

# ESTUDIO DEL TIEMPO DE PERMANENCIA EN EL DESEMPLEO DE LA MUJER EN ARAGÓN

Laura MUÑOZ GARATACHEA

Raquel YUSTE ARA

*Departamento de Métodos Estadísticos. Facultad de C.C.E.E. y E.E.  
Universidad de Zaragoza*

**SUMARIO:** I. Introducción. II. Modelo de supervivencia univariante. III. El paradigma bayesiano y el gibbs sampling. IV. Convergencia de los parámetros. Densidades y momentos a posteriori. V. Conclusiones.

**RESUMEN:** Aplicamos una metodología Bayesiana para la estimación del modelo de supervivencia univariante de Cox, con un proceso correlado para la función de *azar baselina* a datos de desempleo en la Comunidad Autónoma de Aragón. Se proponen distribuciones a priori difusas para los parámetros del modelo: los coeficientes asociados a las<sup>1</sup> covariables, para los cuales se propone una mixtura de normales, en una forma similar al contexto de Macculloh (1993) [2] para la selección de variables en el modelo de regresión lineal, y la función *baselina*. Hemos implementando la estrategia mediante ejecuciones de 10 cadenas del algoritmo del Gibbs Sampling programadas en S-PLUS versión 3.3. Una de las conclusiones más

---

1. Este trabajo ha sido financiado por el proyecto 228-26 del III Plan Nacional de I+D.

relevantes es la no discriminación entre niveles profesionales y educativos en los tiempos de desempleo.

**Palabras clave:** función de azar, análisis Bayesiano, gibbs sampling.

## I. INTRODUCCIÓN

El descenso de natalidad en las sociedades industriales esta afectando también a España. Detrás de este fenómeno existen, sin duda, complejos factores de muy diversa naturaleza, pero sobre todo destacaremos la mayor participación laboral de la mujer.

Probablemente interrelacionado con este fenómeno, se viene desarrollando un fenómeno de mayor cualificación profesional de la mujer española. Las decisiones tomadas por las mujeres sobre su participación laboral (o permanencia en el desempleo) vienen afectadas e influyen a su vez sobre las decisiones tomadas por las familias en cuanto al número de hijos. Puede vislumbrarse que, para mujeres de cualquier categoría profesional con hijos menores, la salida de la situación de desempleo debe compensar la pérdida de lo que se conoce como *tiempo dedicado a la producción de bienes domésticos*. Pero además para mujeres cualificadas este tiempo de permanencia en el desempleo puede ser mayor debido a lo que se conoce como el *salario sombra*, en definitiva la salida del desempleo se produce cuando además hay una remuneración que responde a sus expectativas basadas en su alta cualificación.

La importancia de un estudio serio acerca del descenso actual de la natalidad requiere un análisis profundo de los *factores significativamente relevantes en el tiempo de permanencia en el desempleo de la mujer española*.

El objetivo fundamental de este trabajo es la aplicación de una metodología fundamentada en el *enfoque bayesiano*, que nos va permitir comparar posibles comportamientos diferenciados en cuanto a participación laboral entre hombres y mujeres, así como detectar factores que son influyentes en la *permanencia en el desempleo* -nivel educativo, nivel profesional, experiencia laboral, idiomas, actividad económica desarrollada (según CNAE)-. El conocimiento de este marco socio-laboral será el punto de partida para la realización de un análisis profundo de las diversas interacciones causa-efecto entre el descenso de la natalidad y las decisiones de participación laboral adoptadas por las mujeres.

En la metodología propuesta y en conexión con otros trabajos econométricos realizados, planteamos como modelo básico el modelo de la función de azar proporcional de Cox (1972) [1] e introducimos métodos bayesianos de estimación. Dichos métodos han alcanzado especial relevancia recientemente debido a los avances en las técnicas de computación bayesiana, que permiten abordar de una forma relativamente cómoda los modelos semiparamétricos.

El trabajo se desarrolla como sigue: en la *Sección 2* se plantea el modelo univariante considerado; en la *Sección 3* se introduce brevemente el paradigma bayesiano

y la metodología del Gibbs Sampling. Las distribuciones. En la *Sección 4* presentamos diagnósticos de convergencia y estimaciones a posteriori de magnitudes de interés sobre los parámetros del modelo, finalizando en la *Sección 5* con algunas conclusiones extraídas del estudio empírico y futuras direcciones más ambiciosas de conocimiento del fenómeno.

## II. MODELO DE SUPERVIVENCIA UNIVARIANTE

El estudio que vamos a realizar se enmarca dentro del tipo de modelos de supervivencia univariantes, por lo que nos centraremos en estos modelos teóricos, orientando el estudio a la implementación del paradigma bayesiano asociado.

El modelo del que partimos es el modelo de *azar proporcional* de Cox para el cual se sabe que la función de azar puede especificarse como sigue:

$$h(t) = h_o(t) \exp[b'x]$$

siendo  $x' = (x_1, \dots, x_p)$  el vector de covariables explicativas,  $b' = (b_1, \dots, b_p)$ , el vector de parámetros, que suponemos constantes en el tiempo, y  $h_o(t)$  es la *función de azar baselina* que se supone desconocida. De este modo, la función baselina constituye la parte no paramétrica del modelo.

Suponemos que los datos observados pueden ser exactos o censurados por la derecha (31/5/1997) y nuestro objetivo primordial consistirá en estimar la distribución a posteriori del vector de parámetros del modelo. Desde una perspectiva bayesiana, necesitamos especificar un proceso aleatorio a priori sobre el vector de parámetros así como una distribución a priori sobre el proceso azar baselina.

Con respecto a la función de *baselina*, en la literatura se han propuesto los siguientes:

- *Proceso de Dirichlet*, que fue inicialmente propuesto por Ferguson (1973) [3] y que ha sido ampliamente utilizado en recientes estudios. Dicho proceso hace referencia a la función de distribución de los datos y tiene una natural aplicación en el análisis bayesiano no paramétrico, pero presenta el inconveniente de carecer de una interpretación simple en términos de la función de *azar*.
- *Proceso Gamma* propuesto por Kalbsfleisch (1978) [4], supone que la función de azar acumulada es la realización de un proceso de Levy con incrementos independientes positivos. Si bien es el más utilizado, presenta fundamentalmente dos inconvenientes: en primer lugar se exige que los incrementos sean independientes en intervalos adyacentes, y en segundo lugar, se necesita que el proceso observado sea discreto con probabilidad 1, lo cual implica una probabilidad positiva de empates que no tiene sentido en una variable continua como es el tiempo de permanencia en el desempleo. Una solución para este problema ha sido propuesta por Dykstra y Laud [5] que plantean una generalización del proceso gamma que tiene las propiedades de incrementos indepen-

dientes y función de azar creciente, de forma que la distribución obtenida es siempre absolutamente continua.

– *Proceso Correlado*, introducido por Leonard (1978) [6] para la estimación de densidades. Supone que la función de azar baselina es constante a trozos con:

$$h_0(t) = 1_{I_j}, \text{ si } t \in I_j = [a_{j-1}, a_j), j = 1, \dots, g+1$$

siendo  $a_0 < a_1 < \dots < a_{g+1} < \infty$  fijos y de forma que:  $a_j = a_{j-1} + e_j, j = 1, \dots, g+1$  con  $a_0 = 0$  y  $e_j \sim N(0, s^2)$  y  $a_j = \log I_{j^*}$ .

En nuestro trabajo hemos adoptado a priori un proceso correlado debido a su flexibilidad y a la escasez de restricciones impuestas sobre las funciones baselina.

### III. EL PARADIGMA BAYESIANO Y EL GIBBS SAMPLING

El marco bayesiano de nuestro trabajo establece como elementos iniciales de análisis las distribuciones a priori sobre los parámetros del modelo:  $\Pi(b')$  y  $\Pi(h_0(t))$ , que combinadas con la verosimilitud del mismo:  $L_N(b', h_0)$  nos permiten obtener la distribución de probabilidad a posteriori de los parámetros. Observar que la no información inicial del analista acerca de los parámetros se canaliza a través de las distribuciones a priori llamadas no informativas.

Por su parte, la distribución de probabilidad a posteriori de los parámetros debe entenderse como la variación de las creencias indiciales del analista acerca de los parámetros, una vez los datos han sido observados. Así, la densidad a posteriori para los parámetros, se obtiene integrando la distribución conjunta  $p(t', x', \vartheta) = L(\vartheta|x', t')\Pi(\vartheta)$  con  $\vartheta = (b', h_0)$  sobre  $\Theta$  el espacio paramétrico asociado al modelo semi-paramétrico.

Salvo en el caso conjugado, en el cual las distribuciones a priori y la verosimilitud están en familias conjugadas y el cálculo es analíticamente abordable, la resolución requiere un análisis diferenciado para cada problema que encierra la evaluación de integrales múltiples, cuya dimensión aumenta con la del espacio paramétrico.

Así, se ha hecho necesaria la utilización de métodos aproximados, basados en aproximaciones numéricas o bien en aproximaciones asintóticas de las funciones a integrar. Recientemente han surgido en la literatura los llamados métodos MCMC (Markov Chain Monte Carlo Methods), que abordan el problema desde una fundamentación basada en cadenas de Markov ergódicas que tienen a la distribución a posteriori como su distribución estacionaria, y que proporcionan una muestra a partir de la posteriori cuando las cadenas han convergido. Entre estos métodos destaca el Gibbs Sampling propuesto por Ritter y Tanner (1996) [7] que vamos a utilizar en el trabajo y que describimos brevemente:

Sea  $\pi(x) = \pi(x_1, \dots, x_k), x \in \mathcal{X}^k$ , con  $1 < k < n$  la densidad estacionaria, donde para  $i = 1, \dots, k$   $x_i = (x_{i1}, \dots, x_{in(i)})$ ,  $n(i) \geq 1$ ,  $n(1) + \dots + n(k) = n$  y los  $x_{ij}$  son las componen-

tes escalares de  $x$  y sean denota la densidad condicional inducida para cada componente, dados los valores de las restantes componentes  $x_{-i} = (x_j; j \neq i), i = 1, \dots, k$ .

La forma sistemática del algoritmo del Gibbs Sampling procede como sigue: tomando valores iniciales arbitrarios  $x^0 = (x_1^0, \dots, x_p^0)$ , sucesivamente obtener muestras aleatorias a partir de  $\pi(x_i|x_{-i})$ :

$$\begin{aligned} x_1^l &\sim \pi(x_1|x_{-1}^0); \\ x_2^l &\sim \pi(x_2|x_1^l, x_3^0, \dots, x_k^0) \\ x_3^l &\sim \pi(x_3|x_1^l, x_2^l, \dots, x_k^0) \\ &\vdots \\ x_k^l &\sim \pi(x_k|x_{-k}^l) \end{aligned}$$

Esta forma de actuar define un mecanismo de transición desde  $x^0$  a  $x^l = (x_1^l, \dots, x_k^l)$ . La iteración de este proceso genera una secuencia  $x^0, \dots, x_p, \dots$  que es una realización de una cadena de Markov con probabilidades de transición dadas por:

$$\prod_{i=1}^l \pi(y_i|x_p, j > l, y_p, j < l)$$

Bajo condiciones que garanticen que la cadena así construida es  $\pi$ -irreducible y aperiódica, se garantiza la convergencia del método.

#### IV. CONVERGENCIA DE LOS PARÁMETROS. DENSIDADES Y MOMENTOS A POSTERIORI

Aplicamos la metodología propuesta sobre datos reales que corresponden al tiempo de desempleo (dado en meses) en la provincia de Zaragoza. Los datos muestrales consisten en 64 registros aleatorios del conjunto de registros de demandantes de empleo inscritos en las Oficinas de Empleo de la Provincia de Zaragoza a 31 de Mayo de 1997. La base de datos contiene la información que utilizaremos como covariables en el estudio y que se detallan en el *Apéndice A.1*.

En cuanto a la implementación del paradigma bayesiano asociado al modelo de Cox, utilizamos el algoritmo del *Gibbs Sampling* para obtener las distribuciones a posteriori, realizando en alguna de las iteraciones algunos pasos del algoritmo de Hastings-Metropolis.

##### Distribuciones a priori:

Para el vector de parámetros  $b'$ , en la literatura se suele adoptar una distribución a priori difusa de media 0 como puede ser una normal multivariante, ó una t de Student con pocos grados de libertad. En nuestro trabajo sin embargo, hemos adoptado un tratamiento similar al de George y McCulloch (1993) en el problema de selección de variables en un modelo de regresión lineal con el fin de estimar además la probabilidad a posteriori de la significatividad de los parámetros que intervienen en la función de azar.

Así, para el vector de parámetros asociados a las covariables:  $b'$  supondremos que son independientes con la distribución a priori no informativa:

$$b_i, g_i \sim (1 - g_i)N(0, \sigma_b^2) + g_i N(0, c^2 \sigma_b^2)$$

donde  $g_i = 1$  si la variable correspondiente  $X_i$  entra en la ecuación del modelo y 0 en caso contrario. Supondremos, además que  $P[g_i = 0] = 1/2, i = 1, \dots, p$ .

Para la varianza, establecemos una distribución a priori gamma invertida:

$$t = \frac{1}{\sigma^2} \sim Ga\left(\frac{\nu}{2}, \frac{\tau}{2}\right)$$

con  $\nu, \tau \in \mathfrak{R}^+$  constantes.

Los hiperparámetros de las distribuciones a priori, se han tomado como sigue: en cuanto a la distribución a priori sobre el proceso *baselina* hemos considerado intervalos semestrales con baselina constante a trozos, obteniendo así un vector  $\alpha$  (cuyas componentes son el logaritmo de la función baselina en cada intervalo) de dimensión 23.

Para los parámetros de escala de la mixtura de normales a priori sobre  $b'$  hemos considerado para  $\sigma_b^2$  el máximo de las varianzas de los estimadores m.l.e. del modelo de Cox con  $c$  obtenida para representar la indiferencia a priori acerca de las variables significativas.

Los parámetros de forma y de escala de la gamma invertida correspondiente a la distribución de la varianza se han tomado con valor 1.

### Verosimilitud:

Sean  $t' = (t_1, \dots, t_N)$ ,  $a' = (a_1, \dots, a_{g+1})$ ,  $d' = (d_1, \dots, d_N)$  y  $X' = (x_1, \dots, x_N)$ . Sea  $f(t|a, b)$  la función de densidad del tiempo de permanencia en el desempleo y sea  $S(t|a, b)$  la función de supervivencia de T, así la verosimilitud del modelo viene dada por la siguiente expresión, que combinada con las distribuciones a priori nos proporciona la obtención de forma laboriosa pero usual de las distribuciones completamente condicionales para el Gibbs.

$$L(a, b | X, t, d) = \prod_{i=1}^N \left( h_0(t_i | a, b, x_i)^{\delta} \exp \left[ - \int_0^{t_i} h(t | \alpha, \beta, x_i) dt \right] \right)$$

En la implementación del Gibbs Sampling se han ejecutado 10 Cadenas de Markov en paralelo, tomando como valores iniciales en cinco de ellas  $\gamma = 0$  (no significatividad de las covariables) y en las cinco restantes  $\gamma = 1$  (todas las covariables son significativas). En cada uno de los dos grupos de cinco cadenas inicializamos el correspondiente vector de parámetros  $b$  de cada una de ellas con:  $b_k - 2s_k, b_k - s_k, b_k, b_k + s_k, b_k + 2s_k, k = 1, \dots, p$ , donde  $s_k$  son los errores estándar de la estimación de  $b_k$ .

A partir de las distintas inicializaciones se han realizado 100 iteraciones del Gibbs Sampling con el objetivo de analizar la convergencia de los parámetros  $b'$  mediante dos métodos: gráficamente y utilizando las autocorrelaciones de las series (cadenas de Markov) en paralelo para dichos parámetros. Alguno de los resultados obtenidos se muestran en el *Gráfico 1*, se observa que, para cada parámetro las diez cadenas convergen a su distribución estacionaria y por tanto podremos obtener muestras de la distribución a posteriori de los parámetros.

Calculamos las funciones de autocorrelación muestral para cada una de las diez cadenas, promediando los resultados y obteniendo la función de autocorrelación muestral media. A partir de ella determinamos cada cuantos pasos  $k$  se toman las muestras en las cadenas. Elegimos el tamaño muestral  $N$  deseado y ejecutamos de nuevo el algoritmo de Gibbs Sampling durante  $(N/10)k$  iteraciones, tomando muestras cada  $k$  pasos en cada una de las cadenas. En nuestro estudio, a la vista de los gráficos de autocorrelación hemos tomado  $k=5$ , y elegimos  $N=500$ .

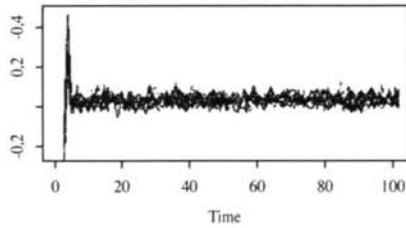
En la *Tabla 1* se muestran las medias y desviaciones típicas de los parámetros correspondientes a las covariables que nuestro algoritmo señala como significativas para el modelo de Cox propuesto. En el *Gráfico 2* representamos las densidades a posteriori y gráficos box-plot para los parámetros más significativos.

*Tabla 1*  
Media y Desviación típica a posteriori de los parámetros  $b$  significativos

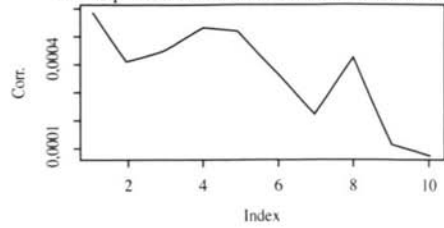
Variable	COVARIABLES SIGNIFICATIVAS		
	Media	Desviación Típica	Prob( $\gamma=1$ )*
SEXODEMA	0,0476	0,027	0,836
NIVEACAD			
NIV1	0,1764	0,086	0,842
NIV2	0,2325	0,083	0,824
NIV3	0,2089	0,082	0,810
NIV4	0,1927	0,084	0,835
NIVPRO			
NIVPRO0	0,1601	0,056	0,828
NIVPRO2	0,1467	0,058	0,822
NIVPRO3	0,1694	0,057	0,806
EXPRO			
EXPRO0	0,1948	0,062	0,818
EXPRO3	0,1720	0,076	0,810
PASLAB			
PASLAB0	0,1694	0,100	0,808
PASLAB3	0,1627	0,078	0,810
ACTIECON			
ACTI1	0,1470	0,054	0,814
ACTI2	0,1238	0,066	0,818
ACTI3	0,1366	0,058	0,806
ACTI4	0,112	0,059	0,816

\* Indica la significatividad del parámetro correspondiente.

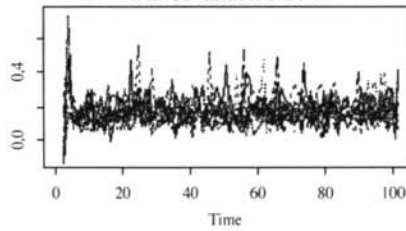
Convergencia del parámetro beta correspondiente a la Covariable SEXODEMA



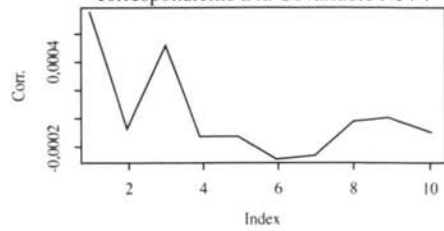
Autocorrelaciones: parámetro beta correspondiente a la Covariable SEXODEMA



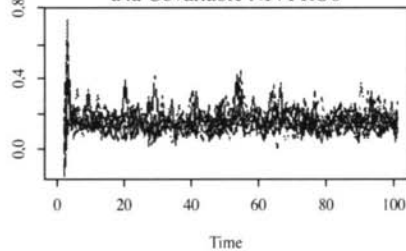
Convergencia del parámetro beta correspondiente a la Covariable NIV4



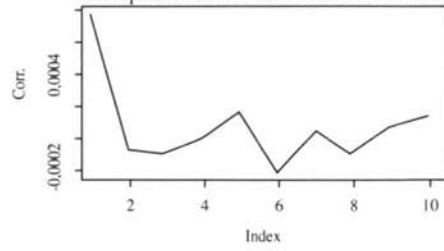
Autocorrelaciones: parámetro beta correspondiente a la Covariable NIV4



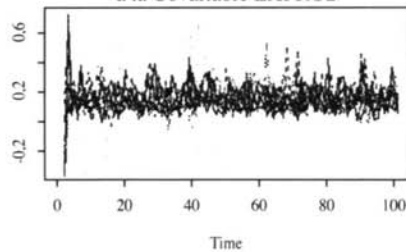
Convergencia del parámetro beta correspondiente a la Covariable NIVPRO0



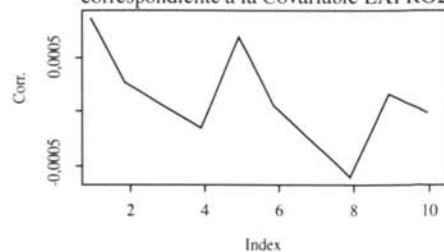
Autocorrelaciones: parámetro beta correspondiente a la Covariable NIVPRO0



Convergencia del parámetro beta correspondiente a la Covariable EXPRO2

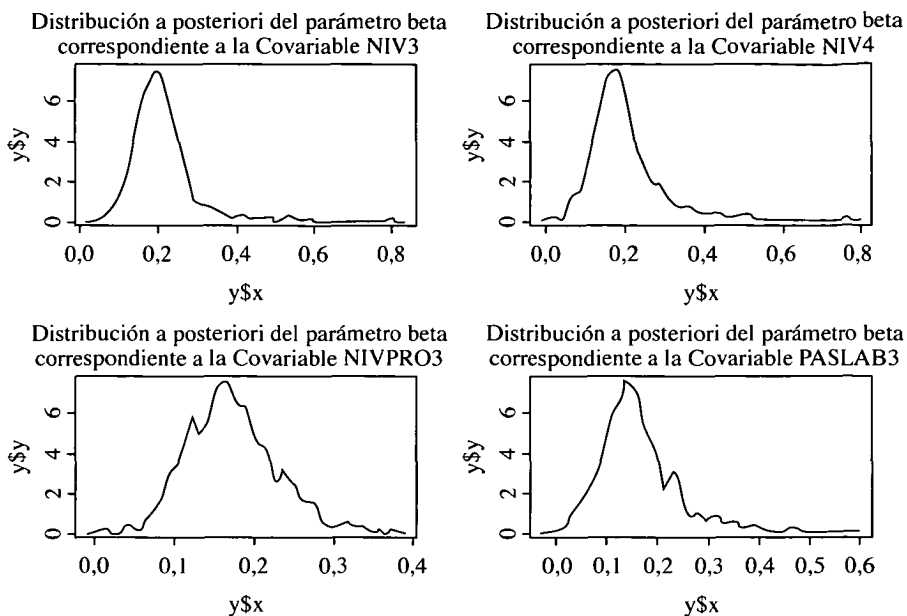


Autocorrelaciones: parámetro beta correspondiente a la Covariable EXPRO2



**Gráfico 1. Gráficos de Convergencia.**





*Gráfico 2. Densidades a posteriori.*

## V. CONCLUSIONES

En este trabajo se han alcanzado dos objetivos: en primer lugar, mostrar la relevancia de la implementación de una metodología bayesiana para estimar modelos de supervivencia y, en segundo lugar, conocer el marco socio-laboral que afecta a las mujeres en situación de desempleo en la Comunidad Autónoma de Aragón.

Hay que tener en cuenta que nuestra variable Tiempo de Permanencia en el Desempleo es el tiempo transcurrido desde la fecha de inscripción en el INEM hasta el 31 de Mayo de 1997, que es una variable proxy del tiempo de permanencia del individuo en la base, ya que no se tiene constancia de entradas y salidas de individuos de la base de datos. Es decir, estamos suponiendo que nuestra variable temporal es censurada (31/5/97) cuando en realidad no lo es. Así pues, para posteriores investigaciones, estamos interesados en conseguir una nueva base en la que conozcamos todo tipo de censura (entradas y salidas de cada individuo de la muestra, pues el comportamiento habitual será que un individuo salga de la muestra un determinado tiempo y vuelva a entrar en la base).

En cuanto a los resultados del análisis empírico, la variable sexo aparece como significativa aunque su efecto es muy débil, como puede apreciarse al observar el valor del parámetro en la *Tabla 1*. Las diferencias en cuanto a las decisiones de participación laboral entre hombres y mujeres no vendrán dadas por la variable sexo en sí misma si no se tienen en cuenta otros factores, como nivel educativo, nivel y experiencia profesional, etc.

Como sospechábamos antes de realizar este estudio, un nivel académico alto no tiene el efecto que cabría esperar en el tiempo de permanencia en el desempleo. Los coeficientes asociados a las distintas categorías de la variable nivel educativo no se diferencian significativamente. Este hecho pone de manifiesto que las expectativas de empleo son quizás más ambiciosas para mujeres con estudios superiores, entendiendo por expectativa el llamado *salario sombra* (dichas mujeres entienden por empleo aquél que de alguna forma remunerare de manera satisfactoria su cualificación), con lo cual permanecen más tiempo en el desempleo que el que se esperaría dada su formación

Respecto a la experiencia profesional resultan significativas las categorías: EXPRO0 (sin experiencia) y EXPRO3 (de 6 a 12 meses) como puede observarse en la *Tabla 1*. El aspecto importante a destacar es que mujeres con una experiencia profesional y que, por tanto, están cualificadas para desempeñar el trabajo que solicitan, rechazan las ofertas de empleo que no satisfacen sus expectativas: su salario sombra-el que perderían por acceder a un puesto de trabajo que no se ajuste a su capacidad- hace que se vean obligadas a permanecer desempleadas.

Finalmente en conexión con el estudio del fenómeno del descenso de la natalidad las conclusiones extraídas acerca del tiempo de desempleo para mujeres cualificadas (cada vez más habituales en el mercado laboral) podría considerarse la influencia en la permanencia en el desempleo del status marital y el numero de hijos. Las deficiencias de nuestra actual base de datos (no disponibilidad de estas covariables) no nos ha permitido evaluar estos factores. No obstante, si un análisis posterior, revela que estos efectos determinan de algún modo las decisiones acerca del número de hijos deseados, podría ser interesante proponer a las empresas una flexibilización de la jornada laboral que paliara la aparente discrepancia entre la decisión de tener hijos y la mayor cualificación de la mujer.

## BIBLIOGRAFÍA

- COX, D.R. (1972): "*Regression Models and Life Tables*". Journal of the
- GEORGE, E.I., y McCULLOCH, R. E. (1993): "*Variable Selection via Gibbs Sampling*". Journal of the American Statistical Association, 88, pp. 881-889.
- FERGUSON, T.S. (1997): "*A Bayesian Analysis of some Non-Parametric Problems*". He Annals of Statistics, 1, pp. 209-230.

KALBFLEISCH (1978): "*Nonparametric Bayesian Analysis of Survival Time Data*". Journal of the American Statistical Association.

DYKSTRA, R.L. y LAUD, P. (1981): "*A Bayesian Nonparametric Approach to Reliability*". The Annals of Statistics, 9, pp. 356-367

LEONARD (1978): "*Density Estimation of Stochastic Processes and Prior Information*". Journal of the Royal Statistical Society, sec. B; 57; pp. 247-262.

TANNER, M.A. (1996): "*Tools for Statistical Inference: Methods for the Exploration of Posterior Distributions and Likelihood Functions*". New York: Springer-Verlag.

SINHA, D. y DEY, D. (1997): "*Semiparametric Bayesian Analysis of Survival Data*". Journal of the American Statistics, 92.

## APÉNDICE A.1. BASE DE DATOS. DESCRIPCIÓN DE LOS CAMPOS

Las covariables en el estudio responden a los campos que se disponen y que detallamos a continuación:

SEXODEMA: Sexo del demandante.

1 → hombre

0 → mujer

CODPOSTA: Procedencia del demandante

1 → Zaragoza

0 → Fuera de Zaragoza

EDADDEMA: Edad del demandante a 31 de Mayo de 1997

NIVEACAD: Nivel académico del demandante. Con cuatro categorías.

NIV1 → Sin estudios

NIV2 → Estudios Elementales (EGB, FP1)

NIV3 → Estudios Intermedios (BUP-COU, FP2)

NIV4 → Estudios Superiores (Universidad)

NIVPRO: Nivel profesional en la ocupación solicitada

NIVPRO0 → Técnicos y varios.

NIVPRO1 → Directores y Jefes de Equipos

NIVPRO2 → Oficial 1ª, 2ª y 3ª

NIVPRO3 → Ayudantes, auxiliares y especialistas

NIVPRO4 → Peones y aprendices

EXPRO: Experiencia profesional en la ocupación solicitada.

EXPRO0 → Sin experiencia

EXPRO1 → De 0 a 3 meses

EXPRO2 → De 3 a 6 meses

EXPRO3 → De 6 a 12 meses

EXPRO4 → Más de 12 meses

**SITESP:** Situaciones especiales

1 → Estudiante < 25 años ó > 25 años demandante de primer empleo

0 → Otras

**PASLAB:** Pasado laboral del demandante de empleo

PASLAB0 → Sin empleo anterior

PASLAB1 → Cese voluntario

PASLAB2 → Fin de contrato

PASLAB3 → Despido

PASLAB4 → Otros

**IDIOMA:** Conocimiento de idioma extranjero del demandante de empleo

1 → Habla al menos un idioma

0 → No habla ningún idioma

**ACTIECON:** Actividad económica de la que procede según CB-82 (Código Nacional de Actividades Económicas)

ACTI1 → Industrias amnufactureras y extractivas

ACTI2 → Distribución de Energía

ACTI3 → Construcción

ACTI4 → Comercio, Transporte

ACTI5 → Actividades con cobertura muy reducida

ACTI0 → Otras

**MES:** Tiempo de permanencia en el desempleo expresado en meses.

**DELTA:** Variable que indica la censura de los tiempos de desempleo. Toma valor 1 si el dato sobre el individuo es exacto y 0 si es censurado.