

Desempleo masculino: probabilidades y diferencias regionales*

A. J. Avila Cano
M. C. Cuerel Baeza
C. Pagés Serra
J. V. Rodríguez Mora

ABSTRACT

One of the most important development in microeconometrics it has taken place in the area of qualitative response models. In these statistical models the endogenous random variables take only discrete values.

This article shows binary probit model's estimations in which dependent variables are the long term and total unemployment. The «Encuesta de Condiciones de Vida y Trabajo en España» data base is used to estimate these models; the data are referred to men with previous labour experience and resident in five spanish regions.

Both models have an interesting goodness of fit and allow for forecasting the probability of being unemployed and long term unemployed within the group specified in the sample.

RESUMEN

Uno de los desarrollos más importantes en microeconomía corresponde al área de los modelos de respuesta cualitativa, en los que la variable endógena toma únicamente valores discretos.

* Este estudio se realizó básicamente en el ámbito del Magister en Análisis Económico de la Universidad Autónoma de Barcelona y el Instituto de Análisis Económico. Los autores agradecen el asesoramiento prestado por J.L. Raymond y los comentarios de A. Pajuelo.

Este artículo presenta estimaciones de un modelo probit binario en el que la variable dependiente ha sido el desempleo, en un caso, y el paro de larga duración, en otro. La base de datos ha sido obtenida a partir de la «Encuesta de Condiciones de vida y Trabajo en España», y se refiere a sujetos varones con experiencia laboral previa y residentes en cinco comunidades autónomas de España.

Ambos modelos presentan una buena capacidad de ajuste y permiten pronosticar las probabilidades de paro y paro de larga duración para el colectivo especificando en la muestra.

1. INTRODUCCION.

Uno de los aspectos del mercado de trabajo que mas interés ha despertado recientemente ha sido el relativo a la duración del desempleo. Cuestiones tales como el desarrollo tecnológico y la reconversión industrial han incidido de forma importante sobre la estructura temporal del paro y las posibilidades de reubicación de los desempleados con experiencia laboral. Todo ello ha derivado en la necesidad de poner en práctica medidas de política económica que afecten de manera especial a aquellos estratos de la población mas incapacitados para volver a encontrar empleo. En consecuencia, se hace imprescindible el estudio y caracterización de este tipo de desempleo, necesidad que ha derivado en la aparición de numerosos estudios que, de una u otra manera, hacen referencia al tema.

Por otra parte, entre los más importantes desarrollos en microeconometría de las ultimas dos décadas, cabe destacar los ocurridos en el área de los modelos de respuesta cualitativa, también llamados de elección discreta. En estos modelos estadísticos, la variable endógena solo toma valores discretos, mientras que las variables explicativas constituyen atributos o características de los sujetos de la muestra.

Dentro de las muchas aplicaciones de este tipo de modelo (Amemiya(1981)), destacan las relativas al mercado de trabajo, en tanto que la variable dependiente puede modelarse en forma discreta, tratando de representar situaciones personales tales como estar o no parado, estar o no en activo, ocupado, etc. El trabajo que presentamos en este artículo tiene como objetivo la estimación de un modelo

dicotómico de elección discreta para la determinación de la probabilidad de paro masculino de larga duración. Asimismo, hemos creído de interés reestimar el mismo modelo tomando como variable endógena el estar o no parado dado que, como se verá más adelante, las variables explicativas son características del sujeto en cuestión; ello permitirá detectar las posibles diferencias en cuanto al condicionamiento que tales atributos ejercen sobre una y otra situaciones. Entre tales atributos figura la Comunidad Autónoma a la que pertenece el sujeto, por lo que podrá inferirse la existencia o no de suficiente evidencia para afirmar comportamientos regionales diferenciados.

El contenido del estudio realizado se estructura como sigue; en la segunda sección se comenta la muestra utilizada en las estimaciones del modelo. En la tercera se describe brevemente el modelo estimado y se especifican las variables-atributos incluídas, así como su proceso de selección, con una especial referencia a aquellas que se han excluído en la especificación final. En el cuarto apartado se analizan algunas de las características mas relevantes del desempleo global y el desempleo de larga duración desde un punto de vista exclusivamente descriptivo relativo a la muestra utilizada. En quinto lugar a la presentación de las estimaciones y contrastes realizados y se comentan brevemente los principales resultados. Después, y para el grupo de referencia, se han pronosticado las probabilidades de paro y paro de larga duración. Asimismo, se han obtenido las diferentes probabilidades de paro variando entre regiones las características del sujeto de referencia. Por ultimo, se presentan unas consideraciones finales a modo de conclusión.

Resulta evidente la posibilidad de realizar interesantes ampliaciones en el presente estudio. Algunas de las que podrían sugerirse se refieren a la especificación de modelos de forma que pudiese contrastarse la distinta influencia de las variables explicativas en ambas variables endógenas, la especificación de modelizaciones más completas que un modelo probit simple, la introducción de nuevas variables caracterizadoras de los sujetos o, incluso, la definición de nuevas categorías de variables, en función de la información muestral disponible, referentes a cuestiones específicas relacionadas con la duración del desempleo, como puede ser la capacitación y formación laboral, la intensidad de búsqueda, etc. En todo caso, creemos habernos acercado al conocimiento comparativo de la variables en cuestión desde una perspectiva más interesante que la mera descripción de las características muestrales, y haber aportado un primer avance sobre el estudio de tal fenómeno laboral.

II. DESCRIPCION DE LA MUESTRA.

La muestra utilizada en las estimaciones se ha obtenido a partir de los datos de la «Encuesta sobre Condiciones de Vida y Trabajo en España» (ECVT), realizada por el Ministerio de Economía y Hacienda en el cuarto trimestre de 1985. Esta encuesta contiene información sobre las características individuales de más de 60.000 personas. A partir de ella se ha obtenido una submuestra que contiene únicamente los datos de varones, que han trabajado alguna vez, pertenecientes a cinco comunidades autónomas: Andalucía, Cataluña, País Vasco, Madrid y la Comunidad Valenciana. Nos hemos centrado en estas cinco comunidades en tanto que suponen más del cincuenta por ciento de la población total española y, además, permiten la realización de comparaciones interregionales dadas las significativas diferencias tanto de estructura productiva como demográfica existentes entre ellas.

La razón por la cual se ha excluido a las mujeres de la muestra seleccionada se basa en las especiales características del desempleo femenino. En efecto, dado que la pretensión era individuos que ya tuviesen alguna experiencia laboral (excluyendo, por tanto, a quienes nunca hubiesen trabajado), el colectivo femenino suponía importantes distorsiones por su mayor propensión de acceso al mercado de trabajo en función de las perspectivas de conseguir empleo. En particular, ello permite excluir el grupo de amas de casa que hubieran trabajado alguna vez y que, en el momento de realización de la encuesta, estaban desempleadas pero no buscaban activamente empleo.

Por otra parte, el hecho de habernos centrado en individuos con alguna experiencia laboral a lo largo de su vida permite una mayor concreción de los resultados en un segmento de la fuerza de trabajo específico que requiere, sin duda, medidas de política diferenciadas. A la vez, se excluyen distorsiones derivadas de la posible consideración del paro juvenil «sin primer empleo», aspecto este del mercado de trabajo que merece por sí solo otra investigación. En todo caso, somos conscientes de la «censura» a que se ha sometido la muestra -truncamiento posterior a la extracción de las observaciones-, lo que puede incidir en la consistencia de los estimadores (Varios,(1988)).

La muestra final que ha servido de base para la estimación la han constituido 6.962 sujetos varones de las cinco comunidades autónomas citadas, que han estado o estaban empleados en el momento de realizarse la encuesta, y

contestaron en ella a las preguntas que permitían especificar las variables que finalmente se consideraron relevantes en el modelo. Para un análisis detallado de las definiciones y criterios utilizados en la ECVT puede consultarse Ministerio de Economía y Hacienda (1986). En cualquier caso, por lo que respecta a la definición de desempleo de larga duración, se ha considerado como tales a aquellos sujetos que llevaban en situación de desempleo más de 18 meses.

III. ESPECIFICACION DEL MODELO Y DEFINICION DE LAS VARIABLES.

El modelo de elección discreta que se ha especificado es un probit binario; en estos modelos (Maddala (1983)), la variable dependiente es dicotómica y sus valores están dados por el signo de una variable latente (Y_i); se supone una distribución normal con media cero y varianza unitaria para las perturbaciones (E_i). Siendo X_i el vector de atributos-variables explicativas y β el vector de parámetros a estimar, el modelo es:

$$Y_i = X_i \beta + \epsilon_i \quad i=1, \dots, n.$$

donde Y_i es inobservable y, en lugar de ella, observamos:

$$D_i = \begin{cases} 0 & \text{si } Y_i \leq 0 \\ 1 & \text{si } Y_i > 0 \end{cases}$$

Con el supuesto de normalidad puede escribirse el logaritmo de la función de verosimilitud para los datos:

$$\log L_i = \begin{cases} \log \text{Prob} (Y_i \leq 0 \mid X_i \beta) & \text{si } D_i = 0 \\ \log \text{Prob} (Y_i > 0 \mid X_i \beta) & \text{si } D_i = 1 \end{cases}$$

Sumando separadamente sobre las observaciones $D_i=0$ y $D_i=1$, tenemos

$$\log L(\beta) = \sum_{D_i=0} \log (1-F(X_i \beta)) + \sum_{D_i=1} \log (F(X_i \beta))$$

donde $F(\cdot)$ es la función de distribución normal. El propósito es 'maximizar este logaritmo de la función de verosimilitud con respecto al vector β mediante

algun algoritmo iterativo del tipo Newton-Raphson. Cada iteración consiste en actualizar el vector de parámetros por un vector de cambio igual al producto de la inversa de la matriz Hessiana del logaritmo de la verosimilitud con su gradiente, evaluados en el valor de los parámetros de la iteración previa. La concavidad global de esta función de verosimilitud está asegurada bajo las condiciones de que no todos los X_i sean nulos y de que el valor de la función de densidad normal evaluada en $X_i\beta$ sea distinto de cero para, al menos, alguna observación [Hall (1981)]. Por tanto, la probabilidad de que un individuo se encuentre en la situación de parado de larga duración viene dada por:

$$P(Y_i=1) = F(X_i\beta^*)$$

siendo, pues, $F(X_i\beta^*)$ el valor de la función de distribución de una variable aleatoria normal de parámetros (0,1) evaluada en el punto de corte $X_i\beta^*$. Por diferencia, la probabilidad de que un individuo no sea parado de larga duración se obtiene como

$$P(Y_i=0) = 1 - F(X_i\beta^*)$$

A partir de la muestra aleatoria antes descrita es posible, entonces, la estimación de los parámetros poblacionales correspondientes a un conjunto dado de atributos del individuo y que constituyen el conjunto de variables explicativas de la «elección» del sujeto. Dado que el modelo especificado ha sido el mismo a la hora de pronosticar las probabilidades de paro y de paro de larga duración, nos centraremos en los comentarios que siguen en uno sólo de ellos.

Así pues, la variable endógena es una variable discreta que toma el valor 1 si el individuo es parado de larga duración (sujeto a las condiciones antes especificadas de la muestra, esto es, llevar más de 18 meses en paro y haber contestado las preguntas relevantes) y 0 si no lo es. Por su parte, el conjunto de atributos-variables explicativas incluidas se detalla a continuación.

a. Variables relativas a la situación familiar.

X_1 . Variable binaria que toma el valor 1 cuando el sujeto está casado y 0 en caso contrario.

X_2 . Variable binaria que toma el valor 1 cuando en la unidad familiar no hay otros miembros -aparte del entrevistado- que estén trabajando de forma remunerada y/o que perciban subsidio de desempleo.

b. Variables relativas a la edad.

X₃. Variable dicotómica con valor 1 si el entrevistado tiene de 16 a 22 años (ambos inclusive) y 0 en caso contrario.

X₄. Variable dicotómica con valor 1 si el entrevistado tiene entre 56 y 64 años (ambos inclusive) y 0 en caso contrario.

c. Variables relativas al nivel de estudios.

X₅. Variable dicotómica que toma el valor 1 cuando el entrevistado tiene estudios primarios (E.G.8.) y 0 en caso contrario.

X₆. Variable dicotómica que toma el valor 1 cuando el entrevistado tiene estudios medios (Bachillerato o Formación Profesional) o superiores (Universidad o Escuela Técnica, de grado medio y superior) y 0 en caso contrario. Se han agrupado estudios medios y superiores dado que en la muestra no aparece ningún parado de larga duración con estos últimos.

d. Variable relativa a experiencia laboral.

X₇. Variable modelizada con una «proxy» dicotómica que vale 1 cuando el entrevistado lleva incorporado al mercado de trabajo menos de 25 años, que es aproximadamente la media muestral de incorporación, definida esta como la edad actual del encuestado menos su edad de entrada al mercado de trabajo.

e. Variable relativa a la movilidad laboral.

X₈. Variable que se aproxima por el número de veces que el entrevistado ha cambiado de trabajo a lo largo de su vida.

f. Variables relativas al sector de actividad.

X₉. Variable dicotómica con valor 1 si el entrevistado pertenece al sector agrícola, pesquero, silvícola o ganadero (grupo O de la Clasificación Nacional de Actividades Económicas), y 0 en caso contrario.

X₁₀. Variable dicotómica que toma el valor 1 cuando el entrevistado pertenece al sector de la construcción (grupo 5 de la C.N.A.E.) y 0 en caso contrario.

X₁₁. Variable dicotómica igual a 1 cuando el entrevistado pertenece al sector servicios (grupos 6,7,8 y 9 de la C.N.A.E.) y 0 en caso contrario.

g. Variables relativas a la Comunidad Autónoma.

X₁₂. Variable dicotómica que toma el valor 1: cuando el entrevistado reside en la Comunidad Autónoma de Andalucía y 0 en caso contrario.

X₁₃. Variable dicotómica que toma el valor 1 cuando el entrevistado reside en la Comunidad Autónoma de Madrid y 0 en caso contrario.

X_{14} . Variable dicotómica con valor igual a 1 cuando el entrevistado reside en la Comunidad Autónoma de Valencia y 0 en caso contrario.

X_{15} . Variable dicotómica con valor igual a 1 cuando el entrevistado reside en la Comunidad Autónoma del País Vasco y 0 en caso contrario.

Por último, cabe hacer algunos comentarios relativos a ciertas variables que, aunque inicialmente fueron consideradas, se optó finalmente por excluirlas del modelo. Así, se planteó incluir una variable relativa al número de hijos **económicamente** dependientes; se definieron tres variables binarias: «Sin hijos», «Con 1,2 o 3 hijos» y «Familia numerosa», tomando como grupo de referencia el segundo. El signo de los coeficientes era positivo para los correspondientes a los grupos extremos aunque no eran estadísticamente significativos. En segundo lugar, también se planteó incluir una variable que reflejase el nivel de renta del sujeto entrevistado; en este sentido, se definieron tres estratos, de renta correspondientes a los ingresos familiares mensuales declarados en la encuesta: «Hasta 75.000 pesetas», «De 75.000 a 150.000» y «Más de 150.000 pesetas», y se tomó como grupo de referencia el primero. Los signos de los coeficientes resultaron positivos, tal y como cabía prever, y también significativos. La eliminación de esta variable se debió a diversas razones, destacando la escasa fiabilidad de los resultados tras la tabulación. Así, nadie en la muestra declaraba ganar lo suficiente como para pertenecer al último estrato; no cabiendo suponer un tamaño de la supuesta «mentira» homogéneo, se decidió eliminar la variable.

Otra variable que se intentó incluir se refería a la estabilidad laboral. Para ello se procedió a la construcción de diferentes «proxys», tales como la duración media de cada empleo, definida como la edad actual menos la edad de incorporación al mercado de trabajo, respecto al número de empleos. Esta variable resultó ser no significativa y se optó por descomponerla en las dos que finalmente se han incluido, esto es, experiencia y movilidad laboral. También se intentó incluir la variable de experiencia laboral por tramos, aunque resultó más conveniente expresarla respecto a la media muestral.

Por último, también se probaron numerosos cruces entre las variables explicativas. Por un lado, la excesiva complejidad de cara a la interpretación final dado el elevado número de cruces resultantes y, por otro, los bajos estadísticos de Student alcanzados en la mayoría de los casos, nos hicieron optar por su exclusión final.

IV. ALGUNAS CARACTERÍSTICAS RELEVANTES DEL PARO SEGUN LA ECVT.

Como resultado de las tabulaciones de las distintas variables consideradas en el modelo, podemos destacar las siguientes características encontradas en la muestra (para una análisis más detallado relativo a la totalidad de la muestra puede consultarse Raymond y Castaner (1988).

Por lo que se refiere a la **situación familiar**, hay un mayor porcentaje de no casados que se encuentra tanto en paro como en paro de larga duración (véase Cuadro 1). Por otro lado, los sujetos que no tienen familiares remunerados y/o subsidiados presentan un porcentaje menor en ambas situaciones.

Por grupos de edad el porcentaje de parados desciende al aumentar la edad, mientras que el porcentaje de parados de larga duración muestra claramente un perfil en forma de U invertida. Esto es debido, en parte, a que si bien es mayor el porcentaje de jóvenes parados, es relativamente mas difícil que éstos hayan permanecido más de 18 meses buscando empleo activamente, dada su corta permanencia en el mercado de trabajo.

En cuanto al **nivel de estudios**, se observa una tendencia decreciente a medida que los individuos incrementan su formación. Respecto a la **formación laboral** o experiencia adquirida en el mercado de trabajo, y tomando como variable de aproximación el exceso o defecto sobre la media de los años de incorporación al mercado de trabajo, se observa que es mayor tanto el porcentaje de parados como de parados de larga duración para aquellos individuos cuya experiencia es inferior a la media, aunque en el caso de los parados de larga duración esta diferencia es poco relevante.

Analizando los datos a nivel de **sector de actividad** en el que trabajan o han trabajado los individuos encuestados, se observa que el sector que cuenta con un mayor número de desempleados es el de la construcción, seguido de agricultura, servicios e industria. Sin embargo, para el colectivo de parados de larga duración, el ranking de altera, siendo la agricultura el sector de actividad que cuenta con menor porcentaje de parados, de larga duración. Este resultado, que por otra parte es esperado, se ha de deber, en parte, al carácter estacional del trabajo agrícola.

Cuadro 1		
TABULACION DE LA MUESTRA (porcentajes del total de cada grupo)		
ATRIBUTO	PARADOS	PARADOS LARGA TEMPORADA
Casados	11,7	4,2
No casados	18,8	6,1
Sin familiares remunerados y/o subsidiados	12,9	4,4
Con familiares remunerados y/o subsidiados	14,9	5,2
De 16 a 22 años	18,9	3,7
De 23 a 55 años	14,2	5,1
De 56 a 65 años	8,8	3,7
Analfabetos	18,4	6,4
Estudios primarios	12,5	4,6
Estudio medios/superiores	9,1	2,8
Baja experiencia laboral	14,9	4,8
Alta experiencia laboral	12,4	4,6
Agricultura	21,1	2,8
Industria	8,3	3,0
Construcción	24,1	10,4
Servicios	15,9	6,2
Andalucía	18,5	5,3
Cataluña	10,2	4,4
Madrid	8,1	3,8
Valencia	11,7	5,0
País Vasco	10,6	4,0

Por lo que se refiere a la variable indicativa de la **Comunidad Autónoma**, es Andalucía la que tiene un mayor porcentaje tanto de parados como de parados de larga duración. Es preciso destacar que ésta es la comunidad de la muestra con una estructura demográfica más joven. Por magnitud de porcentajes en

el caso de parados le siguen Valencia, País Vasco, Cataluña y, en último lugar, Madrid, mientras que en el caso de parados de larga duración, Cataluña y el País Vasco alteran sus posiciones anteriores.

V. ESTIMACION DEL MODELO.

V.1. El paro de larga duración como variable endógena.

El modelo estimado en este caso contiene todas las variables explicativas mencionadas anteriormente. Sin embargo, como puede observarse en el Cuadro 2, los valores del estadístico t-Student relativos a las comunidades autónomas resultan muy bajos, por lo que se realizó una nueva estimación excluyendo estas cuatro variables. De hecho, aplicando un test de la razón de verosimilitudes no se puede rechazar la hipótesis nula de que el conjunto de los coeficientes de estas variables sean nulos (véase Cuadro 3). Por tanto, se optó por retener el modelo en el que no se incluyen estas variables; los resultados de su estimación se presentan en el Cuadro 4.

Teniendo en cuenta que el individuo de referencia es un varón no casado, sin familiares remunerados y/o subsidiados, de entre 23 y 55 años, sin estudios primarios acabados, con experiencia laboral por encima de la media y adscrito al sector industrial (residente en Cataluña, en la primera estimación), observamos que:

* En lo referente a las características familiares, el estar casado incide negativamente en la probabilidad de ser parado de larga duración, mientras que la existencia de otros perceptores, de renta en la unidad familiar incide positivamente en dicha probabilidad.

* En cuanto a la edad, los distintos coeficientes denotan de nuevo un claro efecto de U invertida, es decir, el grupo de mayor riesgo es el de las edades medias.

* Como cabía esperar, la probabilidad de ser parado de larga duración disminuye claramente al aumentar la formación educativa del individuo.

* El poseer menor experiencia laboral que la media incide positivamente en la probabilidad de ser parado de larga duración. Aunque la significación del coeficiente correspondiente es relativamente baja, nos ha parecido conveniente

Cuadro 2		
MODELO PROBIT DE PROBABILIDAD DE PARO DE LARGA DURACION		
VARIABLES EXPLICATIVAS	COEFICIENTE ESTIMADO	ESTADISTICO t-STUDENT
Constante	-1,219	-10,55
Situación familiar		
X ₁	-0,338	-5,17
X ₂	-0,147	-2,57
Edad		
X ₃	-0,284	-2,59
X ₄	-0,160	-1,93
Nivel de estudios		
X ₅	-0,231	-3,53
X ₆	-0,495	-5,72
Experiencia y movilidad laboral		
X ₇	0,080	1,18
X ₈	-0,006	-4,03
Sector de actividad		
X ₉	-0,252	-2,30
X ₁₀	0,450	5,91
X ₁₁	0,223	3,09
Comunidad Autónoma		
X ₁₂	0,108	1,43
X ₁₃	-0,027	-0,24
X ₁₄	0,095	1,07
X ₁₅	0,015	0,15
Número de observaciones:		6.962
Logaritmo función verosimilitud:		-1.238,74
Test de la razón de verosimilitud:		179,06
(Valor Ji-Dos al 99%: 30,6)		

mantener la variable, más aún si tenemos en cuenta que en el modelo en el que la variable endógena es el desempleo, dicho coeficiente resulta ampliamente significativo.

* En cuanto a la movilidad laboral, aunque el estadístico de Student indica que el valor del coeficiente es significativo, éste es prácticamente nulo y negativo, lo que indica una débil incidencia de este signo sobre la probabilidad de paro de larga duración a medida que aumenta la rotación en el empleo.

Cuadro 3

TEST DE LA RAZON DE VEROSIMILITUD (Contraste de significación de las variables referentes a las Comunidades Autónomas)	
<i>VARIABLE ENDOGENA</i>	<i>ESTADISTICO</i>
Paro de larga duración	5,49
Paro	47,54

Valor de Ji-Dos al 99%: 13,3

* Por último, y centrándonos en los sectores de actividad, la pertenencia al sector agrícola incide negativamente en la probabilidad de ser parado de larga duración, mientras que el pertenecer al sector servicios o de la construcción incide de forma positiva respecto al grupo de referencia, que es el sector industrial.

V.2. El paro como variable endógena.

Como se ha dicho, dada la estructuración del modelo en función de las características-atributos del sujeto, se ha creído de interés reestimar el desempleo, independientemente de su duración. Se comprueba que la única comunidad cuyo coeficiente es significativo es Andalucía. Sin embargo, el bloque de las comunidades no se excluye dado que, como vemos en el Cuadro 3, realizando un test de la razón de verosimilitudes, los datos rechazan la hipótesis de que los coeficientes sean nulos para este conjunto de variables. Por lo tanto, nos quedamos con el modelo que incluye todas las variables explicativas: en el Cuadro 5 se presentan los resultados de su estimación. Cabe destacar lo siguiente:

* En lo referente a las características familiares, los coeficientes indican exactamente lo mismo que en el modelo anterior, es decir, ser casado y sin

Cuadro 4		
MODELO PROBIT DE PROBABILIDAD DE PARO DE LARGA DURACION (Estimación sin comunidades autónomas)		
VARIABLES EXPLICATIVAS	COEFICIENTE ESTIMADO	ESTADISTICO t-STUDENT
Constante	-1,162	-11,32
Situación familiar		
X ₁	-0,337	-5,17
X ₂	-0,141	-2,48
Edad		
X ₃	-0,277	-2,54
X ₄	-0,165	-2,01
Nivel de estudios		
X ₅	-0,249	-3,86
X ₆	0,518	-6,05
Experiencia y movilidad laboral		
X ₇	0,083	1,23
X ₈	-0,005	-3,89
Sector de actividad		
X ₉	-0,225	-2,08
X ₁₀	0,456	6,04
X ₁₁	0,229	3,21
Número de observaciones:		6.962
Logaritmo función verosimilitud:		-1.240,57
Test de la razón de verosimilitud: (Valor Ji-Dos al 99%: 27,7)		175,40

familiares remunerados y/o subsidiados, incide negativamente en la probabilidad de ser parado respecto al individuo de referencia que, en este caso, reside en Cataluña.

* En cuanto a la edad, aunque el coeficiente para el grupo de individuos de entre 16 y 22 años no es significativo, destaca la significación del correspondiente al último tramo de edad. Hay que hacer notar que, si bien para este último grupo, como era de esperar, la probabilidad de ser parado disminuye,

en el grupo más joven, pese a no incluir a los desempleados sin empleo anterior, el coeficiente, aunque poco significativo, resulte positivo.

* Los resultados para la variable nivel de estudios son idénticos a los del modelo anterior, esto es, menor probabilidad de ser parado a medida que aumenta la formación educativa.

* La experiencia laboral resulta significativa y su coeficiente es positivo; por tanto, tener una experiencia laboral inferior a la media incide positivamente en la probabilidad de ser parado.

* En cuanto a la movilidad laboral, observamos que en este caso el coeficiente deja de ser significativo, resultado lógico si se tiene en cuenta que el modelo anterior parece indicar que cambiar muchas veces de empleo supone permanecer desempleado pero, siempre, por poco tiempo. Esta claro, pues, que este coeficiente refleja que la posibilidad de ser parado de larga duración disminuye a medida que el individuo ha cambiado más veces de trabajo durante su vida laboral, aunque no resulta significativo a la hora de computar la probabilidad de que está o no en paro.

* Por su parte, la adscripción al sector servicios, de la construcción o primario incide favorablemente en la probabilidad de estar desempleado. Estos resultados son parcialmente distintos a los obtenidos en el modelo anterior, donde el coeficiente referente al sector primario presentaba signo negativo. Ello puede deberse, en parte, al carácter estacional del trabajo agrícola, donde es mayor la probabilidad de estar parado pero no la de permanecer largas temporadas desempleado.

* Por último, este modelo permite hacer un análisis selectivo por comunidades autónomas, aunque las diferencias entre los parámetros no son muy grandes. Es de destacar que los referentes a Andalucía, Comunidad Valenciana y País Vasco sean positivos, lo que indica que estar censado en estas comunidades implica una mayor probabilidad de paro que para los censados en Cataluña. El signo negativo de la Comunidad Autónoma de Madrid comporta, en términos relativos, una menor probabilidad de paro para sus habitantes que para los de Cataluña.

Cuadro 5

MODELO PROBIT DE PROBABILIDAD DE PARO

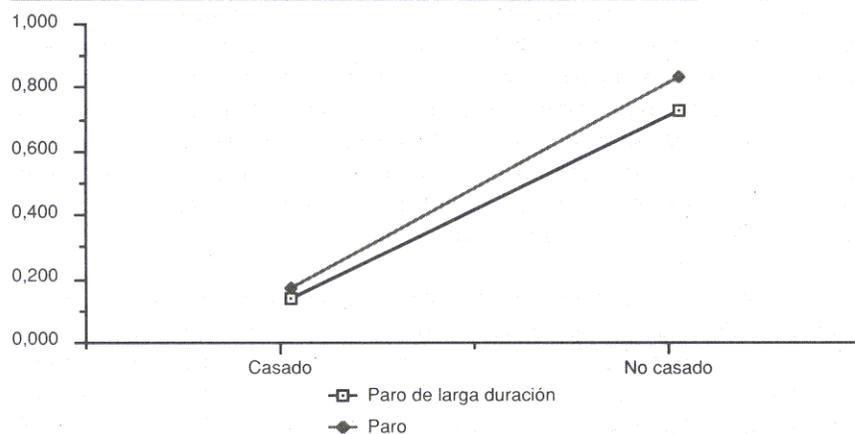
VARIABLES EXPLICATIVAS	COEFICIENTE ESTIMADO	ESTADISTICO t-STUDENT
Constante	-1,029	-11,53
Situación familiar		
X ₁	-0,329	-6,70
X ₂	-0,111	-2,62
Edad		
X ₃	-0,014	0,19
X ₄	-0,325	-5,23
Nivel de estudios		
X ₅	-0,314	-6,34
X ₆	0,508	-8,15
Experiencia y movilidad laboral		
X ₇	0,134	2,65
X ₈	0,006	0,61
Sector de actividad		
X ₉	0,435	6,46
X ₁₀	0,674	11,05
X ₁₁	0,427	7,68
Comunidad Autónoma		
X ₁₂	0,296	5,23
X ₁₃	-0,111	-1,26
X ₁₄	0,100	1,46
X ₁₅	0,097	1,26
Número de observaciones:		6.962
Logaritmo función verosimilitud:		-2.543,50
Test de la razón de verosimilitud:		465,06
(Valor Ji-Dos al 99%: 30,6)		

VI. DIFERENCIAS REGIONALES: PROBABILIDADES DE PARO.

La estimación previa de los dos modelos seleccionados y comentados permite el cómputo de probabilidades para individuos de diferentes características. Obviamente, los dos modelos contienen suficiente número de atributos-variables explicativas como para poder realizar múltiples combinaciones y obtener otras tantas probabilidades para cada tipo de individuo especificado. Este ejercicio se realiza, a modo de ejemplo, solo con algunas combinaciones de características distintas en el Cuadro 6 donde, claro está, los resultados coinciden con los indicado por los parámetros estimados.

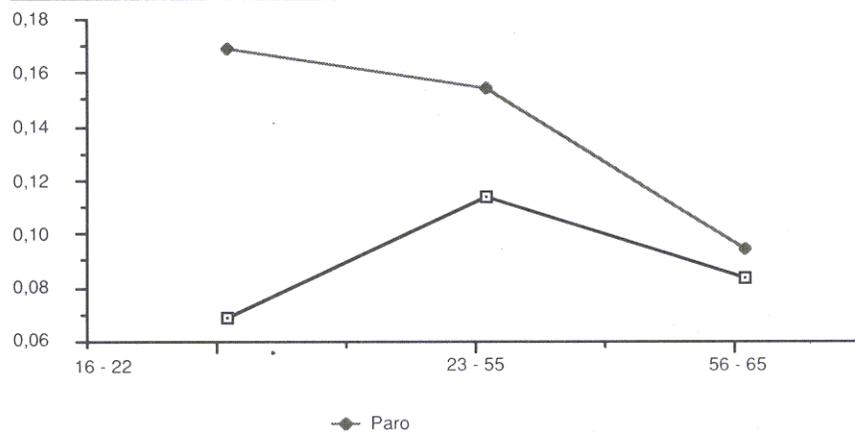
Así, una rápida observación de estas probabilidades expresadas en porcentajes indica que el estar casado supone menor probabilidad de paro para el individuo de referencia: ya citado (gráfico 1). También disminuye la probabilidad de paro a medida que aumenta la edad -aunque no en lo referente a larga duración (gráfico 2), la formación educativa (gráfico 3) y la experiencia laboral. Por sectores, destaca la elevada probabilidad de desempleo -y desempleo de larga duración- en la construcción (gráfico 4); este hecho hay que atribuirlo al bajo nivel relativo de actividad que, en el momento de realización de la encuesta y los años inmediatamente previos, presentaba este sector. Por último, la probabilidad de estar parado en Andalucía es netamente superior al resto de comunidades consideradas (gráfico 5).

Gráfico 1
Probabilidad de paro según estado civil.



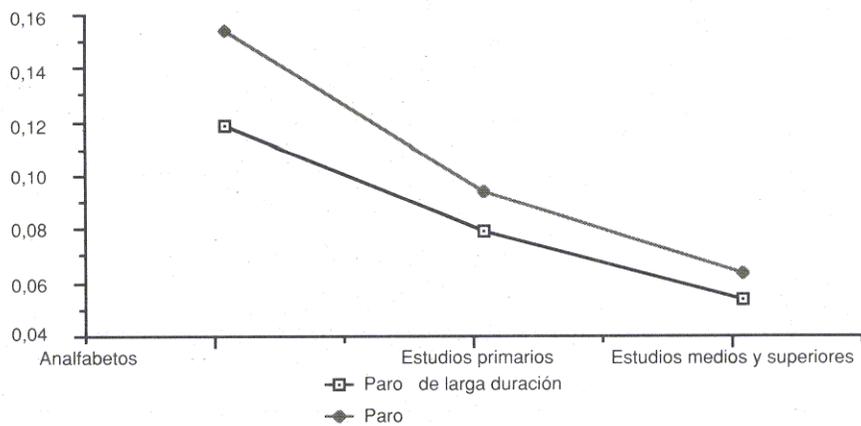
Individuo de referencia: Varón catalán y soltero, de 23 a 54 años, analfabeto con experiencia laboral y adscrito al sector industrial. En el caso de paro de larga duración no hay referencia a Comunidad Autónoma.

Gráfico 2
Probabilidad de paro según la edad.



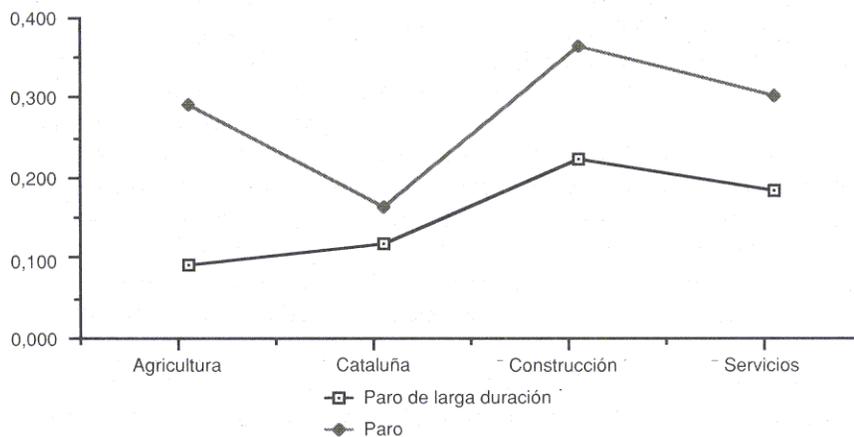
Individuo de referencia: ver gráfico 1.

Gráfico 3
Probabilidad de paro según nivel de estudios.

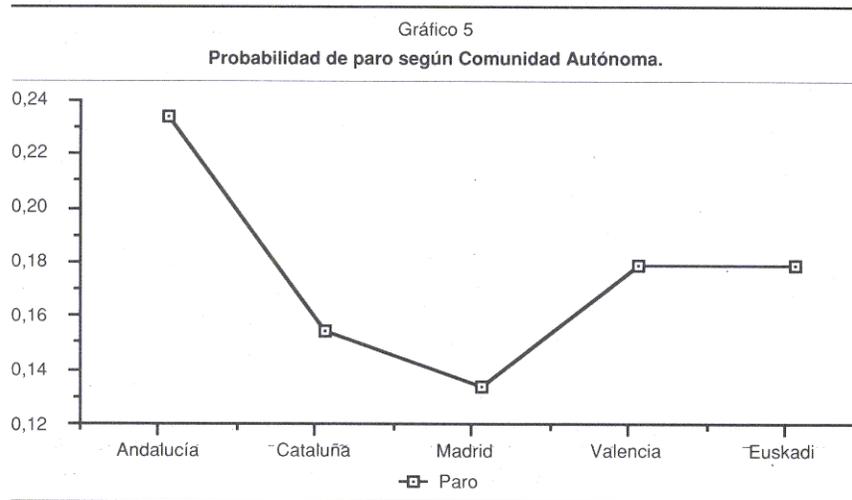


Individuo de referencia: ver gráfico 1.

Gráfico 4
Probabilidad de paro según sector.



Individuo de referencia: ver gráfico 1.



Individuo de referencia: ver gráfico 1.

Uno de los aspectos a destacar, en este sentido, son las diferencias que a nivel regional pueden detectarse en función de los parámetros correspondientes a las Comunidades Autónomas estimados en el modelo de desempleo. Recuérdese que el modelo estimado finalmente con el paro de larga duración como variable endógena excluye las Comunidades Autónomas como variable explicativa, dada la nula significación de los estadísticos, tanto a nivel particular como conjunto (Cuadros 2 y 3). Por el contrario, al estimar el modelo con el desempleo en general como variable endógena, aunque Andalucía es la única comunidad cuyo coeficiente es significativo, el conjunto de variables explicativas regionales no se excluye ya que los datos rechazan la hipótesis nula de que sus coeficientes sean nulos. Por tanto, una primera conclusión que consideramos relevante en el presente trabajo hace referencia a la ausencia de evidencia empírica que permita afirmar la existencia de diferencias entre las regiones consideradas con respecto a la probabilidad de que los varones con experiencia laboral lleguen a encontrarse en situación de parados de larga situación.

Un resultado evidentemente contrario se obtiene para el caso en el que la variable relevante sea el desempleo en general. En efecto, en el Cuadro 7 se han pronosticado las probabilidades de desempleo según Comunidad Autónoma para el grupo de referencia y variaciones, una a una, de las características

Cuadro 7

PROBABILIDADES DE PARO SEGUN COMUNIDAD AUTONOMA ANTE VARIACIONES EN LAS CARACTERISTICAS DEL SUJETO DE REFERENCIA (porcentaje)					
	<i>ANDALUCIA</i>	<i>MADRID</i>	<i>VALENCIA</i>	<i>PAIS VASCO</i>	<i>CATALUÑA</i>
Suejo de referencia	23,2	12,7	17,6	17,6	15,2
Casado	14,5	7,1	10,4	10,4	8,7
De 16 a 22 años	23,6	14,0	17,9	17,9	15,4
De 56 a 64 años	14,5	7,1	10,4	10,4	8,7
Estudios primarios	14,7	7,4	10,8	10,8	9,0
Estudios medios o superiores	10,8	4,9	7,5	7,5	6,2
Baja experiencia laboral	27,4	15,9	21,2	21,2	18,4
Agricultura	38,2	23,9	31,2	31,2	27,8
Construcción	47,6	31,9	39,7	39,7	35,9
Servicios	37,8	23,9	30,8	30,8	27,4

incluidas como variables explicativas dicotómicas. Lógicamente, las diferencias interregionales se mantienen entre grupos de individuos, aunque con distinta intensidad; así, las características personales, educativo-laborales y sectoriales influyen de forma más acusada en las comunidades de Madrid y Cataluña que en el resto.

De lo observado en este último cuadro cabe deducir tanto la similitud de respuesta del desempleo en el País Vasco y Valencia, dados los coeficientes estimados en el Cuadro 5, como la grave problemática que el desempleo supone en Andalucía. Sin pretender animar interpretaciones ligadas al agravio comparativo, puede comprobarse que salvo que medien estudios superiores o medios (y pese a la importante influencia de las distintas características personales), cualquier varón andaluz tenía en el momento de realizarse la encuesta, mayor probabilidad de estar parado que, el individuo de referencia en la Comunidad de Madrid; también, el sujeto de referencia andaluz con estudios

medios o superiores (que reduce de un 0,238 a un 0,108 la probabilidad de estar parado respecto al sujeto sin estudios primarios) tiene mayor probabilidad de estar desempleado que un individuo censado en Madrid o Cataluña que, sin embargo, tenga solo estudios primarios, y análoga a la que presenta este en el País Vasco o Valencia.

Por último, la menor intensidad en las diferencias interregionales se comprueba entre los sujetos adscritos a sectores productivos no industriales, mientras que es especialmente acusada entre los individuos con mayor nivel educativo, de mayor edad y estado civil casado.

VII. CONSIDERACIONES FINALES.

En este trabajo se ha planteado la estimación de dos modelos probit binarios en los que la variable dependiente ha sido el desempleo, en un caso, y el desempleo de larga duración, en otro. La muestra ha sido obtenida a partir de la «Encuesta de Condiciones de Vida y Trabajo en España», realizada en el cuarto trimestre de 1985, y se refiere a sujetos varones residentes en cinco comunidades autónomas y con experiencia laboral previa.

La especificación del modelo, al igual que otros muchos de este tipo, se plantea en términos de atributos-características de la población desempleada. La posible simplicidad teórica de estos modelos resulta ampliamente compensada por su efectividad, al permitir pronosticar las probabilidades, en nuestro caso, de estar desempleado y de ser parado de larga duración respecto al colectivo especificado en la muestra. Por otra parte, es evidente la posibilidad de ampliar el análisis aquí efectuado en términos de modelos más generales (así, se podría efectuar un análisis mediante modelos Tobit y Heckman relativos a la duración del desempleo), a la vez que la mejora en la definición de los estimadores considerando la posible inconsistencia que supone el no asumir la censura efectuada (truncamiento ex-post de la muestra).

Una vez estimados los modelos finalmente asumidos sobre paro de larga duración (Cuadro 4) y desempleo en general (Cuadro 5), los resultados más relevantes pueden resumirse como sigue. La existencia de otros perceptores de renta en la unidad familiar y la pertenencia a un estrato de edad medio (de 23 a 55 años) inciden de forma positiva en la probabilidad de ser parado de

larga duración, mientras el pertenecer al sector agrícola, y variables como la formación educativa, la experiencia laboral y la movilidad en el trabajo, inciden de forma negativa. Nótese que análoga incidencia presenta la variable relativa al estado civil cuando toma el valor correspondiente a estar casado, lo que sin duda está recogiendo efectos indirectos de otros factores, tales como la edad o la misma ausencia de perceptores de renta en la unidad familiar si tenemos en cuenta la relativamente baja tasa de actividad femenina.

El comportamiento es similar en el caso de ser parado en general, aunque con diversas matizaciones. En concreto, mientras que, lógicamente, la pertenencia al tramo de edad juvenil supone una disminución de la probabilidad de ser parado de larga duración, el efecto no está claro en cuanto al desempleo en general por la razón de que se han eliminado de la muestra aquellos sujetos que carecían de empleo anterior. Además, la movilidad laboral no presenta efecto significativo sobre la probabilidad de paro, efecto que es significativamente negativo, aunque débil, en el caso del paro de larga duración. Por su parte, mientras que la adscripción al sector primario supone una incidencia negativa en la probabilidad de paro de larga duración, esta es claramente positiva en cuanto al desempleo en general; en este último caso, el sector que genera menor probabilidad de paro es la industria. Por último, hay que destacar que las variables relativas a comunidades autónomas sí presentan una afectación clara sobre la probabilidad de desempleo, mientras que no inciden de forma significativa en el desempleo de larga duración.

BIBLIOGRAFIA

- AMEMIYA, T. (1981)**, «Qualitative response models: a survey», *Journal of Economic Literature*. vol. XIX, PP. 1483-1536.
- HALL, B.H. (1981)**, «A program for binary probit estimation by the maximum likelihood method», Stanford University, mimeo.
- MADDALA, G.S. (1983)**, *Limited dependent and qualitative variables*, Cambridge University Press.
- MINISTERIO DE ECONOMIA Y HACIENDA (1986)**, *Análisis de las condiciones de vida y trabajo en España*, Secretaría de Estado de Economía.
- RAYMOND, J.L. y CASTANER, J.M. (1988)**, «Algunos rasgos de la actividad y el paro según la E.C.V.T.», *Documentos de Trabajo* núm. 33, FIES.
- VARIOS (1988)**, *Cuadernos Económicos de I.C.E.*, núm 39.