

ANÁLISIS BAYESIANO DE LA DURACIÓN DEL DESEMPLEO*

José D. Bermúdez y Eduardo Beamonte**

WP-EC 2000-21

Correspondencia: Eduardo Beamonte, Universitat de València. Dpto. de Economía Aplicada, Ed. Departamental Oriental. Avda. de los Naranjos, s/n. 46022 Valencia. E-mail: beamonte@uv.es.

Editor: Instituto Valenciano de Investigaciones Económicas
Primera Edición Diciembre 2000
Depósito Legal: V-5170-2000

Los documentos de trabajo del IVIE ofrecen un avance de los resultados de las investigaciones económicas en curso, con objeto de generar un proceso de discusión previo a su remisión a las revistas científicas.

* Este trabajo ha sido financiado parcialmente por el Ministerio de Educación y Ciencia con cargo al proyecto PB96-0776. Agradecemos a Ángel Soler y Rodrigo Aragón, del Instituto Valenciano de Investigaciones Económicas, su desinteresada colaboración en la obtención de los datos de la Encuesta de Población Activa. Asimismo, los comentarios y sugerencias de un evaluador anónimo han mejorado la presentación de este trabajo.

** J.D. Bermúdez y E. Beamonte: Universitat de València.

ANÁLISIS BAYESIANO DE LA DURACIÓN DEL DESEMPLEO*

José D. Bermúdez y Eduardo Beamonte**

WP-EC 2000-21

Correspondencia: Eduardo Beamonte, Universitat de València. Dpto. de Economía Aplicada, Ed. Departamental Oriental. Avda. de los Naranjos, s/n. 46022 Valencia. E-mail: beamonte@uv.es.

Editor: Instituto Valenciano de Investigaciones Económicas
Primera Edición Diciembre 2000
Depósito Legal: V-5170-2000

Los documentos de trabajo del IVIE ofrecen un avance de los resultados de las investigaciones económicas en curso, con objeto de generar un proceso de discusión previo a su remisión a las revistas científicas.

* Este trabajo ha sido financiado parcialmente por el Ministerio de Educación y Ciencia con cargo al proyecto PB96-0776. Agradecemos a Ángel Soler y Rodrigo Aragón, del Instituto Valenciano de Investigaciones Económicas, su desinteresada colaboración en la obtención de los datos de la Encuesta de Población Activa. Asimismo, los comentarios y sugerencias de un evaluador anónimo han mejorado la presentación de este trabajo.

** J.D. Bermúdez y E. Beamonte: Universitat de València.

ANÁLISIS BAYESIANO DE LA DURACIÓN DEL DESEMPLEO

José D. Bermúdez y Eduardo Beamonte

RESUMEN

En este trabajo se propone un modelo jerárquico Gamma para el análisis bayesiano de tiempos de desempleo con covariables. Modelizamos la función de azar o de riesgo de cada individuo con la correspondiente a una distribución Gamma asociada a sus covariables mediante una relación no determinista incluida en el orden más alto de la jerarquía. Desde una perspectiva bayesiana, se realiza el estudio de la distribución final utilizando técnicas de Monte Carlo basadas en cadenas de Markov. Con información procedente de la Encuesta de Población Activa, se aplica el modelo teórico propuesto al análisis de datos de desempleo sobre los parados valencianos que buscan su primer empleo.

Palabras clave: Datos de desempleo, distribución predictiva, Encuesta de Población Activa, Inferencia Bayesiana, técnicas de Monte Carlo.

ABSTRACT

In this paper a hierarchic Gamma model for the Bayesian analysis of unemployment duration with covariables is proposed. We modelize the risk function of each individual with the corresponding to a Gamma distribution associated with its covariables by means of a non-deterministic relationship included in the highest order of the hierarchy. The final distribution using Monte Carlo techniques based on Markov chains is accomplished from a Bayesian perspective. Using data from the Spanish Labour Force Survey, the proposed theoretic model is applied to the analysis of the unemployed of the Valencian region that are looking for their first job.

Key words: Unemployment data, predictable distribution, Labour Force Survey, Bayesian inference, Monte Carlo techniques.

1 Introducción

El análisis de supervivencia o de tiempos de espera entre dos sucesos perfectamente definidos ha sido utilizado por numerosos autores para el estudio de tiempos de desempleo. En este contexto socio-económico cabe situar trabajos como los de Follmann et al. (1990), Beenstock (1996) y Gonzalo (1998), en los que se tratan estos tiempos atendiendo a su particular tipología que se caracteriza porque algunos de ellos pueden estar censurados, esto es, no son conocidos exactamente al no haber sido observado alguno de los sucesos que los definen.

En cualquier análisis de supervivencia es habitual la consideración de las características propias de cada individuo. Es evidente que circunstancias personales como la edad, estado civil o nivel de estudios deben influir 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 azares 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. No obstante, el modelo de Cox presenta el grave inconveniente de su restricción a azares proporcionales y su complicado análisis desde un punto de vista bayesiano.

El modelo paramétrico que proponemos es un modelo jerárquico en el que la función de azar o riesgo es una función de azar Gamma particular para cada individuo y asociada a sus covariables mediante una relación no determinista incluida en el orden más alto de la jerarquía. De este modo, al considerar un modelo en poblaciones se trata de recoger la heterogeneidad propia de la población que frecuentemente no es explicada por las covariables.

El modelo jerárquico Gamma que se propone es una particularización del modelo Gamma-poligonal aditivo (Beamonte, 1998), en el que la función de azar es modelizada mediante la suma de una función de azar base poligonal no completamente especificada y una función de azar Gamma dependiente del vector de covariables de cada individuo. Este modelo Gamma-poligonal es un modelo en poblaciones semiparamétrico aditivo que guarda ciertas analogías con el modelo de Cox al considerar ambos dos partes, una paramétrica y otra no paramétrica, en la función de azar e incorporar las covariables vía la parte

paramétrica. Sin embargo, la principal diferencia entre los dos, aparte de que el propuesto se trata de un modelo en poblaciones, es que el 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.

En consecuencia, el modelo jerárquico Gamma que aquí se presenta admite un razonable tratamiento desde la perspectiva bayesiana y también resulta adecuado en situaciones de supervivencia con azares no proporcionales, por lo que puede ser propuesto como una alternativa al modelo de Cox.

Los grandes avances en materia de computación posibilitan el análisis bayesiano completo de los modelos jerárquicos empleando técnicas de simulación. Desde la introducción del muestreo de Gibbs en el análisis bayesiano (Geman y Geman, 1984), se ha profundizado mucho en las propiedades matemáticas y en la metodología de ésta y otras técnicas de Monte Carlo basadas en cadenas de Markov (MCMC) que han sido ampliamente utilizadas con éxito. 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.

En la siguiente sección se explicita el modelo jerárquico Gamma utilizado para el tratamiento bayesiano de tiempos de desempleo con covariables y se propone una implementación del mismo utilizando el algoritmo de Metropolis dentro de Gibbs para la obtención de una muestra aleatoria a partir de la distribución final. En la sección 3 se aplica el modelo al análisis de unos datos de desempleo de individuos de la Comunidad Valenciana que buscan su primer trabajo, obteniendo distribuciones predictivas para individuos concretos y en la sección 4 se comentan algunas conclusiones interesantes.

2 Un modelo jerárquico Gamma y su metodología de análisis

Sea T el tiempo de espera de un individuo con vector de covariables \mathbf{x} . Para la variable aleatoria T proponemos una función de azar Gamma con parámetros

α y β (media α/β y varianza α/β^2) que, a su vez, son modelizados según la siguiente relación estocástica:

$$\begin{aligned}\beta &\sim N(\log \beta | \mu_\beta, \sigma_\beta^2) \\ \alpha | \beta &\sim N\left(\log \frac{\alpha}{\beta} | \mathbf{b}'\mathbf{x}, \sigma_\alpha^2\right).\end{aligned}$$

Los parámetros α y β son específicos para cada uno de los individuos y están relacionados con el vector \mathbf{x} mediante un segundo nivel del modelo jerárquico. De este modo, el modelo jerárquico Gamma propuesto resulta un modelo en poblaciones o de efectos aleatorios que permite cierta heterogeneidad en la población.

Cualquiera que sea la distribución inicial elegida, la distribución final resulta tan complicada que no parece posible llevar a cabo su estudio de un modo analítico. Este problema, común a casi todo modelo en poblaciones, puede ser abordado utilizando técnicas de simulación. En este trabajo proponemos el algoritmo de Metropolis dentro de Gibbs para la obtención de una muestra a partir de la distribución final, pero cualquier otro método MCMC puede resultar igualmente válido.

El algoritmo de Gibbs (Gelfand y Smith, 1990; Chib y Greenberg, 1996; Geweke, 1999) es, sin duda, el método de simulación dinámica más utilizado en los últimos tiempos. Al igual que una gran parte de ellos, también resulta un caso particular del método de Hastings (1970), en el que se obtienen muestras de las distribuciones condicionales completas componente a componente. Si consideramos el vector paramétrico $\boldsymbol{\theta}$ y denominamos $p(\boldsymbol{\theta} | y)$ a la distribución de interés, la estrategia radica en descomponer el vector en bloques y definir la función de transición de probabilidades como el producto de las densidades condicionales completas, esto es, la densidad condicional de cada bloque dados los datos y el resto de parámetros no incluidos en el bloque.

Básicamente, la implementación del algoritmo de Gibbs para la obtención de una muestra aleatoria a partir de la distribución final del vector paramétrico completo requiere la especificación de un punto inicial de la cadena de Markov con el que comenzar el proceso iterativo y saber muestrear de las distribuciones condicionales completas. Para ello, es fundamental realizar una adecuada descomposición del vector paramétrico completo. El esquema algorítmico de este método es el siguiente:

```

INICIALIZAR
    descomponer  $\boldsymbol{\theta} = (\theta_1, \dots, \theta_k)$ 
     $\boldsymbol{\theta}^{(0)}$ ;  $t \leftarrow 0$ 
REPETIR
    generar  $\theta_1^{(1)} \sim p(\theta_1 | y, \theta_2^{(0)}, \dots, \theta_k^{(0)})$ 
    generar  $\theta_2^{(1)} \sim p(\theta_2 | y, \theta_1^{(1)}, \theta_3^{(0)}, \dots, \theta_k^{(0)})$ 
    .
    .
    .
    generar  $\theta_{k-1}^{(1)} \sim p(\theta_{k-1} | y, \theta_1^{(1)}, \dots, \theta_{k-2}^{(1)}, \theta_k^{(0)})$ 
    generar  $\theta_k^{(1)} \sim p(\theta_k | y, \theta_1^{(1)}, \dots, \theta_{k-1}^{(1)})$ 
     $t \rightarrow (t + 1)$ .

```

Bajo condiciones muy generales, la sucesión $(\boldsymbol{\theta}^{(1)}, \dots, \boldsymbol{\theta}^{(n)})$ construida de este modo es una realización de una cadena de Markov con $p(\boldsymbol{\theta} | y)$ como distribución estacionaria.

El algoritmo de Metropolis dentro de Gibbs (Muller, 1991) es tan sólo un caso particular del algoritmo de Gibbs en cuanto a la forma de muestrear de las distribuciones condicionales completas, cuando no resulta posible un muestreo directo de las mismas. Se trata de un algoritmo de aceptación-rechazo que utiliza una determinada función importante para el muestreo dentro de algunas etapas de actualización del algoritmo de Gibbs. Su esquema algorítmico queda como sigue:

```

INICIALIZAR  $\theta_0$ ;  $t \leftarrow 0$ 
REPETIR
    generar  $\theta^* \sim f(\theta | \theta_t)$  simétrica,  $u \sim Un(0, 1)$ 
    si  $u \leq \alpha(\theta_t, \theta^*)$  entonces  $\theta_{t+1} \leftarrow \theta^*$ 
    si no  $\theta_{t+1} \leftarrow \theta_t$ 
     $t \rightarrow (t + 1)$ ,

```

donde:

$$\alpha(\theta_t, \theta^*) = \min \left[\frac{p(\theta^* | y)}{p(\theta_t | y)}, 1 \right].$$

Una definición formal del modelo Gamma-poligonal aditivo, que generaliza al modelo jerárquico propuesto, así como una descripción técnica del mismo y su análisis e implementación, aparece en Beamonte y Bermúdez (1999) y, con mayor detalle, en Beamonte (1998).

3 La situación de desempleo de los parados sin experiencia laboral en la Comunidad Valenciana

Las técnicas del análisis de supervivencia ya han sido aplicadas con anterioridad al estudio de datos provenientes de la EPA, donde la respuesta de los individuos a una de las preguntas del cuestionario (*¿en qué fecha empezó a buscar empleo?*) puede proporcionar directamente el tiempo de supervivencia. En este sentido, Alba-Ramírez (1999) propone la utilización de un modelo logit multinomial para el estudio de los tiempos de desempleo de hombres con una edad comprendida entre 20 y 59 años, mientras que Bover y Gómez (1999) estiman modelos de duración discreta con salidas del desempleo múltiples, bien hacia el empleo, la inactividad o el estudio.

Los datos aquí utilizados corresponden a una parte de la población activa de la Comunidad Valenciana. En concreto, la población parada, en busca de su primer empleo, entre el tercer trimestre de 1997 y el último de 1998, de modo que se completa un ciclo de entrevistas de año y medio. Utilizando estos seis paneles de encuestas se obtiene la variable objeto de estudio como los meses de búsqueda del citado primer empleo. Además, algunos otros items de la encuesta, posiblemente relacionados con el tiempo de desempleo, son incorporados como covariables. Así, para cada individuo son considerados su sexo (0, mujer y 1, hombre), edad (en años), nivel de estudios, EC=estado civil (0, casado y 1, soltero) y CURSA=si ha seguido durante las cuatro últimas semanas algún tipo de estudios (0, no y 1, sí).

La utilización de esta población responde a sus particulares características que le confieren una enorme importancia desde el punto de vista económico, además de gozar de una mayor sensibilidad social al estar compuesta mayoritariamente por jóvenes. También la propia EPA facilita este tipo de estudio al proporcionar de una forma relativamente sencilla el tiempo de desempleo.

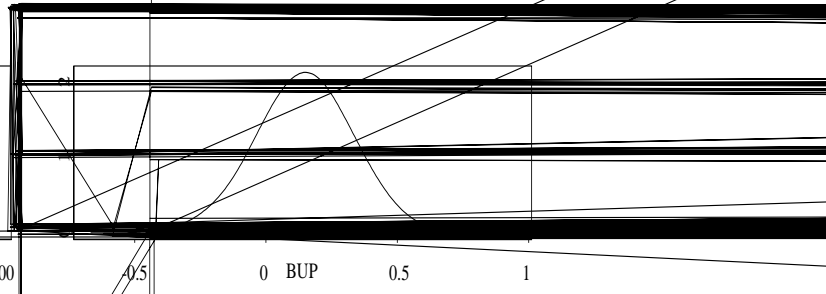
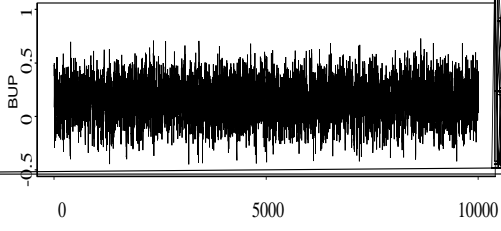
Para la variable *nivel de estudios* consideramos inicialmente seis categorías, las correspondientes a sin estudios (SIN), EGB y ESO (EGB), FPI, BUP, FPII y estudios universitarios (UNI). La codificación efectuada con esta covariable categórica es la habitual mediante cinco *dummies*, donde consideramos como grupo de referencia (todas las *dummies* iguales a cero) el correspondiente a estudios universitarios y respetamos la ordenación anterior.

Consideramos los datos progresivamente censurados por la derecha, en el sentido de que a lo largo de los seis trimestres un dato es censurado si el individuo continúa desempleado en su última entrevista. Tras una necesaria depuración de los datos, la muestra final es de 375 individuos, de los cuales 188 se corresponden con datos censurados, para los cuales disponemos del correspondiente tiempo de desempleo en meses, indicador de censura y nueve covariables (ocho de ellas dicotómicas).

Para realizar el análisis preliminar de esos datos generamos una larga cadena de Markov desechando los 100,000 primeros pasos (para alcanzar convergencia a la distribución final) y registrando uno de cada 100 pasos (para reducir la autocorrelación de la cadena), hasta obtener una muestra de tamaño 10,000. La Figura 1 recoge la evolución de dicha cadena para los valores de algunos coeficientes de las covariables y la estimación Monte Carlo de la correspondiente distribución marginal.

El diagnóstico de convergencia lo hemos realizado con la aplicación CODA (Best et al., 1995). Además de las trazas representadas en la Figura 1, 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 1 muestra los intervalos de confianza Monte Carlo al 95% para los parámetros del modelo. Algunos de ellos, los correspondientes a las covariables SEXO y BUP, contienen al cero por lo que parece conveniente realizar una selección de covariables.



parámetro	2.5%	50%	97.5%
SEXO	-0.3480	-0.1400	0.0642
EDAD	0.0681	0.0894	0.1100
SIN	0.4570	0.8960	1.3100
EGB	0.4260	0.6990	0.9770
FPI	0.2820	0.6620	1.0200
BUP	-0.2130	0.1480	0.4840
FPII	0.0742	0.4270	0.7610
EC	1.0800	1.5700	2.0800
CURSA	0.0097	0.2310	0.4480

Tabla 1: *Intervalos intercuantílicos para los parámetros del modelo inicial.*

La Tabla 1 también proporciona información interesante sobre las categorías de la covariable *nivel de estudios*. Los coeficientes de EGB y FPI son muy parecidos, lo que sugiere que esos dos niveles podrían resultar prácticamente equivalentes para la obtención de empleo, y tampoco son muy distintos del coeficiente de SIN. Otro hecho relevante es que el coeficiente de BUP no es significativamente distinto de cero, por lo que dicho nivel puede ser en la práctica equivalente al nivel de estudios universitarios, que es el grupo de referencia.

Hemos utilizado contrastes para investigar con mayor detalle las relaciones entre los distintos grupos de la variable *nivel de estudios*. Así por ejemplo, la Figura 2 muestra el resultado del estudio de la distribución final conjunta de los contrastes SIN-EGB y EGB-FPI. A partir de la muestra de la distribución final obtenida por Gibbs sólo hay que calcular, para cada vector simulado, la diferencia entre los coeficientes de las variables SIN y EGB y de las variables EGB y FPI. Los datos bivariantes así obtenidos son una muestra de la distribución final de los contrastes SIN-EGB y EGB-FPI. Las campanas dibujadas arriba y a la derecha de la Figura 2 son las dos densidades marginales, obtenidas por el método del Kernel para estimación de densidades, y en ellas se señala una estimación de sus cuantiles 0.025 y 0.975, lo que constituye un intervalo de confianza al 95%. En el centro de la figura se muestra la región de confianza bidimensional, con contenido probabilístico 0.95, obtenida mediante el método de Wei y Tanner (1990). Se observa que ambos intervalos univariantes incluyen al cero y que el punto (0,0) también pertenece a la región bidimensional, por lo que directamente podríamos aceptar la hipótesis de igualdad entre los tres grupos SIN, EGB y FPI. Sin embargo preferimos seguir un procedimiento más

conservador y aceptar en cada etapa sólo una igualdad, reanalizando los datos con el modelo resultante antes de aceptar una nueva reducción del modelo.

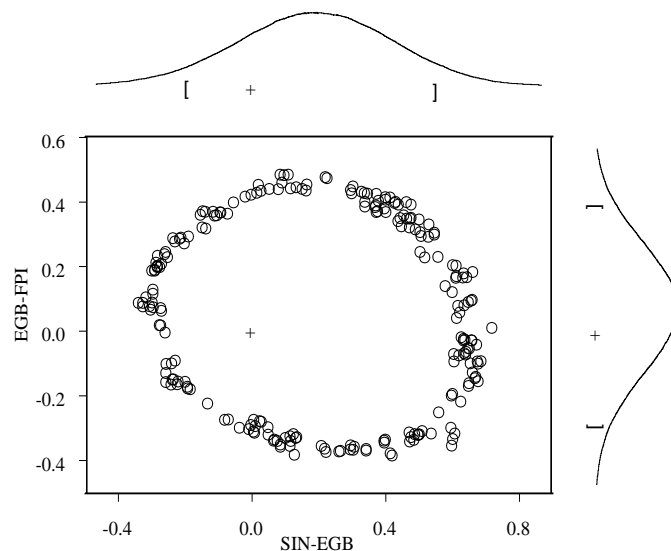


Figura 2: Regiones de confianza al 95% para los contrastes *SIN-EGB* y *EGB-FPI*. En el centro se representa la región bidimensional y los intervalos univariantes se muestran en los laterales de la gráfica. El valor cero se muestra con un signo +.

Así pues, hemos utilizado contrastes analizados de esta forma para realizar una selección de covariables en pasos sucesivos hacia atrás, *backware elimination*, que han proporcionado como resultado final el modelo que incluye las covariables EDAD, EC, CURSA y *nivel de estudios*. Además, las categorías de ésta última han sido reagrupadas en cuatro: G1 constituido por SIN, G2 formado por EGB y FPI, G3 formado por FPII y G4 incluyendo a BUP y UNI.

Simulamos de este último modelo de un modo análogo a como lo hicimos con el modelo completo, obteniendo también unos excelentes resultados en el diagnóstico de convergencia. La Tabla 2 proporciona los intervalos de confianza al 95% para los parámetros del modelo final. El coeficiente asociado a G4 no aparece en la tabla pues es el grupo tomado como referencia. Ninguno de los intervalos incluye al cero, por lo que las covariables EDAD, EC y CURSA no deben eliminarse del modelo. Sin embargo, la Tabla 2 no muestra si es posible seguir reagrupando los niveles de estudio; de ella sólo se desprende que el grupo G4 es distinto de los otros tres, pero no si estos últimos son diferentes entre sí.

parámetro	2.5%	50%	97.5%
EDAD	0.0671	0.0883	0.1090
G1	0.3810	0.8050	1.2000
G2	0.3870	0.6140	0.8430
G3	0.0286	0.3600	0.6850
EC	1.0100	1.5000	2.0100
CURSA	0.0078	0.2240	0.4420

Tabla 2: Intervalos intercuantílicos para los parámetros del modelo definitivo.

La Figura 3 presenta el comportamiento de los contrastes $G1-G2$ y $G2-G3$. En ella se observa que podría aceptarse alguna de las dos igualdades, $G1=G2$ o $G2=G3$, pero difícilmente ambas, pues el punto $(0, 0)$ está prácticamente en la frontera de la región bidimensional de confianza al 95%. En esta situación, preferimos no reducir más el modelo pues el número de covariables presentes ya es pequeño para el volumen de datos disponible. De todas formas, como era de esperar, la Figura 3 muestra que el comportamiento del grupo $G2$ no es muy distinto del $G1$ o el $G3$, encontrándose en una posición intermedia.

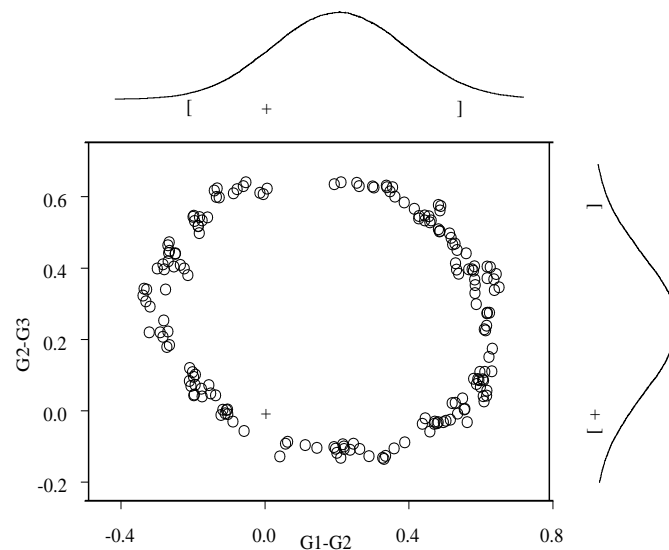


Figura 3: Regiones de confianza al 95% para los contrastes $G1-G2$ y $G2-G3$. En el centro se representa la región bidimensional y los intervalos univariantes se muestran en los laterales de la gráfica. El valor cero se muestra con un signo +.

A partir de la muestra simulada del modelo final pueden estudiarse, además de las distribuciones finales de los parámetros, distribuciones predictivas. Por ejemplo, en la Figura 4 se representan las funciones de distribución predictivas del tiempo de desempleo para personas solteras, de 25 años de edad y que han estudiado recientemente, pero con niveles de estudios distintos. Así, la probabilidad de que una persona con esas características y nivel de estudios BUP encuentre trabajo antes de dos años es 0.38, mientras que si su nivel de estudios es FPII esa probabilidad desciende hasta 0.20.

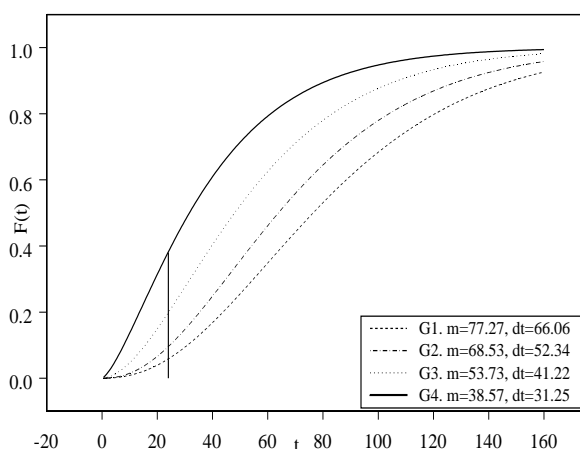


Figura 4: Funciones de distribución predictivas para personas de 25 años, solteras y que han estudiado recientemente. La esperanza y desviación típica del tiempo de desempleo aparecen en la leyenda.

Es de destacar que la Figura 4 muestra una ordenación de las funciones de distribución marcada por la complejidad del nivel de estudios, ordenación que no había sido incluida como restricción en el modelo. Algo similar ocurre con la edad, que influye en la supervivencia de modo que los más jóvenes encuentran con mayor facilidad el primer empleo. En la Figura 5 se representa la función de distribución predictiva para individuos de diferentes edades, solteros, con un nivel de estudios de FPII y que no han seguido ningún tipo de estudios durante las cuatro semanas anteriores a la entrevista.

Por su parte, la covariable CURSA influye positivamente en la supervivencia, esto es, a mayores valores de la misma, se tienen supervivencias más grandes. Por lo tanto, un individuo que no ha seguido ningún tipo de estudios durante las cuatro semanas anteriores a la entrevista presenta un menor tiempo

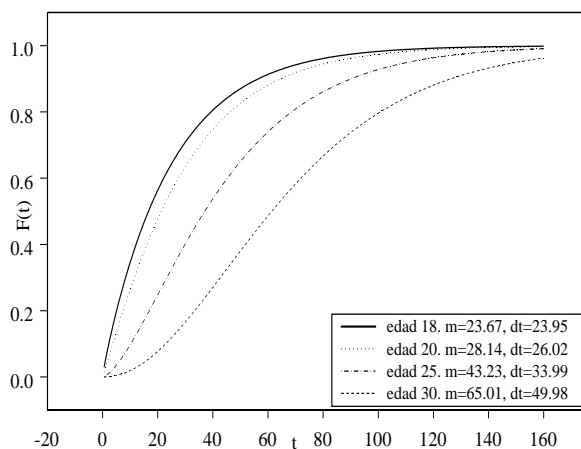


Figura 5: Funciones de distribución predictivas para personas solteras, con FPII y que no han estudiado recientemente. La esperanza y desviación típica del tiempo de desempleo aparecen en la leyenda.

de desempleo que otro con las mismas características pero que sí ha estudiado recientemente. Para la covariable estado civil, EC, se tiene una interpretación análoga, con menores tiempos de desempleo para los individuos casados.

Finalmente, cabe resaltar que de los resultados obtenidos se infiere una coincidencia genérica con gran parte de las conclusiones alcanzadas por Albaramírez (1999). Parece lógico que en la actualidad la covariable sexo no influya significativamente en el tiempo de desempleo, así como que la edad lo haga positivamente y que, por lo tanto, un individuo joven tenga mayor facilidad para encontrar su primer empleo que otro de mayor edad. También resulta coherente que un individuo casado encuentre su primer trabajo antes que uno soltero y que uno que está estudiando demore su entrada en el mercado laboral.

Mención aparte merece la agrupación obtenida para la covariable *nivel de estudios*. Es razonable una consideración diferenciada para los individuos sin estudios y que éstos, a su vez, presenten las mayores supervivencias, como también es lógico agrupar los niveles de EGB y FPI, tratándolos como uno solo. Tampoco puede resultar extraño que los individuos con un nivel de estudios equivalente a BUP tengan tiempos de desempleo prácticamente iguales que aquéllos con un nivel universitario. Si bien el grado de preparación para el mercado laboral suele diferir bastante entre ambos, un individuo que termina

sus estudios de BUP y no quiere ampliarlos con enseñanzas universitarias tiene ya una cierta necesidad económica y una edad muy apropiada para entrar en el mercado laboral. Quizá lo más sorprendente, relativamente, es la consideración del nivel de estudios equivalente a FPPII como un grupo aparte y que se asocie a un mayor tiempo de desempleo que BUP. Este resultado indica que los intentos institucionales de potenciar la Formación Profesional en este país aún no han sido todo lo exitosos que cabía esperar puesto que, contrariamente a lo pretendido, el hecho de seguir unos estudios específicos para la actividad profesional no supone una entrada más rápida en el mercado laboral.

4 Conclusiones

En este trabajo se presenta un modelo paramétrico para el estudio de fenómenos de supervivencia con datos censurados. Es bien conocida la complicación inherente a este tipo de estudios que se ve incrementada considerablemente si se quiere incorporar información inicial acerca de la población y se pretende utilizar para su análisis un modelo paramétrico. En lo referente a los aspectos metodológicos del modelo propuesto, cabe resaltar como relevante la posibilidad de realizar de un modo bastante razonable un análisis bayesiano de este tipo de modelos de duración. En este sentido, los métodos MCMC han devenido en los últimos tiempos en una herramienta imprescindible para el estudio bayesiano de complicados modelos de supervivencia.

Adicionalmente, la consideración de un modelo jerárquico posibilita la captación de parte de la heterogeneidad propia de la población y que no puede ser explicada por las covariables. En contrapartida, el manejo de un considerable número de parámetros aumenta la complicación del modelo bajo estudio que, en cualquier caso, presenta un análisis abordable utilizando técnicas de simulación.

Del análisis empírico de los tiempos de desempleo de los parados valencianos sin experiencia laboral se concluye como principales resultados la no influencia de la covariable sexo en la duración del desempleo, así como menores tiempos de desempleo para los individuos casados, para aquellos que no han seguido ningún tipo de estudios durante las cuatro semanas anteriores a la entrevista y para los más jóvenes. En cuanto a la covariable nivel de estudios, se ha

obtenido una ordenación genérica de la función de distribución predictiva con respecto a la complejidad de dicho nivel. De este modo, se predicen tiempos de desempleo más pequeños para el grupo de universitarios y bachilleres y duraciones mayores para individuos sin estudios. Especialmente interesante es la obtención de mayores tiempos de desempleo para individuos con un nivel de estudios equivalente a FPII que para aquellos que han estudiado BUP.

Referencias

- Alba-Ramírez, A.** (1999). Explaining the transitions out of unemployment in Spain: the effect of unemployment insurance. *Applied Economics*, **31**, 183-193.
- Beamonte, E.** (1998). *Aportaciones al análisis bayesiano semiparamétrico de datos de supervivencia*. Tesis Doctoral. Universitat de València Estudi General.
- Beamonte, E. y Bermúdez, J. D.** (1999). Bayesian semiparametric analysis for additive hazard models with censored observations. Bajo revisión en *Journal of Econometrics*.
- Beenstock, M.** (1996). Training and the time to find a job in Israel. *Applied Economics*, **28**, 935-946.
- Best, N. G., Cowles, M. K. y Vines, S. K.** (1995). *CODA manual version 0.30*. Cambridge: MRC Biostatistics Unit.
- Bover, O. y Gómez, R.** (1999). Nuevos resultados sobre la duración del desempleo: el paro de larga duración y la salida a un empleo fijo. *Documento de trabajo nº 9903*. Servicio de Estudios del Banco de España.
- Cox, D. R.** (1972). Regression models and life tables (with discussion). *Journal of the Royal Statistical Society B*, **34**, 187-220.
- Chib, S. y Greenberg, E.** (1996). Markov chain Monte Carlo simulation methods in Econometrics. *Econometric Theory*, **12**, 409-431.

- Follmann, D. A., Goldberg, M. S. y May, L.** (1990). Personal characteristics, unemployment insurance, and the duration of unemployment. *Journal of Econometrics*, **45**, 351-366.
- Gelfand, A. E. y Smith, A. F. M.** (1990). Sampling-based approaches to calculating marginal densities. *Journal of the American Statistical Association*, **85**, 398-409.
- Geman, S. y Geman, D.** (1984). Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **6**, 721-741.
- Geweke, J.** (1999). Using simulation methods for Bayesian econometric models: inference, development, and communication (with comments). *Econometric Reviews*, **18**, 1-126.
- Gonzalo, M. T.** (1998). Job exit of men: new job destination versus unemployment destination. *Applied Economics*, **30**, 1231-1238.
- Hastings, W. K.** (1970). Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*, **57**, 97-109.
- Muller, P.** (1991). Metropolis based posterior integration schemes. *Technical report*, Department of Statistics. University of Purdue.
- Wei, G. C. G. y Tanner, M. A.** (1990). Calculating the content and boundary of the highest posterior density region via data augmentation. *Biometrika*, **77**, 649-652.