

# **GRADUACIÓN MEDIANTE FUNCIONES GOMPERTZ- MAKEHAM DINAMICAS**

Ana Debón Aucejo  
Universidad Politécnica de Valencia

Francisco Montes Suay  
Ramón Sala Garrido  
Universitat de València

## **RESUMEN**

El objetivo de este trabajo es modelar la influencia del tiempo de calendario y la edad, en las probabilidades de muerte,  $q_x$ . Para ello, se gradúa los datos de mortalidad de la Comunidad Valenciana correspondientes a los años 1980 a 1999.

En primer lugar obtenemos el ajuste de las funciones Gompertz-Makeham con términos correspondientes a las variables edad y tiempo mediante modelos lineales generalizados. La elección del modelo que mejor se ajusta a los datos de cada periodo se lleva a cabo atendiendo a criterios de bondad de ajuste y de complejidad del mismo. Se trata de buscar un equilibrio entre ambos aspectos, para lo cual utilizamos medidas y contrastes implementados en S-plus.

Del modelo obtenido para cada sexo hacemos un análisis más detallado mediante los tests no paramétricos habituales y análisis gráfico, para concluir la adecuación de éstos a la experiencia de mortalidad.

**Palabras clave:** Modelos lineales generalizados, Mortalidad, funciones Gompertz-Makeham, dinamicidad.

## 1. INTRODUCCIÓN

La graduación de datos de mortalidad por diferentes métodos paramétricos y no paramétricos, se centra en la influencia que la edad tiene sobre ésta. Sin embargo, es un hecho reconocido por la literatura actuarial más reciente que la mortalidad evoluciona con el transcurrir de los años, experiencias de mortalidad correspondientes a distintos periodos tienen probabilidades de muerte a una determinada edad diferentes. Por ello, nos ocupamos en este trabajo de analizar además, el efecto del tiempo del calendario.

La importancia de medir con exactitud cambios en la mortalidad a través del tiempo, reside en el hecho de que los beneficios de las pólizas de seguros de vida dependen de la supervivencia. Por este motivo, si la tabla estándar utilizada para calcular tanto anuales y reservas predice unas probabilidades de mortalidad más altas que las realmente experimentadas por los asegurados, se habrá cobrado de menos a éstos, y la compañía incurrirá en pérdidas. Y no sólo eso, sino además las reservas para tal fin serán insuficientes. Una recopilación de algunas de estas técnicas podemos encontrarla en Benjamín y Soliman (1995), posteriormente Felipe y Guillén (1999) generalizan la clasificación anterior para incluir los modelos desarrollados desde entonces.

En los modelos paramétricos se trata de modelizar las medidas de la mortalidad mediante una función  $f(x, \alpha, t) + \varepsilon_x$ , expresión en la cuál las propuestas sobre la forma funcional  $f$  han sido múltiples, en nuestro artículo nos centramos en el modelo que propone como función la de Gompertz-Makeham.

El contenido del trabajo esta estructurado de la siguiente forma. En la sección 2 describimos el modelo. En la sección 3 realizamos una aplicación del mismo, calculamos primero las estimaciones brutas para los datos y posteriormente estimando los parámetros mas adecuados.

Las conclusiones más relevantes se encuentran en la sección final.

## 2. FUNCIONES GOMPERTZ-MAKEHAM CON RESPECTO A LA EDAD Y EL TIEMPO

En su artículo Renshaw et al. (1996) adaptan las funciones Gompertz-Makeham a la nueva situación mediante la inclusión del tiempo como variable. El modelo que se deduce para  $q_{xt}$  de la propuesta de Renshaw et al. (1996) es el siguiente,

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

sujeto a la convención que algunos de los términos pueden ser cero,  $x'$  y  $t'$  son transformaciones de la edad y el año de calendario, respectivamente, de forma que los valores están dentro de intervalo  $[-1,1]$  y  $L_j(x')$  son los polinomios de Legendre generados por,

$$L_{n+1}(x) = xL_n(x) - nL_{n-1}(x)$$

donde  $n \geq 0$ ,  $L_0(x) = 1$  y  $L_1(x) = x$ .

Reescribiendo la ecuación (1) de la forma \

$$\frac{q_{xt}}{1-q_{xt}} = \exp\left(\beta_0 + \sum_{j=1}^s \beta_j L_j(x')\right) \exp\left(\sum_{i=1}^r \left(\alpha_i + \sum_{j=1}^s \gamma_{ij} L_j(x')\right) t^i\right), \quad (2)$$

el primer término puede interpretarse como una función Gompertz-Makeham  $LGM(0,s+1)$  correspondiente a la graduación. El segundo término puede ser interpretado como término de ajuste del efecto del año del calendario, de forma que cuando al menos uno de los  $\gamma_{ij}$  es no nulo depende también de la edad. Un ejemplo de esto sería el incremento de muertes de hombres adultos y jóvenes debido al SIDA.

La forma de proceder para determinar los parámetros  $\alpha_i$ ,  $\beta_j$  y  $\gamma_{ij}$ , considerando el esquema de modelo lineal generalizado, familia *Binomial* y link *logit*, es la es la siguiente,

1. Los valores elegidos de  $r$  y  $s$  son aquellos máximos a partir de los cuales los incrementos de la *Deviance* no resultan estadísticamente significativos.
2. Se completa el modelo mediante la elección de los coeficientes  $\gamma_{ij}$  cuyo incremento de la *Deviance* resulta significativo.
3. Paralelamente a los puntos anteriores se determinan los errores estándar de la estimaciones de los parámetros y su significación mediante la prueba usual t-Student.

4. La validación final del modelo se realiza aplicando los test de Forfar et al. (1988) a los diferentes años.

Sithole et al. (2000) aplican este modelo, pero el objetivo principal de su trabajo no sólo es encontrar un modelo que proporcione un buen ajuste de los datos, sino que también proporcione un buen esquema para realizar proyecciones. Para ello proceden de la siguiente forma,

1. Determinan un modelo que proporcione el mejor ajuste a los datos.
2. Consideran las proyecciones basadas en dicho modelo para un periodo de 20 años.
3. Finalmente, utilizando la información obtenida por ajuste del modelo y las fuerzas de mortalidad proyectadas, revisan el modelo.

Consideran adecuado el modelo para realizar proyecciones si todas las fuerzas de mortalidad predichas progresan suavemente con respecto a la edad y el tiempo, y predice una reducción en los tantos de mejora en la mortalidad a edades muy avanzadas. En definitiva, se trata de alcanzar un equilibrio entre el buen ajuste y el esquema predictivo.

Por último, compara el segundo termino de la ecuación (2), factor de mejora de la mortalidad,  $RF_{GLM}(x,t)$ , con los  $RF_{CMI}(x,t)$  obtenidos mediante el método habitual del CMI Bureau, para analizar la consistencia de unos con otros.

## 2. AJUSTE DE LAS FUNCIONES GOMPERTZ-MAKEHAM CON RESPECTO A LA EDAD Y EL TIEMPO

### 2.1. Obtención de las estimaciones brutas de $q_{xt}$ .

El modelo descrito ha sido utilizados para ajustar datos de mortalidad de hombres y mujeres , por separado, de la Comunidad Valenciana correspondientes al periodo comprendido entre 1980 y 1999.

Las estimaciones brutas de  $q_{xt}$ , necesarias para los modelos, se han obtenido con el procedimiento utilizado por el Instituto Nacional de Estadística (INE),

$$\hat{q}_x = \frac{1/2(D_{xt} + D_{x(t+1)})}{P_{xt} + 1/2D_{xt}}$$

donde  $D_{xt}$  son los fallecidos en el año  $t$  a la edad  $x$ ,  $D_{x(t+1)}$  son los fallecidos en el año  $t+1$  a la edad  $x$ , y  $P_{xt}$  población que a 31 de diciembre del año  $t$  tienen edad  $x$ .

La fórmula puede ser aplicada para todas las edades salvo para la edad cero por la concentración de defunciones en los primeros meses de vida. Para la edad cero hemos aplicado,

$$\hat{q}_x = \frac{0.8D_{xt} + 0.2D_{x(t+1)}}{P_{xt} + 1/2D_{xt}}$$

## 2.2 Obtención del ajuste.

Para seleccionar los valores óptimos de  $r$  y  $s$  mostramos en la Tabla (1) la mejora de la *Deviance* para el ajuste de las funciones  $GM(r,s)$ , para  $r=0,1,2,3$ ,  $s=0,1,2,3,\dots,19$  y donde los coeficientes  $\gamma_{ij}$  se suponen cero, eligiendo aquellos valores,  $r=1$  y  $s=17$ , a partir de los cuales el incremento ya no es significativo.

	HOMBRES				MUJERES			
	r=0	r=1	r=2	r=3	r=0	r=1	r=2	r=3
s=0	934242.90				1044087			
s=1	43411.38	41886.32	41884.82	41883.9	53538.60	50015.33	50007.63	50005.24
s=2	22099.82	20333.49	20330.92	20330.6	19978.97	15689.62	15684.99	15680.32
s=3	17626.23	15997.45	15994.66	15994.0	15378.20	11353.82	11348.59	11344.88
s=4	17080.46	15464.40	15461.17	15460.7	14825.96	10807.09	10802.84	10798.77
s=5	14452.05	12841.23	12837.26	12836.5	10034.61	5967.92	5963.25	5959.14
s=6	11644.78	9951.69	9945.41	9945.10	9196.54	5116.30	5112.35	5108.20
s=7	7426.26	5666.84	5664.23	5664.13	8072.62	3918.12	3913.50	3908.62
s=8	5896.12	4203.75	4201.57	4201.25	7003.32	2917.36	2911.48	2907.19
s=9	5619.03	3914.28	3911.20	3910.92	6672.97	2580.46	2575.74	2571.45
s=10	5577.25	3866.23	3863.55	3863.33	6619.89	2516.03	2510.97	2506.48
s=11	5507.79	3807.84	3805.21	3804.89	6519.14	2428.78	2423.51	2419.28
s=12	5475.78	3774.29	3771.40	3771.10	6468.55	2378.58	2373.58	2369.37
s=13	5350.17	3642.86	3640.31	3640.07	6433.04	2339.92	2334.88	2330.59
s=14	5229.34	3526.01	3523.54	3523.20	6365.52	2273.30	2268.12	2263.89
s=15	5045.51	3343.09	3340.27	3339.95	6327.66	2236.83	2231.75	2227.60
s=16	4947.21	3243.42	3240.65	3240.38	6288.45	2197.12	2192.09	2187.81
s=17	4889.70	3185.35	3182.61	3182.32	6272.15	2180.31	2175.25	2170.99
s=18	4888.45	3183.99	3181.26	3180.96	6266.36	2174.52	2169.45	2165.18
s=19	4875.24	3170.16	3167.41	3167.12	6266.34	2174.44	2169.37	2165.10

Tabla 1: Deviance para las funciones Gompertz-Makeham

A continuación, procedemos a completar el modelo mediante la elección de los coeficientes  $\gamma_{ij}$ . En la Tabla 2 se analiza si el incremento de la *Deviance* al introducir cada uno de estos parámetro resulta ser significativo.

	HOMBRES				MUJERES			
	Dev.	D.f.	Dev (dif)	p-value	Dev.	D.f.	Dev (dif)	p-value
	3185.35	1921			2180.31	1921		
$\gamma_{11}$	3161.86	1920	23.49	0.0000	2104.49	1920	75.83	0.0000
$\gamma_{12}$	3136.43	1919	25.42	0.0000	2006.32	1919	98.17	0.0000
$\gamma_{13}$	2327.51	1918	808.92	0.0000	1434.07	1918	572.25	0.0000
$\gamma_{14}$	2327.13	1917	0.38	0.5396	1433.98	1917	0.08	0.7699

Tabla 2: Diferencia de Deviance cuando introducimos interacciones en los modelos.

Finalmente el modelo elegido para ambos sexos es el mismo,

$$\log\left(\frac{q_{xt}}{1-q_{xt}}\right) = \beta_0 + \sum_{j=1}^{17} \beta_j L_j(x') + \alpha_1 t' + \sum_{j=1}^3 \gamma_{1j} L_j(x') t'^j \quad (3)$$

con 1918 grados de libertad.

Se concluye al completar el modelo que la elección de los coeficientes  $\gamma_{ij}$  indican que el factor de actualización temporal depende de la edad. Factor que puede apreciarse en la formula (3) y que corresponde a la segunda parte del exponente, la primera sería el resultado de graduar  $q_x$  mediante una función LGM(0,18). Sus coeficientes para ambos sexos se muestran en la Tabla 3, siendo todos ellos significativos.

	HOMBRES			MUJERES		
	coef	std error	t-value	coef	std error	t-value
$\beta_0$	-4.9452	0.0046	-1070.50	-5.5457	0.0059	-943.39
$\beta_1$	3.7199	0.0094	396.14	3.8965	0.0113	344.99
$\beta_2$	0.6151	0.0098	62.80	1.1508	0.0114	100.56
$\beta_3$	-0.1839	0.0097	-18.96	-0.3138	0.0117	-26.77
$\beta_4$	0.0927	0.0108	8.54	0.2311	0.0131	17.56
$\beta_5$	-0.4868	0.0126	-38.67	-0.5134	0.0152	-33.86
$\beta_6$	0.5570	0.0133	41.72	0.3377	0.0163	20.74
$\beta_7$	-0.5139	0.0134	-38.23	-0.4008	0.0168	-23.85
$\beta_8$	0.1814	0.0139	13.08	0.2758	0.0177	15.60
$\beta_9$	0.0188	0.0150	1.25	-0.1046	0.0190	-5.51
$\beta_{10}$	-0.0679	0.0154	-4.40	0.0468	0.0196	2.39
$\beta_{11}$	-0.0348	0.0153	-2.28	-0.0828	0.0197	-4.19
$\beta_{12}$	0.1180	0.0147	8.01	0.1198	0.0193	6.20
$\beta_{13}$	-0.2540	0.0145	-17.52	-0.1692	0.0190	-8.93
$\beta_{14}$	0.2753	0.0151	18.22	0.2078	0.0190	10.88
$\beta_{15}$	-0.2884	0.0158	-18.22	-0.1917	0.0199	-9.63
$\beta_{16}$	0.1787	0.0143	12.47	0.1398	0.0190	7.36
$\beta_{17}$	-0.0911	0.0121	-7.53	-0.0565	0.0138	-4.10
$\alpha_1$	-0.0838	0.0054	-15.61	-0.1569	0.0072	-21.72
$\gamma_{11}$	0.0623	0.0093	6.70	0.0384	0.0113	3.39
$\gamma_{12}$	-0.1183	0.0088	-13.47	-0.0330	0.0104	-3.17
$\gamma_{13}$	0.2530	0.0089	28.27	0.2414	0.0101	23.84

Tabla 3: Coeficientes de la función Gompertz-Makeham

Para el diagnóstico del modelo realizamos un gráfico de tallo y hojas para las desviaciones relativas, en él hemos podido apreciar un comportamiento aproximadamente normal. La bondad de ajuste del modelo para cada uno de los años lo hemos analizado a través de los test habituales de Forfar et al. (1988) los resultados están recogidos en la Tabla 4 para hombres y mujeres, respectivamente. En el modelo resultan mejores los resultados en el caso femenino. Realizamos además los test de autocorrelaciones que resultan significativos para el ajuste de los hombres prácticamente en todos los años, mientras que para las mujeres no es así, lo que viene a corroborar que el modelo es mejor en el caso femenino.

El perfil de la evolución de las probabilidades de muerte a lo largo de los años así como su ajuste al modelo lo mostramos gráficamente en la Figura 1 y 2 para hombres y mujeres respectivamente. Las edades entre 20 y 40 para los hombres presentan oscilaciones a lo largo de los años difíciles de recoger por el modelo. Para los años 1991 en adelante la sobremortalidad de la joroba de los accidentes no se ajusta según este modelo, motivo por el cual el ajuste de los hombres resulta peor que el de las mujeres donde aquella no aparece.

Respecto a la predicción para futuros años como es un método no estructural, obtenemos una función del tiempo, en la que basta con sustituir los valores correspondientes a  $t$  para calcular el valor de los futuros  $q_{xt}$ .

HOMBRES									
	$\chi^2$	$z>2$	$z>3$	pos.	neg.	p(pos)	rachas	p(rachas)	EPAM
1980	105.74	7	0	46	51	0.3424	35	0.2735	14.33
1981	96.50	4	0	37	60	0.0125	34	0.2748	14.20
1882	93.13	2	0	37	60	0.0125	24	0.1430	13.28
1983	113.67	7	0	36	61	0.0072	37	0.3284	14.72
1984	89.07	3	0	45	52	0.2713	36	0.2884	12.68
1985	81.83	2	0	45	52	0.2713	39	0.3330	12.57
1986	81.80	3	1	39	58	0.0335	37	0.3157	10.41
1987	85.16	4	1	57	40	0.9665	37	0.3125	10.87
1988	117.88	7	1	61	36	0.9960	35	0.2946	12.05
1989	148.39	7	5	60	37	0.9928	41	0.3934	11.98
1990	146.99	9	3	57	40	0.9665	37	0.3125	11.76
1991	150.63	5	4	54	42	0.9079	40	0.3599	14.82
1992	93.62	6	0	49	46	0.6591	31	0.2284	12.83
1993	116.30	8	1	39	56	0.0501	29	0.2069	13.58
1994	136.22	6	3	43	51	0.2352	36	0.3054	16.63
1995	122.80	8	1	42	52	0.1767	29	0.2081	14.33
1996	142.21	7	2	48	46	0.6214	35	0.2881	16.36
1997	128.45	9	1	32	60	0.0023	45	0.5482	14.57
1998	144.22	8	2	36	56	0.0235	28	0.2073	17.13
1999	126.99	8	1	26	66	0.0000	28	0.2443	19.51
MUJERES									
	$\chi^2$	$z>2$	$z>3$	pos.	neg.	p(pos)	rachas	p(rachas)	EPAM
1980	90.80	0	1	55	42	0.9226	44	0.4207	14.07
1981	59.73	1	0	48	49	0.5000	35	0.2729	11.44
1882	57.57	2	0	42	55	0.1114	42	0.3872	13.40
1983	69.34	3	0	36	61	0.0072	43	0.4376	13.65
1984	69.68	1	0	33	64	0.0011	39	0.3869	13.08
1985	62.45	3	0	44	52	0.2376	43	0.4044	12.79
1986	54.36	1	0	48	47	0.5812	48	0.4916	12.54
1987	67.47	2	0	52	43	0.8476	41	0.3796	12.44
1988	78.68	3	0	64	30	0.9999	41	0.4806	13.41
1989	67.89	2	0	58	35	0.9938	42	0.4478	12.32
1990	68.25	3	0	53	39	0.9413	40	0.3922	INF
1991	68.40	0	0	47	45	0.6227	44	0.4479	13.88
1992	68.61	2	0	43	49	0.3012	43	0.4330	14.60
1993	75.67	5	0	48	44	0.6988	35	0.2995	11.74
1994	74.42	2	1	43	48	0.3377	33	0.2750	12.43
1995	71.28	2	0	41	50	0.2009	39	0.3745	13.08
1996	55.15	0	0	44	47	0.4170	39	0.3700	14.20
1997	60.30	3	0	35	56	0.0177	31	0.2581	19.55
1998	82.34	2	1	43	47	0.3760	33	0.2800	25.93
1999	71.54	1	1	44	46	0.4581	37	0.3431	23.60

Tabla 5: Contrastes no paramétricos de la función Gompertz-Makeham dinámica

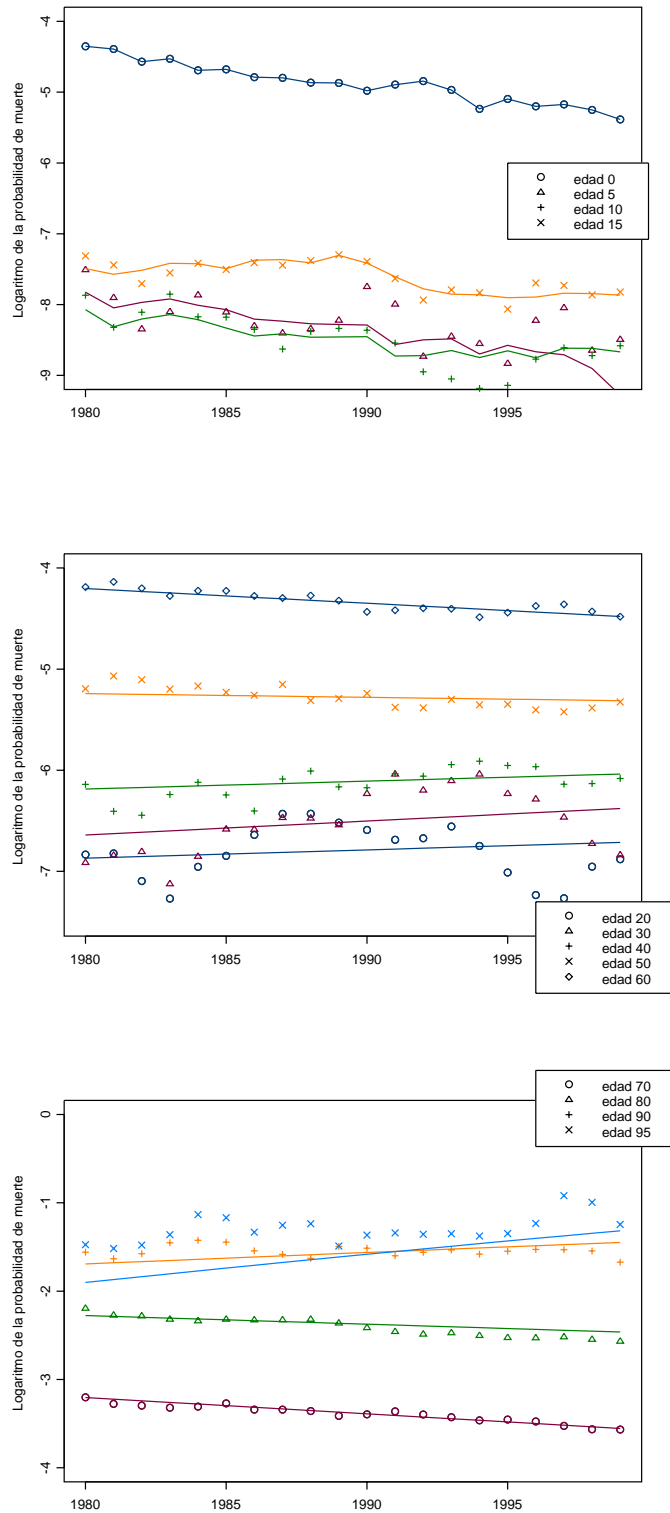


Figura 1: Evolución de  $\log(q_x)$  para algunas edades de la función Gompertz-Makeham dinámica para hombres.

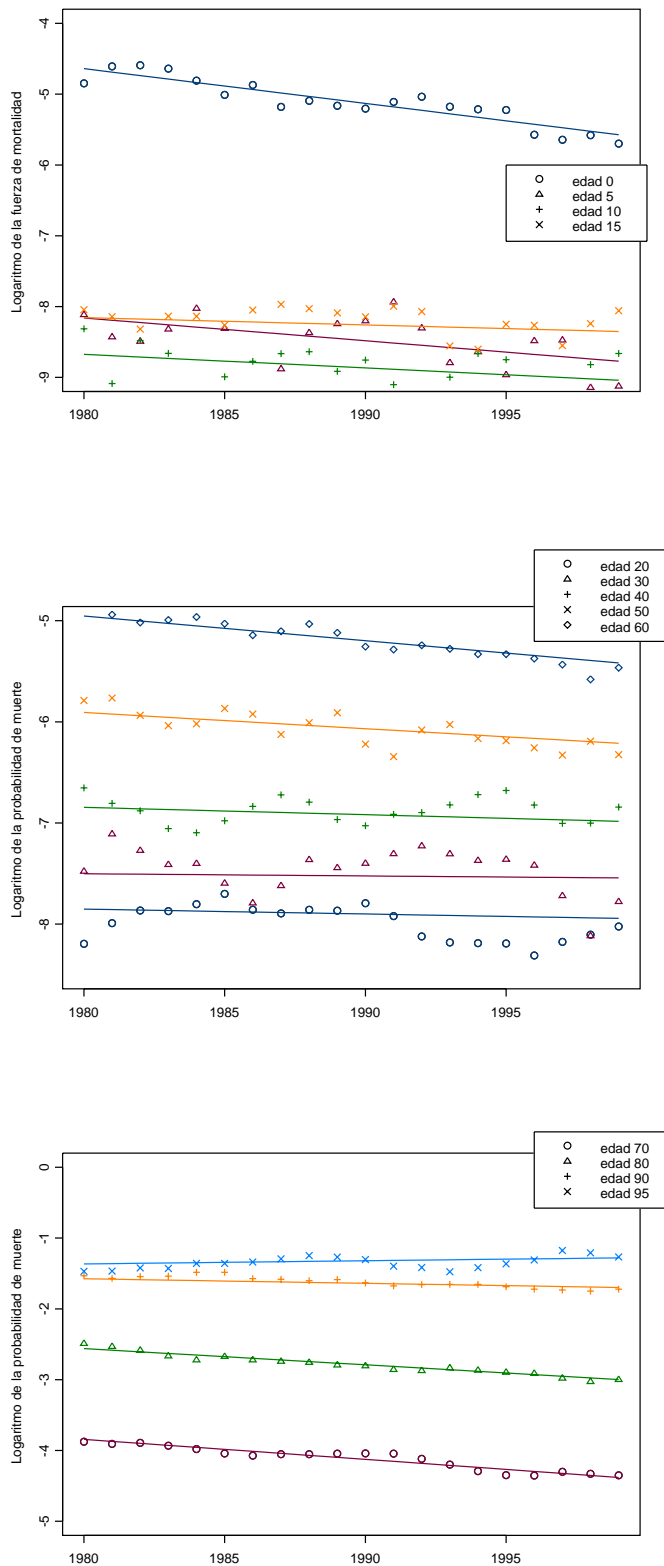


Figura 2: Evolución de  $\log(q_x)$  para algunas edades de la función Gompertz-Makeham dinámica para hombres y mujeres

### **3. CONCLUSIONES**

Hemos de comenzar esta sección de conclusiones señalando que, hasta donde conocemos, las funciones Gompertz-Makeham dinámicas no han sido utilizados en la graduación de datos de mortalidad españoles.

La adecuación del modelo dinámicos se ha llevado a cabo mediante los tests no paramétricos descritos en Forfar et al. (1988) a los que hemos añadido los valores del error porcentual absoluto medio (EPAM) que utilizan Felipe y Guillen (1999) en su trabajo.

De la observación de resultados se deduce que aunque el ajuste es los valores de bondad de ajuste como la  $\chi^2$  son buenos, mejor para las mujeres, la función presenta problemas no superando el test de signos para algunos años, tanto en hombres como en mujeres. Sobreestimando o subestimando para gran parte de edades.

Como posibles mejoras se propone la utilización de otros modelos como puede ser el de Lee-Carter.

### **4. REFERENCIAS BIBLIOGRÁFICAS**

- BENJAMÍN, B. Y SOLIMAN, A. (1993). “Mortality on the Move”. Actuarial Education Service, Oxford.
- FELIPE, A. Y GUILLEN, M. (1999). “Evolución y predicción de las tablas de mortalidad dinámicas para la población española”. Cuadernos de la Fundación Mapfre, Fundación Mapfre Estudios.
- FORFAR, D., MCCUTCHERON, J. Y WILKIE, A. (1988). On the graduation by mathematical formula. Journal of the Institute of Actuaries, 115 part I(459), pp. 1-149.
- RENSHAW, A., HABERMAN, S. Y HATZOPOULOS, P. (1996). The modelling of recent mortality in United Kingdom. British Actuarial Journal, 2(II), pp.449-477.