Antolin and Bover 1997; Maza 2006). Coupled with the significance of return migrations, this seems to explain two important changes in the evolution of net regional migration rates: namely, the reversal in some Spanish regions from a net out-migration to a net in-migration (or vice-versa), and the dramatic drop in overall net migration rates.

3 Internal migration dynamics: Previous approaches based on Markov chains

In what follows, we model the internal migration dynamics using a Markov chain approach. This is appropriate since it makes a temporal system-wide approach to migration possible by estimating the so-called transition probabilities $p_{ij}$. These probabilities, organised in a transition matrix $P$, give the probability of moving out of state $i$ and into state $j$ during a particular time period. In our study, each state corresponds to a different origin or destination region.

Most papers following the above approach have implicitly assumed stationary, time-invariant transition probabilities (for example, see Magrini 1999, Hammond 2004, Ezcurra et al. 2005). Although this is usually not strictly valid, it has nonetheless become a common assumption since it is a necessary condition for the existence of a stationary distribution and because of computational simplicity. Given an observed sequence of origin and destination migration flow matrices at times $t = 0, 1, \ldots, n$, we have $M(t) = (m_{ij}(t): i, j \in S)$ where $S$ is the space of states and $m_{ij}$ represents the migration flow from origin region $i$ to destination region $j$. The most common methods for estimating stationary transition probabilities (‘static methods’) are given by:

\[
\hat{p}_{ij} = \frac{m_{ij}(n)}{m_i(n)}, \tag{1}
\]

\[
\hat{p}_{ij} = \frac{\sum_{t=0}^{n} m_{ij}(t)}{\sum_{t=0}^{n} m_i(t)}, \tag{2}
\]

and

\[
\hat{p}_{ij} = \frac{1}{n+1} \sum_{t=0}^{n} \frac{m_{ij}(t)}{m_i(t)}, \tag{3}
\]

where $m_i = \sum m_{ij}$.

The most used method is Equation (1), the ‘Markov matrix method’, which assumes that the most recent transition matrix is constant in time; thus, only current information is relevant in the future behaviour of transition probabilities. Among the approaches that base estimates on the entire sequence of observed flow matrices, it is worth mentioning that Equation (2), the ‘homogeneous matrix method’ (Anderson and Goodman 1957), pools observed migration flows and obtains constant transition probabilities by maximum likelihood estimation. Equation 3, the

---

4 The study by Maza (2006) obtained evidence about the influence of climate condition and housing price differentials in inter-regional migration during the period 1995–2002.

5 In a migration context, for instance, location preferences of migrants are quite likely to change over time.
‘average matrix method’ (Collins 1972), assumes that the average of the observed transition probabilities yields a decent picture of transition probability dynamics. On the other hand, non-stationary formulas (‘dynamic methods’) have received surprisingly little attention. Some of these approaches extend the temporal series analysis, such as the ‘linear tendency model’ (Rogerson 1979, Rogerson and Plane 1984) and the ‘average moving technique’ (Betancourt 1999). The most popular approach has been the ‘constant causative matrix model’ (CCM model) suggested by Lipstein (1965), which introduced ‘inter-regional dependency effects’ through a constant causative operator. Generally speaking, this means that apart from conditions influencing the transition probabilities between two specific regions, conditions influencing all other ‘competing’ regions are also taken into account (Plane and Rogerson 1986). Time-dependent transition probabilities for this model are given by:

\[ p_{ij}(t+1) = \sum_k p_{ik}(t) \cdot c^R_{kj}, \]  

or, alternatively,

\[ p_{ij}(t+1) = \sum_k c^L_{ik} \cdot p_{kj}(t), \]  

where \( c^R_{kj} \) and \( c^L_{ik} \) correspond to the elements of the right–and left-causative matrices \( C^R \) and \( C^L \). These elements gauge the rate of change of transition probabilities from both a ‘competing destination perspective’ (Equation 4) and a ‘competing origin perspective’ (Equation 5). Accordingly, as indicated by Equation (4), a transition probability tomorrow, \( p_{ij}(t + 1) \), is not only influenced by its value today, \( p_{ij}(t) \), but also by the current transition likelihood from region \( i \) to all the other ‘competing’ regions outside of destination \( j \) (where \( k \neq j \)). Alternatively, from an origin perspective (Equation 5), a transition probability tomorrow is influenced by both its value today and by the current likelihood of transition to region \( j \) from all the other ‘competing’ regions outside of origin region \( i \).

As indicated by Plane and Rogerson (1986), a column sum \( \sum_k c^R_{kj} \) that is greater (lower) than 1 may be interpreted as an increased (decreased) in the relative attractiveness of destination region \( j \). Similarly, a column sum \( \sum_i c^L_{ik} \) greater (lower) than 1 indicates a higher (lower) relative retention of migrants by origin region \( k \) (Plane and Rogerson 1986). The largest eigenvalue of the causative matrix, denoted by \( \lambda^* \), provides a scalar summary index of stability that describes the convergence behaviour of a migration system (Lipstein 1965). In particular, if \( \lambda^* > 1 \), then some disturbances are pushing the system toward disequilibrium. On the other hand, if \( \lambda^* \leq 1 \) and the remaining eigenvalues are lower than 1 as well, then the system is tending toward equilibrium. In this case, it is expected that eventually all regions will achieve stable migration shares from the system; indeed, the closer the second highest eigenvalue \( \lambda^{**} \) is to unity, then the higher the rate of convergence (Plane and Rogerson 1986). In a case of perfect stability, namely, stationary transition probabilities, we would have both \( \lambda^* \) and the remaining eigenvalues equal to unity.

---

6 The performance of these two last approaches is expected to be particularly bad the larger the period and the more complex the dynamics of migration flows are.
7 The common procedure used to estimate these parameters is ordinary least squares (OLS). For more details, see Dent (1973).
8 That is to say, when \( k = j \) in Equation (4).
9 In case of stationarity, all regions would undergo no change in relative attractiveness and retention and, subsequently, \( C^R \) and \( C^L \) would be equal to the identity matrix.
10 The right and left-causative matrices yield the same eigenvalues.
However, there is an important drawback with this model; namely, it assumes that the transition probabilities change at a constant rate. In fact, when we have a long temporal sequence of transition matrices and the value of $\lambda^*$ indicates that the system is under disequilibrium, it is unlikely that a constant causative matrix offers a good description of changes that have occurred to the transition probabilities. In such a case, the hypothesis of non-stationary causative matrices seems to be more realistic.

4 Internal migration dynamics: An alternative proposal

The model we propose here is a non-stationary causative matrix approach that extends the procedure of Lipstein (1965) by allowing the transition probabilities to vary at a variable rate. Given a sequence of transition matrices $\{P(t): t = 0, 1, \ldots, n\}$ with transition elements $p_{ij}(t)$, we define the relationship between present and future transition probabilities as

$$p_{ij}(t+1) = \sum_{k \in S} p_{ik}(t) \cdot c_{kj}^S(t,t+1),$$

or alternatively,

$$p_{ij}(t+1) = \sum_{k \in S} c_{kj}^S(t,t+1) \cdot p_{ij}(t),$$

for all $t = 0, 1, \ldots, n - 1$.

This approach is similar in spirit to the CCM model since it is dynamic in nature. However, the main difference is that changes in the transition probabilities are not necessarily constant, as indicated by the time-dependent notation of the non-stationary causative matrices.

When forecasting the next transition matrix in the sequence, namely, $P(n+1)$, this model assumes that only the most recent changes in a transition matrix have continuity in the following period. Thus, if the changes that occurred between $P(n-1)$ and $P(n)$ also operate in the next period (between times $n$ and $n+1$), then we can obtain a forecast for the matrix $P(n+1)$. This assumption once again distinguishes this approach from the CCM model, which assumes that the changes operating in $P(n)$ that affect $P(n+1)$ are constant changes between pairs of transition matrices over the sample period. Thus, according to the above premises, the forecast procedure of our model is given by:

$$\hat{p}_{ij}(n+1) = \sum_{k \in S} p_{ik}(n) \cdot c_{kj}^S(n-1,n),$$

or alternatively,

$$\hat{p}_{ij}(n+1) = \sum_{k \in S} c_{kj}^S(n-1,n) \cdot p_{ij}(n).$$

Applying this forecast formula generates a sequence of estimated transition matrices $\{\hat{P}(t): t = 2, 3, \ldots, n\}$, which then allows us to obtain two sequences of estimated non-stationary causative matrices: $\{\hat{C}^R(t-1, t): t = 3, 4, \ldots, n\}$ and $\{\hat{C}^L(t-1, t): t = 3, 4, \ldots, n\}$.

11 Thus, this procedure implies a lack of estimation for the first two transition matrices.
In this section, we test the robustness of the models discussed above for modelling the dynamics of internal migration within Spain between 1986 and 2003. We use annual data on gross migration flows divided by region of origin and destination.

First, the $c^2$-test of stationarity proposed by Anderson and Goodman (1957) is applied to the migration data (Table 2). The $c^2$-statistic shows that we can reject the hypothesis of stationarity for both the whole period 1986–2003 and for the 16 time periods of increasing duration (starting in the year 1986 and increasing by one year until the entire 1986–2003 period is obtained). This demonstrates that migrants’ preferences have changed over time.

Given this result, we would expect dynamic modelling techniques to provide a more precise description of migration patterns than static methods. Below we consider both static and dynamic approaches in order to gauge the accuracy of each method. We also consider three error statistics to quantify the differences between forecasting techniques: the root mean square error (RMSE),\(^{12}\) the mean absolute percentage error (MAPE), and the Theil coefficient (TC).

According to the results for the whole period 1986–2003 reported in Table 3, there are appreciable differences in the forecast accuracies of the tested models regardless of the statistic used. Our model provides the most accurate forecast over the period 1986–2003 while the methods that base predictions on the whole sequence of migration flows perform the worse; in particular, the CCM model and the Markov matrix method generate the least accurate forecasts. Thus, we find strong support for an influence of most recent changes in transition probabilities on the dynamics of internal migration flows. Indeed, as suggested by Maza and Villaverde (2004), there is some inertia in the dynamics of migration flows in Spain.

With our results on forecast accuracy in hand, we now examine the dynamics of Spanish migration using the non-stationary CM model. We use the estimates generated by the non-stationary causative matrix model for the period 1988–2003. The stability of the system can be judged by looking at the first and second highest eigenvalues of the estimated non-stationary causative matrices, as demonstrated in Figure 1. The evolution of $\lambda^*$ from 1988 to 2003 shows

\(^{12}\) Given the influence of data size on the magnitude of the RMSE, this error metric has been calculated taking transition probabilities expressed in percentages.

### Table 2. Anderson and Goodman’s $c^2$–test of stationarity

<table>
<thead>
<tr>
<th>Period</th>
<th>$\chi^2$</th>
<th>Prob.</th>
</tr>
</thead>
<tbody>
<tr>
<td>1986–1987</td>
<td>2,625.4</td>
<td>0.000</td>
</tr>
<tr>
<td>1986–1988</td>
<td>8,114.2</td>
<td>0.000</td>
</tr>
<tr>
<td>1986–1989</td>
<td>3,603.0</td>
<td>0.000</td>
</tr>
<tr>
<td>1986–1990</td>
<td>19,859.3</td>
<td>0.000</td>
</tr>
<tr>
<td>1986–1991</td>
<td>25,269.2</td>
<td>0.000</td>
</tr>
<tr>
<td>1986–1992</td>
<td>33,034.0</td>
<td>0.000</td>
</tr>
<tr>
<td>1986–1993</td>
<td>48,370.8</td>
<td>0.000</td>
</tr>
<tr>
<td>1986–1994</td>
<td>63,487.8</td>
<td>0.000</td>
</tr>
<tr>
<td>1986–1995</td>
<td>77,418.2</td>
<td>0.000</td>
</tr>
<tr>
<td>1986–1996</td>
<td>89,762.5</td>
<td>0.000</td>
</tr>
<tr>
<td>1986–1997</td>
<td>103,889.4</td>
<td>0.000</td>
</tr>
<tr>
<td>1986–1998</td>
<td>119,388.6</td>
<td>0.000</td>
</tr>
<tr>
<td>1986–1999</td>
<td>134,608.4</td>
<td>0.000</td>
</tr>
<tr>
<td>1986–2000</td>
<td>177,671.5</td>
<td>0.000</td>
</tr>
<tr>
<td>1986–2001</td>
<td>162,225.7</td>
<td>0.000</td>
</tr>
<tr>
<td>1986–2002</td>
<td>177,671.5</td>
<td>0.000</td>
</tr>
<tr>
<td>1986–2003</td>
<td>194,894.8</td>
<td>0.000</td>
</tr>
</tbody>
</table>
two distinct periods. Prior to 1997, the system fluctuated around a situation of persistent disequilibrium. From 1998 to 2003, the stability indicator reached values slightly higher than 1 but somewhat lower than in the previous period, indicating negligible instability. In addition, values for $\lambda^*$ from 1998 to 2003 were closer to 1, indicating a more rapid approach towards stability, that is, convergent behaviour.

Fig. 1. First and second maximum eigenvalues of the estimated non-stationary causative matrices.

With that in mind, and in order to check the robustness of our estimation results, we additionally present in Table 3 estimations for the sub-periods up to 1997 and from 1998 to 2003. As expected, the non-stationary CM model performs the best over the period 1986–1997. On the other hand, important changes in the ranking of models are observed across the sub-period 1998–2003. Thus, the best estimation corresponds to the CCM model, what could reflect constant changes in transition probabilities over the period 1998–2003, although of little significance according to the maximum eigenvalues. The Markov matrix model once again
performs the worst. With regard to the other methods (included the non-stationary CM model), differences in estimation accuracy are negligible. Thus, it should be pointed out that the model proposed here provides a good description of internal migration dynamics when instability is dominating.

Finally, in order to analyse changes in the relative attractiveness of different regions, dark-shaded areas in Figures 2a–2b represent Spanish regions with a positive column sum of the estimated right-causative matrices over the years 1992 to 1994 (Figure 2a) and 1995 to 2003\textsuperscript{13} (Figure 2b). We chose to compare these two time periods because of the relevant changes between them. During the economic crisis of 1992–1994, there was a decrease in relative attractiveness for regions under traditional focus of in-migration, such as Cataluña, Madrid and Comunidad Valenciana. Indeed, as Devillanova and García-Fontes (2004) point out, high unemployment rates during the recession of the early 1990s reduced inter-regional labour migration. On the other hand, low-income regions such as Galicia, Asturias, Cantabria, Castilla y León, Castilla-La Mancha and Murcia all increased in relative attractiveness, which was likely due to significant return migration. During the 1995–2003 period of economic prosperity (see Figure 2b), the preceding regional map changed drastically. Cataluña, Madrid and Comunidad Valenciana recovered their traditional roles as competitive attractors of migrants, along with new regional focuses of attraction such as Andalucía and Castilla-La Mancha which benefited from the increasing importance of both ‘inverse’ and return migration.

6 Concluding remarks

In this paper, we have looked at internal migration dynamics in Spain between 1986 and 2003. We began by describing new patterns of internal migrations in Spain. After reviewing the most widespread Markov chain models, we proposed and tested a new approach to modelling migration flow dynamics consisting of an extension of the constant causative matrix (CCM) model that addressed the non-stationary problem in causative matrices. This model permitted time-dependent changes in migrants’ transition probabilities and assumed that changes in these transition probabilities were most heavily influenced by changes in the immediately preceding time period.

The application of this new approach yields new insights into the dynamics of gross migration flows in Spain between 1986 and 2003. Our conclusions are stated as follows. First, the lack of stationarity, second, our non-stationary causative matrix model performed better than other static and dynamic forecasting techniques (including the CCM model) in modelling Spanish migration trends, implying that the recent changes in migration patterns seem to play a major role in influencing preferences of Spanish migrants. However, when considering the sub-periods 1986–1997 (period of instability) and 1998–2003 (a more stable period), estimations revealed that our model performed best when high instability was dominating, whereas the CCM method provided the best results when the instability was relatively low. Thus, a clear lesson is that model performance is strongly influenced by stability conditions. Third, some changes in the relative attractiveness of Spanish regions coincide with changes in relative economic conditions in Spain. Thus, where high-income regions such as Cataluña, Comunidad Valenciana and Madrid experienced a reduction in relative attractiveness during the economic recession period 1992–1994, these same regions demonstrated an increase in relative attractiveness during the economic expansion period of 1995–2003. Meanwhile, Andalucía and Castilla-La Mancha (both low-income regions) became attractive locations for migrants as well.

\textsuperscript{13} For this long period, dark-shaded areas correspond to regions with increased relative attractiveness over more than five of the nine years that comprise the period.
Fig. 2. Spanish regions with increased relative attractiveness over the periods: (a) 1992–1994 and (b) 1995–2003

Note: Dark-shaded areas correspond to regions with increased relative attractiveness.
which likely reflects the role played by location-specific amenities such as climate and other attributes (Cushing 1986; Montolio and Solé-Ollé 2009), and by return migration as well. And finally, notwithstanding the above conclusion, we observed a trend toward more stable migration dynamics since the late 1990s, coinciding with a period of high economic stability in Spain.

References


