

MÈTODE PER A L'ESTIMACIÓ DE MIGRACIONS EN ÀREES PETITES A PARTIR DE DADES INCOMPLETES

J.A. SÁNCHEZ CEPEDA

Institut d'Estadística de Catalunya*

La disposició de dades de migració per a àrees petites s'emmarca dins el projecte d'estimacions de població, consistent en determinar la població recent de Catalunya per comarques i municipis majors de 45.000 habitants, desagregada per sexe i edat. La no disposició de dades amb aquest nivell de desagregació per alguns anys ha portat a la implementació d'un mètode per a la seva estimació.

Aquest treball descriu el mètode d'estimació dels efectius migratoris per àrea geogràfica, sexe i edat que s'aplica a l'Institut d'Estadística de Catalunya (Idescat). El mètode consisteix en calcular per separat la intensitat i el perfil tant de l'emigració com de la immigració. El principal problema en l'estimació del nivell o intensitat migratòria és haver de treballar amb àrees menors de 45.000 habitants. En el càlcul del perfil migratori es redueix el nombre de perfils dels 62 inicials a 6, que constitueixen els patrons migratoris observats a Catalunya. Es comenta la tipologia dels patrons migratoris, en base als paràmetres dels mateixos determinats pel model. Finalment, es calcula el nombre d'emigrants i immigrants a partir de la intensitat i el perfil migratori estimats.

Method for estimating small area migration from incomplete data.

Paraules clau: Àrea petita, model migratori, anàlisi cluster, ajust no lineal.

* Institut d'Estadística de Catalunya. Via Laietana, 58. 08003 Barcelona.

– Rebut el novembre de 1997.

– Acceptat l'abril de 1998.

1. INTRODUCCIÓ

El coneixement de l'evolució anual dels efectius demogràfics és necessari, tant per ser una dada bàsica per a l'assignació de recursos com per ser imprescindible per al càlcul d'indicadors demogràfics i socio-econòmics. L'Institut d'Estadística de Catalunya (Idescat) realitza les estimacions de població, consistents en calcular la població catalana recent pel mètode dels components: la població en un any Y es calcula sumant a la població de l'any anterior els naixements i la immigració, i restant les defuncions i l'emigració. La població es desagrega en 2 sexes, 100 grups d'edat i 62 àrees geogràfiques: els 21 municipis de més de 45.000 habitants i les 41 comarques i restes comarcals (entem per resta comarcal el total de la comarca menys els seus municipis de més de 45.000 habitants). De les 62 àrees, 23 tenen menys de 45.000 habitants i 25 tenen entre 45.000 i 100.000.

Un dels objectius d'aquest projecte és poder disposar d'una sèrie temporal coherent i amb un important nivell de desagregació geogràfica i d'edat; aquesta sèrie comença l'any 1986. El registre actual de la migració és l'Estadística de Variacions Residencials (EVR), que permet conèixer les dades amb un bon nivell de detall. Es presenten, però, 2 problemes: en primer lloc, l'origen de l'EVR data de 1988, de manera que manquen dades pels anys 1986 i 1987; en segon lloc, en els anys padronals i/o censals, les EVR registren un significatiu descens, de manera que el seu valor no és coherent dins la sèrie temporal. Aquestes 2 situacions han portat a la creació d'un mètode per a l'estimació de la migració en àrees petites.

A continuació, es defineixen els indicadors i les variables que intervenen en el càlcul de la migració. Les definicions són anàlogues per a l'emigració i la immigració, de manera que es parlarà de taxa de migració i índex sintètic de migració, GMR (gross migraton rate). Això no vol dir, però, que es tracti la migració en conjunt, sinó que tot el mètode s'aplica 2 vegades, 1 per a la immigració i 1 per a l'emigració.

Siguin

a = àrea geogràfica

s = sexe

e = edat

M = migrants

P = població

Y = any

$t_Y(a, s, e) = M(a, s, e)/P_Y(a, s, e)$ taxa de migració per àrea, sexe i edat en un any Y

$t_Y(a, e) = M(a, e)/P_Y(a, e)$ taxa de migració per àrea i edat en un any Y

$GMR_Y(a, s) = \sum_{e=0,1,\dots,99} t_Y(a, s, e)$ índex sintètic de migració per àrea i sexe
 $GMR_Y(a) = \sum_{e=0,1,\dots,99} t_Y(a, e)$ índex sintètic de migració per àrea i total sexes
 $p_Y(a, s, e) = t_Y(a, s, e)/GMR_Y(a, s)$ pes d'una edat dins el conjunt de taxes
 $(p_Y(a, s))_{e=0,1,\dots,99}$ perfil migratori

Notem que $\|(p_Y(a, s))_{e=0,1,\dots,99}\|_1 = 1$, fet que permet comparar diferents perfils a la mateixa escala.

Amb aquestes definicions s'identifiquen i separen els 2 elements constitutius de la migració: el nivell (o intensitat) i el perfil. Donat que el que es vol és estimar les dades en els anys pels quals manca informació, es prendran com a valors de la intensitat i el nivell migratori en un determinat any Y la mitjana aritmètica dels 4 anys més propers (en particular els 4 anteriors quan es vulgui estimar els valors per a l'últim any de la sèrie històrica).

2. L'ESTIMACIÓ DEL NIVELL MIGRATORI

Donada l'existència d'una sèrie històrica amb dades sobre les migracions i per tant sobre l'indicador GMR, es podria pensar en estimar el valor del GMR en els anys pels quals no es troba disponible a partir de les tendències observades (bé per regressió lineal, sèries temporals o altres mètodes). En aquest sentit, cal fer una observació important: en treballar amb àrees que generalment tenen una població inferior als 45.000 habitants (23 de les 62 àrees), el valor del GMR es troba distorsionat per l'aleatorietat de les xifres petites dels numeradors de les taxes, de manera que l'objectiu principal és calcular uns GMR robustos.

L'objectiu d'aquesta etapa és, doncs, agrupar les àrees de població de manera que tots els grups resultants superin la barrera dels 45.000 habitants. El primer criteri en què es pot pensar és el geogràfic (agrupació de comarques veïnes), però aquest criteri no és vàlid perquè, malgrat el veïnatge, hi ha comarques en expansió junt amb altres en retrocés, fet que es reflecteix clarament en la intensitat de la migració. El criteri emprat és realitzar una anàlisi cluster entre les àrees, agrupant cada àrea de població inferior a 45.000 habitants amb l'àrea que presenti distància mínima dins l'anàlisi cluster. Si el conjunt no arriba al mínim poblacional s'agrupa de nou. Si l'agrupació és d'una àrea petita (menor de 45.000 habitants) amb una de gran, el GMR del cluster només s'aplica a la petita, mentre que el càlcul del GMR de la gran no es modifica. Si l'agrupació és entre 2 o més àrees petites, el GMR del cluster s'aplica a cada una d'elles. Per obtenir un GMR més robust, aquesta anàlisi es realitza sobre les

migracions dels 4 anys més propers a l'any a estimar i sobre el conjunt dels 2 sexes, és a dir sobre $GMR_Y(a)$.

En l'anàlisi cluster s'ha emprat un mètode jeràrquic acumulatiu: partint de 62 àrees, en cada pas s'agrupen les 2 àrees o clusters amb distància mínima. El procés s'itera fins a reduir el nombre de clusters a 2.

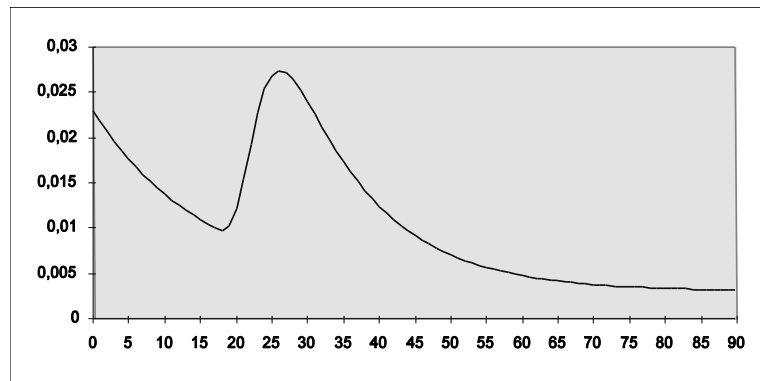
3. L'ESTIMACIÓ DEL PERFIL MIGRATORI

Les dades empíriques mostren que la migració és un fenomen altament selectiu segons l'edat: els joves entre els 20 i els 30 anys exhibeixen les taxes més altes de migració, bàsicament relacionades amb motius laborals i de matrimoni, mentre que els adolescents tenen les taxes de migració més petites. D'altra banda, les taxes migratòries dels fills reflecteixen les dels pares i això explica que els nens tinguin taxes més elevades que els adolescents. S'ha constatat que en algunes àrees pot trobar-se un augment de les taxes a la tercera edat, relacionat amb canvis de residència lligats a la jubilació i, en edats encara més avançades, es poden constatar migracions lligades a canvis en l'estat de salut (ingrés en residències, trasllat a cases dels fills, etc.).

El model matemàtic que descriu la migració a partir de l'edat es coneix com model de Rogers. La seva formulació, prenent l'edat com a variable dependent, és:

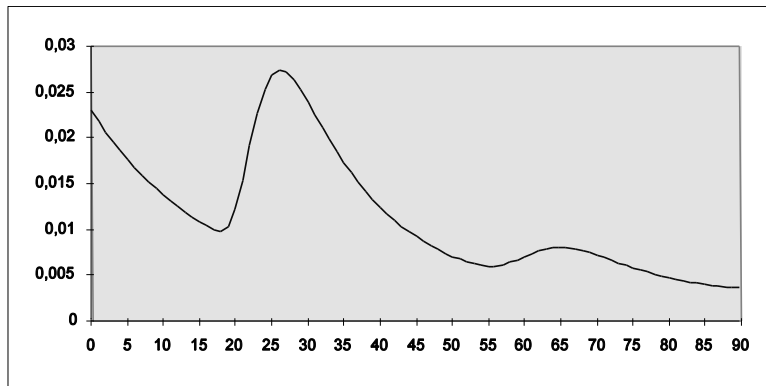
model general:

$$p(x) = a_1 \cdot \exp(-\alpha_1 \cdot x) + a_2 \cdot \exp\{-\alpha_2 \cdot (x - \mu_2) - \exp(-\lambda_2 \cdot (x - \mu_2))\} + c$$



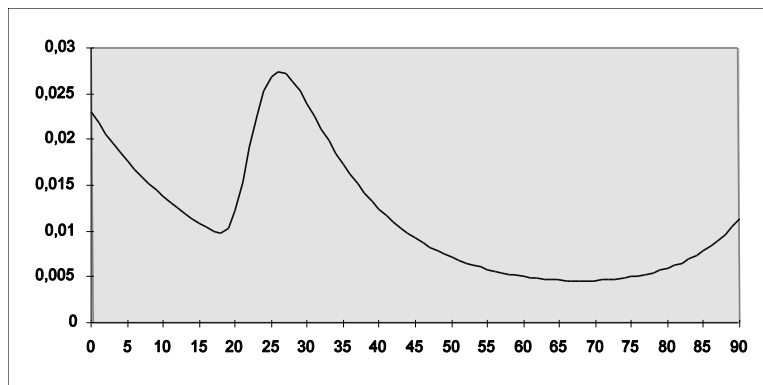
model amb pic de jubilació:

$$p(x) = a_1 \cdot \exp(-\alpha_1 \cdot x) + a_2 \cdot \exp\{-\alpha_2 \cdot (x - \mu_2) - \exp(-\lambda_2 \cdot (x - \mu_2))\} + a_3 \cdot \exp\{-\alpha_3 \cdot (x - \mu_3) - \exp(-\lambda_3 \cdot (x - \mu_3))\} + c$$



model amb pendent de jubilació:

$$p(x) = a_1 \cdot \exp(-\alpha_1 \cdot x) + a_2 \cdot \exp\{-\alpha_2 \cdot (x - \mu_2) - \exp(-\lambda_2 \cdot (x - \mu_2))\} + a_3 \cdot \exp(\alpha_3 \cdot x) + c$$



Per obtenir el perfil migratori de cada àrea i sexe es realitza una anàlisi cluster de les corbes de pesos pel conjunt dels 2 sexes. En aquest sentit, és important treballar amb les corbes normalitzades $p_Y(a)$ en lloc de $t_Y(a)$, per determinar l'efecte de la migració a cada edat, independentment de la intensitat de la migració. Es defineix la distància entre perfils com

$$d(p_Y(a), p_Y(a')) = \|p_Y(a) - p_Y(a')\|_2$$

Un pas previ a l'anàlisi cluster és disposar d'àrees amb prou població com per eliminar l'efecte pertorbador de les poblacions en els numeradors de $t_Y(a, s, e)$. Amb aquest efecte s'agrupen les àrees que no superen els 45.000 habitants, atenent a criteris de veïnatge geogràfic (es considera que el comportament migratori ve influït per factors com l'estructura de la població, el medi geogràfic o el tipus de vida). El nombre d'àrees queda reduït de 62 a 45. A continuació, es realitza l'anàlisi cluster, definint la distància entre 2 clusters com $\|p_Y(k_1) - p_Y(k_2)\|_2$. El resultat d'aquesta etapa és l'obtenció de 6 clusters per a l'emigració i 6 per a la immigració (veure taula 1).

Un cop obtinguts els clusters, per a cada un d'ells se sumen els efectius migratoris i poblacionals de totes les àrees que el componen, obtenint el perfil migratori del cluster per sexe $p_Y(k, s)$ (on k indica el cluster). Aquests vectors tenen com a components valors observats. L'etapa final és procedir a la substitució dels punts $p_Y(k, s)$ per la corba paramètrica $\hat{p}_Y(k, s)$. Aquest pas es realitza amb el paquet estadístic SAS, per aproximació de funcions no lineals; s'empra el mètode de Marquard, que és un compromís entre el de Gauss-Newton i el de màxima pendent.

Els gràfics permeten apreciar la qualitat de l'ajustament de la corba teòrica a les dades observades a diferents comarques i ciutats, amb un efecte de suavització de les corbes. És evident que hi ha un efecte de grandària i que com més població tenen les àrees millor s'ajusten al model. Les gràfiques de la distribució dels errors permeten observar que hi ha diferències en la qualitat de l'ajustament de les dades segons el sexe i l'edat. D'una banda, el model s'ajusta millor al sexe masculí (valors mínims de $\|p_Y(k) - \hat{p}_Y(k)\|_2$) que al femení. Pel que fa a les edats, els valors outliers es concentren als vells i als joves pels homes, mentre que pel sexe femení es troba una major dispersió, apareixent una certa sobrevaloració de la migració en edats laborals i prelaborals.

La introducció de corbes amb pendent de jubilació ha permès millorar la qualitat de l'ajustament en algunes corbes d'immigració, encara que el procediment ha suposat un augment notable del nombre d'iteracions per assolir els criteris de convergència. S'ha introduït també un pic de jubilació en algunes corbes d'emigració amb el mateix resultat positiu (veure gràfics 1 a 4).

De la parametrització dels perfils migratoris es deriva la definició de valors característics que faciliten el seu estudi i comparació. Entre ells destaquem els següents:

$$\begin{aligned}
 x_m &= \Sigma_x \cdot p(x) / \Sigma_p(x) \text{ edat mitjana de la migració} \\
 \delta_{1c} &= a_1/c \\
 \delta_{12} &= a_1/a_2 \text{ índex de dependència infantil} \\
 \beta_{12} &= \alpha_1/\alpha_2 \text{ índex de regularitat parento-infantil} \\
 \sigma_2 &= \lambda_2/\alpha_2 \text{ asimetria laboral}
 \end{aligned}$$

$$\begin{aligned}\delta_{32} &= a_3/a_2 \text{ \u00edndex de depend\u00e8ncia postlaboral} \\ \sigma_3 &= \lambda_3/\alpha_3 \text{ asimetria postlaboral}\end{aligned}$$

El pes de les edats actives i preactives queda reflectit en els par\u00e0metres a_1 i a_2 respectivament. El ratio δ_{21} reflecteix el grau de dominan\u00e7a laboral, mentre que el seu invers δ_{12} mesura el grau de depend\u00e8ncia infantil, o taxa amb la qual els adults migren amb fills. L'asimetria laboral, la forma esbiaixada cap a l'esquerra de la migraci\u00f3 a les edats laborals, es mesura per σ_2 .

Una classificaci\u00f3 de les corbes pot obtenir-se segons el car\u00e0cter familiar o laboral de la migraci\u00f3 (veure taules 2 a 5 i quadres 1 i 2). Les primeres s\u00f3n les corbes migrat\u00f2ries amb depend\u00e8ncia infantil, \u00e9s a dir, amb un gran pes de fam\u00edlies amb nens. Les corbes de car\u00e0cter predominantment laboral s\u00f3n aquelles en les quals es troba un pes relatiu molt important en la migraci\u00f3 de la poblaci\u00f3 en edat de treballar, en comparaci\u00f3 a la migraci\u00f3 de la resta de poblaci\u00f3. Una mesura del car\u00e0cter laboral o familiar de la migraci\u00f3 la d\u00f3na δ_{12} . El valor d'aquest \u00edndex, en els models emp\u00edrics observats, se situa entorn a 1/3 i es considera que valors superiors a 0,4 indiquen un valor predominantment familiar de la migraci\u00f3. Inversament, valors inferiors a 0,4 indiquen un car\u00e0cter predominantment laboral de la migraci\u00f3. De l'estudi dels patrons migratoris es dedueix que les migracions a la Catalunya actual semblen haver substitu\u00eft el motiu laboral pel residencial, aix\u00f2 lligat probablement a les noves facilitats pels transports que permeten un increment de la mobilitat di\u00e0ria per motiu de treball i estudi. L'estudi d'altres par\u00e0metres porta a conclusions com ara l'exist\u00e8ncia d'un pic de jubilaci\u00f3 en un grup de ciutats, que va rebre fortes onades d'immigraci\u00f3 la segona meitat del segle XX i que ara experimenta un fenomen de retorn de migrants cap a altres comunitats aut\u00f2nomes. D'altra banda, el pendent de jubilaci\u00f3 apareix a determinats grups de car\u00e0cter predominantment urb\u00e0, explicable per la migraci\u00f3 de la gent gran lligada a canvis en l'estat de salut i dirigida cap a centres urbans que disposen d'equipaments i/o cap a la llar dels fills de persones que ja no poden viure soles per l'estat de salut.

4. L'ESTIMACI\u00d3 DELS EFECTIUS MIGRATORIS

Un cop s'han obtingut els GMR estimats i els patrons migratoris, s'estima el nombre de migrants de la seg\u00fcent manera:

Sigui ' a ' una \u00e0rea que pertany al cluster ' k_1 ' de GMR i al cluster ' k_2 ' de patr\u00f3 migratori. Llavors

$$M_Y(a, s, e) = P_{Y-1}(a, s, e-1) \cdot \widehat{GMR}_Y(k_1) \cdot \hat{p}_Y(k_2, s, e) \quad e = 0, 1, \dots, 99$$

La raó d'aplicar aquesta fórmula és que si a la part dreta se substituís $P_{Y-1}(a, s, e-1)$ per $P_Y(a, s, e)$, $\widehat{GMR}_Y(k_1)$ per $GMR_Y(a)$ i $\hat{p}_Y(k_2, s, e)$ per $p_Y(a, s, e)$ s'obtidria

$$P_Y(a, s, e) \cdot GMR_Y(a) \cdot p_Y(a, s, e) = P_Y(a, s, e) \cdot GMR_Y(a) \cdot t_Y(a, s, e) / GMR_Y(a) = P_Y(a, s, e) \cdot t_Y(a, s, e) = P_Y(a, s, e) \cdot M_Y(a, s, e) / P_Y(a, s, e) = M_Y(a, s, e)$$

Com que els valors observats de $p_Y(a, s, e)$ estan subjectes a més irregularitats se substitueixen per $\hat{p}_Y(k_2, s, e)$, que són més estables. D'altra banda, no es coneix $GMR_Y(a)$ ni $P_Y(a, s, e)$ i se substitueixen per $\widehat{GMR}_Y(k_1)$ i $P_{Y-1}(a, s, e-1)$.

5. CONCLUSIONS I RESULTATS

En aquest article s'ha descrit el mètode emprat a l'Idescat per estimar dades sobre les migracions, tant d'entrada com de sortida, del territori de Catalunya dividit en 62 àrees: 21 municipis i 41 comarques. La disposició d'un registre complet i coherent de la migració es remunta a 1988, any de la creació de l'Estadística de Variacions Residencials (EVR). D'acord amb l'EVR, tota migració queda registrada tant en el municipi de sortida com en el d'arribada. Abans això no era així, de manera que el nombre d'emigrants i el d'immigrants no era coherent. Un altre tema que es planteja és l'exhaustivitat: mentre que esdeveniments demogràfics com són els naixements, les defuncions o els matrimonis queden registrats en la seva totalitat, no passa el mateix amb les migracions. Sovint, la gent canvia la residència d'un municipi a un altre, però no registra el canvi a l'ajuntament, bé sigui per desconeixement, comoditat o interès personal. Aquest registre aflora, sovint, temps després, per exemple: quan cal inscriure els fills a l'escola o quan es realitza la renovació padronal.

L'estudi de la migració separant-ne els dos factors que la componen, el perfil i el nivell, permet fer una estimació acurada de la migració i, per tant, solucionar els problemes de l'inexhaustivitat: d'una banda s'aplica als anys 1986 i 1987 per poder perllongar 2 anys més la sèrie de dades pel seu inici i, de l'altra, per poder disposar de dades més completes als anys de renovació padronal, 1991 i 1996 fins al moment, en què es produeix un subregistre de la migració per quedar absorbit el seu registre al padró.

L'altre punt important del mètode és que dóna una eina per caracteritzar la migració arreu de Catalunya. S'estudia de manera sistemàtica la migració a nivell comarcal, amb un gran nivell de desagregació d'edat, fet que no s'havia realitzat a Catalunya fins ara. La migració és un fet que varia notablement d'una comarca a una altra, no només per la seva intensitat sinó també per la seva distribució per edats. Per primera vegada es fa una anàlisi que separi aquests factors, determinant 6 patrons d'emigració i 6 d'immigració. Cal fer notar que no s'ha treballat amb saldos migratoris, sinó amb emigració i immigració, de manera que una comarca o municipi queda caracteritzat per

2 patrons, un per a les sortides i un altre per a les entrades. Els patrons són diferenciats per sexe i permeten quantificar els components de la migració: el familiar, el laboral, el de jubilació i el d'edats avançades. Donada la influència dels municipis grans en el seu entorn, aquests s'han separat de les comarques, ja que sovint tenen comportaments diferents i fins i tot oposats, que no quedarien ben registrats si s'estudiés la comarca com un tot.

Pel que fa al nivell de migració es poden distingir 4 menes d'àrees: primer les molt dinàmiques o d'alta mobilitat, que presenten valors de GMR elevats, tant per la immigració com per l'emigració; segon, les de poca mobilitat, en ambdós sentits; tercer, les de caràcter marcadament emigratori, és a dir, que presenten un GMR elevat d'emigració combinat amb un GMR mitjà o baix d'immigració, i, finalment, les de caràcter marcadament immigratori.

Destaca el poc dinamisme migratori d'algunes de les ciutats de més de 45.000 habitants, mentre que les posicions amb els nivells més elevats són ocupades per les comarques i restes comarcals. Ciutats com Terrassa, Sabadell, Mataró, Manresa i Lleida destaquen pel seu nivell particularment baix tant d'entrades com de sortides. El caràcter marcadament emigratori de ciutats com Cornellà de Llobregat, Santa Coloma de Gramenet, L'Hospitalet de Llobregat, Barcelona i Badalona contrasta amb la situació de Rubí, Viladecans, Girona, Cerdanyola del Vallès i Reus, de caràcter marcadament immigratori. Entre les comarques amb caràcter marcadament immigratori destaquen el Garraf, el Baix Penedès, el Tarragonès, el Maresme, el Vallès Occidental i el Vallès Oriental.

En conclusió, l'estudi del GMR permet, en comparació amb estudis ja existents, apreciar el dinamisme migratori de les comarques pirinenques que, independentment del saldo migratori, presenten una mobilitat elevada. A l'extrem oposat, destaca la baixa mobilitat a les àrees de l'interior i del sud-oest de Catalunya, amb signes migratoris generalment negatius. Finalment, destaquen per la seva elevada immigració les restes comarcals de les comarques properes a Barcelona i tota l'àrea de les comarques gironines, així com l'entorn de l'eix Reus-Tarragona.

L'estudi dels patrons migratoris ens permet caracteritzar les àrees des d'un altre punt de vista. El grup de ciutats constituït per Cerdanyola del Vallès, Rubí, Sant Adrià de Besòs, Santa Coloma de Gramenet, l'Hospitalet de Llobregat i Viladecans atrauen treballadors i expulsen famílies: presenten un perfil immigratori de tipus predominantment laboral i, en canvi, una emigració amb un caràcter marcadament familiar. A la banda oposada trobem les ciutats de Reus i Tarragona, atractives per a la immigració familiar i amb una emigració predominantment laboral. Pel que fa als moviments de les comarques, destaca el caràcter altament familiar, tant de la immigració com de l'emigració, als municipis menors de 45.000 habitants del Vallès Occidental, Baix Llobregat, Garraf, Baix Camp i Tarragonès, i a la comarca del Baix Penedès. De la resta de comarques cal esmentar el Segrià i la Garrotxa, caracteritzades per una corba

amb un perfil marcadament laboral en la seva emigració, amb la particularitat d'una elevada sobre-emigració del sexe femení en les edats laborals, típica del medi rural a Catalunya.

Taula 1. Composició dels clusters de perfil migratori, 1988-90, 1992.

| <i>Immigració</i> | <i>Emigració</i> |
|---|---|
| Cluster 1 | Cluster 1 |
| ALT CAMP CONCA DE BARBERA | ALT CAMP CONCA DE BARBERA |
| ALT PENEDES ANOIA Resta BAGES | BAIX EBRE |
| GARRIGUES PRIORAT RIBERA D'EBRE TERRA ALTA | BERGUEDA SOLSONES CERDANYA RIPOLLES |
| BERGUEDA SOLSONES CERDANYA RIPOLLES | NOGUERA PLA D'URGELL |
| Terrassa | SELVA |
| Esplugues de Llobregat | SEGARRA URGELL |
| OSONA | Resta BAIX CAMP Resta TARRAGONES |
| Manresa | Girona |
| SELVA | Manresa |
| Mataró | Granollers |
| Granollers | ALT EMPORDA BAIX EMPORDA |
| BAIX EBRE | ALT URGELL ALTA RIBAGORÇA PALLARS JUSSÀ PALLARS SOBIRA VAL D'ARAN |
| Barcelona | Resta GIRONES PLA DE L'ESTANY |
| Lleida | GARRIGUES PRIORAT RIBERA D'EBRE TERRA ALTA |
| MONTSIA | MONTSIA |
| NOGUERA PLA D'URGELL | Cluster 2 |
| Sabadell | GARROTXA Resta SEGRIA |
| Cluster 2 | |
| Resta BARCELONES Santa Coloma de Gramenet | |
| Badalona | |
| Cornellà de Llobregat | |
| Prat de Llobregat | |
| Viladecans | |
| Hospitalet de Llobregat | |

Taula 1. (cont.) Composició dels clusters de perfil migratori, 1988-90, 1992.

| <i>Immigració</i> | <i>Emigració</i> |
|-----------------------|--------------------------|
| Cluster 3 | Cluster 3 |
| ALT EMPORDA | ALT PENEDES |
| GARROTXA | OSONA |
| Resta GIRONES | ANOIA |
| PLA DE L'ESTANY | Resta BAGES |
| SEGARRA | Resta VALLES ORIENTAL |
| URGELL | Sabadell |
| ALT URGELL | Vilanova i la Geltrú |
| ALTA RIBAGORÇA | Cluster 4 |
| PALLARS JUSSÀ | Resta MARESME |
| PALLARS SOBIRA | Lleida |
| VAL D'ARAN | Reus |
| Girona | Barcelona |
| BAIX EMPORDA | Badalona |
| Resta SEGRIA | Terrassa |
| Cluster 4 | Prat de Llobregat |
| Sant Boi de Llobregat | Tarragona |
| Cerdanyola del Vallès | Cluster 5 |
| Rubí | Resta BAIX LLOBREGAT |
| Cluster 5 | Sant Boi de Llobregat |
| BAIX PENEDES | Mataró |
| Resta GARRAF | BAIX PENEDES |
| Cluster 6 | Resta GARRAF |
| Resta BAIX CAMP | Resta VALLES OCCIDENTAL |
| Resta TARRAGONES | Rubí |
| Resta MARESME | Cerdanyola del Vallès |
| Resta BAIX LLOBREGAT | Cluster 6 |
| Resta VALLES ORIENTAL | Resta BARCELONES |
| Tarragona | Santa Coloma de Gramenet |
| Reus | Hospitalet de Llobregat |
| Vilanova i la Geltrú | Cornellà de Llobregat |
| | Esplugues de Llobregat |
| | Viladecans |

☐ Agrupacions prèvies al cluster per assolir 45.000 habitants.

Taula 2. Paràmetres de la corba migratòria. Immigració.

| | tipus | a1 | $\alpha 1$ | a2 | $\alpha 2$ | $\mu 2$ | $\lambda 2$ | a3 | $\alpha 3$ | $\mu 3$ | $\lambda 3$ | c |
|-------------|-------|--------|------------|--------|------------|---------|-------------|--------|------------|---------|-------------|----------|
| Catalunya H | 0 | 0,0177 | 0,0837 | 0,0430 | 0,1196 | 23,37 | 0,2930 | 0,0000 | 0,0000 | 0,00 | 0,0000 | 0,004713 |
| Catalunya D | 0 | 0,0183 | 0,0847 | 0,0457 | 0,1269 | 22,56 | 0,2955 | 0,0000 | 0,0000 | 0,00 | 0,0000 | 0,004690 |

(*) Tipus model. Valors possibles: 0=normal, 1=amb pic de jubilació, 2= amb pendent de jubilació.

(**) a1 i $\alpha 1$ determinen la migració infantil; a2, $\alpha 2$, $\mu 2$ i $\lambda 2$ determinen la migració laboral; a3, $\alpha 3$, $\mu 3$ i $\lambda 3$ determinen la migració de jubilació; c és el terme independent de l'edat. L'equació del model es troba a l'apartat 3 del text.

Taula 3. Valors característics de la corba migratòria. Immigració.

| | tipus | xm | %0-14 | %15-64 | %65+ | $\delta 1c$ | $\delta 12$ | $\delta 32$ | $\beta 12$ | $\sigma 2$ | $\sigma 3$ |
|-------------|-------|-------|-------|--------|-------|-------------|-------------|-------------|------------|------------|------------|
| Cluster1 H | 2 | 37,89 | 22,95 | 62,61 | 18,72 | 4,46 | 0,3894 | 0,000444 | 0,6436 | 2,21 | 0 |
| Cluster1 D | 0 | 35,97 | 23,02 | 63,10 | 17,32 | 3,53 | 0,2949 | 0,000000 | 0,4967 | 1,75 | 0 |
| Cluster2 H | 2 | 51,66 | 21,09 | 58,18 | 35,84 | 5,18 | 0,2966 | 0,000004 | 0,6834 | 1,73 | 0 |
| Cluster2 D | 2 | 41,37 | 22,91 | 57,80 | 26,02 | 23502,67 | 0,3337 | 0,018828 | 0,5370 | 1,46 | 0 |
| Cluster3 H | 0 | 32,11 | 24,36 | 66,64 | 11,03 | 7,50 | 0,4842 | 0,000000 | 0,7943 | 3,44 | 0 |
| Cluster3 D | 0 | 32,30 | 25,22 | 64,44 | 12,84 | 6,05 | 0,4151 | 0,000000 | 0,6328 | 2,96 | 0 |
| Cluster4 H | 2 | 60,90 | 18,79 | 55,65 | 46,48 | 2,86 | 0,2461 | 0,000001 | 0,5792 | 1,60 | 0 |
| Cluster4 D | 2 | 46,07 | 19,43 | 58,54 | 30,52 | 3,36 | 0,2455 | 0,000670 | 0,5652 | 1,74 | 0 |
| Cluster5 H | 0 | 40,75 | 19,45 | 64,11 | 19,72 | 3,39 | 0,7484 | 0,000000 | 1,0545 | 11,29 | 0 |
| Cluster5 D | 0 | 40,73 | 19,10 | 64,10 | 20,37 | 2,56 | 0,6794 | 0,000000 | 1,0950 | 13,17 | 0 |
| Cluster6 H | 2 | 39,59 | 22,56 | 62,17 | 19,92 | 19014,51 | 0,6193 | 0,032861 | 0,6923 | 4,67 | 0 |
| Cluster6 D | 2 | 39,86 | 23,13 | 60,36 | 21,78 | 20157,32 | 0,5818 | 0,038010 | 0,6487 | 3,80 | 0 |
| Catalunya H | 0 | 35,63 | 22,57 | 60,80 | 16,63 | 3,76 | 0,4095 | 0,000000 | 0,6992 | 2,45 | 0 |
| Catalunya D | 0 | 35,02 | 22,97 | 60,58 | 16,45 | 3,91 | 0,4016 | 0,000000 | 0,6671 | 2,33 | 0 |

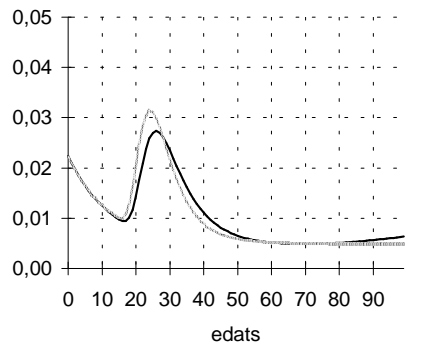
(*) Tipus model. Valors possibles: 0=normal, 1=amb pic de jubilació, 2=amb pendent de jubilació.

(**) La definició dels valors característics es troba a l'apartat 3 del text.

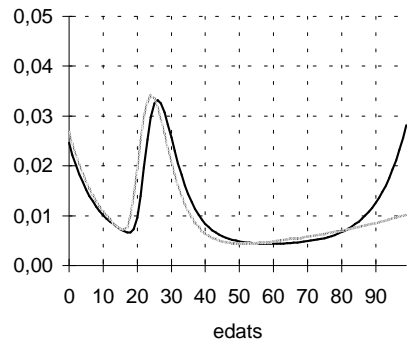
Quadre 1. Classificació de les comarques i restes comarcals segons el tipus d'emigració i d'immigració.

| Immigració Emigració | Laboral Alt | Laboral Mitjà | Familiar Mitjà | Familiar Alt |
|-------------------------|---|---|--|--|
| Laboral Alt | | Segrià* Garrotxa | | Maresme* |
| Laboral Mitjà | Anoia Alt Penedès Bages* Osona | Alt Camp, Conca Barberà Ribera d'Ebre Montsià, Baix Ebre Garrigues, Priorat, Terra Alta Noguera, Pla d'Urgell | Vallès Oriental* | |
| Familiar Mitjà | | Berguedà, Solsonès Ripollès, Val d'Aran Alt Ribagorça, Pallars Jussà Alt Urgell, Pallars Sobirà Cerdanya | Baix Empordà Pla Estany Gironès* Alt Empordà Segarra, Urgell | Tarragonès* Baix Camp* |
| Familiar Alt | | | | Vallès Occidental* Baix Penedès Baix Llobregat* Garraf* |

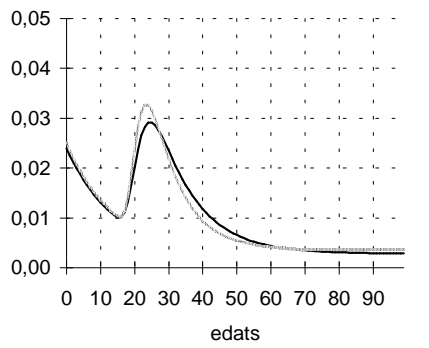
* Restes comarcals



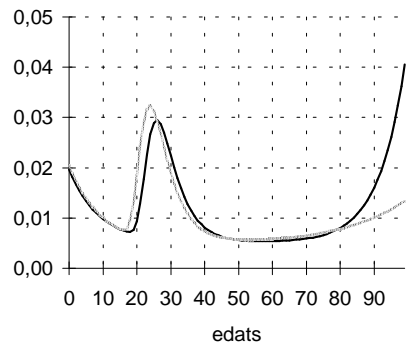
—— clust.1homes clust.1dones



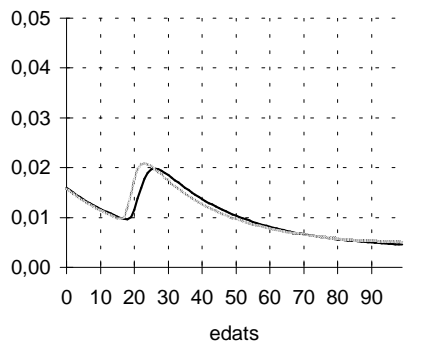
—— clust.2homes clust.2dones



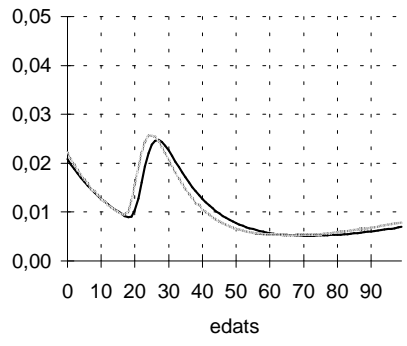
—— clust.3homes clust.3dones



—— clust.4homes clust.4dones

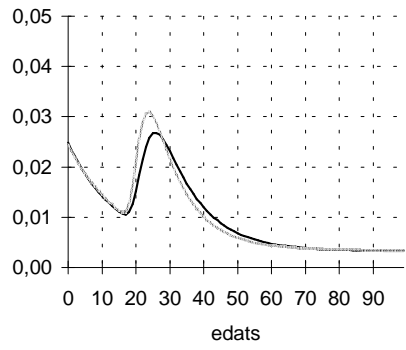


—— clust.5homes clust.5dones

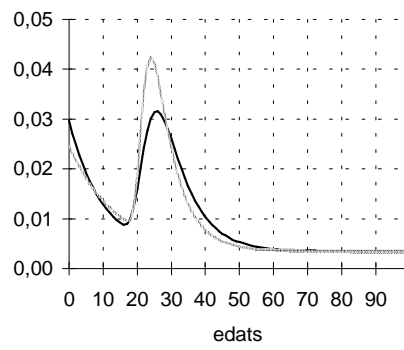


—— clust.6homes clust.6dones

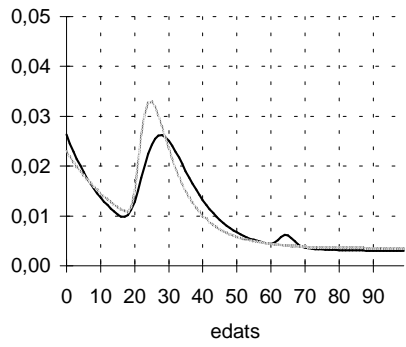
Gràfic 1. Corbes dels clusters* de perfil migratori. Immigració 1988-90, 1992.



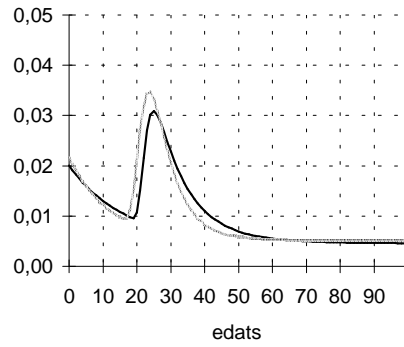
—— clust.1homes clust.1dones



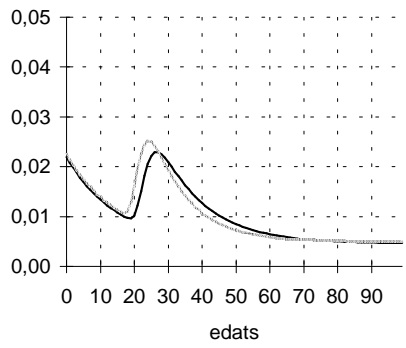
—— clust.2homes clust.2dones



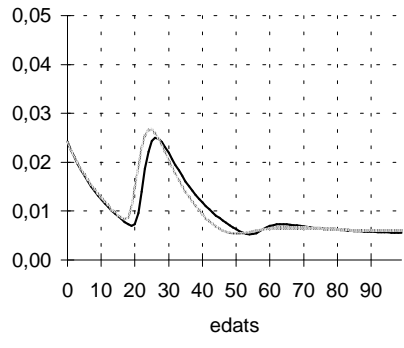
—— clust.3homes clust.3dones



—— clust.4homes clust.4dones

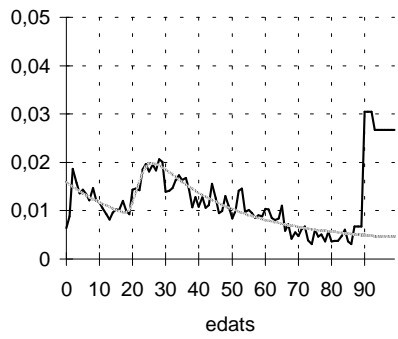


—— clust.5homes clust.5dones

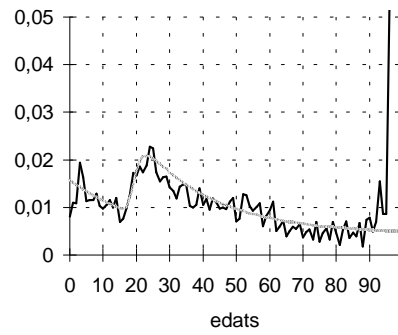


—— clust.6homes clust.6dones

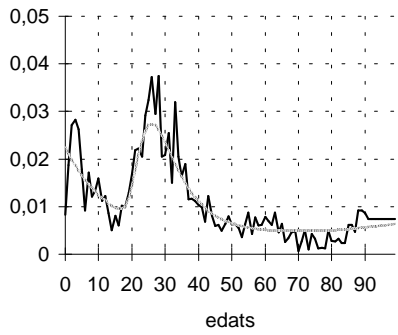
Gràfic 2. Corbes dels clusters de perfil migratori. Emigració 1988-90, 1992.



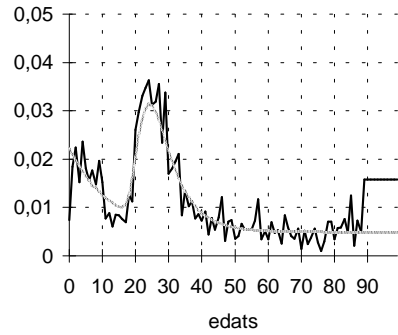
—— B. PenedèsH cluster5H



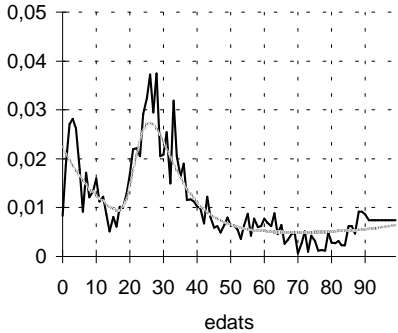
—— B. PenedèsD cluster5D



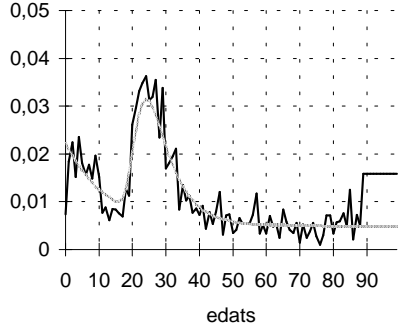
—— MontsiàH cluster1H



—— MontsiàD cluster1D

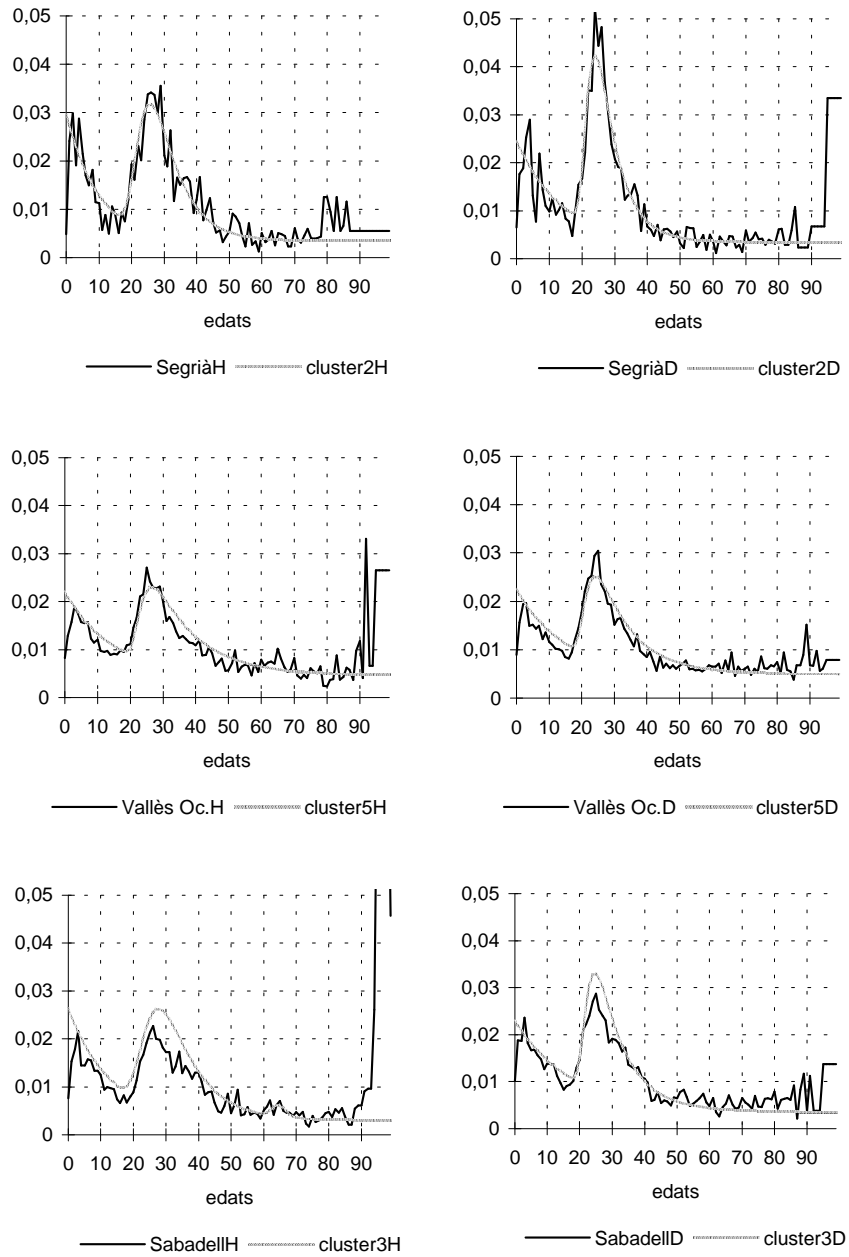


—— MontsiàH cluster1H



—— MontsiàD cluster1D

Gràfic 3. Comparació per edat i sexe de les dades observades amb el model ajustat. Immigració 1988-90, 1992. Baix Penedès, Montsià i L'Hospitalet.



Gràfic 4. Comparació per edat i sexe de les dades observades amb el model ajustat. Emigració 1988-90, 1992. Segrià*, Vallès Occidental* i Sabadell.

REFERÈNCIES

- [1] **Aldenferden** and **Blashfield** (1990). «Cluster analysis». *Quantitative applications in the social sciences*, **44**. Sage publications.
- [2] **Boden, P., Stillwell, J.** and **Rees, P.** (1991). «Internal migration projection in England: the OPCS/DOE model examined» in Stillwell, J. and Congdon, P. *Migration models*. Belhaven Press.
- [3] **Castro, L.** and **Rogers, A.** (1986). «Migration» in Rogers, A. and Willekens, F.J. *Migration and settlement. A multiregional comparative study*. IIASA.
- [4] **Dûchene, J.** (1989). *Téchniques auxiliares en démographie II: Analyse des données en démographie*. Institut de Démographie, Université Catholique de Louvain.
- [5] **Institut d'Estadística de Catalunya.** *Estimacions de població 1986-1995*. Generalitat de Catalunya.
- [6] **Lewis-Beck, M.S.** (1990). «Applied regression». *Quantitative applications in the social sciences*, **22**. Sage publications.
- [7] **Longfod, N.T.** (1996). «Small-area estimation using adjustment by covariates». *Qüestió*, **20**, **2**.
- [8] **SAS/STAT.** *User's guide*, version 6 (1990). SAS Institute.

ENGLISH SUMMARY

METHOD FOR ESTIMATING SMALL AREA MIGRATION FROM INCOMPLETE DATA

J.A. SÁNCHEZ CEPEDA

Institut d'Estadística de Catalunya*

The knowledge of small area migration data is a part of the project of Population estimates, which calculates recent Catalan population for towns above 45.000 inhabitants and small areas, broken down by age and sex. The difficulty, for some years, of having figures with such level of detail has led to the implementation of a method for its estimation.

This article describes the method applied by the Institut d'Estadística de Catalunya (Idescat) for estimating migration broken down by geographic area, sex and age. The method calculates separately the level and the profile of migration. Working with areas below 45.000 inhabitants is the main problem that arises in estimating the level of migration. The number of profiles is reduced from the initial 62 small areas into 6, which define the migration patterns observed in Catalonia. The typology of this migration patterns is commented on the base of parameters determined by the model. Finally, the number of immigrants and emigrants is calculated from the estimated migration levels and profiles.

Keywords: Small area, migration model, cluster analysis, non-linear adjustment.

* Institut d'Estadística de Catalunya. Via Laietana, 58. 08003 Barcelona.

–Received November 1997.

–Accepted April 1998.

The knowledge of the yearly evolution of the population is necessary as it is a basic tool for resources allocation and it is essential for calculating demographic and social-economic indicators. L'Institut d'Estadística de Catalunya (Idescat) estimates recent catalan population using the components method. Population at the end of a year Y is calculated adding births and immigration to population at the end of the previous year, and subtracting deaths and outmigration registered during year Y . Population is broken down by sex, 100 age groups and 62 geographic areas, 23 of which have less than 45.000 inhabitants, 25 have between 45.000 and 100.000 inhabitants and the rest have more than 100.000 inhabitants. For years 1986 and 1987, data for outmigration and immigration are not available and they were estimated from data belonging to other years.

Let ' a ' be a geographic area, ' s ' a sex, ' e ' an age, ' M ' the number of migrations, ' P ' the population and ' Y ' a year. Then,

$$t_Y(a, s, e) = M(a, s, e)/P_Y(a, s, e) \text{ migration rate by area, sex and age in a year } Y$$

$$t_Y(a, e) = M(a, e)/P_Y(a, e) \text{ migration rate by area and age in a year } Y$$

$$\text{GMR}_Y(a, s) = \sum_{e=0,1,\dots,99} t_Y(a, s, e) \text{ gross migration rate by area and sex}$$

$$\text{GMR}_Y(a) = \sum_{e=0,1,\dots,99} t_Y(a, e) \text{ gross migration rate by area}$$

$$p_Y(a, s, e) = t_Y(a, s, e)/\text{GMR}_Y(a, s) \text{ weight of an age in the set of rates}$$

$$(p_Y(a, s))_{e=0,1,\dots,99} \text{ migration profile}$$

Notice that $\|(p_Y(a, s))_{e=0,1,\dots,99}\|_1 = 1$.

The two elements of migration are the level of migration (the gross migration rate) and the profile. They are separately estimated.

The main problem with the level of immigration is its definition: as it is calculated as the sum of rates by age, when populations are under 45.000 inhabitants the value of GMR is statistically poor. In order to calculate better values for GMR, a cluster analysis is carried out. Areas are joined in order to have more than 45.000 inhabitants and $\text{GMR}_Y(a)$ is calculated for them.

Migration rates behave under Rogers' age dependant model,

$$p(x) = a_1 \cdot \exp(-\alpha_1 \cdot x) + a_2 \cdot \exp\{-\alpha_2 \cdot (x - \mu_2) - \exp(-\lambda_2 \cdot (x - \mu_2))\} + c$$

This is known as the general model. The first term is the pre-labor force component, the second one the labour component and the third one a non-age dependant component. In some populations, an additional component for older ages, or retirement

peak, is observed,

$$p(x) = a_1 \cdot \exp(-\alpha_1 \cdot x) + a_2 \cdot \exp\{-\alpha_2 \cdot (x - \mu_2) - \exp(-\lambda_2 \cdot (x - \mu_2))\} + a_3 \cdot \exp\{-\alpha_3 \cdot (x - \mu_3) - \exp(-\lambda_3 \cdot (x - \mu_3))\} + c$$

whereas in other populations a post-retirement slope component is observed,

$$p(x) = a_1 \cdot \exp(-\alpha_1 \cdot x) + a_2 \cdot \exp\{-\alpha_2 \cdot (x - \mu_2) - \exp(-\lambda_2 \cdot (x - \mu_2))\} + a_3 \cdot \exp(\alpha_3 \cdot x) + c$$

A cluster analysis is made for the set of arrays $p_Y(a)$, using the distance between clusters $d(p_Y(a), p_Y(a')) = \|p_Y(a) - p_Y(a')\|_2$. It is important to analyze $p_Y(a)$ instead of $t_Y(a)$, as $\|(p_Y(a, s))_{e=0,1,\dots,99}\|_1 = 1$. The number of profiles is reduced from 62 to 6; these 6 profiles define the migration patterns observed in Catalonia. For every cluster k , $p_Y(k, s)$ is computed from the areas belonging to k . This profile is substituted for a smooth one, $\hat{p}_Y(k, s)$, using nonlinear functions approximation.

Finally, once GMR and profile patterns have been obtained, the number of migrants is obtained as follows:

Let 'a' be an area belonging to the GMR cluster 'k,' and the profile cluster 'K'; then $M_Y(a, s, e) = P_{Y-1}(a, s, e - 1) \cdot \widehat{GMR}_Y(k_1) \cdot \hat{p}_Y(k_2, s, e)$ $e = 0, 1, \dots, 99$.

The reason to use this formula is that substituting $P_{Y-1}(a, s, e - 1)$ for $P_Y(a, s, e)$, $\widehat{GMR}_Y(k_1)$ for $GMR_Y(a)$ and $\hat{p}_Y(k_2, s, e)$ for $p_Y(a, s, e)$,

$$P_Y(a, s, e) \cdot GMR_Y(a) \cdot p_Y(a, s, e) = P_Y(a, s, e) \cdot GMR_Y(a) \cdot t_Y(a, s, e) / GMR_Y(a) = P_Y(a, s, e) \cdot t_Y(a, s, e) = P_Y(a, s, e) \cdot M_Y(a, s, e) / P_Y(a, s, e) = M_Y(a, s, e)$$

Notice that observed values for $p_Y(a, s, e)$ are much more irregular than the ones for $\hat{p}_Y(k_2, s, e)$; besides, $GMR_Y(a)$ and $p_Y(a, s, e)$ are unknown and they are substituted for $\widehat{GMR}_Y(k_1)$ and $P_{Y-1}(a, s, e - 1)$.

As the result of applying this method, a tool to determine migration in Catalonia has been obtained. Migration in small areas is studied by sex and age: 6 immigration and 6 outmigration patterns are obtained. These patterns allow to determine the level of migration in every area due to family, work related, retirement causes. To work with retirement peak and post-retirement slope models has improved the results.