38ᵀᴴ CONGRESS OF THE EUROPEAN REGIONAL SCIENCE ASSOCIATION

28 AUGUST – 1 SEPTEMBER 1998 IN VIENNA

Mixed Markov models applied for population changes in Spain

Perez-Mayo, Jesus

Fajardo, Miguel

University of Extremadura

Spain

jperez@unex.es

Abstract

In this paper, common patterns of transition are tried to find in population. We look for chains in order to group Spanish provinces into clusters. Besides, income changes are analysed by the same procedure. Finally, we combine both studies to compare them Population and income data since 1960 to 1996 are collected form I.N.E. (National Spanish Institute of Statistics) databases.

1. Introduction

The aim of this paper is to study evolution of the population in Spain since 1960. This long period, more or less thirty years, has been one of the most important in Spanish history.

After some years of isolation, the Plan de Estabilización tried to modernise and liberalise Spanish economy in the late 50. This plan and the economic growth in some regions in the '60 caused large migration from the poorest regions to the richest, even abroad.

Besides, Spain became a democratic system in 1976 and joined the EU in 1986. As it can be seen, Spain has changed a lot for these years. Then, we tried to study an aspect of this global change: population evolution.

Moreover, since Spain has begun a decentralisation process of its public administration, we think the regional governments have to be very interested in this topic if they want to think on the long term.

The paper uses data collected from the INE Censuses of Population of 1960, 1970, 1981, 1991 and from the Census Actualisation of 1997.

2. Methodology

Suppose we have T repeated and consecutive measurements of one characteristic x, denoted by x^1, x^2,…, x^t,…, x^T, for instance we have a variable x, measured at T consecutive occasions (panel data). The scores of panel members are limited to a restricted number of states, J, on every occasion. Thus, variable x^t can have i, i=1,..., J, realisations. Instead of one indicator at several occasions, also multiple indicators for one or more latent variables can be analysed, if a Markov chain or a mixture of Markov chains relates these latent variables. Besides, this may be done for one population or for several subpopulations. Analysis with MMLC models can be focused on the types of change that exist in these subpopulations.

The general MMLC model assumes that each subject (province, in our case) belongs to one of one or more subpopulations. Membership of subpopulation h (h=1,..., H) is assume the same for all indicators. The proportion that belong to subpopulation h is

denoted γ_h. All other parameters that will be described below are considered conditional on subpopulation h, i.e. all (or some) parameters can be different in each subpopulation. Each member of subpopulation h can belong to one of one or more (latent or manifest) chains, i.e. parts of the (sub)population that have the same dynamics. A proportion $\pi_{s|h}$ of subpopulation h belong to chain s. Therefore, the proportion in subpopulation h and chain s is $\gamma_h \pi_{s|h}$ The chain is a sort of latent class.

A member of subpopulation h and chain s is assumed to belong to one of A classes. The proportion in class a (a=1,..., A) with latent variable 1 (latent variable measured at occasion 1), for subpopulation h and chain s is denoted $\delta_{a|sh}^{1}$. (4). Hence the proportion in subpopulation h, chain s and class a for latent variable 1 is $\gamma_h \pi_{s|h} \delta_{a|sh}^{1}$.

The probability to answer i for manifest variable 1 (manifest variable measured at occasion 1), given h, s and a, the response probability $\rho_{1|ash}^{11}$, is assumed the same for all subjects in subpopulation h, chain s and class a. Thus the proportion in subpopulation h, chain s, class a with latent variable 1, and category i for manifest variable 1 is $\gamma_h \pi_{s|h} \delta_{a|sh}^{1} \rho_{1|ash}^{11}$. If the variables of the model are manifest, not latent, the latent classes coincide with the manifest categories. Moreover, it is not necessary to have the response probabilities in that manifest model.

If the model does not only involve a latent variable al time 1, but also one at time 2, then, for subpopulation h, each member of chain s is assumed to behave according to the same transition probabilities, $\tau_{b|ash}^{21}$, from class a for time 1 to class b for time 2.

As for the time 1 manifest variable, also for the time 2 manifest variable the probability of answer j, given class b, chain s and subpopulation h, $\rho_{j|bsh}^{22}$, is assumed the same for all subjects in subpopulation h, chain s and class b. So, P_{hsaibj}, the proportion in subpopulation h, chain s, class a for latent variable 1, category for manifest variable 1, class b for latent variable 1 and category j for manifest variable 2, is $\gamma_h \pi_{s|h} \delta_{a|sh}^{1} \rho_{1|ash}^{11} \tau_{b|ash}^{21} \rho_{j|bsh}^{22}$ (1).

We obtain an expression for the observed population table, P_{hij}, of two manifest variables summing over the latent indices s, a and b. The model equation for the latent mixed Markov chain model for two time points is written as

$$P_{hij} = \gamma_h \sum_{s=1}^{S} \sum_{a=1}^{A} \sum_{b=1}^{B} \pi_{s|h} \, \delta_{a|sh}^{1} \, \rho_{i|ash}^{11} \, \tau_{b|ash}^{21} \, \rho_{j|bsh}^{22} \, . \, (2)$$

Several sorts of models may be considered as special cases of MMLC models.

1. The Latent Markov model (Wiggins, 1973) is a model with one manifest variable and one chain.

2. The Mixed Markov model (Poulsen, 1982) is a model 1 without response probabilities.

3. The Mover-Stayer (Blumen, Kogan and McCarthy, 1966) is a Mixed Markov model with two chains, one of which has probability 1 of no change.

4. Others

In our study, models 2 and 3 are going to be used.

Suppose that a sample of size n is available. Then, sample fractions $p_{ij...m}^{12...T}$ are available as estimates of $P_{ij...m}^{12...T}$. Therefore, we can obtain the likelihood of a given set of parameters under the model using formula (1) and (2).

The frequencies $np_{ij...m}^{12...T}$ in the observed T-way table are multinomially distributed with J^T-1 degrees of freedom.

The ML estimates are found by the EM algorithm. It is an iterative procedure for estimating latent or missing data when a specific model is assumed. We can use it for a discrete data model with a latent variable as Mixed Markov model. Each iteration consists of two steps.

In the E-step, we compute the Expectation of the full (T+1)-way table with cell proportions $\theta_{sij...m}^{y12...T}$, given a set of parameter estimates and the observed cell entries $p_{ij..m}^{12...T}$.

The M-step is the maximisation of an auxiliary log-likelihood of the parameters, which would apply if the full table with cell proportions $\theta_{sij...m}^{y12...T}$ would be observed in a sample.

The log-likelihood of the observed table is maximised alternating E-steps and M-steps. If we observed the full table, the problem would be very easy. Therefore, we build an auxiliary table (E-step) and solve the problem (M-step). Then we build a better auxiliary table and solve the problem again. The procedure must be repeated until no further improvement is possible.

We started the EM algorithm with an E-step, using some set of plausible initial parameter estimates. In the first E-step, formula (1) is applied to the present set of parameter estimates.

$$\hat{\underline{\xi}}_{sij...m}^{y12...T} = \underline{\pi}_s^y \, \underline{\delta}_{si}^{y1} \underline{\tau}_{sij}^{y12} ... \underline{\tau}_{slm}^{yT-1T}$$

We denote the result by ξ and not by θ, because the observed table was not yet taken into account. The subscript _ denoted that, before convergence is reached, we do not have the ML estimates yet. Then, the expectation of the full table, given the present parameters and the observed table is found by the following correction:

$$E\left[\theta_{sij...m}^{y12...T} \mid \hat{\underline{\pi}}_s^y, \hat{\underline{\delta}}_{si}^{y1} ..., \underline{\tau}_{slm}^{yT-1T}, p_{ij...m}^{12...T}\right] = \hat{\underline{\xi}}_{sij...m}^{y12...T} \frac{p_{ij...m}^{12...T}}{\sum_s \hat{\underline{\xi}}_{sij...m}^{y12...T}} = \hat{\underline{\theta}}_{sij...m}^{y12...T}$$

The proportions $\hat{\underline{\xi}}_{sij...m}^{y12...T}$ will be the same as the ML proportions $\hat{\underline{\theta}}_{sij...m}^{y12...T}$ after convergence of EM algorithm.

Since we have provisional estimates $\hat{\underline{\theta}}_{sij...m}^{y12...T}$ of cell proportions $\theta_{sij...m}^{y12...T}$ available, we can found maximum likelihood estimated based on the full table easily. The auxiliary log-likelihood for a given set of parameters has this form:

$$\ln(L) = \sum_s \sum_i \sum_j ... \sum_m n\hat{\underline{\theta}}_{sij...m}^{y12...T} \ln\left[\hat{\underline{\pi}}_s^y \, \hat{\underline{\delta}}_{si}^{y1} \, \hat{\underline{\tau}}_{sij}^{y12} ... \hat{\underline{\tau}}_{slm}^{yT-1T}\right].$$

Thus, the parameter estimates of the M-step are in formula (7):

$$\hat{\underline{\underline{\pi}}}_s^y = \frac{\hat{\underline{\theta}}_{s+++++}^{y12...T}}{\hat{\underline{\theta}}_{++++++}^{y12...T}}$$

$$\hat{\underline{\underline{\delta}}}_{si}^{y1} = \frac{\hat{\underline{\theta}}_{si+++}^{y12...T}}{\hat{\underline{\theta}}_{s++++}^{y12...T}}$$

$$\hat{\underline{\underline{\tau}}}_{sij}^{y12} = \hat{\underline{\underline{\tau}}}_{sij}^{y23} = \cdots = \hat{\underline{\underline{\tau}}}_{s\ i\ j}^{yT-1T} = \frac{\hat{\underline{\theta}}_{sij++++}^{y123\cdots T} + \hat{\underline{\theta}}_{s+ij++}^{y123\cdots T} + \cdots + \hat{\underline{\theta}}_{s++++i\ j}^{y12\cdots T-1T}}{\hat{\underline{\theta}}_{si+++++}^{y123\cdots T} + \hat{\underline{\theta}}_{s+i+++}^{y123\cdots T} + \cdots + \hat{\underline{\theta}}_{s++++i\ +}^{y12\cdots T-1T}}$$

The subscript + denotes a summation over the dimension concerned and the subscript = denotes that they are new parameter estimates.

Parameters with a – relate to iteration q and those with a = relate to iteration q+1.

It is possible to find formula 7 from formula 6 setting the derivative of the log-likelihood with respect to each parameter to zero. Formulas 7b and 7c can also be found taking the full table conditional on latent variable y. From formula 6, it is possible to see that the auxiliary log-likelihood is separable on s. Therefore, ML estimates for the parameters of a single chain may be used. If we leave out the sub and superscripts on the latent variable y, formulas 7b and 7c are the ML estimates for the parameters of a single Markov chain.

The EM algorithm for the Mixed Markov model can be started choosing random values in the 0-1 range, under the restriction that $\sum_{c} \hat{\underline{\pi}}_{s}^{y} = 1$, $\sum_{i} \hat{\underline{\delta}}_{si}^{y1} = 1$ and $\sum \hat{\underline{\tau}}_{si\ j}^{ytt+1} = 1$. However, if a parameter is started as 0 or 1, it will remain fixed. In this way, it fixed the stayer chain in the Mover-Stayer model.

The iteration process must be continued until the log-likelihood is increasing less than a small number, for instance 10^{-6}. Even if iterations are carried on very long, we can find a local optimum, especially in Mixed Markov models. For these models, several sets of starting values should be used.

When the number of degrees of freedom is larger than 0 the model can be tested against the saturated model by means of a likelihood ratio test,

$$LR = 2n\sum_{i}\sum_{j}\cdots\sum_{m} p_{ij\cdots m}^{12\cdots T} \ln p_{ij\cdots m}^{12\cdots T} - 2n\sum_{i}\sum_{j}\cdots\sum_{m} p_{ij\cdots m}^{12\cdots T} \ln \theta_{+ij\cdots m}^{y12\cdots T}$$

For large samples, the LR is χ^2-distributed with df degrees of freedom. Under a stationary Mixed Markov model, df is equal to the number of cells in the observed table minus the number of independent parameters.

$$df = J^T - SJ^2 + d_{MS}J(J-1)$$

Here d_{MS} denotes a dummy variable with value 1 if the model is a Mover-Stayer model and with value 0 otherwise. Other parameter restrictions increase the df in a similar way. We can test a more restrictive model (ii) against a less restrictive model (i) with the statistic $LR_{ii}-LR_{i}$, which is asymptotically χ^2-distributed with $df_{ii}-df_{i}$ degrees of freedom.

3. Application

3.1 Introduction

We will study how population has changed in Spanish provinces since 1960. More exactly, we are going to analyse growth rates.

As we have measured the same variable in different times, we are going to work with panel data.

Besides, growth rates values will be categorised in a such way that we will have a 3-classes discrete variable:

- Class 1: negative growth rate

- Class 2: positive but lesser than national growth rate

- Class 3: positive and bigger than national growth rate

Since the national growth rate has been positive for all the years, it is not necessary to distinguish between lesser and bigger then national negative rates.

We will assume that the probability to belong to class i at time t depends only on class j at time t-1 in this discrete-time discrete-space panel data. Hence, we will assume that changes can be explained by conditional probabilities of belonging to class i at time t given class j at time t-1. Therefore, this dynamic process does not take into account its history and, at first sight, it seems to be little useful.

However, the main assumption to be considered is that transition probability matrix is a constant one, and so, we assume a stationary process.

Hence, this model is very useful: it allows to predict which class an individual can belong. This predictive power is its most important feature.

3.2 Evolution 1960-1991

Since the analysis has three waves, the data are collected in a 3x3x3 table.

We will apply the Mixed Markov model to these data. This model estimates transition probability matrix and, besides, allows to think that subjects are heterogeneous, that is, there are different patterns of change.

It can be seen that population data from 1960 until 1991 holds a 1-chain model. Hence, evolution of all provinces can be explained by only one transition matrix. This model has the best goodness of fit (probability 92.52%)

Table 1. The proportion belonging to each chain ($\hat{\pi}_s$), the initial proportions in category i for chain s ($\hat{\delta}_{i|s}$), the transition probabilities $\hat{\tau}^{21}_{j|is}$ between time _t_ and time _t+1_. The 1-chain model

	$\hat{\pi}_s$	$\hat{\delta}_{i|s}$	$\hat{\tau}^{21}_{j|is}$		
			1	2	3
		1			
1		0.440	0.649	0.324	0.027
2		0.220	0.172	0.517	0.310
3		0.340	0.029	0.265	0.706

Source: Panmark 3

Provinces show a great stability. Main diagonal cells have larger than 50% probabilities, and so, more than a half of provinces did not change their trends.

3.3 Evolution 1960-1997

We are going to take into account an additional wave: 1991-1997. Hence, the growth rates we have to analyse are 60/70, 70/81, 81/91, 91/97. Since the last wave is shorter, we decide to use the average growth rate, r, $P_{i+n} = P_i(1+r)^n$, being P_i population at time i and n the final year of the period.

Therefore, we use a measure that is not affected by different sizes of waves.

Table 2. Goodness of fit, with degrees of freedom

	χ^2	g.l.	Prob.
1 Markov chain	84.31947	72	0.15194
2 Markov chains, one of which is a stayer chain	84.31947	72	0.15194
2 Markov chains	59.23972	67	0.73890
3 Markov chains	52.83807	65	0.86039

Source: Panmark 3

The 3-chain model fits the data very well; however, it is necessary to comment some features of the matrix (contingency table) before choosing this model.

It is a sparse matrix because there are more cells than subjects are. Besides, the model has 26 parameters to estimate. Both facts cause that asymptotic χ^2 can show a good fit while we actually do not know if it is so good.

Therefore, we use a bootstrapping measure of fit quality. (See Appendix A)

In the foollowing table, we show bootsretap probabilites of three meausres: G^2, Reand-Cressie statistic and χ^2.

Table 3. Bootstrp measures of goodness of fit

	$\alpha(G^2)$	$\alpha(RC)$	$\alpha(X^2)$
1Markov chain	0.0414	0.0950	0.1796
2 Mover-Stayter Markov cahins	0.0150	0.0654	0.1652
2 Markov chains	0.0242	0.0618	0.1282
3 Markov chains	0.0028	0.0128	0.0492

Sourse: Panmark 3

According to bootstrap probabilities, we must reject the 3-chains model while the rest of models may be accepted.

Thus, we choose the 2-chains model and so we assume that Spanish provinces can be grouped in two chains according to their patterns of change.

According to this model, we expect that a province with a negative growth rate at time t keeps losing population at time t+1 whatever chain it belonged.

On the other hand, depending on the chain provinces belong, we expect a growth acceleration process (chain 1) or a deceleration process (chain 2), even a loss of population.

Finally, we conclude that it is expected that provinces that lose population keep their trend and the rest can be divided in two groups with opposite behaviours.

Effects of heterogeneity can be analysed if 1-chain model and 2-chain model are compared. Both models estimate very similar transition probabilities for class 1. That is, according to both, it ids very likely that provinces that lose population keep doing it.

Heterogeneity actually affects to provinces with positive rates at time t.

Table 4. The proportion belonging to each chain ($\hat{\pi}_s$), the initial proportions in category i for chain s ($\hat{\delta}_{i|s}$), the transition probabilities $\hat{\tau}_{j|is}^{21}$ between time _t_ and time _t+1_. The 2-chains model

	$\hat{\pi}_s$	$\hat{\delta}_{i|s}$	$\hat{\tau}_{j|is}^{21}$		
			1	2	3
Chian 1	0.677				
1		0.442	0.647	0.262	0.091
2		0.236	0.000	0.269	0.731
3		0.322	0.164	0.070	0.766
Cahin 2	0.323				
1		0.436	0.771	0.229	0.000
2		0.186	0.619	0.381	0.000
3		0.378	0.000	0.722	0.278

Source: Panmark 3

3.3 Comparison 60-91 / 60-97

The best model to the first period uses only one matrix to explain expected evolution while the best for 60-97 uses two.

The first difference we can fin between both models is goodness of fit because 3-waves model has a probability of 92.52% to be held by the data and the probability of the 4-waves model is 73.89%. That disparity is because there are more parameters to estimate and more cells in the latter model.

Since both models are different, we may argue that increase of information (an additional wave) shows heterogeneity in provinces with positive rates of growth.

Whereas we said that it is more likely for a province to keep its trend in the 3-waves model, we expect two different behaviours: high stability of extreme categories and increasing growth rate of medium class and, on the other hand, stability for provinces that lose population and decreasing growth rates for the others.

This discrepancy is provoked by two simultaneous facts:

- Population growth in provinces where population had been decreased.

- Decrease of national rate of growth makes growth faster in relative terms.

Therefore, it would be very interesting to repeat this analysis some years later to see if these changes of trends run on the long term.

Studying transition matrices in both models, we may wonder: if both matrices show the same behaviour for provinces in class 1, can we assume that there is a chain formed by provinces that lose population? Can we explain the change of these provinces by means of an unique transition matrix? This is one of the outcomes of the stationary 3-chains model.

Unfortunately, we must remember that this model was rejected according to the bootstrapping measure of fit quality. Hence, it is not possible to use this model and we assume that provinces in class 1 belong to both chains.

3.4 Geographical distribution

Once the model that explains the evolution has been chosen, it is possible to performed the chains in a map. It shows that provinces in the middle of Spain belong to both chains. They lost population in the great countryside-urban area migration.

Figure 1. Map of chians of provinces.
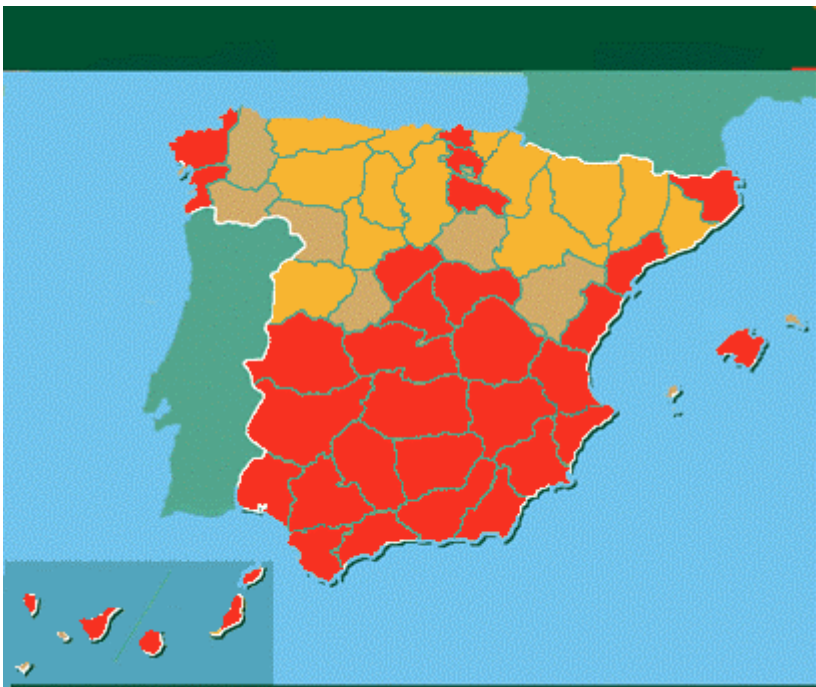
Red colour means belonging to chanin 1, yaellow colour cahin 2 and brown colour uncertainty.

Besides, Mediterranean coastal provinces, except Barcelona, belong to a chain that estimate populations increase if the province is not in class 1 and on the other hand, we can expect than northern provinces, but Coruña and Pontevedra, loses population.

In sum, provinces with a highly specialised economy in agriculture and heavy industry are expected to lose population while those specialised in services are more likely to increase it.

3.5 Trends vs. relative position

Finally, it is interesting to compare both topics. We have analysed evolution trends but we have not known yet if there has been a radical change in distribution of population in Spain. That is why we are going to study the evolution of relative positions.

It is not necessary to estimate a model. Data from the table show that a large change has not happened. On one hand, the less populated provinces are the same every year and on the other, the most populated provinces are the same every year. Hence, evolution of trends is not strong enough to change distribution of population in Spain.

4. Concluding Remarks

After doing this study, we may say that:

- Provinces where population decreases are expected to continue this process. Besides, this process is more likely in the central provinces.
- When the study is carried out until 1997, we can see that there are two different trends for the behaviour of provinces where population increases.
- These different trends consist of an acceleration of increase and a deceleration of growth.
- The coastal provinces are expected, in general, to have a population growth in the Mediterranean. We think it is caused by the importance of tourism in their economies.

- On the other hand, northern coastal provinces are more likely to lose population. This fact can be provoked because they are economies specialised in mining and heavy industry.
- Although a change of trends is seen, the analysis of relative positions shows that it has not done an important change in Spanish population.
- We think a second step looking for the variables involved in population evolution can be interesting.

Appendix A

With sparse data the observed cross-table has too many cells in relation to the number of observations to use the theoretical ji-square distribution which only holds asymptotically. Besides, the choice of a theoretical distribution is difficult when the model expects probabilities zero.

These problems can be solved by of the fit measure used using bootstrap methods.

In this case, the used bootstrap method is the Monte Carlo bootstrap procedure and it has five steps.

1. Assume that the model is true.

2. Treat the fitted proportions $\hat{P}$ under the model as population proportions.

3. Draw samples from this multinomial distribution with known parameters.

4. Estimate the same model for these samples and assess the G2 for each sample.

5. Reject he model if the proportion α of bootsatrap-G^2's that are larger than the original G^2, is very small.

Bibliography

- Langeheine, R., Van de Pol, F., (1994) *Discrete-time mixed Markov latent class models*. Dale, A., Davies, R. eds: Analyzing social and political change: A casebook of methods. Sage. Londres.

- Markus, G. (1979) Analyzing panel data. Sage. Londres.

- Poulsen, C.S. (1982) *Latent structure analysis with choice modelling applications*. Ph. D. Dissertation. Wharton School. Pennsylvania Univ.

- Van de Pol, F., Langeheine, R. (1989) *Mover-stayer models, mixed Markov models and the EM algorithm; with an application to labour market data from the Netherlands Socio Economic Panel*. Coppi, R., Bolasco, S. eds. Multiway data analysis. North Holland. Amsterdam.

- Van de Pol, F., Langeheine, R. (1990) M*ixed Markov latent class models*. Clogg, C.C., ed. Sociological Methodology. Blackwell. Oxford.