

## Mètodes estadístics per a les assegurances de vida

ANA M. DEBÓN, FRANCISCO MONTES

**Resum** L'estudi i evolució de la mortalitat, per les repercussions econòmiques i socials que presenta, és un tema d'elevat interès tant per als actuaris i els estadístics com per als demògrafs. En aquest treball revisem les diferents alternatives per a graduar les probabilitats de mort mitjançant models dinàmics, aquells que prenen en compte simultàniament la influència de l'edat i el temps del calendari, i en mostrem l'aplicació a dades de mortalitat d'Espanya. Finalment, es comenten les tendències futures en l'anàlisi dinàmica de la mortalitat.

**Paraules clau:** taules de mortalitat dinàmiques, predicció de la mortalitat, longevitat.

**Classificació MSC2000:** 62P05.

### 1 Introducció

Una de les derivacions més importants obtingudes a partir de les dades censals són les taules de mortalitat de la població. No només perquè es poden aplicar en la pràctica al càlcul actuarial per a obtenir el valor de primes i/o indemnitzacions, sinó, a més, perquè es tracta d'un instrument rellevant per a estudiar l'evolució de la població i els moviments d'aquesta. El primer pas, i potser una de les parts fonamentals en què intervé l'estadística, és la graduació de les probabilitats de mort,  $q_x$ . Definim la graduació [41] com el conjunt de principis i mètodes pels quals les probabilitats observades (o brutes) s'ajusten per a proporcionar una base allisada que permeti fer inferències i càlculs pràctics de primes i reserves.

El professional de l'assegurança de vida ha de ser capaç d'assignar primes suficients per a cobrir les quantitats que haurà de pagar la companyia en cas de mort de l'assegurat, per a la qual cosa utilitza les taules de mortalitat graduades. Els mètodes de graduació recollits en la literatura i que són emprats en la pràctica poden classificar-se en dos tipus fonamentals, paramètrics i no paramètrics, segons que ajusten les dades a una funció matemàtica o que només realitzen un allisatge. La graduació de la mortalitat mitjançant mètodes

paramètrics ha sigut un tema tractat àmpliament en [34, 67, 26]; de la no paramètrica s'ocupen [35, 36, 37, 80, 81], i pot trobar-se'n una revisió i comparació en [28]. Tots aquests articles s'ocupen de la graduació estàtica, és a dir, de la influència que l'edat té en la graduació de la mortalitat, per a la qual cosa s'analitzen dades d'un any en particular o, en el cas de diversos anys, es treballa amb les xifres agregades. L'experiència mostra que les taules de mortalitat estàtiques sobreestimen la probabilitat de mort i, com a conseqüència d'aquesta sobreestimació, les primes d'annualitats, de pensions i d'assegurances de vida no són les que haurien de ser, la qual cosa provoca efectes negatius sobre la companyia o sobre els assegurats. La raó de la dita sobreestimació és que les taules estàtiques han sigut avaluades per a un període de temps específic i no tenen en compte la tendència decreixent de la mortalitat al llarg del temps. Una solució a aquest problema són les taules de mortalitat dinàmiques, que el resolen incorporant la influència del calendari quan es gradua la mortalitat, particularitat que les ha convertides en una eina molt adequada per a la ciència actuarial.

Encara que en la literatura actuarial el coneixement de l'evolució de la mortalitat data de les primeres dècades del segle XX, és recentment que els càlculs actuarials han utilitzat les taules de mortalitat projectades o dinàmiques. En conseqüència, la predicció adequada de les probabilitats de mort mitjançant aquest tipus de taules s'ha convertit en un dels eixos centrals de la reducció del risc que s'assumeix. Tradicionalment, els mètodes utilitzats han sigut majoritàriament paramètrics, perquè aquest tipus de graduació afavoreix la comparació al llarg del temps i de l'espai [20], en estudiar l'evolució dels paràmetres que cal estimar o incorporar el temps com una nova variable explicativa.

Per a entendre millor la notació que emprarem d'ara endavant hem de definir la taula de mortalitat. Segons [60], es tracta de una sèrie temporal que indica la reducció gradual d'un grup inicial d'individus a causa de les defuncions, i ha sigut construïda a partir de les observacions d'un col·lectiu real. L'estructura bàsica, tal com ens la descriu [82], està constituïda per les columnes següents,

$x$	$l_x$	$d_x$	$q_x$	$p_x$
-----	-------	-------	-------	-------

$x$ : l'edat de l'individu,  $0 \leq x \leq \omega$ , on  $\omega$  és l'edat límit

$l_x$ : nombre dels qui sobreviuen a l'edat  $x$

$d_x$ : nombre dels qui moren entre les edats  $x$  i  $x + 1$ ,

$$d_x = l_x - l_{x+1}$$

$q_x$ : ràtio anual de defuncions a l'edat  $x$ ,

$$q_x = \frac{d_x}{l_x},$$

$p_x$ : ràtio anual de supervivència a l'edat  $x$

$$p_x = \frac{l_{x+1}}{l_x}.$$

A partir de les taules de mortalitat, pot obtenir-se l'esperança de vida

$$e_x = \frac{T_x}{l_x},$$

on  $l_x$  i  $T_x$  representen, respectivament, el total de persones que han complert  $x$  anys i el total d'anys que totes elles esperen viure,  $T_x = \sum_{i \geq x} L_i$ , sent  $L_x = l_{x+1} + d_x/2$  el nombre corresponent de persones-anys. El càlcul del valor de la vida mitjana o vida probable ha sigut i és, sens dubte, un indicador de la capacitat de supervivència d'una societat, i el fet que s'incremente suposa, en tots els aspectes, una millora de les condicions de vida d'una societat, raó per la qual ens ocuparem també de l'esperança de vida i de l'estudi d'aquesta.

El contingut d'aquest article està estructurat de la manera següent. En la secció 2 es descriu l'evolució de la mortalitat espanyola en el darrer segle. La secció 3 es dedica a la descripció dels models dinàmics. En l'apartat 3.1 descriu els models estructurals [43, 48]. En l'apartat 3.2 estudiem la influència del temps del calendari mitjançant els models no estructurals: les funcions Gompertz-Makeham ajustades per mitjà de models lineals generalitzats (d'ara endavant GLM) i els models amb factors de reducció de la mortalitat. La secció 4 està dedicada a la predicció de la mortalitat emprant intervals de confiança. En la secció 5 es descriuen els models edat-període-cohort. La darrera secció està dedicada a tendències futures en la graduació de la mortalitat.



FIGURA 1: Gràfic de descens de la mortalitat per a algunes edats.

## 2 Evolució de la mortalitat

La secció està dedicada a descriure l'evolució de la mortalitat a Espanya durant el segle passat, com a justificació a la introducció dels models dinàmics i la necessitat d'una predicció adequada. Les dades utilitzades per a mostrar

l'evolució de la mortalitat a Espanya provenen de [42], i corresponen al període 1908-2002 i a un rang d'edats de 0 a 110. En la figura 1 pot observar-se com, en general, les probabilitats de mort han disminuït al llarg del temps, tot i que ho han fet de manera diferent per als distints grups d'edat. A Espanya, igual que en altres països desenvolupats, cal destacar el descens que s'ha produït en la mortalitat infantil, l'augment de mortalitat en l'última dècada per a les edats intermèdies, al voltant dels trenta anys, i l'estabilitat, i fins i tot un lleuger augment, per a les edats elevades, noranta-cinc i més anys, a causa de l'augment de la població longeva que s'ha produït en els darrers temps.

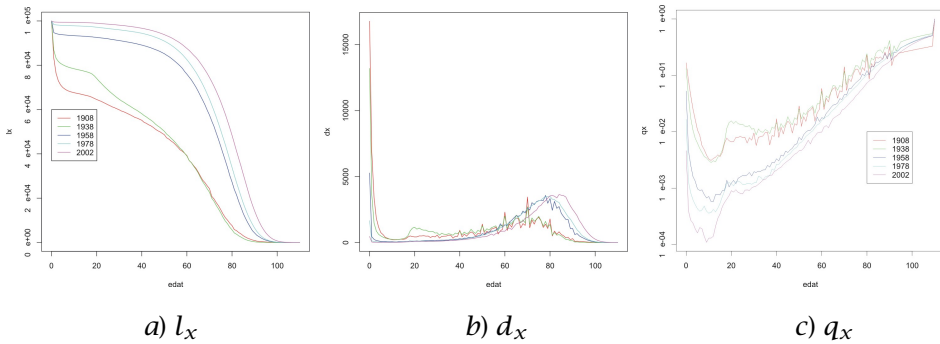


FIGURA 2: Rectangularització i expansió per als homes.

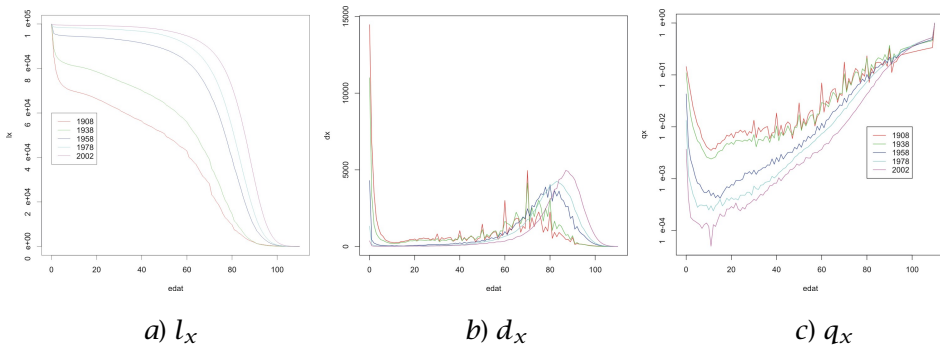


FIGURA 3: Rectangularització i expansió per a les dones.

Les darreres tendències pel que fa a l'anàlisi de la mortalitat han estat descrites entre d'altres per [59], que defineix els processos de *rectangularització* i de *expansió* de la corba de supervivents. L'expansió consisteix en el desplaçament de la corba de supervivents cap a les edats elevades, tal com s'aprecia en les figures 2 a) i 3 a), la qual cosa implica un desplaçament de la moda (figures 2 b) i 3 b)) cap a aquestes edats. La rectangularització és deguda a

la concentració de les morts al voltant de la dita moda i té com a conseqüència una transformació de la corba de supervivència, que sembla un rectangle. El fenomen s'observa en les figures 2 a) i 3 a).

Adicionalment, apareixen nivells de mortalitat alts i gran dispersió en les edats joves i intermèdies en les figures 2 c) i 3 c), particularment per als homes, fenomen que es coneix com la *gepa dels accidents* perquè es produeix a causa dels accidents de trànsit.

Tot aquest procés ha anat acompanyat d'un increment de les esperances de vida, que pot observar-se en la figura 4.

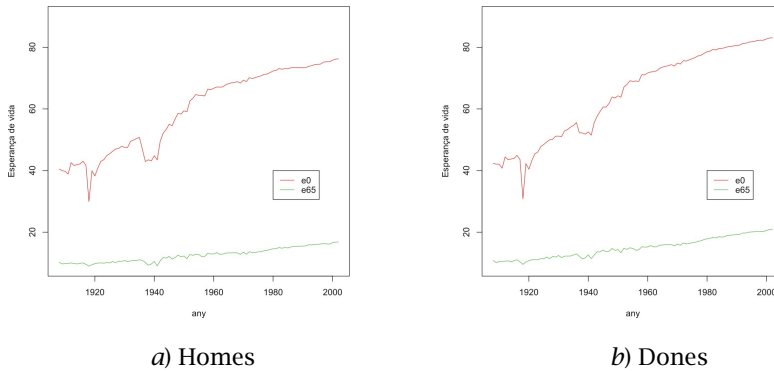


FIGURA 4: Evolució de l'esperança de vida.

### 3 Models dinàmics

Considerem un conjunt de ràtios de mortalitat en forma de taula dinàmica, és a dir, una taula de doble entrada les cel·les de la qual contenen el valor de  $q_{xt}$ , ràtio de mortalitat corresponent a l'edat  $x$  i l'any  $t$ . El nostre objectiu és obtenir  $\hat{q}_{xt}$ , estimacions suaus de les desconegudes i vertaderes probabilitats de mort  $q_{xt}$ , a partir de les estimacions brutes,  $\hat{q}_{xt}$ . Aquestes estimacions brutes s'obtenen a partir del nombre de morts observades,  $d_{xt}$ , i el nombre dels inicialment exposats al risc,  $E_{xt}$  [27].

Una recopilació de les tècniques d'anàlisi i de construcció de taules dinàmiques podem trobar-la en [3, 78, 30, 76, 63, 27]. Para una atenció especial al risc que la longevitat té per a l'assegurador [63], a més de recopilar les contribucions més recents. Una mostra de l'actualitat del tema pot veure's en [6], en què es descriuen les propostes més actuals.

Els mètodes de graduació de taules dinàmiques poden classificar-se en dos grans grups:

1. *Models paramètrics*. Són models que ajusten a les mesures de la mortalitat una funció  $f$  dependent d'uns paràmetres. Els tractaments, bàsicament, són dos:

- a) considerar que la influència del temps del calendari tan sols afecta els paràmetres (models que denominem *estructurals*) o bé
  - b) incorporar el temps del calendari com a variable  $t$  en la funció (són els models que denominem *no estructurals*).
2. *Models no paramètrics*. Són generalitzacions de les tècniques d'allisatge (*smoothing*) que depenen de l'edat i del temps.

En la construcció de les taules dinàmiques per a l'estudi de la mortalitat, els models són majoritàriament paramètrics, perquè faciliten la comparació al llarg del temps [20]. Els models no paramètrics s'han utilitzat habitualment per a l'anàlisi exploradora de les dades, prèvia a la graduació paramètrica. Un exemple de modelització no paramètrica prèvia és l'article [31], en què es combinen idees bàsiques de processos estocàstics [52, 53, 54] amb l'ús d'un allisatge basat en el nucli (*smoother kernel*) ja descrit per al cas univariant en [35, 36, 37]. Més recentment, un treball del *Continuous Mortality Investigation (CMI) Bureau* [56] recomana la utilització al Regne Unit de mètodes semiparamètrics basats en *p-splines* [22].

### 3.1 Models estructurals

Els models estructurals segueixen els dos passos següents:

1. ajusten la mesura de mortalitat emprant el mateix model per a tots els anys, amb la qual cosa s'obté una sèrie temporal de paràmetres estimats, i a continuació
2. analitzen la sèrie temporal de cada paràmetre.

Les sèries temporals ajustades s'utilitzen per a obtenir estimacions dels paràmetres per a anys futurs, que ens permeten fer prediccions si substituïm, en la llei de mortalitat, els paràmetres per aquestes estimacions. Un model d'aquestes característiques és el model *logit*<sup>1</sup> proposat per [9], la descripció i aplicació del qual pot trobar-se en [4]. Es tracta, tanmateix, de metodologia merament empírica [30], de la qual no ens ocuparem. Sí que ho farem, en canvi, de propostes més depurades, com les dues que venen a continuació.

**3.1.1 Model de Heligman i Pollard.** Les lleis de Heligman i Pollard, [43], han sigut utilitzades àmpliament per diferents països del nostre entorn europeu (Anglaterra, Suècia, Alemanya i Espanya) i per altres països desenvolupats (els Estats Units d'Amèrica i Austràlia) des que l'ONU va promoure l'ajust de la mortalitat mitjançant la primera llei de Heligman i Pollard [30]. Els autors, inspirant-se en [77], ajusten una nova llei de mortalitat en l'Austràlia de la

---

<sup>1</sup> La funció *logit* va ser introduïda per Joseph Berkson el 1944 com  $\text{logit}(p) = \log(p/(1-p))$ , la inversa de la funció logística que apareix en els primers models de creixement de població. (Nota de l'editor).

postguerra, l'expressió genèrica de la qual és

$$\frac{q_x}{1 - q_x} = \sum_{i=1}^n A_i \exp(-B_i(f_i(x) - C_i)^{D_i}),$$

on  $A_i, B_i, C_i, D_i$ ,  $i = 1, 2, \dots, n$ , són els paràmetres que cal estimar. Normalment amb  $n = 3$  s'obtenen bons ajustos. Per a  $f_i(x)$  sol utilitzar-se  $x$  o  $\ln(x)$ .

Les tres expressions que realment es van ajustar a la mortalitat australiana van ser:

- Primera llei de Heligman i Pollard

- versió 1

$$\frac{q_x}{1 - q_x} = A^{(x+B)^C} + D \exp(-E(\ln x - \ln F)^2) + GH^{(x-x_0)},$$

expressió que no pot distingir-se, consideren els autors, de la

- versió 2

$$q_x = A^{(x+B)^C} + D \exp(-E(\ln x - \ln F)^2) + \frac{GH^x}{1 + GH^x}.$$

- Segona llei de Heligman i Pollard

$$q_x = A^{(x+B)^C} + D \exp(-E(\ln x - \ln F)^2) + \frac{GH^x}{1 + KGH^x}. \quad (1)$$

- Tercera llei de Heligman i Pollard

$$q_x = A^{(x+B)^C} + D \exp(-E(\ln x - \ln F)^2) + \frac{GH^{x^k}}{1 + GH^{x^k}}.$$

Cada un dels tres termes de l'equació bàsica representa un component diferent de la mortalitat: el primer terme la mortalitat infantil, el segon la *gapa dels accidents* i el tercer la mortalitat natural causada per senectut [43]. El gràfic de la figura 5 mostra aquesta descomposició. La interpretació dels paràmetres és la següent:  $A$  representa la ràtio de mortalitat infantil;  $B$  representa la probabilitat de morir per a un infant d'un any d'edat;  $C$  està relacionat amb l'adaptació dels individus a l'entorn. Els tres prenen valors en l'interval  $[0,1]$ .  $D$ ,  $E$  i  $F$  es refereixen a la *gapa dels accidents*,  $D$  indica la severitat de la *gapa* i prenen valors en  $[0,1]$ ;  $E$ , amb valors entre  $[0, \infty[$ , indica la concentració de la *gapa*, i  $F$ , des de quinze fins a edats avançades, indica la localització del màxim de la *gapa*. Finalment,  $G$  indica el nivell base de la mortalitat senil, i  $H$  és la taxa de creixement de la dita mortalitat senil, i els dominis respectius són  $[0,1]$  i  $[0, \infty[$ .

Els paràmetres s'estimen per a cadascun dels anys pel mètode dels mínims quadrats ponderats no lineals per a tot el rang d'edats,

$$\sum_x \omega_x (q_x - F(x))^2,$$

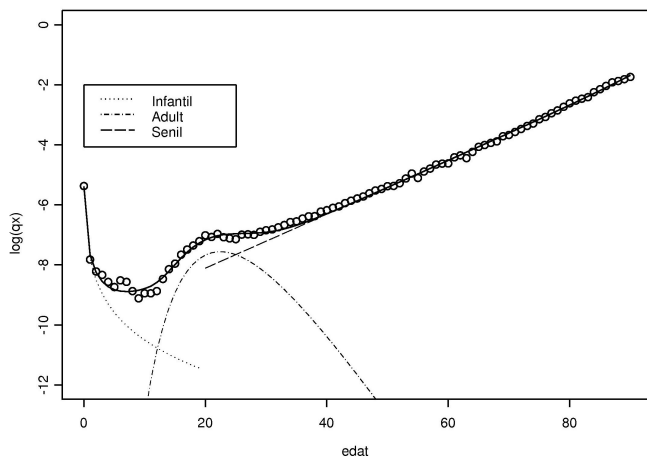


FIGURA 5: Descomposició de la llei de Heligman i Pollard.

on  $\omega_x^{-1}$  és proporcional a la variància de l'observació a l'edat  $x$ ,  $F(x)$  és la funció (1) que es vol ajustar i  $q_x$  són les probabilitats de mort observades. La necessitat d'introduir pesos en l'ajust es deu a la desigualtat de variàncies per a  $q_x = F(x)$ , ja que com que hem triat el model binomial,  $var(q_x) = \frac{q_x(1 - q_x)}{E_x}$ . Per aquesta raó els pesos proposats són  $1/q_x$  o alguna potència seua, com per exemple es fa en [30] i ja van fer Heligman i Pollard. L'últim pas per a la predicció és ajustar una sèrie temporal per a cadascun dels paràmetres.

Una aproximació bayesiana per a les lleis de Heligman i Pollard utilitzant mètodes de MCMC per a evitar els problemes numèrics que tenen els mètodes clàssics pot trobar-se en [29].

La utilització que es fa en [55] del model de Heligman i Pollard ha sigut adaptada al model espanyol per [30] i [32], que n'empren només la segona llei.

**3.1.2 Model de Lee-Carter.** Des que Lee i Carter, [48], van proposar el seu model de projecció el 1992 i el van utilitzar per a l'estudi de la mortalitat als EUA, han sigut nombrosos els treballs que l'han emprat per fer prediccions de la mortalitat en altres països desenvolupats, com ara el Canadà [50], Xile [49], el Japó [86], Bèlgica [11], Àustria [16], Anglaterra i Galles [69], Austràlia [8] i Espanya [40] o [25], alguns d'ells en la versió original i d'altres millorant-lo.

El model de Lee-Carter consisteix a ajustar a la mesura de mortalitat la funció

$$q_{xt} = \exp(a_x + b_x k_t + \epsilon_{xt})$$



o, equivalentment, la funció

$$\ln(q_{xt}) = a_x + b_x k_t + \epsilon_{xt},$$

al seu logaritme.

En les dues expressions anteriors, el doble subíndex es refereix a l'edat,  $x$ , i a l'any o la unitat de temps,  $t$ ;  $a_x$  i  $b_x$  són paràmetres dependents de l'edat; i  $k_t$  és un índex específic per a cada any o unitat de temps. Els errors  $\epsilon_{xt}$ , amb mitjana zero i variància  $\sigma_\epsilon^2$ , reflecteixen influències històriques que no són capturades pel model.

Quant a l'ajust, observem que, donada una solució qualsevol  $(a_x, b_x, k_t)$ , són també solucions,

$$(a_x, b_x/c, ck_t) \text{ o } (a_x + cb_x, b_x, k_t - c). \quad (2)$$

Perquè el model tinga una solució única els paràmetres han de ser normalitzats,  $\sum_x b_x = 1$  i  $\sum_t k_t = 0$ , i llavors s'obté,

$$\check{a}_x = \frac{\sum_t \ln q_{xt}}{T}.$$

Els valors  $b_x$  i  $k_t$  s'estimen pel mètode *singular value decomposition* (d'ara endavant SVD), aplicat a la matriu  $\ln(q_{xt}) - \check{a}_x$ ,

$$\ln(q_{xt}) - \check{a}_x = \sum_{i=1}^{\min(n,T)} s^i u_x^i v_t^i,$$

amb  $n$  nombre d'edats, i  $s^i$ ,  $u_x^i$  i  $v_t^i$  els valors singulars i els vectors singulars esquerre i dret respectius, on el superíndex indica l'ordre [70]. Així,  $\hat{k}_t^{SVD} = s^1 v_t^1$  i  $\hat{b}_x^{SVD} = u_x^1$ . Com que  $\sum_x \hat{b}_x^{SVD} = c \neq 1$ , normalitzem amb la transformació

$$\hat{b}_x = \frac{\hat{b}_x^{SVD}}{c} \text{ i } \check{k}_t = c \hat{k}_t^{SVD}.$$

Posteriorment,  $k_t$  es reestima per a ajustar el nombre total de morts observades i estimades, fent servir l'equació

$$D_t = \sum_x (N_{xt} \exp(\hat{a}_x + k_t \hat{b}_x)), \quad (3)$$

on  $D_t$  és el nombre total de morts l'any  $t$  i  $N_{xt}$  és la població d'edat  $x$  l'any  $t$ . La solució de (3) és  $\check{k}_t$ , que, amb la translació de (2),

$$\hat{k}_t = \check{k}_t - \text{mitjana}(\check{k}_t) \quad \text{i} \quad \hat{a}_x = \check{a}_x + \hat{b}_x \text{mitjana}(\check{k}_t),$$

ens dona la solució final  $(\hat{a}_x, \hat{b}_x, \hat{k}_t)$ .

En [85] s'estimen els paràmetres per mínims quadrats ponderats, en [11] per màxima versemblança i en [23] mitjançant GLM. Tots aquests s'han comparat i implementat per a dades espanyoles en [25]. Més recentment, [24] i [61] introdueixen una estimació bayesiana dels paràmetres, en el segon cas fent ús de models *state-space*.

L'últim pas del mètode de Lee-Carter consisteix a trobar un model de sèrie temporal per als índexs de mortalitat,  $\{\hat{k}_t\}$ , utilitzant metodologia Box-Jenkins. El model ajustat permet predir valors futurs per a  $k_t$  i amb aquests, al seu torn, obtenir prediccions per a  $q_{xt}$ . En moltes de les aplicacions ha donat bon resultat el model

$$\hat{k}_t = p + \hat{k}_{t-1} + u_t,$$

on  $p$  és una constant i  $u_t$  un soroll blanc.

La principal crítica al model de Lee-Carter és que els paràmetres  $a_x$  i  $b_x$  són els mateixos al llarg del temps i que la predicció de futurs valors de la mortalitat es basa només en  $k_t$ , la qual cosa suposa admetre que no hi ha interacció entre l'edat i el temps. Els avantatges que presenta són, entre d'altres, la fàcil interpretació dels paràmetres i la parsimònia (vegeu-ne la descripció detallada en [47] o [7]). El model, tanmateix, gaudeix actualment de molta popularitat gràcies als bons resultats que proporciona i a la seua simplicitat, per la qual cosa hi ha una àmplia literatura dedicada a implementar-lo i millora-lo.

L'estudi realitzat per [7] introdueix tres modificacions al model de Lee-Carter. Concretament, una millora de l'ajust del model, la identificació del període òptim per a l'ajust del model ARIMA i la incorporació d'interaccions edat-any en la base del model utilitzant els dos primers vectors singulars esquerres i drets de la matriu  $\ln(q_{xt}) - \hat{a}_x$ , com també es fa en [70]. L'anàlisi es du a terme utilitzant un codi específic de R [66]. L'article [51] considera modificacions al mètode de Lee-Carter per a predir la mortalitat per a països tot considerant que formen part d'un grup enlloc de considerar-los individualment.

Diguem, per acabar, que un bon ajust no sempre és garantia d'una bona predicció, com s'assenyala en [76]. Així, el model amb vuit paràmetres desenvolupat per [43] ajusta bé a un gran nombre de patrons de mortalitat, però la gran quantitat de paràmetres que requereix limita el model per a potencials projeccions [7].

**Aplicació del model de Lee-Carter a dades de mortalitat espanyoles.** En la figura 6 es mostra l'estimació del model de Lee-Carter per a dades espanyoles de mortalitat corresponents al període 1980-1999 i un rang de edats de 0 a 105. La comparació dels paràmetres  $a_x$  per a ambdós sexes mostra que la mortalitat per a les dones és inferior. La gepa en la figura 6 *a*) revela un increment de la mortalitat en el rang d'edat d'11 a 40 per als homes que alguns autors [40] atribueixen a la mortalitat accidental.

Els valors negatius per als paràmetres  $b_x$ , figura 6 *b*), per a les edats intermèdies (de 24 a 40 per a homes i de 28 a 32 per a dones) indiquen que la mortalitat en aquests grups de edat s'incrementa amb els anys. Per al rang d'edats de 24

a 40, aquest increment pot explicar-se per l'efecte de l'epidèmia de la sida, com assenyalen [32] i [40]. El gràfic dels valors estimats de  $k_t$ , figura 6 c), evidencia que el decreixement és més pronunciat per als homes que per a les dones.

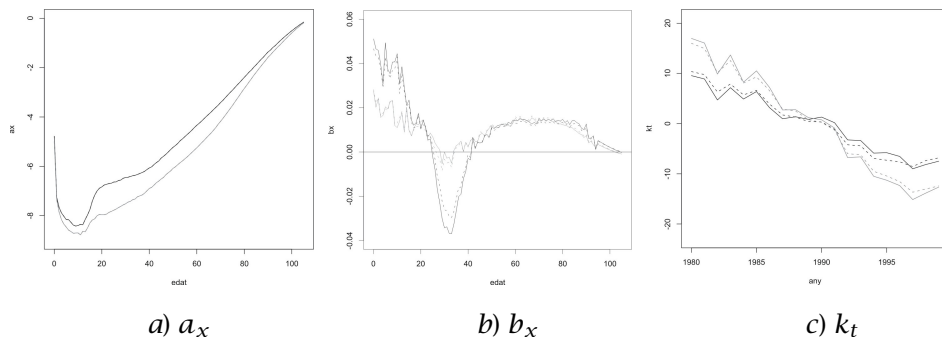


FIGURA 6: Valors estimats dels paràmetres de Lee-Carter (línia contínua) i Lee-Carter-GLM (línia discontinua) per a Espanya (homes en negre i dones en gris).

## 3.2 Models no estructurals

Una representació de l'evolució de la mortalitat, diferent de l'anterior, és la que plantegen els mètodes no estructurals, en els quals el temps està explícitament inclòs en la funció de mortalitat. Un model d'aquestes característiques és el model logarítmic descrit i utilitzat en [4], del qual no ens ocuparem aquí, ja que ens estímem més parlar de propostes més actuals.

**3.2.1 Funcions Gompertz-Makeham respecte a l'edat i el temps.** En [73] es proposa adaptar les funcions Gompertz-Makeham a les taules dinàmiques mitjançant la inclusió del temps com a variable. El nou model adaptat per a  $q_{xt}$  és

$$\ln\left(\frac{q_{xt}}{1-q_{xt}}\right) = \beta_0 + \sum_{j=1}^s \beta_j L_j(x') + \sum_{i=1}^r \alpha_i t'^i + \sum_{i=1}^r \sum_{j=1}^s \gamma_{ij} L_j(x') t'^i, \quad (4)$$

i està subjecte a la convenció que alguns dels termes  $\gamma_{ij}$  poden ser zero. En (4)  $x'$  i  $t'$  són transformacions de l'edat i l'any del calendari, respectivament, de manera que els seus valors estiguen entre  $[-1, 1]$ , i  $L_j(x')$  són els polinomis de Legendre generats per  $L_{n+1}(x) = xL_n(x) - nL_{n-1}(x)$ , on  $n \geq 1$ ,  $L_0(x) = 1$  i  $L_1(x) = x$ .

Reescrivint l'equació (4) de la manera

$$\frac{q_{xt}}{1-q_{xt}} = \exp\left[\beta_0 + \sum_{j=1}^s \beta_j L_j(x')\right] \exp\left[\sum_{i=1}^r \left(\alpha_i + \sum_{j=1}^s \gamma_{ij} L_j(x')\right) t'^i\right], \quad (5)$$

el primer terme pot interpretar-se com una funció logit Gompertz-Makeham corresponent a la graduació mitjançant l'edat,  $LGM(0, s + 1)$  (vegeu-ne la descripció detallada en [34] o [26]). El segon terme pot ser interpretat com un terme d'ajust de l'efecte de l'any del calendari, de manera que quan almenys un dels  $\gamma_{ij}$  és no nul depèn també de l'edat. Un exemple d'això seria l'increment de morts d'homes adults i joves a causa de la sida.

Per a ajustar la funció (4) sota l'esquema de GLM, hem de suposar que  $D_{xt}$ , nombre de morts a l'edat  $x$  en l'any  $t$ , segueix una distribució binomial,  $D_{xt} \sim Bi(E_{xt}, q_{xt})$ , amb valors observats  $d_{xt}$ . Els resultats del model dinàmic seran avaluats amb la *deviance*.<sup>2</sup> La versemblança val

$$\begin{aligned} L(q) &= \prod_t \prod_x \binom{E_{xt}}{d_{xt}} q_{xt}^{d_{xt}} (1 - q_{xt})^{E_{xt} - d_{xt}} \\ &\propto \prod_t \prod_x q_{xt}^{d_{xt}} (1 - q_{xt})^{E_{xt} - d_{xt}} \end{aligned}$$

i el logaritme de la versemblança és

$$\ln L(q) = \sum_t \sum_x (d_{xt} \ln(q_{xt}) + (E_{xt} - d_{xt}) \ln(1 - q_{xt})).$$

A partir del logaritme de la versemblança podem calcular la *deviance*,

$$D(\hat{q}) = 2 \ln L(\hat{q}_{xt}) - 2 \ln L(\hat{q}_{xt}).$$

Quan comparem dos models, la diferència de *deviance* segueix aproximadament una distribució  $\chi^2$  que té com a graus de llibertat la diferència entre els que corresponen a cada model; així doncs hem de triar aquells models que produïsquen millores significatives de la *deviance*.

La manera de procedir per a determinar els paràmetres  $\alpha_i$ ,  $\beta_j$  i  $\gamma_{ij}$  és considerar l'esquema de GLM amb família binomial i enllaç logit, per a això,

1. els valors triats de  $r$  i  $s$  són aquells a partir dels quals els increments de la *deviance* no són significatius,
2. els coeficients  $\gamma_{ij}$  es trien de manera que l'increment de la *deviance* resulte significatiu,
3. paral·lelament es determinen els errors estàndard de les estimacions dels paràmetres i la significació d'aquestes mitjançant la prova usual  $t$  de Student.

Com s'assenyala en [75], que aplica aquest model, l'objectiu principal no sols és trobar un model que proporcione un bon ajust de les dades, sinó trobar-ne un que també proporcione un bon esquema per a realitzar projeccions. Amb aquest objectiu en [87] s'arriba a la conclusió

<sup>2</sup> El terme *deviance* fa referència al logaritme de la versemblança, multiplicat per dos i canviat de signe. D'aquesta manera s'obté una variable que segueix una distribució  $\chi^2$ .

1. els valors òptims de  $r$  i  $s$  no sempre generen tendències plausibles per a l'objectiu de projectar, els ordres dels polinomis han de ser baixos encara que calga sacrificar, si escau, una major bondat de l'ajust;
2. per a tots els conjunts de dades que van analitzar, els resultats més satisfactoris els van obtenir amb  $r = 1$  i  $s = 3$ , amb interacció d'ordre 1.

Reescriuint l'equació (5) d'acord amb aquestes conclusions, tenim

$$\frac{q_{xt}}{1 - q_{xt}} = \exp \left[ \beta_0 + \sum_{j=1}^3 \beta_j L_j(x') \right] \exp [(\alpha_1 + \gamma_{11} x') t']. \quad (6)$$

**Aplicació de les funcions Gompertz-Makeham a dades de mortalitat espanyoles.** Les estimacions dels coeficients del model (6) per a les dades espanyoles de mortalitat esmentades abans es mostren en la taula 1. Com pot comprovar-se, tots són significatius.

Espanya						
Homes			Dones			
	coef.	error estàndard	prova t	coef.	error estàndard	prova t
$\beta_0$	-4,379241	0,001592	-2751,625	-4,965508	0,002020	-2458,30
$\beta_1$	3,696105	0,002917	1267,210	3,959519	0,003151	1256,74
$\beta_2$	1,043690	0,002586	403,668	1,687126	0,002918	578,19
$\beta_3$	-0,476037	0,003090	-154,077	-0,681732	0,003318	-205,46
$\alpha_1$	-0,118890	0,001868	-63,651	-0,244256	0,002539	-96,19
$\gamma_{11}$	0,015907	0,003794	4,193	0,146082	0,004395	33,24

TAULA 1: Coeficients per a la funció Gompertz-Makeham.

Hi ha també una versió estructural per a funcions Gompertz-Makeham que consisteix a estimar els paràmetres del model per a cada any i després ajustar un model ARIMA a cadascuna de les sèries dels paràmetres, aquesta versió pot consultar-se en [74]. També hi ha una versió bayesiana els detalls de la qual poden trobar-se en [45].

Assenyalem, finalment, que aquest model, en les diferents versions que presenta, ha sigut utilitzat per a valorar l'efecte de la millora de la mortalitat sobre anualitats [45] i sobre opcions [2].

**3.2.2 Models basats en factors de reducció de la mortalitat** Els *factors de millora de la mortalitat*,  $RF(x, t)$ , són valors que permeten projectar les taules de mortalitat al llarg del temps perquè tenen en compte els canvis que la mesura de mortalitat considerada ha patit amb el pas dels anys. El procediment es du a terme en dos passos:

1. les dades d'un període determinat són graduades per a l'obtenció de les taules base mitjançant algun dels procediments exposats en [34, 67],
2. es construeixen les taules de mortalitat projectades aplicant els factors de reducció  $RF(x, t)$ , per a un individu que ha assolit l'edat  $x$  en el temps  $t$ , on  $t$  es mesura en anys a partir d'un origen apropiat,  $t = 0$ , situat en el centre del període base.

La probabilitat de mort projectada al temps  $t$  ve donada per

$$q_{xt} = q_{x0}RF(x, t), \quad (7)$$

on  $q_{x0}$  és el tant de mortalitat de la taula base. Aquesta expressió caracteritza els factors de reducció per a  $q_{xt}$  sota les restriccions

$$\begin{aligned} RF(x, 0) &= 1 \quad \forall x \geq 0 \\ 0 < RF(x, t) &\leq 1 \quad \forall x \geq 0, \quad \forall t \geq 0. \end{aligned}$$

Els models emprats per a representar el comportament dels factors de reducció que en permeten també l'estimació i la predicció per a anys futurs han sigut recopilats en diversos treballs. Considerem necessari citar, entre aquests, la secció 4 de l'informe 10 del CMI [12], que conté una descripció completa del mètode. Posteriorment el CMI Bureau ha proposat un nou model per a la projecció de la mortalitat per a pensionistes i anualitats (informe 17, [13]), amb taules de mortalitat basades en experiències del període 1991-1994. En [75] comparen el segon terme de l'equació (5),  $RF_{GLM}(x, t)$ , amb els  $RF_{CMI}(x, t)$  obtinguts mitjançant el mètode habitual del CMI Bureau, per a analitzar la consistència de l'un amb l'altre. Citem, finalment, els treballs de Renshaw i Haberman [68, 70, 71]. En el primer d'aquests es presenta un esquema alternatiu per a modelitzar i predir factors de reducció de la mortalitat basat en GLM. Segons aquest esquema, en l'equació (7) considerem el model binomial amb enllaç logit per a ajustar a  $q_{xt}$  el predictor,

$$\eta_{xt} = g(q_{x0}) + \beta_x t, \quad (8)$$

on  $g(q_{x0}) = \log\left(\frac{q_{x0}}{1 - q_{x0}}\right)$ , que es calcula aplicant l'enllaç als valors graduats del període base. En els dos següents s'adapta la metodologia de Lee-Carter a la construcció de factors de reducció de la mortalitat. Entre les últimes aportacions cal esmentar el treball [72], que, utilitzant el model de Lee-Carter, estudia l'extensió dels models basats en factors de reducció de la mortalitat [71, 70], per a incloure-hi l'efecte cohort, que són els individus nascuts el mateix any. De l'efecte cohort parlarem amb detall més endavant.

**Aplicació dels models amb factors de reducció a dades de mortalitat espanyols.** En aquesta secció apliquem el càlcul i la modelització dels factors de reducció de la mortalitat segons [68] a les anteriors dades espanyoles de mortalitat. Prenem 1980 com a origen ( $t = 0$ ) perquè és l'any inicial. No graduarem les dades, ens interessa tan sols estudiar l'evolució dels  $q_{xt}$  suposant linealitat.

Considerem la probabilitat de mort  $q_{xt}$  per als anys del 1980 al 1999 amb  $t = \text{any} - 1980$ . Els factors de reducció estan caracteritzats per l'equació (7), i per a modelitzar-los segons GLM hem de considerar el model binomial amb enllaç logit per a ajustar a  $q_{xt}$  el predictor,

$$\eta_{xt} = g(q_{x0}) + \beta_x t,$$

on  $g(q_{x0}) = \log\left(\frac{q_{x0}}{1 - q_{x0}}\right)$  perquè es calcula aplicant l'enllaç als valors del període base 1980. En la figura 7 mostrem els valors dels  $\beta_x$  per a cada edat i cada sexe.

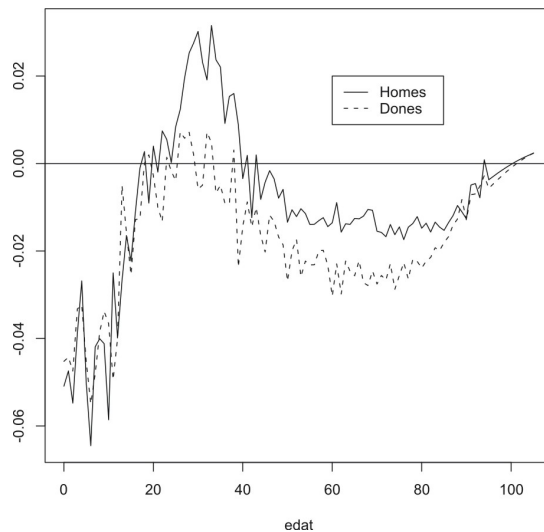


FIGURA 7: Paràmetres  $\beta_x$  per a homes i dones en les dades espanyoles de mortalitat.

La reducció de la mortalitat és semblant en ambdós sexes per a les edats de 0 a 20, després apareixen clares diferències. En el cas dels homes de 20 a 40 hi ha un increment de la mortalitat, mentre que en les dones l'increment té lloc per a edats puntuals. Per a l'interval de 40 a 100 hi ha una major reducció de la mortalitat per a les dones, i en les edats finals se n'observa un lleuger increment.

#### 4 Predicció de la mortalitat per intervals de confiança

La principal crítica als models de predicció utilitzats oficialment és la sistemàtica sobreestimació de la mortalitat i la falta de mesures de sensibilitat i d'incertesa. En [61], entre d'altres, s'argumenta que les mesures d'incertesa en la predicció de la mortalitat són necessàries i es puntualitza que han de dur-se

a terme mitjançant intervals de confiança per tal d'expressar la incertesa com una probabilitat.

En el seu article original Lee i Carter construeixen intervals de confiança per a les prediccions de l'esperança de vida a partir tan sols de l'error en la predicció de l'índex de mortalitat  $k_t$ , de manera que no tenen en compte altres fonts d'error, com ara la que introdueix l'estimació dels paràmetres del model. Una manera de combinar aquestes dues fonts d'incertesa és utilitzar procediments de *bootstrap*, com es fa en [10] i [46]. En el primer treball els autors utilitzen un *bootstrap* paramètric, mentre que en el segon s'utilitza un *bootstrap* no paramètric.

En el cas espanyol aquesta metodologia ha sigut utilitzada per [25] per a prediccions obtingudes a partir del model de Lee-Carter amb un i dos termes SVD i comparant els dos tipus de *bootstrap*. En ambdós casos s'utilitza la distribució binomial, mentre que en els treballs [10] i [46], s'utilitza la distribució de Poisson. El procediment emprat és el següent: a partir de les observacions  $(E_{xt}, d_{xt})$ , se simulen  $N$  mostres *bootstrap*  $(E_{xt}, d_{xt}^n)$ ,  $n = 1, 2, \dots, N$ , on  $d_{xt}^n$  són realitzacions d'una binomial de paràmetres  $(E_{xt}, \hat{q}_{xt})$ . Per a cada mostra *bootstrap* s'estimen els  $a_x$ ,  $b_x$  i a continuació es projecten els  $k_t$  a partir del model ARIMA ajustat amb les dades originals. Tenim així  $N$  realitzacions d' $a_x^n$ ,  $b_x^n$ ,  $k_t^n$  i  $k_t^n$  projectats que s'empren per a projectar els  $q_{xt}$  i calcular l'esperança de vida. L'interval de confiança és el que deriva dels percentils,  $IC_{95} = [p_{0.025}, p_{0.975}]$ .

Per al cas no paramètric, les  $N$  mostres *bootstrap* se simulen a partir dels residus

$$\hat{\epsilon}_{xt} = \ln \left( \frac{q_{xt}}{1 - q_{xt}} \right) - \left( \hat{a}_x + \sum_{i=1}^r \hat{k}_t^i \hat{b}_x^i \right),$$

obtinguts amb les dades originals. Cada mostra proporciona unes estimacions  $\widehat{\text{logit}}(q_{xt})^n$  a partir de la fórmula inversa

$$\widehat{\text{logit}}(q_{xt})^n = \text{logit}(\hat{q}_{xt}) - \hat{\epsilon}_{xt}^n,$$

on els  $\text{logit}(\hat{q}_{xt})$  s'obtenen de les observacions inicials  $(E_{xt}, d_{xt})$ . A partir d'aquest punt es procedeix igual que en el cas paramètric. L'extensió dels mètodes *bootstrap* a altres models és immediata.

**Aplicació de tècniques *bootstrap* a dades de mortalitat espanyoles.** En la figura 8 es mostren els intervals de confiança per a la predicció, en el període 2000-2011, de l'esperança de vida residual als seixanta-cinc anys d'edat,  $e_{65t}$ , obtinguts mitjançant les tècniques de *bootstrap* descrites. Pot observar-se en el gràfic que l'esperança de vida residual és major per a les dones que per als homes i que hi ha un clar augment d' $e_{65t}$  al llarg del temps. Pel que fa al mètode paramètric, l'amplitud dels intervals és menor en aquest cas que la dels no paramètrics. Més detalls i l'extensió a altres models, com el Lee-Carter amb dos SVD, pot consultar-se en [25].



Hem de destacar que els intervals obtinguts per a l'esperança residual amb tots dos mètodes són sorprenentment estrets, fet que ha cridat també l'atenció de diferents investigadors com ara [48, 47, 7, 46], que tracten de trobar-hi alguna explicació. En el nostre cas, els resultats estan d'acord amb l'evolució de valors que publica l'Institut Nacional d'Estadística (INE).

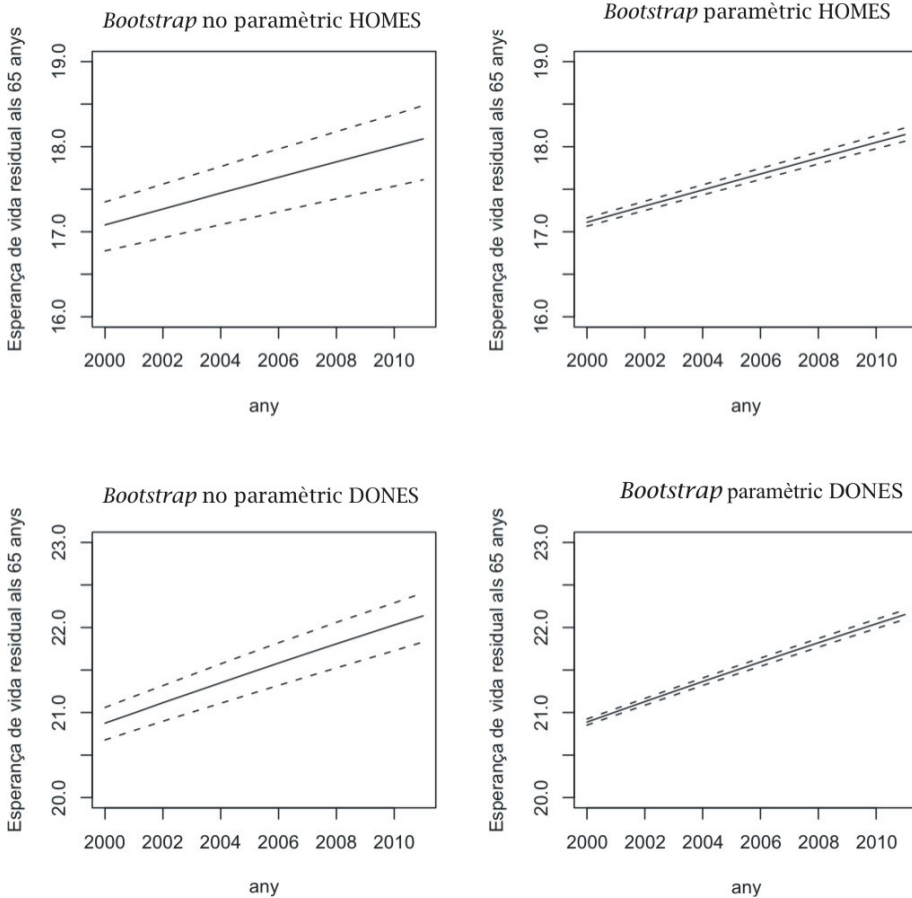


FIGURA 8: Intervals de confiança per a l'esperança de vida residual als seixanta-cinc anys.

## 5 Efecte cohort

Tabeau *et al.* [76] descriuen en el seu llibre els models de graduació i predicció desenvolupats recentment i conclouen la necessitat d'integrar tècniques de

distintes disciplines amb l'objectiu d'aconseguir prediccions satisfactòries. En [57] s'aconsella que els actuaris adoptin els models edat-període-cohort (d'ara endavant APC), [17, 18, 44], que han mostrat bons resultats en el camp de l'epidemiologia. Aquests models constitueixen una evolució natural dels models dinàmics, perquè incorporen l'any de naixement (cohort). Els models APC són, d'altra banda, una extensió natural dels models edat-període, AP, i edat-cohort, AC, en considerar l'efecte simultani dels tres factors. L'efecte de l'edat representa els riscos associats als diferents grups d'edat. L'efecte del període representa la variació de les probabilitats de mort deguda a l'any de defunció. L'efecte cohort s'associa a canvis en aquestes probabilitats per als grups d'individus amb el mateix any de naixement.

El model AP té  $1 + (A - 1) + (P - 1)$  paràmetres; si hi afegim el factor cohort, trobarem  $C - 2$  paràmetres i no pas  $C - 1$  com esperaríem. De manera anàloga, si comencem pel model AC, el qual té  $1 + (A - 1) + (P - 1)$  paràmetres, i hi afegim el factor període, tindrem  $P - 2$  nous paràmetres i no pas  $P - 1$ . Això és deu a la relació que lliga els tres factors,  $a = p - c$ . L'exacta dependència lineal entre els tres factors que aquesta igualtat suposa és el major problema que plantegen els models APC, i és conegut com *el problema d'identificabilitat*. Hi ha diferents solucions al problema; la que proposa [44] consisteix a ajustar primer el model amb qualsevol parametrització dels efectes, i després dur a terme una regressió de cada conjunt d'estimacions sobre el factor corresponent. Vegem com procedir.

Suposem que s'ajusta un model amb un cert conjunt de paràmetres

$$\text{logit}[q(a, p)] = \alpha_a + \beta_p + \gamma_c. \quad (9)$$

L'equació és equivalent si afegim a cada paràmetre qualsevol valor,  $\mu_a$ ,  $\mu_p$  i  $\mu_c$  a condició que  $\mu_a + \mu_p + \mu_c = 0$ . Cal tenir en compte que en el model APC es verifica la relació  $a = p - c$  i, per tant, per a qualsevol constant  $\delta$  tindrem  $\delta(a - p + c) = 0$ . Incorporant totes dues igualtats a (9) obtindrem

$$\begin{aligned} \text{logit}[q(a, p)] &= \alpha_a - \mu_p - \mu_c + \delta a \\ &\quad + \beta_p + \mu_p - \delta p \\ &\quad + \gamma_c + \mu_c + \delta c. \end{aligned} \quad (10)$$

Aquesta expressió comporta una descomposició dels efectes en una part lineal i una altra no lineal. Tal com se suggereix en [44] efectuem una regressió de les estimacions de cada efecte sobre el seu factor per a poder, una vegada substituïdes en (9) les regressions calculades, obtenir una nova expressió semblant a (10),

$$\begin{aligned} \text{logit}[q(a, p)] &= \tilde{\alpha}_a + \hat{\mu}_a + \hat{\delta}_a a \\ &\quad + \tilde{\beta}_p + \hat{\mu}_p + \hat{\delta}_p p \\ &\quad + \tilde{\gamma}_c + \hat{\mu}_c + \hat{\delta}_c c. \end{aligned} \quad (11)$$

El que esperem és que tots els pendents siguin iguals i estimen el pendent teòric comú  $\delta$ . Aquest problema pot resoldre's fent alguna hipòtesi sobre la importància relativa dels efectes. Per exemple, si decidim que l'edat és la més important i la del període la menys important, podríem triar una reparametrització basada en les hipòtesis següents:

- els efectes període hauran de tenir mitjana zero,
- els efectes cohort hauran de ser un risc relatiu a alguna cohort central, i
- els efectes edat hauran de representar les probabilitats de mort per a cada edat en la cohort central de referència, després de la correcció perquè els efectes període tinguin mitjana zero.

Començant pel primer punt, a partir de l'equació (11) podem utilitzar com a efecte període

$$g(p) = \tilde{\beta}_p = \beta_p - \hat{\mu}_p - \hat{\delta}_p p,$$

precisament perquè els residus tenen zero com a mitjana. En substituir en (11)

$$\begin{aligned} \text{logit}[q(a, p)] &= \tilde{\alpha}_a + \hat{\mu}_a + \hat{\delta}_a a + g(p) + \hat{\mu}_p + \hat{\delta}_p p + \tilde{y}_c + \hat{\mu}_c + \hat{\delta}_c c \\ &= \alpha_a + g(p) + \hat{\mu}_p + \hat{\delta}_p p + \gamma_c, \end{aligned} \quad (12)$$

sumant i restant la cohort de referència,  $\hat{\delta}_p c_0$ , i l'efecte d'aquesta,  $\gamma_{c_0}$ , i tenint en compte que  $p = c + a$ , tindrem

$$\begin{aligned} \text{logit}[q(a, p)] &= [\alpha_a + \hat{\mu}_p + \hat{\delta}_p(a + c_0) + \gamma_{c_0}] + g(p) + [\gamma_c - \gamma_{c_0} + \hat{\delta}_p(c - c_0)] \\ &= f(a) + g(p) + h(c), \end{aligned} \quad (13)$$

on

$$h(c) = \gamma_c - \gamma_{c_0} + \hat{\delta}_p(c - c_0)$$

representa l'efecte de la cohort, que val zero per a  $c_0$  i té el pendent correcte, i

$$f(a) = \alpha_a + \hat{\mu}_p + \hat{\delta}_p(a + c_0) + \gamma_{c_0}$$

representa l'efecte de l'edat.

Una altra solució és la que proposa el mètode seqüencial: ajusta primer el model AC i tot seguit ajusta als residus un model únicament amb el període. De manera diferent de la proposta de [44], aquest mètode permet construir intervals de confiança per als efectes estimats. En [15] poden trobar-se detalls de tots dos mètodes.

**Aplicació dels models APC a dades de mortalitat espanyoles.** En la figura 9 hem representat en color negre les estimacions de màxima versemblança obtingudes amb [44] a partir de les dades espanyoles de mortalitat. Hem pres l'edat com al factor més important i el període com al menys important, i el 1937 com a cohort de referència segons l'expressió (13). Les línies de color gris corresponen a les estimacions obtingudes per mitjà de l'ajust pel mètode

seqüencial, que hem acompanyat dels intervals de confiança corresponents. En la gràfica, totes dues estimacions es confonen perquè els seus resultats són molt semblants.

S'observa que l'efecte de la cohort en la mortalitat dels dos sexes a Espanya mostra un increment en les generacions a partir, aproximadament, del 1960, efecte que és menys destacat per a les dones.

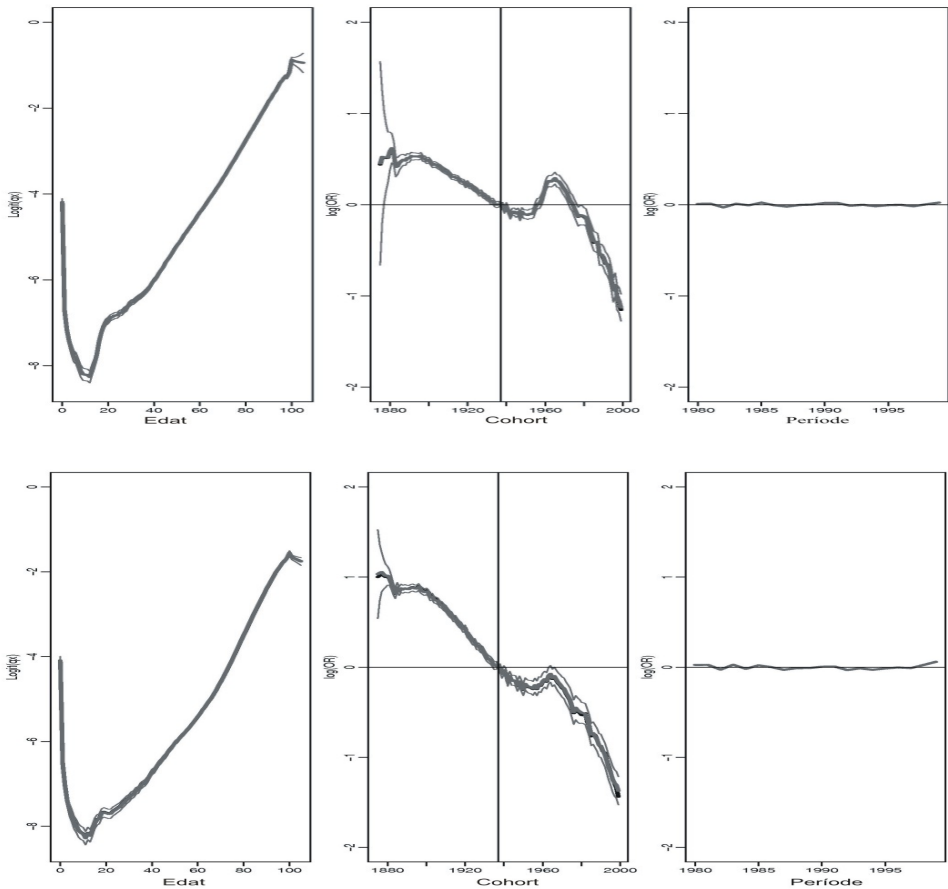


FIGURA 9: Model APC aplicat a Espanya per a homes (dalt) i dones (baix).

Un punt dèbil dels models APC presentats en l'expressió (9) és que assumeixen efectes independents per a l'edat i per al període, mentre que alguns autors com Pitacco i Olivieri, [64], opinen que l'impacte de les millors en la mortalitat al llarg del temps pot ser variable amb l'edat.

## 6 Tendències futures

### 6.1 Graduació no paramètrica

L'anàlisi de dades de mortalitat mitjançant models no paramètrics ha atret l'atenció d'actuaris, demògrafs i estadístics al llarg del passat segle xx. La diferència fonamental amb els models paramètrics és que els primers no necessiten suposar una funció dependent de l'edat, la qual cosa implica un avantatge quan no tenim informació del model subjacent.

Entre els mètodes no paramètrics de graduació ocupen un lloc important les tècniques d'allisatge (*smoothing*), que allisen les probabilitats brutes obtingudes directament de les dades. Una revisió dels mètodes d'allisatge aplicats a taules de mortalitat pot trobar-se en [84, 83]. En [58] es revisen articles sobre d'allisatge i predicció, la majoria amb consideracions teòriques, i se'n discuteixen aplicacions a la ciència actuarial, a la bioestadística i a les finances. Una altra revisió, però aplicada a la medicina, pot trobar-se en [88].

En [28] es revisen les diferents alternatives tenint en compte només l'efecte de l'edat  $x$  i se'n mostra l'aplicació a dades de mortalitat de la Comunitat Valenciana. Una línia de treball futura seria estendre els mètodes revisats en [28] a la graduació de dades de mortalitat al llarg del temps amb l'objectiu d'obtenir taules de mortalitat dinàmiques. Els treballs [31, 39, 33] van en aquesta direcció usant una funció nucli (*kernel*) bivariant. En [19] s'utilitzen models additius generalitzats (GAM) per a modelitzar i predir ràtios de càncer de pulmó, la qual cosa pot suposar una alternativa interessant en la graduació de la mortalitat.

### 6.2 Modelització dels residus

Els mètodes exposats fins ara adapten lleis tradicionals i hi incorporen la influència del temps, però cap d'aquests no té en compte l'estructura de dependència que hi pugua haver entre els residus, que alguns autors consideren necessari modelitzar [7, 69]. En tots els mètodes precedents s'assumeix independència de les observacions, una hipòtesi difícil de mantenir quan fem una representació com la de la figura 10, que mostra una gràfica dels residus obtinguts amb el model de Lee-Carter en funció de l'edat i el temps [25]. S'hi pot observar l'existència d'un patró per als residus, que entre d'altres coses posa de manifest l'existència d'un efecte cohort (diagonals amb valors semblants).

Una solució al problema pot consistir a modelitzar la dependència dels residus mitjançant tècniques adequades, la qual cosa obri la possibilitat de línies de treball futures que exploren el problema en profunditat. En [1] s'estudia la solució que la geoestadística pot aportar al problema, analitzant el comportament temporal de la mortalitat i fent prediccions per a anys futurs, tot això amb les dades de mortalitat espanyoles que hem esmentat. Els resultats obtinguts s'han comparat amb els dels models Lee-Carter i Gompertz-Makeham dinàmics, la bondat dels quals ja ha sigut provada.

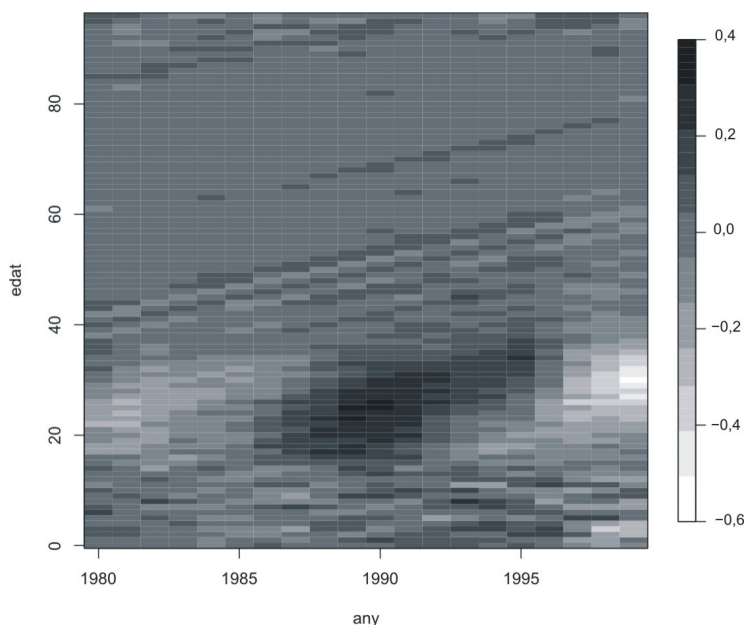


FIGURA 10: Residus per al model de Lee-Carter.

Dues aproximacions són possibles, les que es deriven de considerar la taula dinàmica com un conjunt d'observacions espacials sobre un reticle rectangular o bé com un conjunt de dades espaciotemporals [21].

### 6.3 Risc de la longevitat

Com assenjala [78], el comportament segons l'edat i el temps de la mortalitat en edats avançades, vuitanta-cin i més anys, és un aspecte clau en el canvi recent i l'evolució futura de la mortalitat. Tanmateix, l'existència d'inexactituds en les dades disponibles i la variabilitat deguda als escassos individus exposats al risc per a aquestes edats són les principals dificultats a l'hora de millorar la modelització de la mortalitat en aquests grups d'edat, com assenyalen [87] en un treball que arreplega també les solucions més interessants proposades.

La tendència de la mortalitat en les edats avançades introdueix un risc important per a l'assegurador, que és conegut com a *risc de la longevitat* (*longevity risk*), del qual s'ocupa [63]; també hi trobem una recopilació de les contribucions més recents en l'estudi i la projecció de la mortalitat. En un treball de revisió més recent, [65], els autors presenten una extensa introducció a alguns temes de gran interès relatius al risc de la longevitat en l'àrea de les anualitats de vida (els beneficis de les pensions).

#### 6.4 Models fràgils (*frailty models*)

Una font d'incertesa en l'estudi i la predicció de la mortalitat és l'heterogeneïtat de les poblacions. De la influència de l'heterogeneïtat en la modelització de la supervivència i en problemes actuuarials es té consciència des de fa molt de temps [62]. Ja en [38] s'hi fa referència amb relació a les ràtios d'embaràs, i en [5] se n'estudia el seu efecte sobre la migració. En [79] els autors analitzen com influeix sobre la mortalitat i introdueixen el concepte de *frailty*, representat com una variable no observable, i en descriuen les implicacions per als mètodes estàndard de taules de mortalitat. En [14] es consideren aspectes pràctics per a seleccionar la família de models *frailty* aplicable a assegurances basades en dades de mortalitat, i la possible interpretació d'aquests models.

#### Agraïments

Aquest treball està parcialment subvencionat per un projecte del MEC (Ministeri d'Educació i Ciència, Espanya, projecte MTM-2004-06231). Gràcies també a l'Àrea de Promoció i Normalització Lingüística de la Universitat Politècnica de València pel finançament que ha aportat per a aquest treball.

#### Referències

- [1] ABRIL, M. L. *Aportaciones de la geostatística a la predicción de la mortalidad*. Memòria d'investigació de 3er cicle, Universitat de València. Departament d'Estadística i Investigació Operativa, 2006.
- [2] BALLOTTA, L.; HABERMAN, S. «The fair valuation problem of guaranteed annuity options: The stochastic mortality environment case». *Insurance: Mathematics & Economics*, 38 (1) (2006), 195-214.
- [3] BENJAMIN, B.; POLLARD, J. *The Analysis of Mortality and Other Actuarial Statistics*. 6a ed., Londres: Butterworth-Heinemann, 1992.
- [4] BENJAMIN, B.; SOLIMAN, A. *Mortality on the Move*. Oxford: Actuarial Education Service, 1993.
- [5] BLUMEN, I.; KOGAN, M.; MCCARTHY, P. J. *The industrial mobility of labour as a probability process*. Ithaca, N.Y.: Cornell University Press, 1955.
- [6] BOOTH, H. «Demographic forecasting: 1980 to 2005 in review». *International Journal of Forecasting*, 22 (3) (2006), 547-582.
- [7] BOOTH, H.; MAINDONALD, J.; SMITH, L. «Applying Lee-Carter under conditions of variable mortality decline». *Population Studies*, 56 (3) (2002), 325-336.
- [8] BOOTH, H.; TICKLE, L. «The future aged: new projections of Australia's elderly population». *Population Studies*, 22 (4) (2003), 38-44.
- [9] BRASS, W. «A generation method for projecting death rates». A: *Population Growth and the Brain Drain*. Edimburg: University Press, 1969, 75-91.

- [10] BROUHNS, N.; DENUIT, M.; KEILEGOM, I. V. «Bootstrapping Poisson log-bilinear model for mortality forecasting». *Scandinavian Actuarial Journal*, 2005 (3) (2005), 212–224.
- [11] BROUHNS, N.; DENUIT, M.; VERMUNT, J. «A Poisson log-bilinear regression approach to the construction of projected lifetables». *Insurance: Mathematics & Economics*, 31 (3) (2002), 373–393.
- [12] BUREAU, C.M.I. *Continuous Mortality Investigation Reports CMIR 10*. The Institute of Actuaries and Faculty of Actuaries, UK, 1990.
- [13] BUREAU, C.M.I. *Continuous Mortality Investigation Reports CMIR 17*. The Institute of Actuaries and Faculty of Actuaries, UK, 1999.
- [14] BUTT, Z.; HABERMAN, S. «Application of frailty-based mortality models using generalized linear models». *Astin Bulletin*, 34 (1) (2004), 175–197.
- [15] CARSTENSEN, B.; KEIDING, N. *Age-Period-Cohort models: Statistical inference in the Lexis diagram*. 2005.  
URL <http://www.pubhealth.ku.dk/bxc/APC/notes.pdf>
- [16] CARTER, L.; PRSKAWETZ, A. «Examining structural shifts in mortality using the Lee-Carter method». Publisher: MPIDR Working Paper 2001-2007 [Center for Demography and Ecology Information, University of Wisconsin-Madison (2001)]
- [17] CLAYTON, D.; SCHIFFLERS, E. «Models for temporal variation in cancer rates. I: Age-period and age-cohort models». *Statistics in Medicine*, 6 (4) (1987), 449–467.
- [18] CLAYTON, D.; SCHIFFLERS, E. «Models for temporal variation in cancer rates. II: Age-period-cohort models». *Statistics in Medicine*, 6 (4) (1987), 469–481.
- [19] CLEMENTS, M.; ARMSTRONG, B.; MOOLGAVKAR, S. «Lung cancer rate prediction using generalized additive models». *Biostatistics*, 6 (4) (2005), 576–589.
- [20] CONGDON, P. «Statistical graduation in local demographic analysis and projection». *Journal of the Royal Statistical Society A*, 156 (2) (1993), 237–270.
- [21] CRESSIE, N. *Statistics for Spatial Data*. Nova York: John Wiley, 1993. [Edició revisada]
- [22] CURRIE, I.; DURBAN, M.; EILERS, P. «Smoothing and forecasting mortality rates». *Statistical Modelling*, 32 (4) (2004), 279–298.
- [23] CURRIE, I.; KIRKBY, J.; DURBAN, M.; EILERS, P. «Smooth Lee-Carter models and beyond». A: *Workshop on Lee-Carter Methods*, [www.ma.hw.ac.uk/iain/workshop/workshop.html](http://www.ma.hw.ac.uk/iain/workshop/workshop.html) (4 de març de 2005), 2004.
- [24] CZADO, C.; DELWARDE, A.; DENUIT, M. «Bayesian Poisson log-bilinear mortality projections». *Insurance: Mathematics & Economics*, 36 (3) (2005), 260–284.
- [25] DEBÓN, A.; MONTES, F.; PUIG, F. «Modelling and forecasting mortality in Spain». *European Journal of Operation Research*. [Acceptat]



- [26] DEBÓN, A.; MONTES, F.; SALA, R. «A comparison of parametric models for mortality graduation. Application to mortality data of the València region (Spain)». *Statistics and Operations Research Transactions*, 29 (2) (2005), 269-287.
- [27] DEBÓN, A.; MONTES, F.; SALA, R. «A comparison of models for dynamical life tables. Application to mortality data of the València region (Spain)». *Lifetime data analysis*, 12 (2) (2006), 223-244.
- [28] DEBÓN, A.; MONTES, F.; SALA, R. «A comparison of nonparametric methods in the graduation of mortality: Application to data from the València region (Spain)». *International Statistical Review*, 74 (2) (2006), 215-233.
- [29] DELLAPORTAS, P.; SMITH, A.; STAVROPOULOS, P. «Bayesian analysis of mortality data». *Journal of the Royal Statistical Society A*, 164 (2) (2001), 275-291.
- [30] FELIPE, M.; GUILLÉN, M. *Evolución y predicción de las tablas de mortalidad dinámicas para la población española*. Madrid: Cuadernos de la Fundación, Fundación Mapfre Estudios, Madrid, 1999.
- [31] FELIPE, A.; GUILLÉN, M.; NIELSEN, J. «Longevity studies based on kernel hazard estimation». *Insurance: Mathematics & Economics*, 28 (2) (2001), 191-204.
- [32] FELIPE, A.; GUILLÉN, M.; PÉREZ-MARÍN, A. «Recent mortality trends in the Spanish population». *British Actuarial Journal*, 8 (4) (2002), 757-786.
- [33] FLEDELIUS, P.; GUILLÉN, M.; NIELSEN, J.; PETERSEN, K. «A comparative study of parametric and nonparametric estimators of old-age mortality in Sweden». *Journal of Actuarial Practice*, 11 (2004), 101-126.
- [34] FORFAR, D.; MCCUTCHEON, J.; WILKIE, A. «On graduation by mathematical formula». *Journal of the Institute of Actuaries*, 115, part I (459), (1988), 1-149.
- [35] GAVIN, J.; HABERMAN, S.; VERRALL, R. «Moving weighted average graduation using kernel estimation». *Insurance: Mathematics & Economics*, 12 (2) (1993), 113-126.
- [36] GAVIN, J.; HABERMAN, S.; VERRALL, R. «On the choice of bandwidth for kernel graduation». *Journal of the Institute of Actuaries*, 121 (1994), 119-134.
- [37] GAVIN, J.; HABERMAN, S.; VERRALL, R. «Graduation by kernel and adaptive kernel methods with a boundary correction». *Transactions. Society of Actuaries*, XLVII (1995), 173-209.
- [38] GINI, C. «Premières recherches sur la fécondabilité de la femme». *Proceedings of the International Mathematical Congress*, 2 (1924), 889-892.
- [39] GUILLÉN, M.; NIELSEN, J.; PÉREZ-MARÍN, A. «Multiplicative hazard models for studying the evolution of mortality». *The Annals of Actuarial Science*, 1 (2006).
- [40] GUILLÉN, M.; VIDIELLA-I-ANGUERA, A. «Forecasting Spanish natural life expectancy». *Risk Analysis*, 25 (5) (2005), 1161-1170.

- [41] HABERMAN, S.; RENSHAW, A. «Generalized linear models and actuarial science». *The Statistician*, 45 (4) (1996), 407-436.
- [42] HMD. «Human Mortality Database University of California, Berkeley (USA), and Max Planck Institute for Demographic Research (Germany)». [www.mortality.org](http://www.mortality.org) o [www.humanmortality.de](http://www.humanmortality.de) (2005).
- [43] HELIGMAN, L.; POLLARD, J. «The age pattern of mortality». *Journal of the Institute of Actuaries*, 107 (1980), 49-80.
- [44] HOLFORD, T. «The estimation of age, period and cohort effects for vital rates». *Biometrics*, 39 (1983), 311-324.
- [45] KHALAF-ALLAH, M.; HABERMAN, S.; VERRALL, R. «Measuring the effect of mortality improvements on the cost of annuities». *Insurance: Mathematics & Economics*, 39 (2) (2006), 231-249.
- [46] KOISSI, M.; SHAPIRO, A.; HÖGNÄS, G. «Evaluating and extending the Lee-Carter model for mortality forecasting confidence interval». *Insurance: Mathematics & Economics*, 38 (1) (2006), 1-20.
- [47] LEE, R. «The Lee-Carter method for forecasting mortality, with various extensions and applications». *North American Actuarial Journal*, 4 (1) (2000), 80-91.
- [48] LEE, R.; CARTER, L. «Modelling and forecasting U. S. mortality». *Journal of the American Statistical Association*, 87 (419) (1992), 659-671.
- [49] LEE, R.; ROFMAN, R. «Modelación y proyección de la mortalidad en Chile». *Notas Población*, 22 (59) (1994), 182-213.
- [50] LEE, R. D.; NAULT, F. «Modeling and forecasting provincial mortality in Canada». Montreal: World Congress of the International Union for Scientific Study of Population, 1993.
- [51] LI, N.; LEE, R. «Coherent mortality forecast for a group of populations: an extension of the Lee-Carter method». *Demography*, 42 (3) (2005), 575-593.
- [52] MACDONALD, A. «An actuarial survey of statistical models for decrement and transition data I: Multiple state, Poisson and Binomial models». *British Actuarial Journal*, 2 (1) (1996), 129-155.
- [53] MACDONALD, A. «An actuarial survey of statistical models for decrement and transition data II: Competing risks, non-parametric and regression models». *British Actuarial Journal*, 2 (2) (1996), 429-448.
- [54] MACDONALD, A. «An actuarial survey of statistical models for decrement and transition data III: Counting process models». *British Actuarial Journal*, 2 (3) (1996), 703-726.
- [55] MCNOWN, R.; ROGERS, A. «Time series Analysis Forecasts of a Parameterised Mortality Schedule». A: *Advances in Regional Demography: Forecasts, Information, Models*, Londres: Belhaven Press, 1989, 107-123.
- [56] MORTALITY COMMITTEE. «Projecting future mortality: Towards a proposal for a stochastic methodology». Continuous Mortality Investigation (2005).

- (Working Paper, 15).  
URL [www.actuaries.org.uk/files/pdf/cmi/wp15/wp15.pdf](http://www.actuaries.org.uk/files/pdf/cmi/wp15/wp15.pdf)
- [57] MORTALITY SUB-COMMITTEE. «Projecting future mortality: A discussion paper». Continuous Mortality Investigation (2004). (Working Paper 15).  
URL [www.actuaries.org.uk/files/pdf/cmi/cmiwp3.pdf](http://www.actuaries.org.uk/files/pdf/cmi/cmiwp3.pdf)
- [58] NIELSEN, J. «Smoothing and prediction with a view to actuarial science, biostatistics and finance». *Scandinavian Actuarial Journal*, (1) (2003), 51-74.
- [59] OLIVIERI, A. «Uncertainty in mortality projections: an actuarial perspective». *Insurance: Mathematics & Economics*, 29 (4) (2001), 231-245.
- [60] PALACIOS, H. *Introducción al cálculo actuarial*. Madrid: Mapfre, 1996.
- [61] PEDROZA, C. «A bayesian forecasting model: predicting U.S. male mortality». *Biostatistics*, 7 (4) (2006), 530-550.
- [62] PITACCO, E. «From Halley to “frailty”: a review of survival models for actuarial calculations». *Giornale dell'Istituto Italiano degli Attuari*, 67 (2004), 17-48.
- [63] PITACCO, E. «Survival models in dynamic context: a survey». *Insurance: Mathematics & Economics*, 35 (2) (2004), 279-298.
- [64] PITACCO, E.; OLIVIERI, A. «Forecasting mortality: an introduction.» Milà: Università Bocconi, 2005. (Working Paper;35).
- [65] PITACCO, E.; OLIVIERI, A. «Life annuities dynamics». Milà: Università Bocconi, 2006. (Working Paper;36).
- [66] R DEVELOPMENT CORE TEAM. *R: A Language and Environment for Statistical Computing*. Viena: R Foundation for Statistical Computing, 2005. URL <http://www.R-project.org>
- [67] RENSHAW, A. «Actuarial graduation practice and generalised linear models». *Journal of the Institute of Actuaries*, 118 (II) (1991), 295-312.
- [68] RENSHAW, A.; HABERMAN, S. «Modelling for mortality reduction factors». Londres: City University, 2000. (Actuarial Research Paper; 127).
- [69] RENSHAW, A.; HABERMAN, S. «Lee-Carter mortality forecasting: a parallel generalized linear modelling approach for England and Wales mortality projections». *Journal of the Royal Statistical Society C*, 52 (1) (2003), 119-137.
- [70] RENSHAW, A.; HABERMAN, S. «Lee-Carter mortality forecasting with age specific enhancement». *Insurance: Mathematics & Economics*, 33 (2) (2003), 255-272.
- [71] RENSHAW, A.; HABERMAN, S. «On the forecasting of mortality reduction factors». *Insurance: Mathematics & Economics*, 32 (3) (2003), 379-401.
- [72] RENSHAW, A.; HABERMAN, S. «A cohort-based extension to the Lee-Carter model for mortality reduction factors». *Insurance: Mathematics & Economics*, 38 (3) (2006), 556-570.

- [73] RENSHAW, A.; HABERMAN, S.; HATZOPOULOS, P. «The modelling of recent mortality in United Kingdom male assured lives». *British Actuarial Journal*, 2 (II) (1996), 449–477.
- [74] SITHOLE, T. *Projection of mortality rates with specific reference to immediate annuitants and life office pensioners*. Tesi Doctoral. Londres: City University, 2004.
- [75] SITHOLE, T.; HABERMAN, S.; VERRALL, R. «An investigation into parametric models for mortality projections, with applications to immediate annuitants' and life office pensioners' data». *Insurance: Mathematics & Economics*, 27 (3) (2000), 285–312.
- [76] TABEAU, E.; BERG JETHS, A. VAN DEN ; HEATHCOTE, C. [ED.]. *A Review of Demographic Forecasting Models for Mortality. Forecasting in Developed Countries: From description to explanation*. Kluwer Academic Publishers, 2001.
- [77] THIELE, P. «On a mathematical formula to express the rate of mortality throughout the whole of life». *Journal of the Institute of Actuaries*, 16 (1972), 313–329.
- [78] TULJAPURKAR, S.; BOE, C. «Mortality change and forecasting: How much and how little do we know». *North American Actuarial Journal*, 2 (4) (1998), 13–47.
- [79] VAUPEL, J. W.; MANTON, K. G.; STALLARD, E. «The Impact of Heterogeneity in Individual Frailty on the Dynamics of Mortality». *Demography*, 16 (1979), 439–454.
- [80] VERRALL, R. «Graduation by dynamic regression methods». *Journal of the Institute of Actuaries*, 120 (1993), 153–170.
- [81] VERRALL, R. «A unified framework for graduation». Londres: City University, 1996. (Actuarial Research Paper; 91).
- [82] VILLALÓN, J. G. *Manual de matemáticas financiero-actuariales*. Madrid: Fernández Ciudad, 1994.
- [83] WANG, J. «Smoothing Hazard Rates». *Encyclopedia of Biostatistics*. 2a ed. Wiley, 2005.
- [84] WANG, J.; MÜLLER, H.; CAPRA, W. «Analysis of oldest-old mortality lifetables revisited». *The Annals of Statistics*, 26 (1) (1998), 126–163.
- [85] WILMOTH, J. «Computational methods for fitting and extrapolating the Lee-Carter model of mortality change». Technical report, Department of Demography, University of California, Berkeley (1993).
- [86] WILMOTH, J. «Mortality projections for Japan: A comparison of four methods». *Health and Mortality among Elderly Populations*. Oxford University Press, 1996, 266–287.
- [87] WONG-FUPUY, C.; HABERMAN, S. «Projecting mortality trends: Recent developments in the United Kingdom and the United States». *North American Actuarial Journal*, 8 (2) (2004), 56–83.

- [88] ZHANG, H. «Mixed effects multivariate adaptative splines model for the analysis of longitudinal growth curve data». *Statistical Methods in Medical Research*, 13 (1) (2004), 63-82.

ANA M. DEBÓN  
DEPARTAMENT D'ESTADÍSTICA I INVESTIGACIÓ OPERATIVA  
APLICADES I QUALITAT  
UNIVERSITAT POLITÈCNICA DE VALÈNCIA  
E-46022 VALÈNCIA  
andeau@eio.upv.es

FRANCISCO MONTES  
DEPARTAMENT D'ESTADÍSTICA I INVESTIGACIÓ OPERATIVA  
UNIVERSITAT DE VALÈNCIA  
E-46100. BURJASSOT, VALÈNCIA  
montes@uv.es

