An application of statistical methods of indirect estimation and projection of internal migration flows within the Portuguese mainland

Maria Filomena Mendes*, Sandra Lagarto*, Filipe Ribeiro*, António Caleiro*

Abstract: The study of migration flows is always problematic, essentially because there is not a systematic process of collecting background information. In the case of internal migration, restrictions on the available data are even more problematic and make it totally unfeasible to measure directly those movements. When the data on (regional) migration is incomplete, inadequate or unavailable, the estimation or quantification of regional migration flows are made possible by the application of indirect methods of estimation. Andrei Rogers, along with several others, developed and tested over several years methodologies that allow us to analyze and quantify indirectly the different behaviors of regional migration. These methodologies are applied in the paper, considering the case of Portugal.

1 Introduction

Mainly due to its temporary nature, migrations are always difficult to study, and the demographic lethargy that characterizes the mortality does not happen in migrations. Furthermore, the lack of register in changes of residence, especially at internal level, does not allow knowing the behavior of the migrants that could help to support the demographic projections, especially at regional level.

Maria Filomena Mendes
Universidade de Evora, Departamento de Sociologia, Largo dos Colegiais nº2, 7000 – 803, Evora, Portugal e-mail: mmendes@uevora.pt

Antonio Caleiro
Universidade de Evora, Portugal e-mail: caleiro@uevora.pt

Sandra Lagarto
Universidade de Evora, Portugal e-mail: sdlagarto@gmail.com

Filipe Ribeiro
Universidade de Evora, Portugal e-mail: flipjribeiro@hotmail.com

Demographic projections for the country and separate at regional level, or project for the regions and aggregate for the country. However, in one case or another, the estimation of internal migration is essential not only in demographic terms, but also in terms of planning, because information about the attractiveness of regions must be taken into account in the design of future, balanced and sustainable, development of regions and country.. Not taking the demographic lethargy as a scale of the scenery of the near 20 years (mortality without great variation and fertility with a great break down of 1980 for 2009), we still have to realize in which measure we can find a standard of internal migratory behaviors in function of gender and ages that could provide us a reasonable future visibility.

Also, the exchanges between regions are extremely important for the demographic projections, since in terms of migratory flows it is important to know from where the migrant that concentrate on a determined region are from, or up to knowing if the relation of exchanges between two regions has been remaining stable along the last decades.

In addition to the volumes of migration flows and its movements (origin/destiny), it is still crucial to know the causes leading to the decision of migrating of a region to another one, as well as defining a profile for ages that allows to us to know the ages where migrations trends are more evdient.

2 Context and data

Data for this paper were collected through the Census of Portuguese Population in1991 and 2001, and made available by the IPUMS International Database (Integrated Public Use Microdata Series)/INE (National Statistics Institute Portugal). Such as in most countries, the questionnaire used for the Portuguese census contained a question about the location (region) of residence at present time, and in a given period of time in the past (in this case, 5 years before the census).

We used the resident population by age group and gender, at the level of the geographical area of residence according to the Mainland Portuguese NUTS II (North, Center, Lisbon and Tejo Valley, Alentejo and Algarve) in 31 of December of 1985 and 1995, and resident population, by age groups and gender, according to the NUT II of residence of the mother at date of the birth. This analysis excluded those who living or born abroad (the latter were not considered relevant to the analysis of individuals in the age group 0-4 years old).

3 Method

A close observation of migration flows allows us to detect the existence of some specific features of its own, such as the differences in the odds of migrating associated with age, and a linear relationship between those at ages 0–4 and their parents.

Furthermore, given the existence of real observed values, the associated error predictions can be measured.

3.1 Age-specific regularities

We found that the higher odds occur early in adult life, when leaving home to attend a degree of higher education, join the military service, a family or simply get to work. This is a often result known to be called as ”labor peak” (Rogers et al., 2007). Moreover, the odds of migrating lower rates occur in late adolescence and usually after entering the labor market until the beginning of retirement.

As for the probability of migrating children, it appears that this reflects the migration of parents, usually young adults. Despite migration childhood reaches values higher than in adolescence, the retirement age, especially in developed countries leads to an increased likelihood of migration resulting in a ”return peak” close to 65 years old.

According to Rogers et al. (2007), the complete model of migration schedule has four components: (1) the pre-labor force stage (children), (2) the labor force (adults), (3) the post-labor force stage (elderly), and (4) a constant curve, which can be translated by the following expression:

$$m(x) = N_1(x) + N_2(x) + N_3(x) + c$$

$$m(x) = a_1 \exp(-\alpha 1 x + a_2 \exp -\alpha 2 ^{(} x - \mu 2 ^{)} - \exp[-\lambda 2 ^{(} x - \mu 2 ^{)]} + a_3 \exp -\alpha 3 (x - \mu 3) - \exp[-\lambda 3 (x - \mu 3)] + c \qquad (1)$$

where m(x) is the migration probability at age x; N_1 the pre-labor force stage (children); N_2 the labor force (adults); $N_3$the post-labor force stage (elderly); c the constant curve; α and μ, the parameters; and x the age.
Its implementation is based on the assumption that migrations of adults are linearly linked to migration of children between 0 and 4 years old, which provides the identification of tree key assumptions:

1. Regardless the size or intensity of migration flows and regions of origin/destination, the rates associated with migration have a very similar pattern when analyzed by age groups;
2. The question on region of birth is present in almost all censuses of population for children from the age group of 0-4 years old, and because it is a five year period is representative of a recent pattern in relation to migration;
3. As children migrate always (or mostly) with parents, young adults, their migration reflect, in a larger scale, the migration of other age groups.

3.2 The linear relationship

This method uses the technique of linear regression between the proportions of children aged 0-4 years old, which were born in region i and living in region j at the time of census, and the proportions of people in each age group that lived in region i five years before the census, and at time census is living in region j. This relation can explain the specific probability of migration, by age, from a parameter estimate based on information from the child migration.

The above-mentioned assumptions, that in similarity with the indirect estimates of mortality are based on a single infant mortality rate to estimate the mortality curve at all ages, result in estimates of the ”survival rates” of migrants in a given age x, represented by $S_{i\,j}(x)$:

$$Si\ j(x) = \frac{Migrants_{i\,j}(x)}{TotalPopulation_{i}(x)},\ x = age \quad (2)$$

In this sense, $Migrants_{i\,j}(x)$ denotes the number of individuals that at the time of the census are living in a location j, but 5 years before were living in i, and $TotalPopulation_{i}(x)$ represents the sum of all individuals, aged x, who lived in i,
5 years before the census date. In this way, $S_{i\,j}(x)$ is a measure that translates, for a person with age x, and lives in i, the probability of survival in j, t years later (in this case t = 5).

To estimate the specific survival rates for migration, we first take a ratio of child migration - $r_{i\,j}(x,-5)$, also known as ATI (age infant-to-migration ratio), which reflects the ratio of the probability of migrating at any age and probability of migration for children aged 0 to 4 years (i.e., all those born to 5 years before):

$$r_{i\,j}(x,-5) = \frac{S_{i\,j}(x)}{S_{i\,j}(-5)},\ x = 0, 5, 10, ..., 70+ \quad (3)$$

This ratio allows us to obtain estimates $S_{i\,j}(x)$ for 10 years later (the usual interval between censuses):

$$St\hat{}_{i\,j}(x) = rt_{i\,j}{-10}_{St}i\ j^{(}{-5)} \quad (4)$$

That results in an approach to a simple linear relationship type: $\hat{S}i\ j(x)=a+b\ Si\ j(-5)+\varepsilon$, where the estimated values $Si\ j(x)$ are explained in terms of $Si\ j(-5)$ through the regression line (4) and its associated error (ε).

3.3 Measuring error

Once all estimates are associated with a certain level of error and that we had access to the data from two censuses, it is convenient to use a measure of goodness-of-fit, like mean absolute percentage error (MAPE), to evaluate the results:

$$\mathrm{MAPE} = \frac{\sum_x \frac{\lVert \hat{S}_{ij}(x) - S_{ij}(x) \rVert}{S_{ij}(x)}}{N} \times 100 \text{ (for a particular flow)} \quad (5)$$

$$\mathrm{MAPE}_{ij} = \frac{\sum_{i=1}^{n} \sum_{j \neq i}^{n} \sum_x \frac{\lVert \hat{S}_{ij}(x) - S_{ij}(x) \rVert}{S_{ij}(x)}}{n(n-1)N} \times 100 \text{ (for all the flows)} \quad (6)$$

4 Analysis

Turning to examine the explanatory power of the estimates made by age groups (Table 1) revealed that in all groups the recorded R^2values very high. On the whole, the estimates have an explanatory power more than 90.0%, except for the age group 55-59, that only explains about 86.0% of the values actually observed.

However, these numbers only indicate the suitability of the model to each age group, and so, it is also essential to evaluate the associated error, which varies between 28.8% and 38.9%.

Note also that the method used has an explanatory power of 96.0% and an associated error of 31.7% for all the movements, or an analysis by region or by age group in Portugal (except islands).

Table 1 Regression Statistics for full samples by ages

Age Group	α	β	S.E.	R^2	MAPE(%)
0–4					
5–9	-0,00005	1,57279	0,06102	0,97	34,02
10–14	0,00030	1,40662	0,06579	0,96	31,35
15–19	-0,00033	1,57709	0,05386	0,98	29,88
20–24	-0,00145	1,73924	0,06062	0,98	30,15
25–29	-0,00158	1,73977	0,07231	0,97	32,21
30–34	-0,00066	1,78629	0,09220	0,95	38,85
35–39	0,00007	1,67198	0,06608	0,97	39,54
40–44	0,00025	1,53799	0,10456	0,92	37,51
45–49	0,00011	1,41712	0,08149	0,94	30,88
50–54	0,00014	1,40917	0,10944	0,90	35,27
55–59	0,00028	1,33062	0,12396	0,86	37,15
60–64	0,00040	1,11801	0,05143	0,96	36,20
65–69	0,00045	0,95372	0,07001	0,91	28,76
70 +	0,00018	1,23467	0,03602	0,98	33,48
Totals	-0,00039	1,64892	0,01892	0,96	31,68

A similar situation can be seen in the observation of table 2, where there is no overall explanatory power below 70.0%. Of relevance also are some differences with respect to the associated error, which vary between 13.3% and 49.6%.

Table 2 Observed and predicted flows, $R^2$and MAPE

Reg. 91	Reg. 01	Predicted	Observed	R^2	MAPE(%)
North	Center	8627	11669	0,96	21,22
North	L.V.T.	9990	14376	0,98	27,21
North	Alentejo	596	1015	0,94	38,50
North	Algarve	1427	2446	0,99	46,83
Center	North	6826	10247	0,98	29,06
Center	L.V.T.	12568	19933	0,98	32,28
Center	Alentejo	1037	1335	0,95	21,70
Center	Algarve	1236	2112	0,94	38,02
L.V.T.	North	10466	13788	0,82	23,59
L.V.T.	Center	18056	20644	0,88	13,32
L.V.T.	Alentejo	6666	11032	0,96	38,84
L.V.T.	Algarve	6201	8832	0,94	30,27
Alentejo	North	971	985	0,95	18,03
Alentejo	Center	902	1400	0,96	35,93
Alentejo	L.V.T.	7362	11421	0,99	3063
Alentejo	Algarve	19,22	3380	0,91	39,66
Algarve	North	689	1334	0,78	49,64
Algarve	Center	1025	1132	0,71	24,87
Algarve	L.V.T.	3802	5660	0,98	29,14
Algarve	Alentejo	945	1688	0,84	44,91

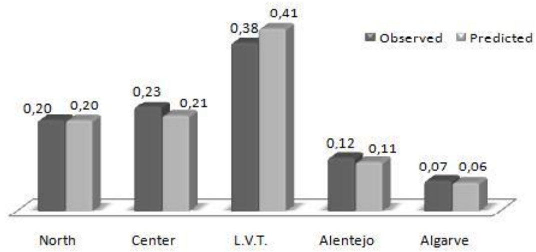
5 Results

Taking into account the total estimated and observed migration flows (Figure 1), we observe that they are mainly from the region of Lisbon and Tejo Valley, which contributes with approximately 40.0% of the total. Of relevance are the North and Center areas, with approximately 20.0% each, and finally the regions of Alentejo and Algarve with 12.0% and 7.0% respectively. These values are obviously related to the size of populations resident in each of the Portuguese NUTS II.

Considering now, the observed and predicted migration flows between all regions, it was found that, of all NUTS II, the region of Lisbon and Tejo Valley was the one that attracted more migrants, registering very close to 50.0% or more of the total (Figure 2).

In the case of Lisbon and Tejo Valley, we found that although about 40.0% of the migrants moving to the Center region, the distribution of these migrations flows occurred more evenly.

Fig. 1 Migrations flows by the out-coming region



At this point we tend to identify from the outset a migratory pattern, the geographical proximity, in that, firstly, the majority of registered movements have always had in common the same fate as the preferred region of Lisbon and Tejo Valley, and moreover, the second option, even in terms of preferential movement, were to the regions that are geographically closer.

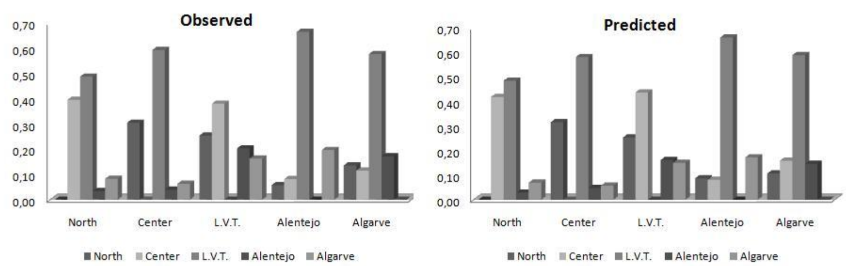


Fig. 2 Migration flows between all regions

By age, as can we see in figure 3, the results allowed to identify three distinct phases, where the first corresponds to the children had ages up to 9 years old; the second identifies individuals aged between 20 and 34 years old; and finally, the third, consisting of those aged 70 years old or even more.

The analysis of the differences between the estimates made and the observed values, shows that, despite the existence of a lag, denotes a good approximation to the behavior patterns actually recorded.

In any of the presented situations, the difference between the estimates and the actual values, results in an underestimation of the proportion of migration flows by age and regions.

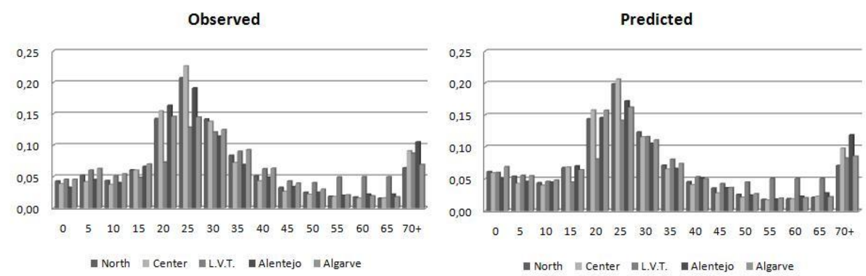


Fig. 3 Migrations flows by age

6 Conclusion

The main conclusion of this work is that it is possible identify a pattern of migration in Portugal, taking into account the economic attractiveness, based upon geographic proximity, in that, firstly, most movements were recorded having the most rich region of Lisbon and Tejo Valley as the preferred destination and, in the second place, that the nearby regions are also important.

According to this methodology, one can only indirectly estimate migration when the data is regular, which we assumed by considering that the migration observed from 2001 did not suffer from structural changes. Only based on this assumption was possible to determine the standards in relation to age structure of internal migration in Portugal.

Spatial analysis of the migration flows is one of the next steps, within the possible lines of investigation to follow.

Once that we treating an original approach that fits into a broader effort aimed at identifying the best methodology for estimation of internal migration within those using a recently developed indirect methods, another via for further examination in future work will deepen the application of indirect methods, including those linked to the work of A. Rogers and J. Raymer.

References

1. Little, J. S., Rogers, A.: What Can the Age Composition of a Population Tell Us About the Age Composition of Its Out-Migrants?. Population, Space and Place. 13(1), 23–39 (2007)
2. Raymer, J., Rogers, A.: Using Age and Spatial Flow Structures in the Indirect Estimation of Migration Streams. Demography. 44(2), 199–223 (2007)
3. Rogers, A.: Demographic Modeling of the Geography of Migration and Population: A Multiregional Perspective. Geographical Analysis. 40(3), 276–296 (2008)
4. Rogers, A., Jordan, L.: Estimating Migration Flows from Birthplace-specific Population Stocks of Infants. Geographical Analysis. 36(1), 38–53 (2004)
5. Rogers, A., Raymer, J., Jordan, L.: Inferring Migration Flows from Birthplace-Specific Population Stocks. In: Population Program, Institute of Behavioral Science, University of Colorado, Working Paper POP 2003-0002, 55p, Colorado (2003)
6. Rogers, A., Castro, L. J., Lea, M.: Model Migration Schedules: Three Alternative Linear Parameter Estimation Methods. Mathematical Population Studies. 12(1), 17–38 (2005)
7. Rogers, A., Jones, B., Partida, V., Muhidin, S.: Inferring Migration Flows from the Migration Propensities of Infants: Mexico and Indonesia. The Annals of Regional Science. 41, 443–465 (2007)