

COMUNICACIÓN

**Análisis del desempleo en la Comunidad Valenciana
con datos de la EPA [†]**

Eduardo Beamonte¹ y Santiago Murgui²

¹Departamento de Economía Aplicada
Universitat de València Estudi General
Profesor Titular de Universidad

²Departamento de Economía Aplicada
Universitat de València Estudi General
Catedrático de Universidad

Dirección para correspondencia

Eduardo Beamonte Córdoba
Departamento de Economía Aplicada
Av. Naranjos s/n
46022 Valencia
e-mail: beamonte@uv.es
Tfno. 963828618/963828428

[†]Este trabajo ha sido financiado parcialmente por el Instituto Valenciano de Estadística con cargo al proyecto 4438.

1 Introducción

En este trabajo nos planteamos el análisis cuantitativo de los tiempos de desempleo de los individuos de las Comunidades Autónomas de Andalucía y Valencia. El objetivo principal es la comparación de la estructura del mercado de trabajo en ambas comunidades, el conocimiento de posibles factores que influyen en el desempleo y la predicción de comportamientos futuros de los individuos desempleados de una u otra comunidad.

Para alcanzar el triple objetivo marcado en este estudio se han empleado técnicas del *análisis de supervivencia* con la finalidad de realizar un adecuado tratamiento de los datos censurados.

El tiempo de desempleo de un individuo está definido por el momento de entrada en la situación de desempleo y por el momento de salida de dicha situación a la de empleo o inactividad. Como quiera que el final del estudio queda marcado por la referencia temporal de la última Encuesta de Población Activa (EPA) considerada, aquellos individuos que continúen clasificados como desempleados en su última encuesta presentarán un tiempo de desempleo *censurado* por la derecha. Dicho tiempo será igual a la suma del número de meses de búsqueda de empleo inicialmente declarados más el tiempo que esos individuos han permanecido en el estudio. Realmente, en estos casos sólo se sabe que el verdadero tiempo de desempleo es igual al tiempo censurado observado más una cantidad desconocida y que no es observada.

Por el contrario, cuando para un individuo entrevistado se observa un cambio en su clasificación laboral en el sentido del desempleo hacia el empleo o inactividad, entonces se conoce exactamente su tiempo de desempleo y se dice que éste es *exacto o de fallo*.

Los *modelos de duración*, utilizados para el tratamiento de datos de supervivencia, suelen estar basados en la *función de riesgo*. La sencilla e intuitiva interpretación de esta función motiva su utilización en lugar de la habitual función de densidad. La función de riesgo en el punto t , o tasa instantánea de fallo en dicho punto, se define como la probabilidad instantánea de fallo dado que ha habido supervivencia

hasta t ,

$$h(t) = \lim_{\Delta t \rightarrow 0} \frac{p(t \leq T < t + \Delta t | T \geq t)}{\Delta t} = \frac{f(t)}{S(t)},$$

donde $f(t)$ y $S(t) = 1 - F(t)$ son las funciones de densidad y *supervivencia*, respectivamente. Esta última se define como uno menos la función de distribución.

El análisis de datos de supervivencia debe tener en cuenta su particular tipología y considerar las peculiares características de la censura. Así por ejemplo, la función de verosimilitud de unos datos continuos es igual al producto de las funciones de densidad correspondientes a los tiempos exactos por el producto de las funciones de supervivencia de los tiempos censurados. Además, la función de verosimilitud asociada a los datos observados puede ser construida de un modo sencillo a partir de $h(t)$ pues se tiene que:

$$\begin{aligned} S(t) &= \exp \left[- \int_0^t h(s) ds \right], \\ f(t) &= h(t) S(t) = h(t) \exp \left[- \int_0^t h(s) ds \right]. \end{aligned}$$

La consideración de las características propias de cada individuo resulta obligado en cualquier análisis de supervivencia. Es evidente que circunstancias personales como la edad, estado civil o nivel de estudios pueden tener cierta influencia en el correspondiente tiempo de desempleo, por lo que resulta imprescindible la incorporación de covariables al estudio. En este sentido, los modelos de regresión - y en particular el *modelo de riesgos proporcionales de Cox* (1972) - son los comúnmente utilizados por la mayoría de los autores. Esto es debido tanto a su sencillez a la hora de incorporar covariables, como a sus generalmente buenos resultados prácticos. En efecto, Cox propone modelizar la función de riesgo de la variable de supervivencia como el producto de una función de riesgo base y una función exponencial que incorpora las covariables. No obstante, el modelo de Cox presenta el grave inconveniente de su restricción a riesgos proporcionales y su complicado análisis desde un punto de vista bayesiano.

Como alternativa al modelo de Cox, en este trabajo también se utiliza un modelo jerárquico aditivo en el que la función de riesgo es modelizada mediante la suma de una función de riesgo base poligonal no completamente especificada y una función de riesgo Gamma dependiente del vector de covariables de cada individuo. De este

modo, al considerar un modelo en poblaciones es posible recoger la heterogeneidad propia de la población que frecuentemente no es explicada por las covariables.

El *modelo Gamma-poligonal aditivo* (Beamonte, 1998) resulta un modelo de efectos aleatorios semiparamétrico que guarda ciertas analogías con el modelo de Cox al considerar ambos dos partes, una paramétrica y otra no paramétrica, en la función de riesgo e incorporar las covariables vía la parte paramétrica. Sin embargo, la principal diferencia entre los dos, aparte de que el Gamma-poligonal se trata de un modelo en poblaciones, es que al considerar el vínculo entre las dos partes mediante una adición, en lugar de una multiplicación, simplifica bastante todo el aparato matemático del modelo y permite el análisis bayesiano del mismo.

El modelo de Cox se halla implementado en la mayoría del *software* estadístico que se encuentra en el mercado y su análisis también aparece referenciado en los principales manuales de supervivencia. Por su parte, el análisis bayesiano completo del modelo Gamma-poligonal sólo puede ser llevado a cabo empleando *técnicas de Monte Carlo basadas en cadenas de Markov*. Básicamente, se trata de obtener una muestra a partir de la distribución final con la cual aproximar, por Monte Carlo, cualquier característica desconocida de la misma. En Geweke (1999) puede encontrarse una completa y actualizada revisión de dichos métodos y su aplicación práctica dentro de un contexto econométrico. Para una consulta más detallada del modelo Gamma-poligonal y su implementación al análisis de datos de desempleo puede consultarse Beamonte (1998).

En la siguiente sección se comenta el banco de datos finalmente analizado y su obtención a partir de la EPA. En la sección 3 se utiliza tanto el modelo de Cox como el modelo Gamma-poligonal aditivo para el análisis de los tiempos de desempleo. Finalmente, en la última sección se recogen las conclusiones más importantes derivadas de este estudio.

2 Los datos

El Instituto Nacional de Estadística es el organismo encargado de elaborar la EPA y de realizar posteriormente el oportuno tratamiento de la información que contiene. Se trata de encuestas con una periodicidad trimestral en las que, de una forma continuada a lo largo de las doce semanas del trimestre, se entrevista aproximadamente a 200000 personas que habitan unas 64000 viviendas familiares de todo el territorio español. Una sexta parte de la EPA es renovada trimestralmente, de modo que cada hogar - y por lo tanto cada uno de sus individuos componentes - forma parte de la muestra durante un máximo de seis períodos trimestrales consecutivos.

La EPA proporciona una amplia información sobre la población potencialmente activa y su principal finalidad es el análisis de la actividad económica en lo relativo a su componente humano. De este modo, posibilita el estudio del mercado laboral español y, por consiguiente, de uno de los más importantes factores de la estructura económica y social de un país.

Los datos utilizados para la realización de este trabajo corresponden a una cohorte de la población activa de las Comunidades Autónomas de Andalucía y Valencia. En concreto, los individuos que habitan dichas comunidades y que son entrevistados por primera vez y clasificados como desempleados por el encuestador en el primer trimestre de 2000. Para la obtención de la variable de supervivencia *tiempo de desempleo*, objeto del estudio, se efectúa el seguimiento de estos individuos a lo largo de seis paneles consecutivos de encuestas, hasta el segundo trimestre de 2001 incluido. De este modo se completa un ciclo de entrevistas de año y medio, al mismo tiempo que todos los encuestados comparten la misma coyuntura social, laboral y económica del país.

Así, utilizando la propia variable de la EPA *tiempo en meses buscando empleo*, DTBUS, considerada en la primera de las entrevistas y teniendo en cuenta la observación, a lo largo de las siguientes, del número de meses que permanece en situación de desempleo desde el primer trimestre de 2000, es estimado el tiempo de desempleo de cada individuo de la cohorte.

Tras una necesaria depuración de los datos, eliminando aquellos que presentaban evidentes incongruencias y, por ejemplo, considerando sólo individuos que han sido entrevistados al menos dos veces, con el fin de garantizar un mínimo seguimiento, obtuvimos una muestra de 841 datos para Andalucía, con 571 de ellos desempleados al final del estudio y 237 individuos de la Comunidad Valenciana, con un porcentaje de censura del 55.7 %. Estos dos grupos de datos conforman los dos estratos objeto del estudio comparativo.

Además, para cada uno de los individuos se han considerado siete covariables obtenidas directamente de la EPA y posiblemente relacionadas con su tiempo de desempleo. Así, hemos creído conveniente tener en cuenta variables como la *edad (años cumplidos por grupos quinquenales)*, *sexo*, *estado civil*, *nivel de formación alcanzado (estudios terminados)*, *¿ha cursado estudios en las 4 últimas semanas?*, *tipo de empleo que busca o ha encontrado* y *situación en la que se encontraba antes de buscar empleo*. En la Tabla 1 se muestran las covariables utilizadas, sus posibles valores y el significado que de ellas recoge la EPA.

Tras un análisis preliminar de los datos y para evitar categorías casi vacías, agrupamos las correspondientes a *viudo* y *separado o divorciado* para la covariable ECIV1; *educación secundaria: segunda etapa y formación e inserción laboral correspondiente con formación e inserción laboral post-secundaria, no superior*, así como *educación superior, excepto doctorado* con la categoría de *doctorado* para la covariable NFORM1. Para CURSA también agrupamos las categorías *en vacaciones* con *no* y para la covariable ANTBUS, las categorías *cumplía el servicio militar o civil sustitutorio* con *se ocupaba del hogar o tenía responsabilidades familiares* y con *otra situación*. Además, como no encontramos ningún individuo clasificado en ellas, las categorías correspondientes a *blanco* para ANTBUS y EMBUS fueron eliminadas.

Consideramos la variable EDAD11 como continua y el resto como covariables categóricas. La codificación efectuada para estas últimas es la habitual mediante variables *dummies*, donde el grupo de referencia (todas las *dummies* iguales a cero) es el correspondiente al último de los valores y respetamos la ordenación de los mismos. En definitiva, la muestra final es de 1078 individuos, de los cuales 703 se corresponden con datos censurados, para los cuales disponemos del correspondiente

covariable	valor	significado
EDAD11	16	<i>16-19</i>
	20	<i>20-24</i>
	25	<i>25-29</i>
	30	<i>30-34</i>
	35	<i>35-39</i>
	40	<i>40-45</i>
	45	<i>45-49</i>
	50	<i>50-54</i>
	55	<i>55-60</i>
	60	<i>60-64</i>
	65	<i>65 o más</i>
SEXO1	1	<i>varón</i>
	6	<i>mujer</i>
ECIV1	1	<i>soltero</i>
	2	<i>casado</i>
	3	<i>viudo</i>
	4	<i>separado o divorciado</i>
NFORM1	1	<i>educación primaria</i>
	2	<i>educación secundaria: primera etapa y formación e inserción laboral correspondiente</i>
	3	<i>educación secundaria: segunda etapa y formación e inserción laboral correspondiente</i>
	4	<i>formación e inserción laboral post-secundaria, no superior</i>
	5	<i>educación superior, excepto doctorado</i>
	6	<i>doctorado</i>
	8	<i>analfabetos</i>
CURSA	1	<i>sí</i>
	2	<i>en vacaciones</i>
	3	<i>no</i>
EMBUS	”	<i>blanco</i>
	1	<i>de jornada completa únicamente</i>
	2	<i>de jornada completa o, en su defecto, de jornada parcial</i>
	3	<i>de jornada parcial o, en su defecto, de jornada completa</i>
	4	<i>de jornada parcial únicamente</i>
ANTBUS	5	<i>de cualquier tipo</i>
	”	<i>blanco</i>
	1	<i>trabajaba</i>
	2	<i>era estudiante</i>
	3	<i>cumplía el servicio militar o civil sustitutorio</i>
	4	<i>se ocupaba del hogar o tenía responsabilidades familiares</i>
	5	<i>otra situación</i>

Tabla 1: Covariables utilizadas en el banco de datos.

tiempo de desempleo en meses, indicador de censura y quince covariables, todas ellas dicotómicas excepto EDAD11.

3 Análisis

3.1 El modelo de Cox

Primeramente, realizamos el análisis de los datos utilizando el conocido modelo de Cox. Considerando como estratos los dos grupos referidos a las Comunidades Autónomas, comenzamos el estudio con el modelo saturado, es decir, incorporando todas las covariables al mismo y procedimos a realizar una selección de variables por pasos hacia atrás, utilizando el método de la razón de verosimilitud basada en la estimaciones condicionales de los parámetros.

De este modo, la primera covariable en salir del modelo es ECIV1, en segundo lugar es eliminada la variable que hace referencia al nivel de estudios terminados, NFORM1, y finalmente la covariable CURSA. En la Tabla 2 se recogen los estadísticos y significatividades asociadas a los valores de las covariables.

Así, el modelo finalmente considerado cuenta con las covariables EDAD11, SEXO, EMBUS y ANTBUS; si bien puede apreciarse claramente en la columna de la significatividad de la Tabla 2 que las dos primeras categorías de la covariable EMBUS pueden ser consideradas iguales a la de referencia y tener, por tanto, tan sólo tres categorías dicha covariable.

La Tabla 2 también contiene los coeficientes del modelo de Cox. Su signo nos indica el sentido de la influencia de las covariables. Por ejemplo, el coeficiente correspondiente a EDAD11 vale -0.023 revelando una influencia positiva o directa de la edad, esto es, a mayores valores de EDAD11 se tienen mayores tiempos de desempleo. Por el contrario, la influencia de SEXO1 es inversa, con mayores tiempos de desempleo para la categoría de referencia, las mujeres. Para la covariable EMBUS, la pertenencia a las categorías *de jornada parcial o, en su defecto, de jornada completa* y *de jornada parcial únicamente*, aumenta el tiempo de desempleo. Finalmente,

		b	ET	Wald	gl	sig.	exp(b)
Paso 1	EDAD11	-0.027	0.007	15.898	1	0.000	0.973
	SEXO1	0.731	0.113	41.625	1	0.000	2.078
	ECIV1			1.082	2	0.582	
	ECIV1(1)	-0.294	0.287	1.049	1	0.306	0.745
	ECIV1(2)	-0.253	0.260	0.945	1	0.331	0.777
	NFORM1			5.012	4	0.286	
	NFORM1(1)	-0.134	0.405	0.109	1	0.742	0.875
	NFORM1(2)	-0.19	0.414	0.219	1	0.640	0.824
	NFORM1(3)	-0.371	0.430	0.745	1	0.388	0.690
	NFORM1(4)	0.034	0.425	0.007	1	0.936	1.035
	CURSA	-0.332	0.206	2.590	1	0.108	0.718
	EMBUS			6.211	4	0.184	
	EMBUS(1)	0.142	0.153	0.866	1	0.352	1.153
	EMBUS(2)	-0.001	0.135	0.000	1	0.996	0.999
	EMBUS(3)	-0.846	0.590	2.055	1	0.152	0.429
	EMBUS(4)	-0.735	0.424	2.999	1	0.083	0.480
	ANTBUS			9.689	2	0.008	
	ANTBUS(1)	0.630	0.206	9.310	1	0.002	1.877
	ANTBUS(2)	0.477	0.276	2.989	1	0.084	1.611
Paso 2	EDAD11	-0.025	0.005	22.160	1	0.000	0.975
	SEXO1	0.721	0.110	42.937	1	0.000	2.056
	NFORM1			4.866	4	0.301	
	NFORM1(1)	-0.098	0.403	0.059	1	0.808	0.907
	NFORM1(2)	-0.150	0.411	0.133	1	0.715	0.861
	NFORM1(3)	-0.339	0.428	0.626	1	0.429	0.713
	NFORM1(4)	0.064	0.423	0.023	1	0.880	1.066
	CURSA	-0.335	0.206	2.651	1	0.103	0.715
	EMBUS			6.386	4	0.172	
	EMBUS(1)	0.153	0.152	1.011	1	0.315	1.166
	EMBUS(2)	0.002	0.135	0.000	1	0.989	1.002
	EMBUS(3)	-0.843	0.588	2.054	1	0.152	0.430
	EMBUS(4)	-0.735	0.423	3.012	1	0.083	0.480
	ANTBUS			9.857	2	0.007	
	ANTBUS(1)	0.634	0.206	9.438	1	0.002	1.885
	ANTBUS(2)	0.476	0.275	2.995	1	0.084	1.609
Paso 3	EDAD11	-0.024	0.005	23.861	1	0.000	0.977
	SEXO1	0.714	0.109	43.283	1	0.000	2.042
	CURSA	-0.298	0.200	2.213	1	0.137	0.743
	EMBUS			6.729	4	0.151	
	EMBUS(1)	0.156	0.152	1.061	1	0.303	1.169
	EMBUS(2)	-0.016	0.134	0.014	1	0.907	0.984
	EMBUS(3)	-0.843	0.586	2.067	1	0.151	0.431
	EMBUS(4)	-0.766	0.423	3.279	1	0.070	0.465
	ANTBUS			9.857	2	0.008	
	ANTBUS(1)	0.634	0.205	9.517	1	0.002	1.884
	ANTBUS(2)	0.502	0.270	3.459	1	0.063	1.651
Paso 4	EDAD11	-0.023	0.005	22.401	1	0.000	0.978
	SEXO1	0.724	0.108	44.721	1	0.000	2.063
	EMBUS			7.352	4	0.118	
	EMBUS(1)	0.159	0.151	1.097	1	0.295	1.172
	EMBUS(2)	-0.013	0.134	0.009	1	0.924	0.987
	EMBUS(3)	-0.852	0.586	2.111	1	0.146	0.427
	EMBUS(4)	-0.821	0.422	3.791	1	0.052	0.440
	ANTBUS			10.094	2	0.006	
	ANTBUS(1)	0.625	0.205	9.281	1	0.002	1.869
	ANTBUS(2)	0.416	0.265	2.474	1	0.116	1.516

Tabla 2: Coeficientes, errores típicos, estadísticos de Wald, grados de libertad y significatividades de las variables en el modelo de Cox.

aquellas personas que con anterioridad a la búsqueda de empleo estaban cumpliendo el servicio militar o se dedicaban a las tareas del hogar presentan mayores tiempos de desempleo.

A continuación, se ilustra el análisis con las distribuciones de supervivencia de algunos individuos concretos. En la Figura 1 se representa la función de supervivencia de varones andaluces y valencianos de 20 años que buscan un empleo de jornada parcial o, en su defecto, completa y que antes de la búsqueda no trabajaban, ni estudiaban. La notable diferencia aquí observada entre sus tiempos de desempleo es genérica para cualquier par de individuos que compartan las mismas características, excepto la Comunidad Autónoma.

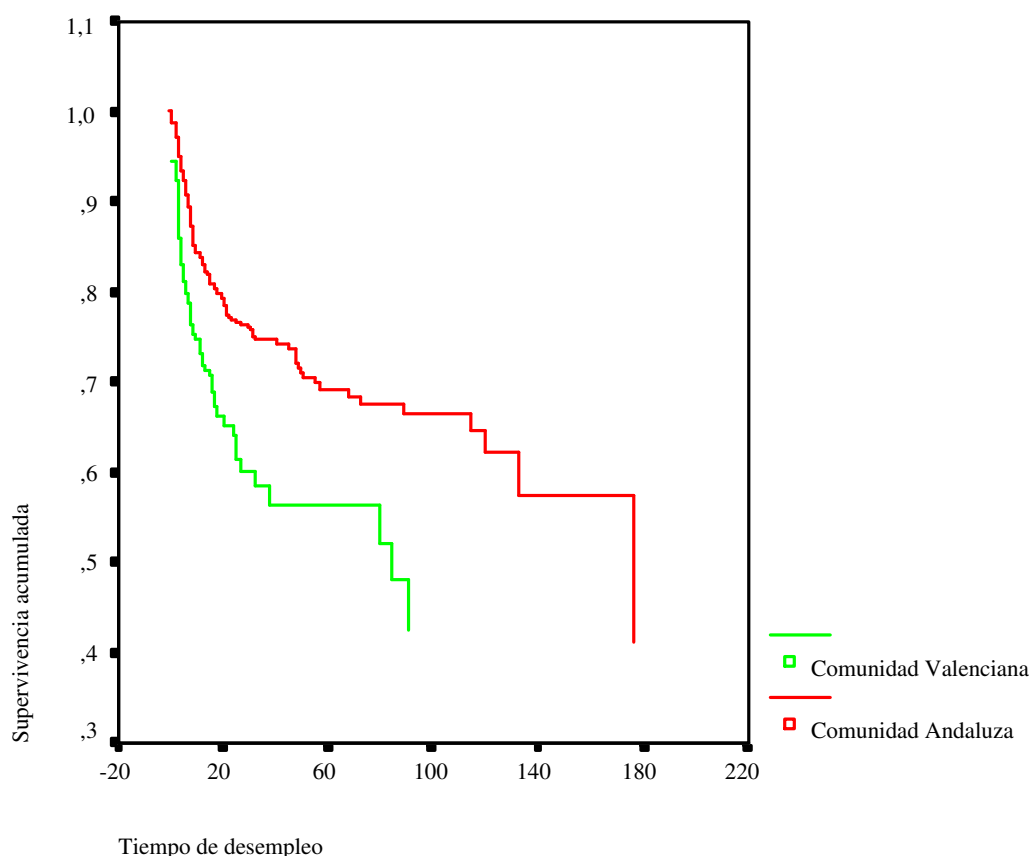


Figura 1: *Funciones de supervivencia para varones de 20 años que buscan un empleo de jornada parcial o completa y que antes no trabajaban, ni estudiaban.*

Por su parte, la Figura 2 muestra la función de supervivencia de una mujer valenciana de 30 años que antes de buscar empleo estudiaba, para las distintas categorías de la covariable EMBUS.

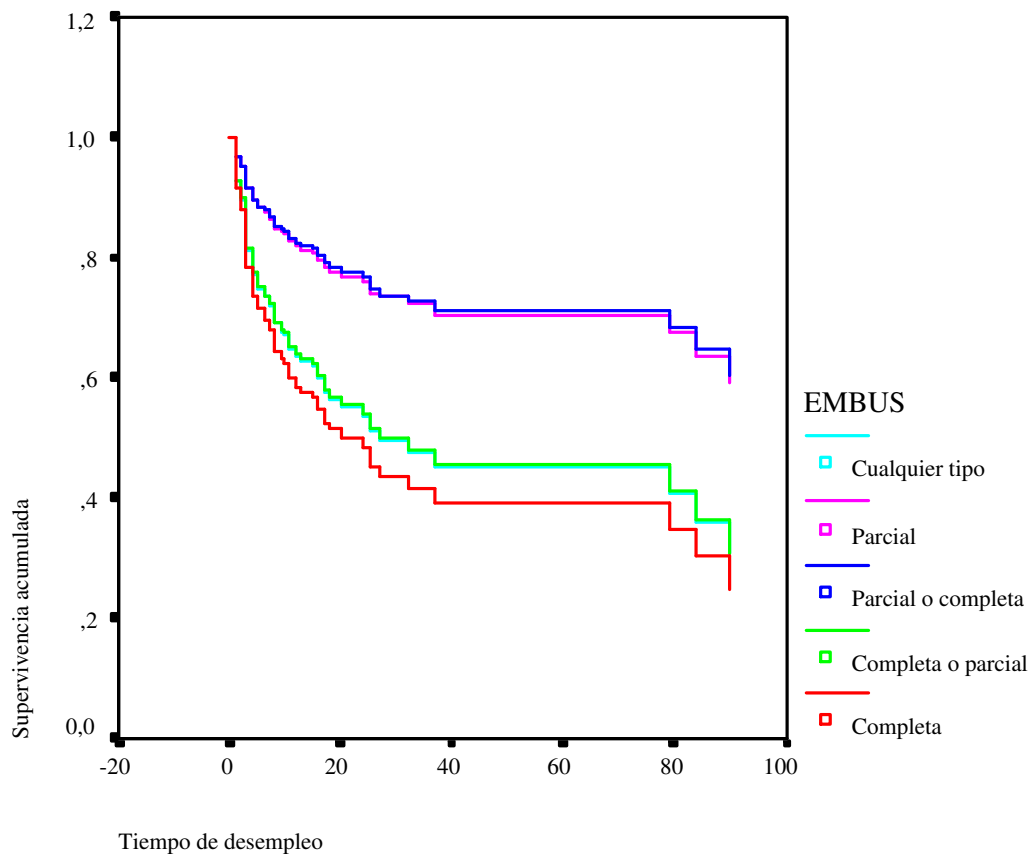


Figura 2: Funciones de supervivencia para una mujer valenciana de 30 años que antes estudiaba.

Es de destacar la práctica coincidencia de las categorías *de jornada parcial o, en su defecto, de jornada completa* y *de jornada parcial únicamente*, por una parte y *de jornada completa o, en su defecto, de jornada parcial* y *de cualquier tipo*, por otra, con menores tiempos de desempleo para éstas últimas y todavía menores para la restante categoría *de jornada completa*.

3.2 El modelo Gamma-poligonal aditivo

Para el análisis bayesiano de los tiempos de desempleo con el modelo Gamma-poligonal proponemos distribuciones iniciales propias pertenecientes a las familias conjugadas habituales y con varianzas muy grandes. De este modo, el estudio se centra básicamente en la información proveniente de los datos, introduciendo la menor información inicial posible.

El análisis preliminar de los tiempos de desempleo exige la generación de una larga cadena de Markov con la que estudiar la convergencia a la distribución final. La Figura 3 recoge la evolución de una cadena de Markov para los valores de algunos coeficientes de las covariables, obtenida desechando los 500000 primeros pasos (para acelerar la convergencia) y registrando uno de cada 50 pasos (para reducir su autocorrelación), hasta obtener una muestra de tamaño 10000. También se muestra la estimación Monte Carlo de la correspondiente distribución marginal.

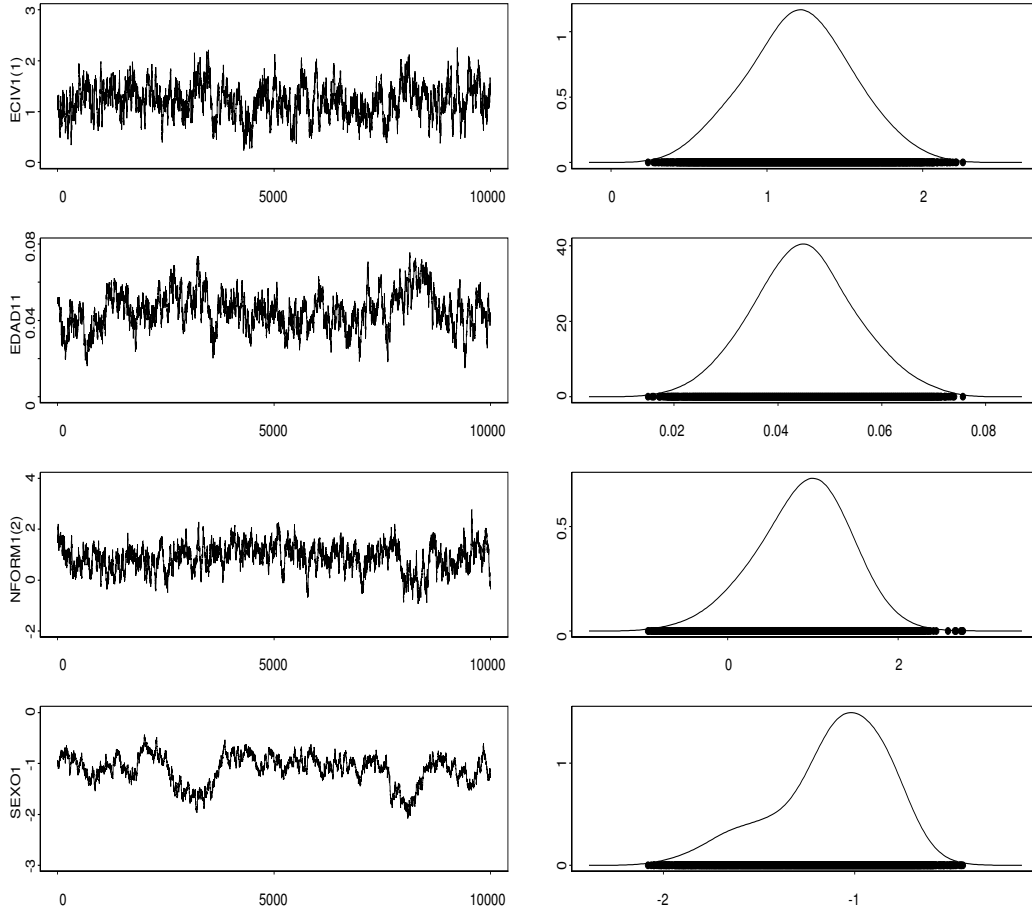


Figura 3: *Evolución de la cadena de Markov asociada a algunos parámetros del modelo, en la izquierda, y su densidad predictiva estimada, en la derecha.*

El diagnóstico de convergencia lo hemos realizado con la aplicación CODA (Best et al., 1995). Además de las trazas representadas en la Figura 3, los tests de Raftery y Lewis y de Geweke también muestran un buen comportamiento. Todo hace suponer que se ha alcanzado la estacionariedad y que, por tanto, ya disponemos de una muestra de la distribución final.

La Tabla 3 muestra los intervalos de confianza Monte Carlo del 95 % para los

parámetros del modelo correspondientes a las covariables. Algunos de ellos contienen al cero por lo que parece conveniente realizar una selección de covariables.

covariable	intervalo de confianza
EDAD11	[0.0264,0.065]
SEXO1	[-1.77,-0.694]
ECIV1(1)	[0.572,1.85]
ECIV1(2)	[0.359,1.48]
NFORM1(1)	[-0.419,1.66]
NFORM1(2)	[-0.248,1.82]
NFORM1(3)	[-0.384,1.93]
NFORM1(4)	[-0.643,1.5]
CURSA	[-0.382,1.56]
EMBUS(1)	[-0.499,0.781]
EMBUS(2)	[-0.293,0.497]
EMBUS(3)	[1.91,14]
EMBUS(4)	[0.0134,11]
ANTBUS(1)	[-0.651,0.407]
ANTBUS(2)	[-0.701,5.11]

Tabla 3: Intervalos de confianza del 95 % para algunos parámetros del modelo inicial.

Al igual que en el estudio anterior con el modelo de Cox, decidimos seguir un procedimiento *backward elimination* conservador y proceder a la eliminación de las covariables de una en una, reanalizando los datos con el modelo resultante antes de aceptar una nueva reducción del modelo. Así, la primera covariable en ser eliminada es CURSA y del modelo sin ella simulamos una cadena de Markov de forma análoga a como lo hicimos con el modelo completo, obteniendo también unos excelentes resultados en el diagnóstico de convergencia. En la Tabla 4 se proporcionan los intervalos de confianza Monte Carlo del 95 % para los parámetros de las covariables que quedan en el modelo.

Razonando de forma totalmente similar, eliminamos secuencialmente del modelo las covariables ANTBUS(2), EDUCA(2), EDUCA(3), EMBUS(2), EDUCA(1), EDUCA(4) y EMBUS(1). La Tabla 5 proporciona los intervalos de confianza del 95 % para los parámetros del modelo final. Aunque algunos de ellos contienen al cero, la alta correlación de la cadena de Markov asociada a dichas covariables y el hecho de que la inclusión sea casi tangencial, motivó que decidiéramos no eliminarlas del modelo. Consecuentemente, el modelo finalmente considerado cuenta con las covariables EDAD11, SEXO1, ECIV1, EMBUS con tres categorías: *de jornada*

covariable	intervalo de confianza
EDAD11	[0.0145,0.0359]
SEXO1	[-0.676,-0.325]
ECIV1(1)	[0.145,1.11]
ECIV1(2)	[0.041,0.895]
NFORM1(1)	[-0.0394,1.33]
NFORM1(2)	[0.0376,1.42]
NFORM1(3)	[0.133,1.55]
NFORM1(4)	[-0.203,1.2]
EMBUS(1)	[-0.293,0.271]
EMBUS(2)	[-0.053,0.401]
EMBUS(3)	[-0.282,0.949]
EMBUS(4)	[0.191,12.7]
ANTBUS(1)	[-0.449,0.0687]
ANTBUS(2)	[-0.439,0.321]

Tabla 4: Intervalos de confianza del 95 % para algunos parámetros del modelo.

parcial o, en su defecto, de jornada completa, de jornada parcial únicamente y resto (considerada ésta como la categoría de referencia); y la covariable ANTBUS, que hace referencia a la situación del individuo antes de buscar empleo, con sólo dos categorías: *trabajaba* y *resto*.

covariable	intervalo de confianza
EDAD11	[0.0568,0.132]
SEXO1	[-2.32,-1.07]
ECIV1(1)	[1.43,3.59]
ECIV1(2)	[0.821,2.86]
EMBUS(1)	[-1.15,3.02]
EMBUS(2)	[-0.359,8.45]
ANTBUS	[-1.22,0.281]

Tabla 5: Intervalos de confianza del 95 % para los parámetros del modelo final.

De un modo análogo al seguido en pasos anteriores, simulamos una muestra a partir de la distribución final con la cual poder obtener distribuciones predictivas de algunos individuos. Por ejemplo, en la Figura 4 se representan las funciones de supervivencia predictivas del tiempo de desempleo para varones andaluces de diferentes edades, solteros, que buscan cualquier tipo de trabajo y que antes han trabajado. En la Figura 5 se muestran las correspondientes predictivas para varones valencianos con las mismas características y en ambas se recoge en la leyenda la media y la desviación típica del tiempo de desempleo. Así, la probabilidad de que

un individuo valenciano con esas características y de 30 años de edad encuentre trabajo antes de dos años es 0.55, mientras que si dicho individuo es andaluz la probabilidad desciende hasta 0.41.

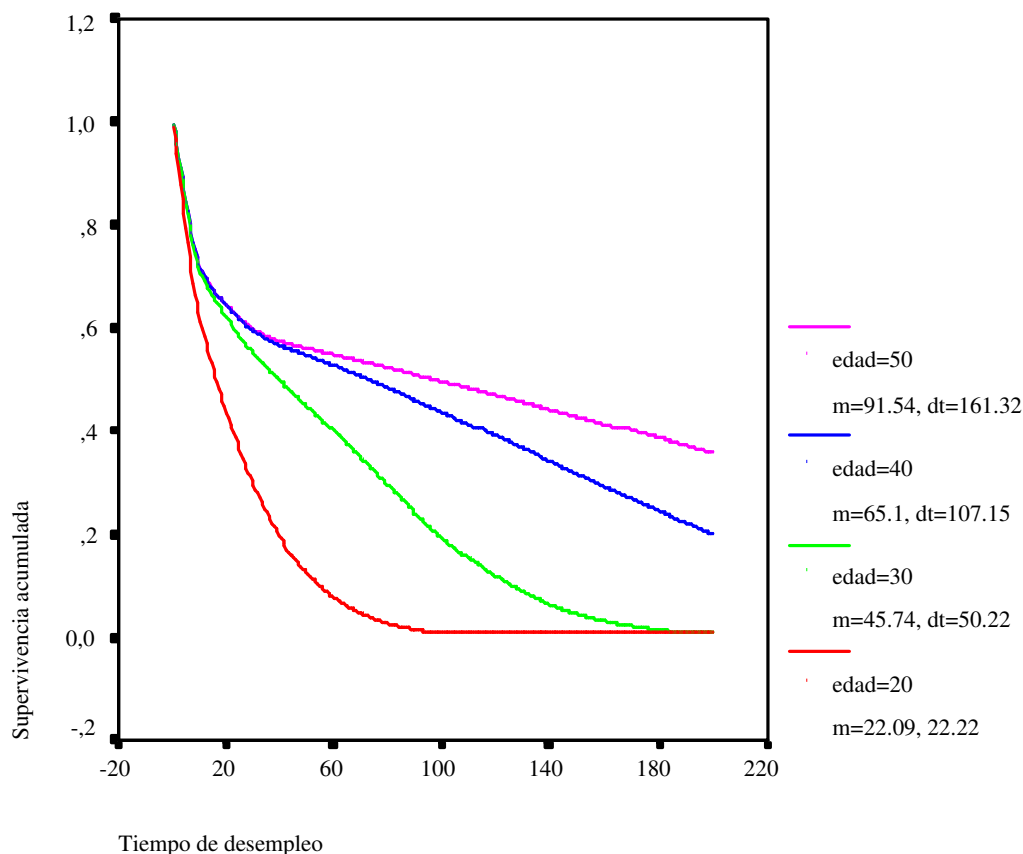


Figura 4: Funciones de supervivencia predictivas para varones andaluces solteros, que buscan cualquier tipo de trabajo y que antes han trabajado.

Cabe destacar de ambas que la edad influye de un modo importante en la supervivencia de modo que los más jóvenes encuentran con mayor facilidad empleo. Además, se tienen mayores tiempos de supervivencia genéricos para la población andaluza que para los habitantes de la Comunidad Valenciana.

Las Figuras 6 y 7 muestran, respectivamente, las distribuciones de supervivencia predictivas de hombres y mujeres que buscan un trabajo de jornada parcial o completa y que antes han trabajado, para diferentes estados civiles y para ambas Comunidades Autónomas. Se proporcionan las medias y desviaciones típicas en las leyendas.

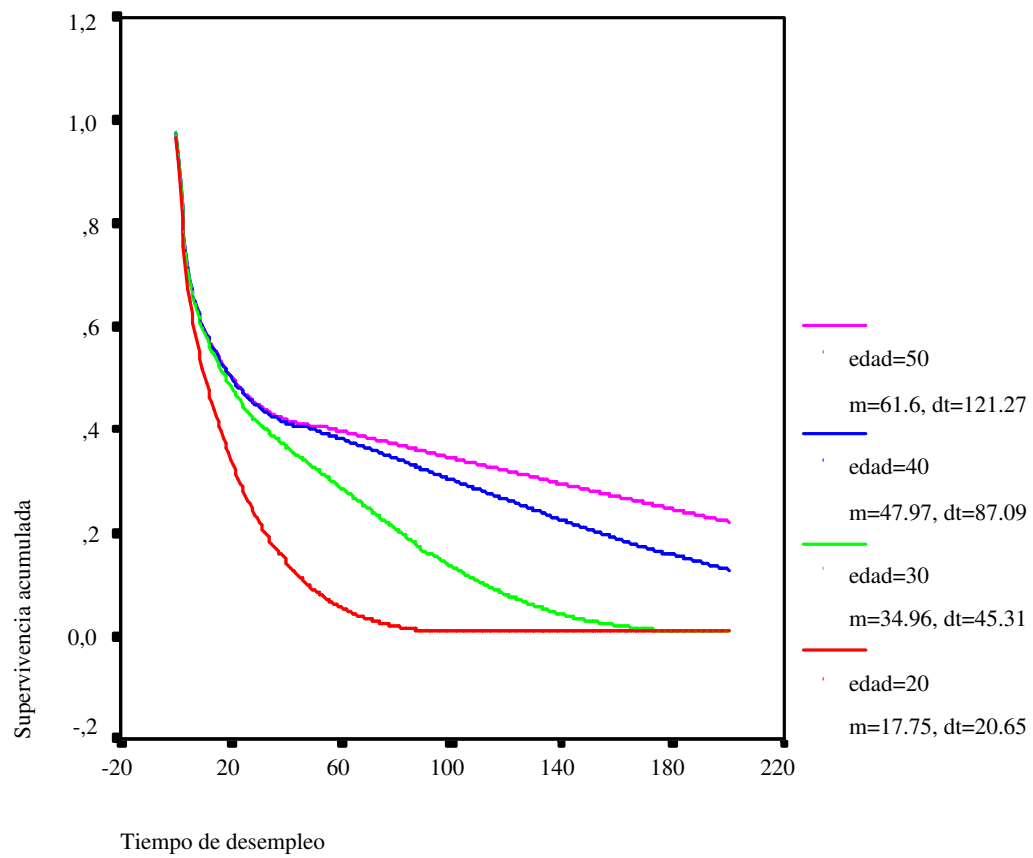


Figura 5: *Funciones de supervivencia predictivas para varones valencianos solteros, que buscan cualquier tipo de trabajo y que antes han trabajado.*

De ellas puede concluirse que los hombres desempleados encuentran trabajo antes que las mujeres y que la ordenación para el estado civil es la siguiente: las personas separadas salen de la situación de desempleo antes que las casadas y éstas antes que los individuos solteros.

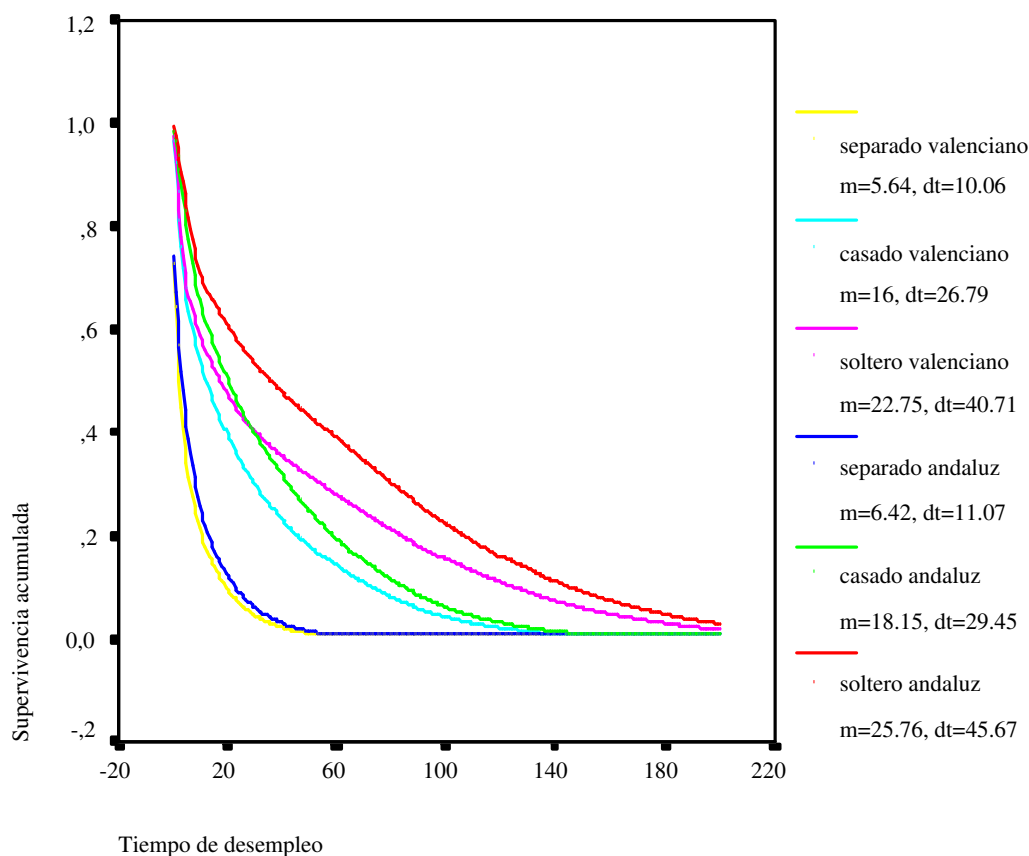


Figura 6: *Funciones de supervivencia predictivas para varones que buscan un trabajo de jornada parcial o completa y que antes han trabajado.*

Las diferencias más notables entre ambas comunidades se dan, además de para sexos, para las personas solteras, que se corresponden con los mayores tiempos de desempleo. Genéricamente, para los primeros meses se tienen curvas de supervivencia muy similares para individuos andaluces y valencianos con las mismas características personales, acentuándose las diferencias con el transcurso del tiempo en la situación de desempleo.

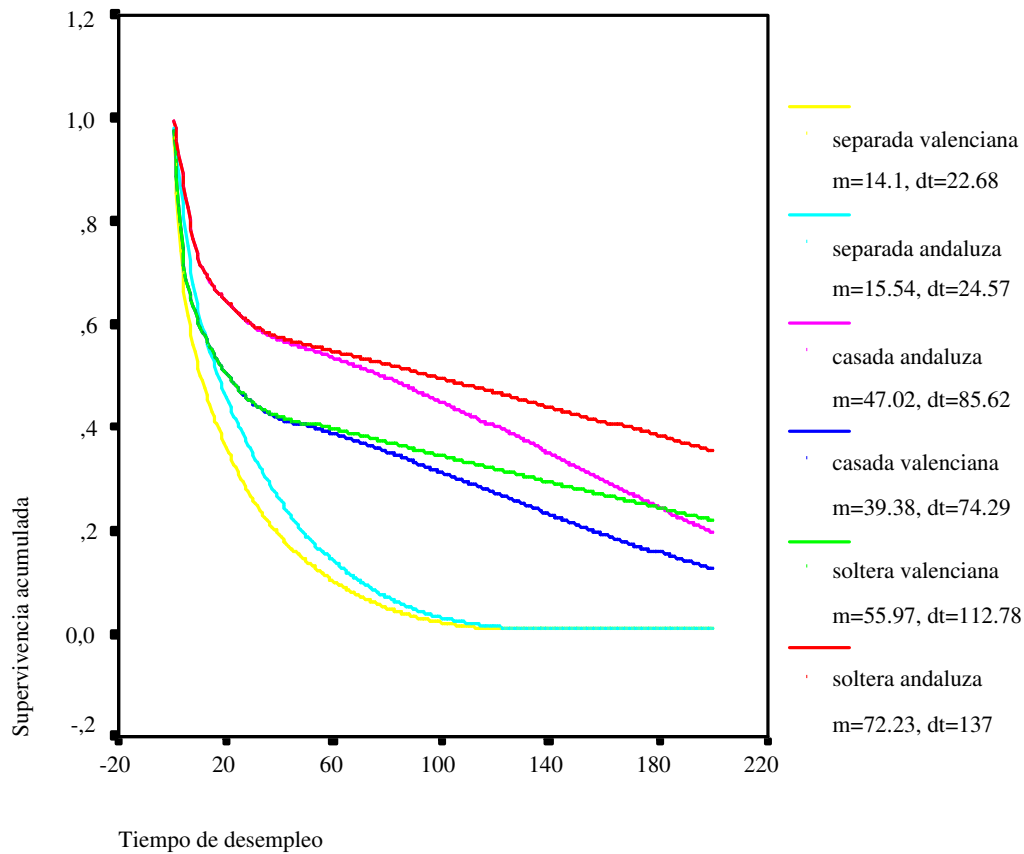


Figura 7: Funciones de supervivencia predictivas para mujeres que buscan un trabajo de jornada parcial o completa y que antes han trabajado.

4 Conclusiones

El estudio comparativo de los datos de desempleo en Andalucía y en la Comunidad Valenciana ha proporcionado supervivencias menores para esta última Comunidad Autónoma. Es decir, genéricamente los andaluces presentan tiempos de desempleo mayores que los individuos de la Comunidad Valenciana. Estas diferencias han sido obtenidas utilizando tanto el modelo de riesgos proporcionales de Cox como el modelo Gamma-poligonal, resultando tanto más apreciables cuanto más se demora la entrada en el empleo del individuo. Si la persona desempleada encuentra trabajo en un período relativamente breve de tiempo, entonces no existe mucha diferencia entre la Comunidad Autónoma de residencia.

La investigación llevada a cabo supone una actualización del trabajo de Beaumont y Bermúdez (2002), en el que fueron tratados datos de la EPA de los años

1997 y 1998. Aunque aquél consideraba tan sólo datos de la Comunidad Valenciana, las principales diferencias encontradas con el paso del tiempo han sido la clara influencia actual de la covariable sexo en el tiempo de desempleo (al contrario de lo concluido en el estudio anterior), que la covariable CURSA ya no influye significativamente en el tiempo de desempleo y que algo similar ocurre con el nivel de formación, para el que no se han encontrado diferencias significativas según sus niveles.

Como elementos comunes a ambos estudios y, por lo tanto, mantenidos en el tiempo, destacamos la influencia en el tiempo de desempleo de las covariables edad y estado civil. La primera de ellas se erige, para las dos referencias temporales, en la variable con mayor poder explicativo del tiempo de desempleo de un individuo, manteniéndose, obviamente, su clara influencia positiva en el mismo. Esto es, los jóvenes, tanto andaluces como valencianos, encuentran empleo antes que las personas de más edad. La ordenación de las supervivencias en función del estado civil también se conserva a lo largo del tiempo, obteniéndose mayores tiempos de desempleo para los individuos solteros, sin compromiso familiar alguno.

Finalmente, cabe destacar la concordancia entre las conclusiones alcanzadas utilizando el modelo de Cox y las obtenidas con el modelo Gamma-poligonal. La principal diferencia entre ambos es la inclusión en éste último del estado civil, mientras que Cox encuentra que esta covariable no tiene una influencia significativa en el tiempo de desempleo de andaluces y valencianos.

Referencias

Beamonte, E. (1998). *Aportaciones al análisis bayesiano semiparamétrico de datos de supervivencia*. Tesis Doctoral. Valencia: Universitat de València.

Beamonte, E. y J.D. Bermúdez (2002). Un análisis del mercado laboral relativo a la población valenciana que busca su primer empleo. *Estudios de Economía Aplicada* (en prensa).

Best, N.G., M.K. Cowles y S.K. Vines (1995). *CODA manual version 0.30*. Cam-

bridge: MRC Biostatistics Unit.

Cox, D.R. (1972). Regression models and life tables (with discussion). *Journal of the Royal Statistical Society B* 34, pp. 187-220.

Geweke, J. (1999). Using simulation methods for Bayesian econometric models: inference, development, and communication (with comments). *Econometric Reviews* 18, pp. 1-126.