

La Educación Ocupacional y la Salida del Paro en España. *

J. Ignacio Murillo Arcos[†]

Universidad del País Vasco

Fac. de Ciencias Económicas

Lehendakari Agirre 83, E48015 BILBAO, Spain.

Tfno. 94 601 3844

Octubre de 1999

*Agradezco a Namkee Ahn y a Marta Regúlez su excelente labor de supervisión en la realización de este trabajo. A Sara de la Rica y Luis Toharia por facilitarme los datos. También agradezco los útiles comentarios recibidos por una versión previa de este artículo en las II Jornadas de Economía Laboral y en el XXII Simposio de Análisis Económico. Finalmente quisiera agradecer la ayuda económica del proyecto SEC97-1249 de la DGICYT. Por supuesto, todos los errores que haya son culpa mía exclusivamente.

[†]Dpto. de Econometría y Estadística. e-mail: im@alcib.bs.ehu.es

Resumen

En este artículo se estudia el efecto de la formación no reglada sobre la probabilidad de encontrar un empleo utilizando datos de la Encuesta de Población Activa (EPA) enlazada desde 1992 a 1997. Al hacerlo se tiene en cuenta la posible existencia de sesgo de selección muestral y de heterocedasticidad, encontrándose evidencia de esta última únicamente. El modelo estimado finalmente es un probit con heterocedasticidad multiplicativa, y en él se encuentra que los parados que realizan estudios ocupacionales durante más de una entrevista de la EPA consiguen incrementos significativos en la probabilidad de encontrar un empleo antes de la última entrevista. Estos incrementos son mayores para las mujeres que para los hombres. Dado que el problema del paro en España afecta con mayor gravedad a las mujeres, las posibilidades de salida del desempleo que les abre la realización de estudios no reglados debe ser destacada.

Palabras clave: mercado de trabajo, búsqueda de empleo, educación ocupacional.

Abstract

This paper considers the estimation of the effect of vocational education on the probability of employment using Spanish data from the linked Active Population Survey (EPA) since 1992 to 1997. After allowing for possible sample-selection bias and heteroskedasticity in estimation, only evidence of the later is found. Finally a probit model with multiplicative heteroskedasticity is estimated and it is found that those unemployed who receive vocational education for more than one survey achieve significant increases in the probability of finding a job before their last survey from the EPA. These increases are greater for women than for men. Since unemployment among women is specially severe in Spain, the greater chances of going out of unemployment that they get due to the realization of vocational education must be emphasized.

Key words: labour market, job search, vocational education.

JEL codes: C35, J24.

1 Introducción.

España es un país donde el problema del paro es muy grave. Algunas cifras –obtenidas de la Encuesta de Población Activa, EPA– pueden dar idea de la magnitud del problema: entre 1992 y 1997 (que será nuestro periodo muestral) la tasa de paro masculina siempre ha sido superior a un 13 % y la tasa de paro femenina nunca ha bajado del 25 % . El problema no es reciente: en los 17 últimos años, la tasa de paro agregada ha estado entre el mínimo del 16 % de 1982 y el máximo del 24 % de 1994.

Dadas la gravedad y la persistencia del desempleo en España, es muy importante la cuestión de qué pueden hacer los propios parados para salir del desempleo. En este artículo se va a investigar una de las teorías, la de la “competencia por los puestos de trabajo,” según la cual una persona puede mejorar su puesto en la “cola del paro” aumentando su capacitación con educación adicional, y gracias a ello encontrar un trabajo antes. Hay muchos estudios con datos de otros países en los que se ha contrastado esta teoría, por medio del estudio del efecto que ha tenido la participación en algún programa específico de formación sobre el historial de empleo de las personas. Así, es abundante la literatura sobre las consecuencias del YTS (Young Training Scheme) en Gran Bretaña –ver por ejemplo O’Higgins (1994)–, o del CETA (Comprehensive Employment and Training Act) en los EE.UU. -ver por ejemplo Card y Sullivan (1988). En España las referencias son escasas y se puede citar el artículo de Saez y Toledo (1996); estos autores estudian las posibilidades de encontrar un empleo de participantes en el FIP (Plan Nacional de Formación e Inserción Profesional). Mucho menos numerosos son los trabajos en los que no se estudian las repercusiones de haber participado en algún programa gubernamental concreto de formación, sino de haber recibido formación en un sentido amplio, permitiendo que varíen las características de la enseñanza recibida. Un trabajo interesante en esta última línea es Gritz (1993); en él se estudian los efectos de haber participado en algún programa de formación, distinguiendo si era público o privado, sobre la duración y la frecuencia de los episodios de empleo de los jóvenes, y se encuentran implicaciones distintas dependiendo del sexo y de si el programa de formación era privado o no.

Alternativamente se han utilizado las teorías del capital humano –Levhari y Weiss (1974), Kodde (1988)– como fundamento del análisis de los efectos de la formación sobre la situación laboral de los individuos. En general, estas teorías están planteadas en un marco de asignación óptima de los recursos, sin paro involuntario, y probablemente este marco no es el más adecuado para analizar una situación como la del mercado laboral español. Además, dentro de un modelo de capital humano las decisiones de los individuos sobre su educación dependen del flujo descontado de ganancias *futuras* que se obtengan gracias a ella, tanto en términos de salario como de probabilidad de empleo; como yo pretendo estimar la dependencia de la probabilidad de empleo respecto de las decisiones *pasadas* de educación de las personas, se produciría una redundancia si se trataran de modelar los dos procesos. Por estas dos razones, plantearé el modelo de acuerdo con la teoría de la “competencia por los puestos de trabajo existentes” en vez de como un modelo de capital humano.

En concreto, el objetivo principal de este trabajo es el estudio del efecto de la formación

no reglada¹ en el sentido de facilitar la salida del estado de desempleo de los jóvenes.² Los estudios empíricos que analizan el éxito en la búsqueda de empleo son bastante exigentes en términos de datos, en dos sentidos. Por una parte, es aconsejable que el período durante el cual se recogen datos de cada individuo sea largo, ya que la búsqueda de empleo lleva tiempo y es interesante conocer no sólo cuánto tardan las personas en conseguir trabajo, sino también cuál es la duración de los episodios de empleo. Por otra parte, algunos factores que son cruciales en la probabilidad de que el individuo encuentre empleo, como su habilidad, o su ética de trabajo, son muy difíciles de medir. En este trabajo se van a utilizar datos desde principios de 1992 hasta el segundo trimestre de 1997 de la EPA de stocks (enlazadas las seis entrevistas a través del número de identificación del cuestionario). Esta fuente no es demasiado rica en los dos sentidos mencionados. En primer lugar, el período de seguimiento de los encuestados es de año y medio únicamente, y por ello es posible que haya muchos casos de personas que reciben formación durante la encuesta pero se colocan después de terminado el período de seguimiento de la EPA. Por otra parte, la única variable algo relacionada con la habilidad del individuo que hay en la EPA es el nivel de estudios. La variable dependiente de nuestro modelo es binaria, y se define como 1 si el individuo estaba trabajando en su sexta –y última– entrevista y 0 si estaba parado. No se va a realizar un estudio longitudinal, considerando también la misma variable dependiente en las encuestas cuarta o quinta, para minimizar en lo posible el problema del corto período de duración de los encuestados en la muestra.

Una pregunta de la EPA es: “¿Ha seguido durante las 4 últimas semanas algún tipo de estudios?”; a los que responden que sí se les piden algunos detalles sobre los mismos, entre ellos “¿Qué tipo de estudios ha seguido?”. Las 9 últimas opciones en que se clasifican las respuestas a esta pregunta corresponden a estudios no reglados. Estos estudios no suelen conducir a la obtención de un título oficial, sino que se realizan con los objetivos de completar la educación formal o reglada y de ajustarse a las demandas del mercado de trabajo. Una razón adicional para seguir en especial a las personas que han realizado estudios no reglados es que este tipo de estudios suelen ser de duración más corta que los estudios reglados, y más encaminados hacia la consecución a corto plazo de un trabajo. La corta duración de este tipo de cursos hace más probable el hecho de que una persona que está haciendo estudios no reglados en sus primeras entrevistas los finalice antes de la sexta y última encuesta. Y al estar encaminados hacia la consecución a corto plazo de un empleo, es menos probable el

¹Como su nombre indica, la formación no reglada alude a toda aquella formación que no es reglada, y es por tanto una categoría definida por defecto. La formación o educación reglada recibe su nombre de que está regulada por el Ministerio de Educación y Cultura, o las consejerías correspondientes de las Comunidades Autónomas allí donde las competencias en materia educativa estén transferidas. Estas regulaciones comprenden las materias que se enseñan y los horarios y fechas de los calendarios escolares. En términos amplios, la educación reglada abarca desde la enseñanza básica obligatoria hasta las carreras universitarias, pasando por la educación secundaria no obligatoria. Cuando la enseñanza no reglada se refiere a materias que pueden resultar útiles en cualquier trabajo, se denomina educación o formación ocupacional (En este sentido, gran parte de la enseñanza no reglada es formación ocupacional, aunque abundan las excepciones, como cualquier curso sobre actividades de ocio). No obstante, en la jerga oficial se suele reservar el término de formación ocupacional para los cursos impartidos a los parados, denominándose formación continua cuando los receptores son ocupados.

²Tomaremos como jóvenes a los menores de 35 años; entre las personas de más edad, muy pocas reciben algún tipo de formación.

hecho de que a su finalización el parado formado siga estudiando.³

No vamos a llevar a cabo un análisis de duración de los episodios de paro debido a que la información disponible sobre la formación ocupacional, que es el centro de nuestro interés, sería dependiente de la duración. Sirva un ejemplo como ilustración de este punto: supongamos que se restringe la muestra a entrantes al paro durante la primera encuesta de la EPA; si uno de estos parados encuentra empleo en la entrevista final, podemos saber si ha hecho formación no reglada o no durante los últimos 6 trimestres; en cambio si el empleo lo consigue en la segunda encuesta, únicamente tendremos esta información para los últimos 2 trimestres, mientras que desconoceremos si ha recibido ese tipo de formación el año anterior a éstos.⁴ El modelo que finalmente estimaremos será un probit con heterocedasticidad. Más habitual en la literatura –LaLonde (1986), O’Higgins (1994)– es emplear un modelo bivalente, que tenga en cuenta la endogeneidad de la formación, debida a la autoselección –que se origina porque la decisión sobre si recibir una formación o no recibirla se realiza empleando criterios económicos–; sin embargo, con la muestra que se va a emplear no se rechazará la hipótesis de que no hay dependencia entre los factores inobservables que influyen en la probabilidad de obtener un empleo y los que afectan a la decisión de formación, y en este caso el prescindir de la ecuación que describa la participación en la formación no supone una pérdida de consistencia ni de eficiencia de los estimadores en el modelo que analiza los determinantes de la probabilidad de empleo. Debido a que la situación en el mercado de trabajo de los jóvenes varía mucho según el sexo, vamos a realizar análisis separados para hombres y para mujeres.

Nuestro principal resultado es que la duración y con cuánta anterioridad a la sexta encuesta se hayan cursado los estudios ocupacionales afectan de manera importante al aumento en la probabilidad de empleo que se consigue con ellos. Se estima que el aumento en la probabilidad de haber conseguido un trabajo para la sexta entrevista es de un 26% en los hombres y de un 37% en las mujeres si se han cursado estudios ocupacionales durante las cinco primeras encuestas de la EPA. Si se ha estado haciendo estos estudios durante un número de encuestas de la EPA inferior los efectos estimados son de menor cuantía, aunque alcanzan el 10% de mejora en la probabilidad de empleo entre los hombres y el 18% entre las mujeres. El efecto de haber hecho estudios ocupacionales en alguna encuesta concreta de la EPA es más modesto, no superando el 6% entre los hombres ni el 8% entre las mujeres, aunque como veremos no resulta despreciable cuando lo comparamos con las repercusiones sobre la salida del paro de las enseñanzas regladas, cuya duración suele ser bastante mayor.

La organización del resto del trabajo es la siguiente. En la segunda sección se describe la muestra, se discute la especificación del modelo probit para la probabilidad de tener

³En cuanto a la formación ocupacional realizada *antes* de empezar a ser seguidos por la EPA, es difícil que sea recogida en la misma, porque ésta pregunta acerca de los estudios orientados a una actividad o profesión, pero impone los requisitos de que su duración sea superior a un año y de al menos 300 horas anuales; en un estudio que hemos hecho sobre la formación ocupacional en el País Vasco, encontramos que la mayoría de los cursos duraban 5 meses o menos, aunque el total de horas a menudo superaba las 300.

⁴Sería posible el limitar este problema con restricciones adicionales de la muestra, por ejemplo eliminando a los que hayan cursado formación no reglada después de la segunda entrevista de la EPA, pero nuestros resultados mostrarán que los cursos que se realizan durante menos tiempo pueden ser los menos eficaces en términos de facilitar la salida del paro, así que este tipo de restricciones podrían afectar seriamente a las conclusiones del estudio.

un empleo en la sexta encuesta y se explica qué variables explicativas incluiremos en la estimación inicial. En la tercera sección se estima nuestro modelo final, que es un modelo probit con heterocedasticidad multiplicativa para la probabilidad de salir del paro durante la EPA, y se comentan sus resultados. Finalmente en la cuarta sección se presentan las conclusiones.

2 El trabajo empírico.

2.1 Los datos.

Para estimar el efecto de las variables de educación del individuo sobre la probabilidad de encontrar empleo, utilizo la Encuesta de Población Activa (EPA) enlazada. La EPA entrevista a una misma persona durante un máximo de 6 trimestres consecutivos, y así se pueden seguir los cambios en la situación laboral de cada entrevistado durante año y medio como máximo. El período que yo he escogido es desde principios de 1992 hasta mediados de 1997, con lo cual es posible seguir el historial de empleo de 17 grupos de personas: para el primer grupo hay entrevistas desde el primer trimestre de 1992 hasta el segundo trimestre de 1993, y para el último hay entrevistas desde el primer trimestre de 1996 hasta el segundo de 1997. Partir de un número tan grande de cohortes de la EPA es conveniente porque el estar recibiendo educación en un trimestre dado es un suceso poco frecuente;⁵ El porcentaje es mucho menor para los mayores de 35 años, y por esta razón nos vamos a quedar únicamente con personas entre 16 y 34 años. Con esto la muestra queda restringida a un grupo que vamos a llamar de personas jóvenes. En general, la situación de los jóvenes en el mercado de trabajo español es más desfavorable que la del resto de la población adulta, por tres razones: porque sus tasas de paro son mucho más altas que para el conjunto de la población activa,⁶ porque tardan más tiempo que el resto en conseguir un empleo y porque una vez que lo hacen éste suele ser más frecuentemente de carácter temporal.

Se supone que un medio que tienen los jóvenes para mejorar sus perspectivas de ingresos futuras es el de mejorar su formación -si no esperaran ningún beneficio futuro, no gastarían tiempo y dinero en ello-. Básicamente, la mejora de ingresos futura puede ocurrir por dos vías: gracias a un aumento del salario cuando encuentren trabajo, y gracias a un aumento de la probabilidad de encontrar trabajo. Nosotros únicamente vamos a investigar la segunda de estas vías, y de una manera muy de corto plazo: como la EPA sólo sigue a los individuos durante año y medio, cualquier mejora en la situación laboral de un individuo que suceda después de ese período es inobservable por este medio. Debido a la brevedad del período de

⁵La desventaja es que la situación del mercado de trabajo en el que intentan introducirse va variando: el inicio de la muestra coincide con una recesión, que hace que entre 1992 y 1994 se destruyan 878.000 puestos de trabajo netos en España, mientras que a partir de 1995 se invierte la tendencia, creándose más de 300.000 empleos netos cada año. Intentaremos controlar esta disparidad mediante regresores, y también tendremos en cuenta de este modo su situación previa en el mercado de trabajo, el nivel de estudios terminados, la edad y la posición en la familia.

⁶Como ilustración de este hecho, en el segundo trimestre de 1995 las tasas de paro en la cohorte de 20 a 24 años eran, tanto para los hombres como para las mujeres, 15 puntos porcentuales por encima de las del total de la población activa según sexo.

Tabla 1.— **Personas que realizan estudios no reglados, por grupo de edad y sexo**

CURSANDO ESTUDIOS NO REGLADOS				
EDAD	HOMBRES		MUJERES	
	n²	%	n²	%
16-19	81	4.3	117	6.3
20-24	295	7.5	527	12.9
25-29	249	8.6	434	13.9
30-34	77	4.0	169	7.5
SUMA	702	6.6	1247	11.0

Muestra: parados en su primera entrevista, con edades entre 16 y 35 años

seguimiento de los individuos en la EPA, vamos a restringir el estudio a las consecuencias de la educación ocupacional. La razón es que es muy raro que alguien emprenda unos estudios secundarios, o universitarios si piensa permanecer en ellos durante menos de 2 años. Como consecuencia, si en la primera entrevista una persona declara que ha seguido algún tipo de educación reglada en el último mes, esperaremos que los beneficios de estos estudios se trasladen a su probabilidad de empleo después de que le realicen la última entrevista de la EPA. Por el contrario, estudios no reglados del tipo de cursillos de informática, de idiomas, u otros más técnicos es más probable que los realicen los individuos con la idea de completar su formación para encontrar un trabajo más a corto plazo. La duración de este tipo de estudios también suele ser más breve, y por ello es más probable que si los estaban cursando en una de las primeras entrevistas de la EPA, los terminen antes de la sexta encuesta, comparado con el caso de que lo que estuvieran haciendo durante las entrevistas iniciales de la EPA fueran estudios reglados.

En los 17 grupos de personas entrevistadas por la EPA durante nuestro periodo muestral, hay un total de 21980 que estaban paradas en su encuesta inicial, que tenían edades entre 16 y 35 años y que respondieron a las 6 encuestas. De ellos, un 8.9 % estaba realizando algún tipo de estudios ocupacionales en su primera encuesta. Como se aprecia en la tabla 1, eran muchas más las mujeres que los hombres que los cursaban. Las edades en las que con más frecuencia se estaban haciendo estos estudios no reglados eran entre los 20 y los 30 años.

Si los jóvenes que realizan estudios ocupacionales y los que no los hacen tuvieran en términos medios las mismas características, se podría medir el efecto de la formación sobre la salida del paro de una manera sencilla: bastaría con controlar quienes han cursado este tipo de estudios en un periodo dado del tiempo y quienes no, y comparar el porcentaje de ellos que están trabajando con posterioridad.⁷ Con los datos de los que nosotros disponemos

⁷Esta es la base del enfoque experimental en la evaluación de programas de formación, que cuenta con múltiples partidarios. En este tipo de experimentos, los candidatos a recibir la formación son asignados aleatoriamente a uno de dos grupos: los que efectivamente son aceptados en los cursos, y los que no lo son, que pasar a constituir el grupo de control. Una descripción reciente de las ventajas e inconvenientes de este tipo de evaluación de los efectos de la formación, así como de las voces más destacadas que han contribuido a su popularidad en los EE.UU. se puede encontrar en Friedlander y otros (1997).

Tabla 2.— Porcentajes de empleados, según si hicieron estudios no reglados y el sexo.

<u>Formación en $t = 1$</u>	<u>% Ocupados</u>			
	<u>HOMBRES</u>		<u>MÚJERES</u>	
	NO	SI	NO	SI
2 ^a trimestre	18.5	12.5	10.8	10.3
3 ^{er} trimestre	27.3	21.2	16.8	15.5
...
6 ^a trimestre	37.1	36.3	25.4	29.3

Muestra: parados en su primera entrevista,
con edades entre 16 y 35 años

este tipo de comparación es criticable, por razones que comentaremos a continuación, pero la hemos realizado como una parte interesante del análisis descriptivo de la muestra. Para ello, hemos diferenciado a los encuestados dependiendo de si habían cursado estudios no reglados en las 4 semanas previas a la primera entrevista de la EPA, y hemos comparado el porcentaje de empleados de uno y otro grupo en encuestas sucesivas. Los resultados aparecen en la tabla 2. En esta tabla se aprecia que el resultado de la formación es relativamente más favorable para las mujeres que para los hombres. En el primer trimestre todos los componentes de la muestra están parados. En las siguientes entrevistas, las mujeres que realizaron la formación ocupacional tienen unas tasas de empleo ligeramente peores que las de las mujeres que no la hicieron, pero para la sexta encuesta la situación se ha invertido, y el porcentaje de ocupadas es superior entre las formadas. En cambio, la inserción laboral de los hombres formados es más lenta que la de los no formados, aunque para la encuesta final casi se ha cerrado la brecha.

No obstante, este es un medio demasiado simple de mirar el efecto de la educación no reglada sobre la probabilidad de encontrar empleo, por varios motivos. En primer lugar, no se está controlando si los individuos estaban haciendo algún tipo de estudios en las encuestas posteriores a la primera. Por otra parte, no se tiene en cuenta que las características de cada individuo pueden afectar al hecho de que consiga empleo o no, y al de que decida formarse o no. Es evidente que circunstancias como el historial de empleo, o el nivel educativo influyen en la probabilidad de encontrar trabajo, y unas simples tablas bidimensionales de frecuencias no pueden controlar estos efectos. Además, como señalan Card y Sullivan (1988), la decisión de iniciar una formación no es independiente de la situación presente en el mercado de trabajo, sino que es más probable que entren en un programa de formación aquellos con unas perspectivas de empleo a corto plazo peores. Una manera de corregir este efecto es tener en cuenta cuál ha sido la situación previa en el mercado laboral de los encuestados, como haremos en la próxima sección.

Finalizaremos este apartado exponiendo los criterios con los que se ha seleccionado la muestra, además de los ya descritos. Como se ha mencionado con anterioridad, el periodo de seguimiento de los entrevistados por la EPA es demasiado breve como para permitir el estudio de la probabilidad de la salida del paro de aquellos individuos que no han terminado su

educación reglada. En consecuencia, hemos eliminado a estas personas de la muestra, dejando en ella únicamente a los que no realizaban ningún tipo de estudios durante las 6 encuestas y a los que declaraban que lo que cursaban eran estudios no reglados. También excluyo de la muestra los individuos que seguían estudiando en el último trimestre. La justificación para ello es que si entonces estaban estudiando, es muy posible que sus esfuerzos de búsqueda de otras situaciones en el mercado de trabajo fueran limitados. También es posible pensar que una parte de las personas así eliminadas de la muestra pueden ser opositores a algún empleo en el sector público. En general, la preparación de unas oposiciones requiere un período de estudio bastante largo, y por ello, dadas las características de la muestra utilizada, sería difícil estudiar el éxito en la salida del desempleo de esta gente. Hay que señalar que se han excluido todas las personas que declaraban estar cursando estudios en su sexta entrevista, no sólo aquellos que lo hacían y además estaban inactivos o parados. El objetivo de este estudio es analizar la salida del paro hacia el empleo, y por ello hemos prescindido de las observaciones de los que están inactivos en la sexta encuesta y de los que inician el servicio militar durante el periodo de seguimiento de la EPA. Con la misma idea, se ha quitado de la muestra a las personas que ya habían encontrado un empleo en la encuesta inicial y que están ocupadas en la siguiente encuesta⁸; las quitamos porque consideramos que estas personas no han encontrado trabajo durante el periodo de seguimiento de la EPA, sino antes. Tampoco se deja en la muestra a los que declaran en su primera encuesta que son incapacitados permanentes⁹. La alternativa de construir una variable explicativa que recoja esta condición e incluirla en la estimación no es viable, porque ninguna de estas personas sale del paro hacia un empleo. Finalmente, hay 144 personas que en alguna entrevista declaran que sí estaban realizando estudios no reglados, y a continuación dicen que el número habitual de horas de formación a la semana era cero; creemos que esta respuesta no es coherente, y por este motivo los hemos suprimido de la muestra.

2.2 Especificación del modelo para la probabilidad de encontrar trabajo.

Una vez que hemos expuesto cuál es la composición de la muestra que será objeto de nuestro análisis, pasemos a considerar el planteamiento del modelo econométrico que nos permita realizar dicho análisis. Un factor que complica nuestro estudio, respecto a la mayoría de los trabajos previos –ver por ejemplo Friedlander, Greenberg y Robins (1997) ó Saez y Toledo (1996)– es que los cursos de formación no reglada de todos los participantes no son simultáneos, es decir, los inician y terminan en encuestas de la EPA distintas. Cuando sí hay homogeneidad en cuanto a la fecha de finalización de la formación, el modelo más simple que se puede plantear en el que la probabilidad de empleo depende de la formación es del siguiente tipo:

⁸Estas personas son paradas cuando aún no ha comenzado su empleo.

⁹La EPA no hace directamente una pregunta del estilo a “¿Es usted incapacitado permanente o no?” y por este motivo es posible que algunos de los individuos que dejamos en la muestra sean incapacitados permanentes. La pregunta H1 de la EPA es : “¿En qué situaciones de las siguientes se encontraba la semana pasada?” De las 7 opciones que se ofrecen una es ‘Incapacitado permanente’, pero como máximo se pueden señalar 3 opciones, así que es posible que un incapacitado permanente indique otras 3 opciones, y no esa.

$$y_i^* = \alpha' \mathbf{x}_i + \beta' \mathbf{h}_{i,-t} + \gamma D_i + u_i \quad i = 1, \dots, n. \quad (1)$$

En este modelo D_i es una variable binaria, con un valor igual a 1 si el individuo realizó estudios ocupacionales y 0 si no lo hizo. El parámetro que acompaña a D_i trataría de medir el efecto de estos estudios. A continuación vamos a describir con detalle el resto de los elementos de este modelo, porque aparecerán también en nuestro análisis empírico.

En primer lugar, y_i^* es una variable latente no observable; en nuestro caso, puede considerarse como un índice de la posición potencial del i -ésimo individuo en la competencia por las vacantes que van surgiendo. Cuando y_i^* sea positiva, el i -ésimo individuo conseguiría un empleo en la sexta entrevista. Es posible dar una interpretación económica a la ecuación (1); en España, la tasa de paro es alta y no hay trabajo para todos los que desean trabajar. Si α , β y u_i fueran observables, con el término derecho de la ecuación (1) se podría poner a los individuos en un orden, y los situados en los primeros lugares conseguirían ocupar los puestos de trabajo. En realidad α , β y u_i son desconocidos y lo que se observa es

$$y_i = \begin{cases} 1 & \text{si } y_i^* > 0 \\ 0 & \text{si } y_i^* < 0 \end{cases}$$

Por lo tanto, $y_i = 1$ quiere decir que el individuo estaba empleado en su sexta encuesta e $y_i = 0$ quiere decir que estaba parado en ella.

En el vector \mathbf{x}_i se recogen las características personales de los individuos de la muestra sobre las que la EPA proporciona información y que a priori consideramos que pueden tener influencia sobre la variable endógena. La elección del número de entrevista en el que se miden estas características se ha hecho de una forma un tanto arbitraria, aunque es un factor que no altera los resultados. Vamos a indicar entre paréntesis, a continuación del nombre de cada variable, el número de entrevista en que se ha medido. **Edad** (1) es la edad en años de cada persona. La variable **estudios** (1) ha sido agrupada en los 8 niveles que se muestran en la tabla 7 al final del artículo y hemos definido una variable ficticia para cada uno de estos 8 niveles. En cuanto a la **formación complementaria** (1) hemos explicado en la introducción –concretamente en una nota a pie de página– que su definición en la EPA es muy estricta, lo que provoca que muchos estudios que en principio podrían mejorar las perspectivas de empleo de los individuos no queden recogidos en la EPA. A pesar de ello creamos una variable ficticia que indique si el individuo declaró que hizo algún tipo de formación complementaria tal y como la define la EPA. Finalmente, utilizamos la escasa información sobre la situación familiar que proporciona la EPA¹⁰ para definir cuatro variables ficticias, que sirven para controlar si el individuo es el **cabeza de familia** (1), el **cónyuge** (1), un **hijo** (1), o si su situación es **otra distinta** (1). Además hemos creado una variable ficticia que recoja si cada encuestado estaba **soltero** (6) o no.

$\mathbf{h}_{i,-t}$ es un vector de variables que describen la historia observable de empleo del individuo. Introduciendo esta información en el modelo de regresión se pretende controlar en la medida de lo posible las diferentes situaciones iniciales en el mercado de trabajo de los componentes de la muestra, ya que éstas pueden hacer que sus probabilidades de empleo en la sexta

¹⁰Decimos que es escasa porque no sabemos nada sobre las rentas de ningún miembro de la familia, ni tampoco sobre su riqueza.

encuesta sean distintas. Para comenzar, un hecho que se encuentra en prácticamente todos los estudios sobre la salida del desempleo es que en general cuanto más tiempo lleve una persona en el paro más difícil es que encuentre trabajo en un periodo de tiempo dado. La forma que hemos escogido para controlar la dependencia de la duración es introduciendo las variables **años en paro** (1) y **años en paro al cuadrado** (1) en el modelo.¹¹ La situación laboral de los individuos entre la encuesta inicial y la final se controla introduciendo en $\mathbf{h}_{i,-t}$ las variables endógenas retardadas $\mathbf{y}_{i,-1}, \dots, \mathbf{y}_{i,-4}$ que indican si el individuo estuvo empleado o no en las cuatro entrevistas previas a la sexta. Consideramos que la inclusión de estas variables es esencial para que las estimaciones tengan sentido. Para entender la razón de ello, piénsese que parte de los miembros de la muestra estarán trabajando en su quinta entrevista y el resto no. La omisión de la variable $\mathbf{y}_{i,-1}$ supondría no tener en cuenta esta información, y por lo tanto estaríamos comparando la probabilidad de retener un empleo que tiene un trabajador con la probabilidad de encontrar un empleo que tiene un parado. Esto sería incorrecto, porque la primera probabilidad es mucho mayor que la segunda.

Otra de las variables que forma parte de $\mathbf{h}_{i,-t}$ es **paro provincial** (5), que es la tasa de paro provincial por sexo. Con esta variable se quiere recoger las distintas condiciones del mercado de trabajo local a las que se enfrentan los individuos de la muestra. Las demás componentes de $\mathbf{h}_{i,-t}$ son variables ficticias. **Con experiencia** (1) indica si en la primera encuesta el parado respondió que había trabajado con anterioridad. **Con subsidio** (5) vale 1 si el individuo estaba recibiendo subsidio o prestación por desempleo en la semana previa a la quinta encuesta. Otra pregunta de la EPA a los que buscan empleo es “¿Qué tipo de empleo busca o ha encontrado?” Las opciones se refieren a empleos de jornada completa ó de jornada parcial. Más de la mitad de los individuos en nuestra muestra respondía que buscaba un empleo “de cualquier tipo”; para tener en cuenta otras respuestas a esta pregunta hemos creado dos variables ficticias, **sólo jornada completa** (1) que vale 1 si la respuesta es que buscaba empleo “de jornada completa únicamente” y **prefiere jornada completa** (1) que indica si la respuesta era que buscaba empleo “de jornada completa o, en su defecto, de jornada parcial”. Ceteris paribus, sería de esperar que en promedio la salida del paro fuera más rápida para los que están dispuestos a aceptar cualquier tipo de empleo que para los que desechan algunas ofertas por ser trabajos a jornada parcial. Sin embargo, también es posible que los que muestran una preferencia por trabajar a jornada completa sean los individuos cuyas características les hacen más propensos a conseguir este tipo de empleos, y en este caso su salida del paro no tendría por qué ser más lenta. Finalmente, el vector $\mathbf{h}_{i,-t}$ contiene información sobre el tipo de método de búsqueda que utilizaron los individuos durante las cuatro semanas previas a la encuesta inicial. En la EPA se pregunta: “¿De qué forma ha buscado empleo en las cuatro últimas semanas?” y el individuo puede señalar un máximo de tres de las 11 alternativas que se ofrecen. Siguiendo nuestro criterio personal, hemos dividido los métodos que declararon en pasivos, medios y activos.¹² La información

¹¹Una manera alternativa de terminar con los problemas que pueden surgir debido a que personas con duraciones en el paro distintas tengan en general probabilidades de empleo diferentes hubiera sido quedándonos en la muestra con entrantes al paro únicamente. Hemos descartado esta opción porque entonces habríamos centrado el estudio de los efectos de la formación ocupacional sobre un grupo demasiado limitado de personas.

¹²Hemos considerado como pasivos 4 métodos: el estar inscrito en una oficina de empleo de la Adminis-

que hemos utilizado es de qué tipo era el método más activo que declaraban. De este modo, **método activo** vale 1 si entre los métodos citados había al menos uno que era activo y 0 en caso contrario; **método medio** vale 1 si señalan algún método de intensidad media pero ninguno activo; y **método pasivo** es el indicador de que todos las alternativas respondidas eran pasivas.

Como se ha dicho, la especificación realizada en la ecuación (1) no es adecuada debido a que es demasiado simple como para recoger la variedad que se registra en la EPA sobre la duración y la fecha de finalización de los estudios no reglados. Esta es una información adicional que proporciona nuestra muestra, y para introducirla en el modelo empírico definimos cinco variables ficticias de estudios ocupacionales en vez de solamente una, y lo hacemos de dos maneras alternativas:

Alternativa cuántas: en este caso se mira el número total de encuestas anteriores a la sexta en que el individuo estaba cursando estudios no reglados. Puede estar entre 0 y 5, y se define una variable ficticia para cada posibilidad (sus nombres serán **una, dos, tres, cuatro y cinco**). La categoría de referencia será cero. Por ejemplo, si un individuo estuvo cursando estudios no reglados en sus entrevistas primera y tercera, entonces $dos_i = 1$ y el resto de variables ficticias de estudios no reglados valen 0.

Alternativa cuándo: aquí se tiene en cuenta el número de la encuesta en la que cursó estudios no reglados. Las variables ficticias se denominan en este caso **primera, segunda, tercera, cuarta y quinta**. Por ejemplo, si un individuo estuvo cursando estudios no reglados en sus entrevistas primera y tercera, entonces $primera_i = 1$, $segunda_i = 0$, $tercera_i = 1$, $cuarta_i = 0$ y $quinta_i = 0$.

La idea es que, por una parte, la *cantidad* de estudios no reglados puede influir en la probabilidad de encontrar un empleo; pero por otro lado, *cuándo* se hayan realizado los estudios también puede ser importante.¹³ Por ejemplo, ceteris paribus, una persona con $cuarta_i = 1$ y $quinta_i = 0$ habrá tenido más tiempo para buscar empleo durante el quinto trimestre que otra persona con $cuarta_i = 0$ y $quinta_i = 1$, y debido a ello su probabilidad de tener empleo en el sexto trimestre puede que sea mayor. Sea \mathbf{E}_i el vector que contiene las 5 variables ficticias de estudios ocupacionales, definidas de una de las dos formas, para el individuo i -ésimo de la muestra. Entonces el modelo a estimar queda:

$$y_i^* = \alpha' \mathbf{x}_i + \beta' \mathbf{h}_{i,-t} + \gamma' \mathbf{E}_i + u_i \quad i = 1, \dots, n. \quad (2)$$

En principio haremos sobre la perturbación el supuesto de que sigue una distribución

tración, el estar inscrito en una oficina de empleo privada, el esperar los resultados de solicitudes anteriores y el esperar la llamada de una oficina de empleo. De intensidad media hemos catalogado el buscar a través de la prensa, el preparar o realizar oposiciones y otros métodos. Finalmente hemos clasificado dentro de los métodos activos el dirigirse a empresarios, el utilizar relaciones personales y el hacer gestiones para establecerse por cuenta propia.

¹³Obviamente, no se pueden incluir estos dos grupos de variables ficticias a la vez en un modelo de regresión porque hay multicolinealidad perfecta entre ellas: no son más que dos formas distintas de mirar la realización de estudios no reglados.

normal estándar, es decir, $u_i \sim \mathcal{N}(0, 1)$. En este caso un modelo probit relaciona la variable y_i con las variables en \mathbf{x}_i , $\mathbf{h}_{i,-t}$ y \mathbf{E}_i :

$$Pr(y_i = 1) = Pr(y_i^* > 0) = Pr(-u_i < \alpha' \mathbf{x}_i + \beta' \mathbf{h}_{i,-t} + \gamma' \mathbf{E}_i) = \Phi(\alpha' \mathbf{x}_i + \beta' \mathbf{h}_{i,-t} + \gamma' \mathbf{E}_i) \quad (3)$$

donde $\Phi(\cdot)$ es la función de distribución Normal estándar. Llamaremos a éste el modelo básico.

Hay varios motivos que pueden causar que el modelo básico esté mal especificado, y en consecuencia que los estimadores de los parámetros sean inconsistentes. Sin duda el problema potencial al que mayor atención se ha prestado en la literatura relativa a los efectos de la formación sobre la situación laboral de los parados es el del sesgo de selección muestral. El origen de este problema radica en que las variables de \mathbf{E}_i , relativas a la realización o no de estudios ocupacionales, no son exógenas, y es posible que estén correlacionadas con el término de error de (2), en cuyo caso los estimadores por máxima verosimilitud del modelo básico serían inconsistentes.

El sesgo de selección muestral es un problema que habrá que corregir en el caso de que las personas que realicen estudios no reglados tengan en términos medios características, no recogidas en \mathbf{x}_i ni en $\mathbf{h}_{i,-t}$, capaces de causar que su probabilidad de empleo hubiera sido diferente a la del resto de los componentes de la muestra incluso en ausencia de esos estudios. Como ilustración, en los trabajos ya citados de LaLonde (1986), Card y Sullivan (1988) y O'Higgins (1994) se ha encontrado que los que entran en programas de formación públicos son las personas con peores perspectivas laborales –bien porque su nivel educativo es bajo, porque pertenecen a minorías étnicas o porque su historial de empleo reciente es malo–, y deciden participar porque creen que esta es una buena manera para ellos de insertarse en el mercado de trabajo. Nosotros hemos hecho lo posible por controlar en \mathbf{x}_i y $\mathbf{h}_{i,-t}$ este tipo de factores. Además las condiciones generales del mercado de trabajo en España son distintas que en los países para los que se realizaron esos estudios –EE.UU y Gran Bretaña–, principalmente porque en España el problema del paro de larga duración afecta a un segmento muy amplio de los jóvenes, así que el inscribirse en un curso de formación de duración corta no parecerá una pérdida de ingresos esperada para la mayoría de ellos. Por estas razones no creemos que sea fuerte un sesgo de selección adverso que siga el siguiente mecanismo: se inscriben en cursos de educación ocupacional personas a las que, por motivos inobservables para nosotros, les sería más difícil encontrar trabajo de todos modos; cuando terminan la formación, siguen teniendo características no observables que hacen que su colocación sea en promedio más difícil que la de los demás. Si esto fuera así, entonces en el modelo probit inicial se estaría infraestimando el efecto de la educación ocupacional sobre la probabilidad de empleo, esto es, los coeficientes estimados de esas variables tendrían un sesgo negativo.

El sesgo de selección muestral puede actuar en el sentido contrario en el caso de que las personas con mayor “propensión al aprendizaje” sean las que más a menudo realicen estudios no reglados, debido a que gracias a ellos “saltan” más puestos en la cola de parados que los individuos menos capacitados para absorber conocimientos. Como la capacidad para aprender cosas nuevas es algo que también pueden valorar positivamente los empleadores, en la medida en que ellos sean capaces de observar esa cualidad darán prioridad a la contratación

de este tipo de trabajadores y esto se reflejará en una correlación positiva entre las variables de estudios no reglados y el término de error en (2), ya que en él se agrupan los factores no observables para el investigador, como es el caso de la capacidad para el aprendizaje y la motivación, y que influyen en la probabilidad de encontrar empleo. Si algo así ocurre y lo ignoramos estimando el modelo (3), los efectos estimados de las variables de estudios ocupacionales estarán sesgados al alza.

Una manera de detectar y corregir si existiera el sesgo de selección muestral en el caso español consiste en estimar un modelo probit bivalente, formado por una ecuación para explicar la probabilidad de estar trabajando en la sexta encuesta y otra para la probabilidad de haber realizado estudios no reglados en alguna de las 5 primeras encuestas. Hay varias razones por las cuales se ha elegido esta segunda variable a explicar, en vez de, por ejemplo, estimar otras 5 ecuaciones, una para cada variable “Probabilidad de haber realizado estudios no reglados en el trimestre i -ésimo”. En primer lugar, actualmente la estimación de un modelo probit con 6 ecuaciones simultáneas es inmanejable. Además, hay una fuerte correlación entre las variables “Realizar estudios no reglados en un trimestre” y “Realizarlos en el trimestre siguiente”. Y por último, hay poquísimos casos en los que estas variables toman el valor ‘Sí’ respecto a los casos en los que son ‘No’.

Un segundo problema que puede afectar al modelo básico es la presencia de heterocedasticidad. En tanto en cuanto la varianza de la perturbación aleatoria sea constante, el supuesto de que la $\text{Var}(u_i) = 1$ para todo i hecho en (3) es inocuo, ya que en realidad lo que está identificado en este modelo son los ratios $\frac{\alpha_j}{\sigma_u}$, $\frac{\beta_j}{\sigma_u}$ y $\frac{\gamma_j}{\sigma_u}$, y no cada uno de estos coeficientes por separado. En cambio, si la $\text{Var}(u_i)$ no es constante y no lo tenemos en cuenta, los estimadores que se obtengan a partir de (3) son inconsistentes y las inferencias que hagamos usando este modelo no son válidas.

La heterocedasticidad puede surgir si hay algunos valores de ciertas variables asociados a una mayor variabilidad en la probabilidad de obtener un empleo que otros. Como ilustración, este podría ser el caso de las variables del nivel de estudios y también de las de haber realizado estudios ocupacionales. El razonamiento sería que el hacer unos estudios es una inversión con un rendimiento incierto a la que unos individuos le sacan un gran partido y en cambio otros obtienen un beneficio prácticamente nulo. En este caso personas que si no hubieran adquirido esa formación habrían ocupado puestos cercanos en la “cola del paro” tendrán tras la finalización de sus estudios probabilidades de empleo más distintas. En el modelo básico, esta idea de “la educación como una inversión arriesgada” se reflejaría en que la varianza de las perturbaciones sería mayor para las observaciones de los individuos más formados. El fenómeno contrario puede ocurrir si predomina el efecto de “la educación como igualadora de oportunidades”. Este efecto se produciría si las personas con una situación inicial peor son los que más provecho extraen al recibir un determinado tipo de formación. Razonamientos similares se podrían hacer para justificar la existencia de heterocedasticidad causada por otras variables explicativas.

Con el objeto de tener en consideración que en el modelo básico puede tener tanto un problema de selección muestral como de heterocedasticidad, hemos seguido los siguientes pasos en la estimación del modelo. En primer lugar, hemos detectado mediante un contraste que el modelo básico es probable que tenga un problema de heterocedasticidad, y hemos

delimitado las variables que pueden ser sus causantes. En cambio, en el modelo probit que especificamos para la probabilidad de haber realizado estudios ocupacionales en al menos una de las cinco primeras entrevistas de la EPA se concluye que hay homocedasticidad. A continuación admitimos que la endogeneidad de las variables de estudios ocupacionales puede provocar un problema de selección muestral. En concreto lo recogemos como un caso de selección en inobservables –ver Friedlander, Greenberg y Robins (1997) ó Heckman y Hotz (1989)– y para ello estimamos cuatro modelos probits bivariantes, con heterocedasticidad en la ecuación de empleo (siguiendo la forma a la que se ha llegado en la fase anterior) y con varianza constante en la ecuación de formación. El coeficiente de correlación entre las perturbaciones de las dos ecuaciones resulta no ser significativamente distinto de cero al nivel del 5 % (aunque de un modo marginal en el modelo de las mujeres con las variables de estudios ocupacionales incluidas de la forma que hemos denominado Alternativa cuántas), por lo que concluimos que no existe con nuestra muestra un problema relevante de selección muestral y en consecuencia adoptamos como especificación definitiva la de un modelo probit con heterocedasticidad multiplicativa en el que la variable endógena es la probabilidad de estar empleado en la sexta encuesta de la EPA. En el siguiente apartado se describen con mayor detalle los diversos modelos estimados.

2.3 Modificaciones por heterocedasticidad y sesgo de selección muestral.

Como hemos mencionado, los estimadores máximo-verosímiles del modelo básico son inconsistentes bajo heterocedasticidad de las perturbaciones. Para decidir si éste era un problema en nuestro caso hemos realizado el contraste que Davidson y MacKinnon (1984) denominan LM_2 . Una ventaja de este contraste es que, debido a que se trata de un test de los multiplicadores de Lagrange, para realizarlo sólo hace falta estimar el modelo básico (3), y no el modelo bajo la hipótesis alternativa de heterocedasticidad. En nuestro caso el modelo probit con heterocedasticidad multiplicativa puede escribirse como

$$y_i^* = \alpha' \mathbf{x}_i + \beta' \mathbf{h}_{i,-t} + \gamma' \mathbf{E}_i + u_i \quad i = 1, \dots, n \quad (4)$$

$$u_i \sim \mathcal{NID} \left(0, [e^{\theta' \mathbf{w}_i}]^2 \right) \quad i = 1, \dots, n. \quad (5)$$

Los parámetros de este modelo son identificables si \mathbf{w}_i no incluye un término independiente. Claramente, el modelo restringido -con varianza de u_i constante- se consigue si el vector $\theta = 0$, así que esta es la forma de especificar la hipótesis nula de homocedasticidad. En cuanto a la composición del vector \mathbf{w} , hemos decidido considerar como una primera aproximación que la varianza de u_i dependa de las mismas variables que y_i^* , salvo el término independiente. La tabla 3 recoge los resultados del contraste de heterocedasticidad de Davidson y MacKinnon (1984) realizado con nuestros datos.

A pesar de que incluso para niveles de significación muy pequeños se rechaza la hipótesis nula en favor de la alternativa de heterocedasticidad, conviene tomar estos resultados con cautela. La razón es que es bien conocido que este contraste puede estar recogiendo algún otro tipo de mala especificación en el modelo básico. No obstante, procederemos como si el problema que origina el resultado de los contrastes hechos fuera efectivamente la

Tabla 3.— Resultados de los contrastes de heterocedasticidad.

<u>LM_2</u>	<u>ALT. CUANTAS</u>	<u>ALT. CUANDO</u>
HOMBRES	121.22 *	120.63 *
MUJERES	123.76 *	121.85 *

Notas: La distribución bajo la H_0 es $\chi^2(31)$; * = significativo al 1 % .

heterocedasticidad, y abandonaremos el modelo básico en beneficio del modelo probit con heterocedasticidad (4)-(5).

Antes de complicar más este modelo con el tratamiento de la posible selección muestral conviene reducir la dimensión del vector \mathbf{w}_i , descartando algunas variables como causantes de que la varianza de u_i no sea constante. La razón es que la especificación del modelo (4)-(5) de la que partimos, con 31 elementos en el vector \mathbf{w}_i , tiene demasiados parámetros, hecho que se manifiesta en la lenta convergencia de los algoritmos de estimación no lineal que hemos empleado para maximizar la verosimilitud¹⁴ y en el elevado número de parámetros que resultan no ser significativamente distintos de cero. Esto nos lleva a pensar que la estimación de este modelo inicial con heterocedasticidad junto a otra ecuación que represente la posible selección muestral sería impracticable.

La reducción del número de variables en el vector \mathbf{w}_i que recoge la heterocedasticidad se hace de forma gradual, empezando por las variables que tienen asociados unos estadísticos t más próximos a 0, y no pretendemos que sea exhaustiva: dejamos en la ecuación (5) algunas variables que no son significativas al nivel del 5 % (al igual que no eliminamos en este paso ninguna variable de la ecuación (4)) con la idea de que tiempo habrá de suprimirlas más adelante, cuando alcancemos la especificación definitiva.¹⁵ El vector \mathbf{w}_i restringido contiene entre 7 y 9 variables, dependiendo de si se emplean los datos de los hombres o de las mujeres y de cómo se introducen las variables de las enseñanzas no regladas. Llamaremos al vector que contiene estas variables $\mathbf{w}_{\mathbf{R}i}$ y a los parámetros que lo acompañan $\theta_{\mathbf{R}}$.

Una vez que hemos reducido la dimensión del vector \mathbf{w}_i , estamos en condiciones de tener en cuenta el posible sesgo de selección muestral del modelo. Lo vamos a tratar como un problema de selección en inobservables, es decir, como un caso de correlación entre la perturbación de la ecuación (4) y la perturbación de una ecuación que describa la probabilidad de haber realizado estudios ocupacionales durante alguna de las cinco primeras encuestas de la EPA.¹⁶ Esta última probabilidad la vamos a modelar mediante una ecuación con variable endógena latente:

$$a_i^* = \alpha_2' \mathbf{x}_i + \beta_2' \mathbf{h}_{2i,-t} + v_i \quad i = 1, \dots, n. \quad (6)$$

¹⁴Para la estimación hemos utilizado el programa LIMDEP 5.0. Para la estimación del modelo con heterocedasticidad multiplicativa se ha aplicado el algoritmo de Davidon, Fletcher y Powell, tomando como valores iniciales las estimaciones del modelo probit básico mediante el método de Newton. Una descripción de este algoritmo se puede encontrar, por ejemplo, en Greene (1993, p. 350).

¹⁵Lo que pretendemos con ello es no incurrir en el error de omitir variables relevantes.

¹⁶En el apartado anterior se han dado algunos argumentos sobre cómo podría surgir esta correlación.

Se puede pensar en a_i^* como la propensión del individuo i -ésimo a realizar estudios no reglados durante los 5 primeros trimestres en que le sigue la EPA. Siguiendo esta interpretación, cuando esa propensión supera un cierto umbral, que sin pérdida de generalidad fijamos en cero, el individuo realiza los estudios y en caso contrario no lo hace. La variable a_i^* no es observable, aunque sí conocemos su signo, y conforme a él podemos crear una variable binaria que valga 1 si el individuo realizó estudios ocupacionales en alguna de las 5 primeras entrevistas y 0 en caso contrario. Denominaremos a esta variable de “haber realizado estudios no reglados en al menos una de las 5 primeras encuestas” como A_i . Los valores que toma esta variable en la misma muestra con la que anteriormente se han estimado las ecuaciones de la probabilidad de haber encontrado empleo para la sexta encuesta son:

	<u>1 (= SI)</u>	<u>0 (= NO)</u>
HOMBRES	663	7119
MUJERES	904	6278

Se puede apreciar que son pocas las personas que realizaron estudios no reglados en alguna de las 5 primeras encuestas. Este es un serio inconveniente a la hora de estimar un modelo de elección discreta, porque un ajuste muy bueno (con aproximadamente un 90 % de predicciones correctas) se puede conseguir simplemente con que el modelo estimado establezca que nadie tiene las características precisas para realizar estudios no reglados. Otra dificultad es que la EPA no es muy útil para analizar las decisiones de estudios de los individuos. En efecto, no contiene información sobre variables relativas a la riqueza de la familia, ni a la capacidad de las personas -o sea, inteligencia, motivación, calificaciones obtenidas en la enseñanza obligatoria, ... - para sacar partido a los estudios; en buena lógica, ambos tipos de variables deberían influir directamente en la decisión de realizar estudios no reglados.

Dada la información disponible, las variables que forman parte del término derecho de la ecuación (6) son las siguientes. El vector \mathbf{x}_i contiene las mismas características personales de los individuos muestrales que en el modelo básico. Por su parte en $\mathbf{h}_{2i,-t}$ hemos incluido las mismas variables que en el vector $\mathbf{h}_{i,-t}$ del modelo básico pero hemos añadido además las variables **tasa de paro por sexo, grupo de edad y nivel de estudios terminados** en el trimestre de la quinta entrevista y en los 6 trimestres anteriores a ésta. Con estas variables se quiere recoger el efecto sobre a_i^* de las condiciones del mercado de trabajo a las que tuvieron que enfrentarse los individuos en el momento en el que iniciaron los estudios ocupacionales.¹⁷ Nos hubiera gustado que con esta variable también se distinguiera por lugar de residencia, para medir mejor las perspectivas de empleo de los distintos individuos, pero la EPA no encuesta a suficientes personas en cada provincia como para permitirnos construir la tasa de paro provincial con ese nivel de desagregación. La manera de tener en cuenta que la situación del mercado de trabajo local puede afectar a la decisión de realizar o no

¹⁷Recuérdese que esta fecha de inicio de los estudios ocupacionales cambia de unas personas a otras, y puede ser anterior a la primera encuesta; en este caso, no se sabe cuándo se iniciaron esos estudios. Esta es la razón que nos ha motivado para tomar la estrategia conservadora de incluir inicialmente seis retardos de la variable ‘tasa de paro por grupo de edad, sexo y estudios terminados’.

estudios ocupacionales ha sido incluyendo en $\mathbf{h}_{2i,-t}$ las tasas de **paro provinciales** durante las encuestas primera y quinta.

A continuación nos hemos preguntado si había heterocedasticidad en el término de error de la ecuación (6), y tanto para el caso de los hombres como el de las mujeres hemos llegado a la conclusión de que no era así. En consecuencia el siguiente paso ha sido estimar un modelo probit bivalente, con heterocedasticidad en la ecuación de la probabilidad de empleo y sin ella en la ecuación de la probabilidad de estudios no reglados:

$$\left\{ \begin{array}{l} y_i^* = \alpha_1' \mathbf{x}_i + \beta_1' \mathbf{h}_{1i,-t} + \gamma' \mathbf{E}_i + u_i \quad u_i \sim \mathcal{NID} \left(0, [e^{\theta' \mathbf{R} \mathbf{w}_{Ri}}]^2 \right) \quad i = 1, \dots, n \\ a_i^* = \alpha_2' \mathbf{x}_i + \beta_2' \mathbf{h}_{2i,-t} + v_i \quad v_i \sim \mathcal{NID} (0, 1) \quad i = 1, \dots, n \\ \text{Corr} (u_i, v_i) = \rho. \end{array} \right. \quad (7)$$

Hay selección muestral en la ecuación de la probabilidad de empleo a través de factores inobservables si $E(u_i v_i) \neq 0$. Como se ha dicho anteriormente, la estimación se realiza de manera separada para los hombres y para las mujeres, y se emplean dos formas alternativas de introducir las variables de estudios no reglados en la ecuación de y_i^* . En cambio, las variables en la ecuación de a_i^* son siempre las mismas, así que se estiman dos modelos probit bivariantes para los hombres y dos para las mujeres.

El resultado más notable de la estimación del modelo (7) es que en ningún caso se rechaza que el coeficiente de correlación entre las perturbaciones de las dos ecuaciones sea cero (a un nivel del 5 %) y por lo tanto debemos concluir que el problema del sesgo de selección muestral no es significativo en nuestro caso. En consecuencia, no son mejores las propiedades estadísticas de los estimadores del modelo (7) respecto a los de los dos probits univariantes. A esto hay que añadir que, por los motivos comentados, la especificación de la ecuación de a_i^* es bastante insatisfactoria y en cualquier caso el punto principal de este trabajo no es analizar por qué unas personas realizan estudios ocupacionales y otras no, sino ver si aquellos que sí los realizan se colocan con mayor facilidad.

En definitiva, dado que se ha rechazado la hipótesis de selección muestral en la ecuación de la probabilidad de estar trabajando en la sexta encuesta, elegimos como especificación final el modelo univariante con heterocedasticidad. No obstante, atendiendo a la importancia que tienen los modelos que corrigen la selección muestral en la literatura que aborda los resultados de la formación, presentamos en las tablas 4 y 5 los coeficientes estimados de las variables de estudios no reglados ($\hat{\gamma}$) y los coeficientes de correlación estimados ($\hat{\rho}_{uv}$) en los cuatro modelos (7). Se puede comentar sobre ellos que para las mujeres la evidencia en favor de la importancia de haber hecho estudios no reglados durante la EPA como una vía exitosa para salir del paro es incluso más fuerte que la que se obtendrá en el modelo final. Así, en el modelo (7) estimado con las variables ficticias de **cuántos** estudios no reglados, todas estas variables afectan directamente a la probabilidad de empleo; en cambio en el modelo final dos de estas variables no son significativas. La explicación de esta diferencia se halla en que en este modelo probit bivalente se ha estimado $\hat{\rho} = -0,255$, lo que implica que es posible que el no tener en cuenta la selección muestral (como ocurre en el modelo univariante final) sesgue los estimadores de las variables de estudios ocupacionales a la baja. Este modelo (7) para las mujeres con las variables de cuántos estudios ocupacionales es en el que más difícil

Tabla 4.— Estimación MV del modelo bivalente con heterocedasticidad.
Alternativa cuántas.

VARIABLE	HOMBRES		MUJERES	
	Coeficiente	Estad. t	Coeficiente	Estad. t
cursando estudios no reglados durante...				
una encuesta de la EPA	0.418	1.198	0.723	2.238
dos encuestas de la EPA	0.678	2.098	1.05	3.208
tres encuestas de la EPA	0.525	1.459	0.836	2.431
cuatro encuestas de la EPA	0.46	1.248	1.08	3.202
cinco encuestas de la EPA	1.223	3.049	1.714	4.848
$\hat{\rho}_{uv}$	-0.099	-0.808	-0.255	-1.904

Tabla 5.— Estimación MV del modelo bivalente con heterocedasticidad.
Alternativa cuándo.

VARIABLE	HOMBRES		MUJERES	
	Coeficiente	Estad. t	Coeficiente	Estad. t
cursando estudios no reglados en...				
la primera encuesta de la EPA	0.018	0.14	0.33	2.446
la segunda encuesta de la EPA	0.141	1.134	0.221	1.718
la tercera encuesta de la EPA	0.116	0.953	0.308	2.285
la cuarta encuesta de la EPA	0.164	1.309	0.2	1.454
la quinta encuesta de la EPA	0.056	0.427	-0.006	-0.042
$\hat{\rho}_{uv}$	0.011	0.212	-0.071	-1.386

es rechazar la existencia de selección muestral, ya que el estadístico t para $H_0 : \rho = 0$ vale -1,904.

3 Estimación del modelo univariante bajo heterocedasticidad.

Como consecuencia de los resultados obtenidos en la sección anterior, hemos decidido que la mejor especificación del modelo para la probabilidad de haber encontrado empleo antes de la sexta entrevista es la compuesta por las ecuaciones (4)-(5), esto es, un modelo probit con heterocedasticidad multiplicativa. De acuerdo con él, la probabilidad de que un individuo

tenga un empleo en la sexta encuesta de la EPA es :

$$Pr(y_i = 1) = \Phi \left(\frac{\alpha' \mathbf{x}_i + \beta' \mathbf{h}_{i,-t} + \gamma' \mathbf{E}_i}{e^{\theta' \mathbf{w}_i}} \right). \quad (8)$$

Inicialmente las variables que componían los vectores \mathbf{x}_i , $\mathbf{h}_{i,-t}$, \mathbf{E}_i y \mathbf{w}_i eran las descritas en la sección 2. Después hemos ido reduciendo progresivamente la dimensión de estos vectores, atendiendo a la no significatividad de algunas variables. En la especificación final del vector \mathbf{w}_i no queda ninguna variable cuyo coeficiente estimado no sea significativo a un nivel del 5 %; de \mathbf{x}_i y de $\mathbf{h}_{i,-t}$ hemos eliminado algunas variables que no eran significativas en ninguno de los cuatro modelos estimados, como las que describían la **situación familiar** o el **tipo de empleo** que buscaban, pero hemos dejado otras, como por ejemplo **con subsidio** o **con experiencia**, a efectos de comparar los diferentes resultados que se obtienen según el sexo o por la importancia que se ha dado a estos factores en la literatura de la salida del paro. Finalmente, en ningún caso hemos intentado quitar del modelo estimado las variables relativas a los estudios aunque no fueran significativas.

Vamos a comenzar los comentarios de las estimaciones de nuestro modelo final presentando en forma de gráficos los resultados principales de este trabajo, que son los efectos¹⁸ que tiene la realización de estudios no reglados durante la EPA sobre la salida del paro.

Quizás sea prudente iniciar los comentarios de estos gráficos con una nota de cautela. Por un amplio margen, el efecto más fuerte sobre la probabilidad de empleo es el que consiguen tanto los hombres como las mujeres que han cursado estudios ocupacionales durante las cinco primeras encuestas, pero también es éste el efecto estimado con menor precisión debido a que hay pocas personas que hayan realizado estudios no reglados en las cinco primeras encuestas de la EPA y no continúen haciéndolos en la sexta.¹⁹ Como las variables ficticias de tener trabajo o no en las encuestas segunda a quinta forman parte del vector $\mathbf{h}_{i,-t}$, básicamente lo que miden los parámetros γ de la ecuación (4) son los incrementos en la probabilidad de encontrar un empleo en la sexta encuesta debido a la realización de estudios ocupacionales si se estaba sin trabajo en la quinta encuesta y los incrementos en la probabilidad de retener el empleo en la sexta encuesta si se estaba trabajando en la quinta encuesta. En nuestra muestra, de los 26 hombres que habían cursado estudios ocupacionales en las cinco primeras encuestas de la EPA, 9 encontraron empleo en la sexta encuesta y nadie lo perdió, y de 52 mujeres en esta situación 17 encontraron empleo en la sexta encuesta y ninguna lo perdió. La ausencia de transiciones en el sentido opuesto –esto es, del empleo al desempleo– entre estas personas explica el fuerte efecto estimado, pero su escaso número comparado con el tamaño de la muestra nos hace ser precavidos en cuanto a la interpretación económica del

¹⁸Denotemos $\mathbf{z}'_i \equiv [\mathbf{x}'_i \quad \mathbf{h}'_{i,-t} \quad \mathbf{E}'_i]$ y $\delta' \equiv [\alpha' \quad \beta' \quad \gamma']$. El efecto de una variable ficticia cualquiera X_j sobre la probabilidad de que y_i sea 1 en el modelo (8) se puede calcular como $\Phi \left(\frac{\delta' \bar{\mathbf{z}} - \delta_j \bar{x}_j + \delta_j}{e^{\theta' \bar{\mathbf{w}}}} \right) - \Phi \left(\frac{\delta' \bar{\mathbf{z}} - \delta_j \bar{x}_j}{e^{\theta' \bar{\mathbf{w}}}} \right)$ porque esta es la diferencia en la probabilidad de empleo, estimada en el punto de medias muestrales, entre que la variable ficticia valga 1 y que valga 0. Si X_j es una variable continua, su efecto se mide como $\phi \left(\frac{\delta' \bar{\mathbf{z}}}{e^{\theta' \bar{\mathbf{w}}}} \right) \cdot \frac{\delta_j}{e^{\theta' \bar{\mathbf{w}}}}$, ya que a esto es igual ahora la $\frac{\partial \widehat{Pr}(y_i=1)}{\partial x_{ji}}$.

¹⁹Recuérdese que a los que estaban estudiando en la sexta entrevista de la EPA los hemos quitado de la muestra bajo el razonamiento de que es posible que sus deseos o esfuerzos de cambiar de situación en el mercado de trabajo sean menores respecto a los que no hacían estudios en su entrevista final.

mismo.

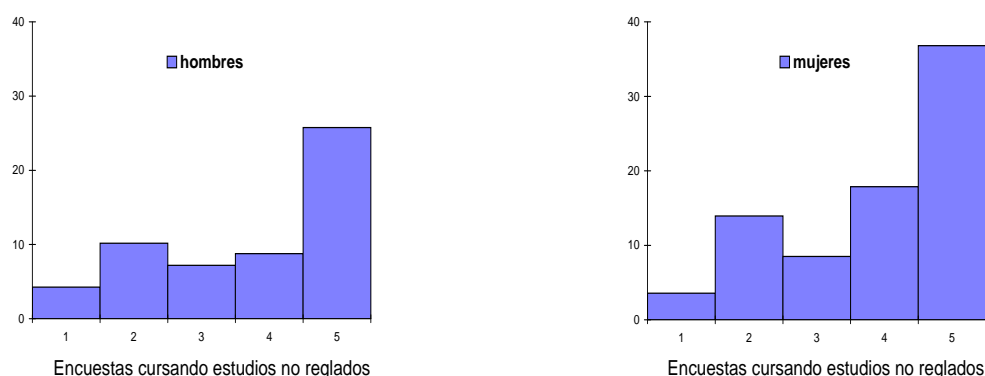


Gráfico 1: Incremento en la probabilidad de empleo, dependiendo de en CUANTOS trimestres de los 5 últimos se han realizado estudios no reglados (base: ninguno).

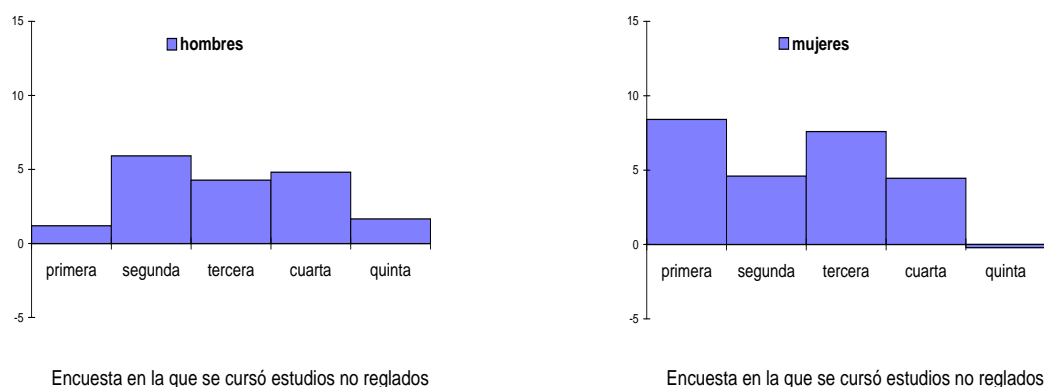


Gráfico 2: Incremento en la probabilidad de empleo, dependiendo de CUANDO se estuvo realizando estudios no reglados (base: nunca).

Es por estas razones por las que preferimos basar los comentarios de los efectos de realizar enseñanzas no regladas en los demás coeficientes estimados, ya que están obtenidos con un número de observaciones mayor. Comparando los gráficos anteriores, es fácil notar que las mujeres consiguen mayores beneficios en términos de empleo por la realización de estudios no reglados que los hombres. Este resultado concuerda con los que se han obtenido para otros países –ver por ejemplo la recopilación de efectos estimados que hacen Friedlander, Greenberg y Robins (1997) o LaLonde (1995)– que en general muestran que tanto en términos de probabilidad de salida del paro como de ganancias salariales una vez que logran un trabajo, las mujeres consiguen mejorar su situación laboral gracias a la participación en cursos de formación en mayor medida que los hombres. También se aprecia en nuestros gráficos que

la duración de los cursos de formación ocupacional puede ser un factor más importante que cuándo se hayan realizado (nótese que las escalas verticales no son iguales en los gráficos 1 y 2). Tanto para los hombres como para las mujeres, los cursos más breves, en el sentido de que únicamente se cursan durante una encuesta de la EPA, son los que menor incremento en la probabilidad de empleo proporcionan. En cambio, los individuos que realizan este tipo de estudios durante dos a cuatro encuestas de la EPA mejoran notablemente su probabilidad de empleo: entre un 7 y un 10% los hombres y entre un 8 y un 18% las mujeres. Este resultado de que cuanto más dure la formación no reglada mayores ganancias en términos de empleo se consiguen es lógico, ya que mayores conocimientos se adquirirán, y sin embargo ha sido sistemáticamente ignorado en la literatura sobre los efectos de la formación no reglada: en ella se tratan los diversos cursos como si fueran homogéneos y por tanto como si todos debieran aportar la misma ganancia.

Otro resultado esperado y que se manifiesta en el gráfico 2 es que, probablemente debido al poco tiempo que transcurre entre los dos sucesos, es poco o nulo el incremento en la probabilidad de salida del paro en la sexta encuesta que se consigue por el hecho de haber estado haciendo estudios ocupacionales en la quinta encuesta. También hemos querido encuadrar los efectos estimados que se muestran en el gráfico 2 dentro de la literatura relacionada con este tema. Comparados con los trabajos que utilizan un enfoque no experimental para realizar el tipo de estudio que nosotros estamos haciendo, los resultados que se muestran en el gráfico 2 se pueden catalogar de modestos. Como mostramos en la tabla 6, otros autores han obtenidos estimaciones más altas de los efectos sobre la probabilidad de obtener un empleo por haber participado en programas de formación. No obstante los partidarios del enfoque experimental suelen acoger con reservas efectos estimados tan grandes como los que obtienen Saez y Toledo (1996) y O'Higgins (1994) –véase al respecto LaLonde (1995, p. 158)– porque en la mayoría de las evaluaciones experimentales las ganancias de los programas de formación son más pequeñas y porque consideran que el rendimiento de la educación debe estar relacionado con la inversión realizada en ella y en este sentido es poco razonable que la formación no reglada proporcione rendimientos mucho más altos que los de las enseñanzas regladas.

Analizadas individualmente, hemos de concluir que para los hombres las variables ficticias de en qué entrevista hicieron estudios ocupacionales no afectan a su probabilidad de salida del paro, ya que ninguno de sus coeficientes resulta ser significativamente distinto de cero. Para las mujeres sí son significativos al nivel del 5% los coeficientes de haber realizado enseñanzas no regladas durante las encuestas primera y tercera de la EPA. Estos dos efectos estimados están en torno al 8%, lo cual nos parece una mejora importante en las perspectivas de empleo de las paradas que cursan estudios no reglados durante la EPA respecto a las que no lo hacen.

Los efectos de las enseñanzas regladas sobre la probabilidad de salida del paro también son marcadamente diferentes dependiendo del sexo. La categoría base, respecto a la que hacemos las comparaciones, es el tener la E.G.B. terminada. Respecto a ella, los hombres parados que no tienen estudios tienen una probabilidad de encontrar trabajo durante el periodo de seguimiento de la EPA que es un 10% menor que para los que han acabado la E.G.B.; en cambio las mujeres sin estudios parece que no tienen esta desventaja, ya que el efecto estimado es inferior al 1% y no resulta significativo. Para ninguno de los dos sexos

Tabla 6.— Efectos estimados de participar en programas de formación no reglada, por sexo (Incremento en la probabilidad de salida del paro).

Autor(es)	Hombres	Mujeres	Al estimar tienen en cuenta...	
			Sesgo de sel.	Heterocedasticidad
Card y Sullivan ^(a) A ^(b)	3.6%	—	sí	no
Card y Sullivan B ^(c)	8-8.5%	—	sí	no
O'Higgins ^(d)	21%	28%	sí	sí
Sáez y Toledo ^(e)	12.3%	14.1%	no	no
Míos (medias)	3.6%	5%	no significativo	sí

(a) Datos: EE.UU, Programa CETA 1976

(b) Método de emparejamiento de los historiales de empleo previos

(c) Modelo logit de efectos aleatorios

(d) Datos: Gran Bretaña, Programa YTS 1985

(e) Datos: España, Programa FIP 1994.

los efectos estimados de tener solamente estudios primarios muestran que ello implique una menor probabilidad de salida del paro que si se hubiera completado la E.G.B. En sentido contrario tampoco el incremento en la probabilidad de salida del paro por el hecho de haber completado BUP o COU es significativamente distinto de cero. Cuando miramos al efecto de tener terminada la FP-I ó la FP-II/III observamos de nuevo que es diferente para los hombres que para las mujeres. Para los hombres el efecto es gradual, en el sentido de que el aumento en la probabilidad de empleo es menor por haber hecho FP-I (un 3%) que por haber hecho FP-II o III (un 8%); la pauta de los efectos estimados de las mujeres es sorprendente, y no tenemos una explicación para ella. El resultado es que las mujeres con FP-I tienen una probabilidad un 6% mayor de encontrar empleo durante la EPA que las que han terminado la E.G.B., y en cambio las que han completado FP-II o FP-III no tienen ninguna ventaja sobre el grupo base. Según este resultado, cualquier mejora que proporcione el seguir estudiando tras la finalización de la FP-I debería ser por la vía de un salario mayor, y además este incremento salarial tendría que ser lo suficientemente alto como para compensar la menor probabilidad de empleo. Finalmente, el haber completado una diplomatura universitaria o 3 cursos de universidad implica un incremento en la probabilidad de empleo respecto a los que sólo tienen la E.G.B. terminada de un 9-10% para los hombres y de un 6% para las mujeres; el rendimiento si lo que se ha terminado es una licenciatura es levemente menor para los hombres (un 8%)²⁰ y bastante mayor para las mujeres. En líneas generales, se puede resumir diciendo que cuanto más alto sea el nivel educativo mayor es la probabilidad de salir del paro, con una penalización superior por no tener estudios en el caso de los hombres y con unas ganancias en términos de mejores perspectivas de empleo que son máximas para los diplomados universitarios y para las licenciadas universitarias.

Comparando los efectos estimados de las variables de estudios reglados y de las de estudios

²⁰Pero, en contraste con lo que ocurría para las mujeres con Formación Profesional terminada, el incremento en la probabilidad de empleo no es significativamente diferente entre los licenciados universitarios y los diplomados.

no reglados y teniendo en cuenta que la duración de los primeros es en general mucho mayor, se puede pensar que la realización de estudios no reglados es más provechosa en términos de facilitar la salida del paro. Nosotros creemos que este hecho se debe a que mediante los estudios ocupacionales se consigue una preparación específica más acorde con las demandas del mercado de trabajo: hay jóvenes que así obtienen un currículum más ajustado a lo que las empresas están buscando. Este comentario no quiere decir que los estudios ocupacionales sean un sustitutivo de los estudios técnicos o de los estudios universitarios: de hecho, buena parte de las personas que han realizado estudios no reglados ya tenían un título de Formación Profesional reglada (24%) ó universitario (25%).

Como término de esta sección comentaremos los efectos estimados de otras variables incluidas en la ecuación (4). Para empezar, en ningún caso tiene un efecto significativo el hecho de tener **formación complementaria**; no podemos considerar que este resultado sea demasiado extraño ya que, como comentamos al describir las variables del modelo, la definición que hace la EPA de la formación complementaria es muy estricta. Otra posibilidad que no podemos descartar pero tampoco confirmar con nuestros datos y que ayudaría a explicar este resultado es que la ventaja que se adquiera con esta formación complementaria no sea duradera; es decir, que o se aprovecha en los primeros meses tras la conclusión del curso o pierda su eficacia en cuanto a suponer una ayuda para salir del paro. Para avanzar en esta idea deberíamos conocer cuándo se realizó esa formación complementaria y no lo sabemos. Entrando a comentar los coeficientes estimados de las variables que recogen la situación de los individuos en el mercado de trabajo, hemos de decir que en general tienen los signos esperados aunque varias de ellas resultan ser no significativas. Así, el tener experiencia laboral previa se estima que aumenta la probabilidad de salida del paro en aproximadamente un 4%, tanto para los hombres²¹ como para las mujeres; es remarcable que este efecto estimado no sea más fuerte. Los coeficientes estimados de las variables **duración** y **duración al cuadrado**, que miden el tiempo que llevaban parados en la encuesta inicial, indican que hay una ‘dependencia de la duración positiva’, esto es, que cuanto más tiempo se haya permanecido en paro menor es la probabilidad de salir de esta situación. Por otro lado la variable ficticia **con subsidio** no es significativa en ningún caso, aunque si habría que decir algo sobre sus coeficientes estimados sería que los indicios con nuestra muestra de jóvenes son que para los hombres el haber estado cobrando subsidio o prestación por desempleo en la semana previa a la quinta encuesta no retrasa su salida del paro y en cambio para las mujeres sí es posible que haya un cierto efecto de desincentivo a la búsqueda de empleo. Finalmente las variables que indican si se ha utilizado algún **método medio** o algún **método activo** de búsqueda de empleo tampoco son significativas.

4 Conclusión.

En este artículo se han utilizado datos sobre jóvenes españoles que estaban parados en la primera entrevista de la EPA y se ha analizado su situación en el mercado de trabajo 5 trimestres más tarde. En especial, se ha estudiado cómo afectan a la probabilidad de obtener

²¹Aunque en este caso los coeficientes solamente son significativamente distintos de cero a un nivel del 10%.

un empleo el nivel educativo y el hecho de haber realizado estudios ocupacionales durante el período de seguimiento de la EPA. Al realizar las estimaciones de estos efectos, se ha tenido en cuenta en la medida que es posible hacerlo con la EPA la situación familiar de los jóvenes, su historial laboral previo y su intensidad de búsqueda de empleo.

Tras la estimación inicial del modelo especificado, que era un probit, se ha contrastado la posible presencia de problemas de heterocedasticidad y de selección muestral, habiéndose encontrado evidencia clara de la primera pero no de la segunda. Como consecuencia de ello, se ha estimado finalmente un modelo probit permitiendo la presencia de heterocedasticidad multiplicativa. Se ha hallado que realizar estudios ocupacionales mientras se está buscando empleo puede incrementar considerablemente la probabilidad de encontrarlo. El período durante el cual la EPA enlazada recoge datos sobre cada individuo es de 6 trimestres; dependiendo de cuántos de estos trimestres se estén cursando estudios ocupacionales, el aumento en la probabilidad de conseguir un empleo puede ser de un 26% para los hombres y un 37% para las mujeres si han hecho estos estudios en las cinco encuestas iniciales y de hasta un 9% para los hombres y hasta un 18% para las mujeres si los estuvieron haciendo durante un número de encuestas menor. Dadas las elevadas tasas de paro que padece España, estos son unos resultados muy destacables, y sería interesante estudiar si se confirman con una muestra en la que estén delimitados con mayor precisión qué tipos de estudios se realizan y cuánto duran en total, para poder establecer cuáles conducen más eficazmente a la salida del desempleo. Estos hallazgos también son importantes por su novedad: en la literatura empírica sobre los efectos de la formación apenas se ha prestado atención a la relación entre la duración de los cursos y su eficacia en términos de salida del paro.

También se ha encontrado que el grado en el que el nivel educativo afecta a la probabilidad de encontrar un empleo es distinto para los hombres que para las mujeres. Para los hombres, el hecho de no tener estudios reduce la probabilidad de encontrar un trabajo en un 10% respecto a los que tienen la E.G.B. terminada, mientras que las mujeres sin estudios parecen no tener esta desventaja. Los hombres tienen una probabilidad de encontrar trabajo durante la EPA mayor si han hecho F.P. II ó III que si solamente hicieron F.P. I, mientras que en las mujeres ocurre lo contrario. Entre los hombres los que tienen una mayor probabilidad de encontrar un empleo son los diplomados universitarios, y entre las mujeres son las licenciadas universitarias.

En definitiva, era bien conocido que en España los jóvenes con mayor nivel educativo son los que más probabilidad tienen de salir de las enormes bolsas de paro existentes. En este trabajo se ha puesto de manifiesto que completar la enseñanza reglada con cursos de formación no reglada es una manera eficaz de conseguir una preparación más específica a las exigencias del mercado de trabajo y gracias a ella se obtienen probabilidades de encontrar empleo notablemente más altas.

REFERENCIAS BIBLIOGRÁFICAS

Card, D. y D. Sullivan (1988), 'Measuring the Effect of Subsidized Training Programs on Movements in and out of Employment,' *Econometrica* 56, 497-530.

- Davidson, R. y J.G. MacKinnon (1984), 'Convenient Specification Tests for Logit and Probit Models,' *Journal of Econometrics* **25**, 241–262.
- Friedlander, D., D.H. Greenberg y P.K. Robins (1997), 'Evaluating Government Training Programs for the Economically Disadvantaged,' *Journal of Economic Literature* **35**, 1809–1885.
- García Pérez, J.I. (1996) 'Las Tasas de Salida del Empleo y del Desempleo en España (1978-1993),' *Tesina CEMFI n. 9601*.
- Greene, W.H. (1993), 'Econometric Analysis,' Second Edition. Englewood Cliffs: Prentice Hall.
- Gritz, R.M. (1993), 'The Impact of Training on the Frequency and Duration of Employment,' *Journal of Econometrics* **57**, 21-51.
- Heckman, J. y V. Joseph Hotz (1989), 'Choosing among Alternative Nonexperimental Methods for Estimating the Impact of Social Programs: The Case of Manpower Training,' *Journal of the American Statistical Association* **84**, 862–874.
- Kodde, D.A. (1988), 'Unemployment Expectations and Human Capital Formation,' *European Economic Review* **32**, 1645–1660.
- LaLonde, R.J. (1986), 'Evaluating the Econometric Evaluations of Training Programs with Experimental Data,' *American Economic Review* **76**, 604–620.
- LaLonde, R.J. (1995), 'The Promise of Public Sector-Sponsored Training Programs,' *Journal of Economics Perspectives* **9**, 149–168.
- Levhari, D. y Y. Weiss (1974), 'The Effect of Risk on the Investment in Human Capital,' *American Economic Review* **64**, 950–963.
- MacKinnon, J.G. y L. Magee (1990), 'Transforming the Dependent Variable in Regression Models,' *International Economic Review* **31**, 315–339.
- O'Higgins, Niall (1994), 'YTS, Employment, and Sample Selection Bias,' *Oxford Economic Papers* **46**, 605–628.
- Sáez, F. y M. Toledo (1996), 'Formación, Mercado de Trabajo y Empleo: una Evaluación,' *Economistas* **69**, 351–357.

Tabla 7.— Medias aritméticas de las variables de los modelos estimados

	HOMBRES			MUJERES		
	Todos	Con trabajo en su 6 ^a encuesta	Sin trabajo en su 6 ^a encuesta	Todas	Con trabajo en su 6 ^a encuesta	Sin trabajo en su 6 ^a encuesta
Con trabajo en la...						
segunda encuesta	0.21	0.345	0.103	0.129	0.289	0.053
tercera encuesta	0.312	0.511	0.152	0.201	0.437	0.09
cuarta encuesta	0.356	0.61	0.153	0.248	0.577	0.092
quinta encuesta	0.39	0.725	0.122	0.274	0.72	0.062
sexta encuesta	0.445	1	0	0.322	1	0
Rel. con el cabeza de familia						
cabeza de familia	0.169	0.201	0.143	0.027	0.031	0.025
cónyuge	0.006	0.007	0.006	0.258	0.201	0.285
hijo/-a	0.769	0.736	0.794	0.676	0.731	0.65
otras	0.056	0.056	0.057	0.039	0.037	0.04
soltero/-a	0.785	0.751	0.813	0.659	0.724	0.628
Edad en la encuesta inicial	25.12	25.11	25.12	24.91	24.32	25.19
Estudios terminados						
ninguno	0.038	0.023	0.05	0.018	0.012	0.021
primarios	0.271	0.251	0.287	0.161	0.129	0.176
EGB	0.44	0.44	0.44	0.437	0.413	0.45
BUP / COU	0.057	0.058	0.056	0.08	0.083	0.078
FP-I	0.076	0.084	0.069	0.098	0.103	0.095
FP-II o III	0.068	0.083	0.056	0.096	0.107	0.091
univ. 3 cursos	0.022	0.029	0.017	0.056	0.079	0.045
univ. > 3 cursos	0.028	0.032	0.025	0.054	0.074	0.044
Formación complementaria	0.028	0.031	0.026	0.052	0.053	0.051
n^o de encuestas anteriores en las que estaba cursando estudios no formales						
una	0.039	0.042	0.036	0.059	0.076	0.051
dos	0.024	0.029	0.021	0.031	0.045	0.024
tres	0.012	0.013	0.012	0.019	0.025	0.016
cuatro	0.007	0.007	0.006	0.01	0.013	0.009
cinco	0.003	0.004	0.003	0.007	0.011	0.005
n^o de encuesta en la que cursó estudios no formales						
primera	0.043	0.055	0.034	0.064	0.102	0.046
segunda	0.039	0.046	0.033	0.059	0.086	0.046
tercera	0.033	0.036	0.031	0.053	0.071	0.044
cuarta	0.031	0.032	0.03	0.044	0.052	0.04
quinta	0.021	0.018	0.023	0.033	0.034	0.033
Situación e historial laboral						
años en paro	1.56	1.38	1.71	2.06	1.7	2.22
con experiencia	0.797	0.829	0.772	0.662	0.709	0.639
con subsidio	0.126	0.078	0.165	0.085	0.038	0.107
Método de búsqueda						
pasivo(s)	0.383	0.386	0.379	0.403	0.372	0.418
medio	0.056	0.056	0.057	0.079	0.092	0.073
activo	0.561	0.558	0.564	0.518	0.536	0.509
Tipo de empleo que busca						
de cualquier tipo	0.604	0.578	0.624	0.613	0.593	0.622
prefiere jornada completa	0.279	0.284	0.275	0.305	0.305	0.305
sólo jornada completa	0.117	0.138	0.101	0.082	0.102	0.073
tasa de paro provincial	20.29	19.19	21.17	32.41	31.37	32.91
Tamaño muestral	7782	3460	4322	7182	2311	4871

Tabla 8.— Estimación MV del modelo (4)-(5). HOMBRES: Alternativa cuántas.

VARIABLE	Coeficiente	Estad. t	Efecto Marginal	Varianza	
				Coeficiente	Estad. t
Edad en la encuesta inicial	-0.008	-1.447	-0.003		
Soltero	-0.247	-4.035	-0.082		
Estudios terminados					
ninguno	-0.308	-2.524	-0.099		
primarios	-0.066	-1.299	-0.022		
EGB		(referencia)			
BUP / COU	0.051	0.522	0.017		
FP-I	0.099	1.31	0.033	-0.132	-2.01
FP-II o III	0.233	2.634	0.077		
univ. 3 cursos	0.286	2.103	0.095		
univ. > 3 cursos	0.246	2.182	0.081	-0.395	-3.964
Formación complementaria	-0.064	-0.522	-0.021		
n^o de encuestas anteriores					
en las que estaba cursando					
estudios no reglados					
una	0.128	1.261	0.042		
dos	0.308	2.639	0.102		
tres	0.218	1.269	0.072		
cuatro	0.265	1.272	0.088		
cinco	0.814	2.944	0.258		
Situación e historial laboral					
años en paro	-0.051	-3.109	-0.02		
(años en paro) ²	0.001	2.687	0.0003		
con experiencia	0.116	1.836	0.038		
con subsidio	0.062	0.963	0.02		
tasa de paro provincial	-0.021	-5.702	-0.008	0.006	2.307
con trabajo en t=2	0.591	6.633	0.194	0.372	6.211
con trabajo en t=3	0.333	5.786	0.11		
con trabajo en t=4	0.395	6.922	0.13		
con trabajo en t=5	1.703	15.628	0.52		
constante	-0.399	-2.314			
- ln función de verosimilitud			3510.25		
- ln f. de verosimilitud restringida			5339.46		
% de predicciones correctas			81.09		
Tamaño muestral			7772		

Tabla 9.— Estimación MV del modelo (4)-(5). HOMBRES: Alternativa cuándo.

VARIABLE	Coeficiente	Estad. t	Efecto Marginal	Varianza	
				Coeficiente	Estad. t
Edad en la encuesta inicial	-0.007	-1.49	-0.003		
Soltero	-0.215	-4.166	-0.081		
Estudios terminados					
ninguno	-0.269	-2.605	-0.099		
primarios	-0.063	-1.423	-0.024		
EGB		(referencia)			
BUP / COU	0.037	0.432	0.014		
FP-I	0.09	1.359	0.034	-0.145	-2.204
FP-II o III	0.204	2.64	0.077		
univ. 3 cursos	0.254	2.082	0.096		
univ. > 3 cursos	0.222	2.247	0.084	-0.437	-4.254
Formación complementaria	-0.044	-0.411	-0.017		
nº de encuesta en la que cursó estudios no reglados					
primera	0.032	0.303	0.012		
segunda	0.156	1.374	0.059		
tercera	0.113	1.047	0.043		
cuarta	0.127	1.083	0.048		
quinta	0.044	0.35	0.017		
Situación e historial laboral					
años en paro	-0.044	-3.183	-0.018		
(años en paro) ²	0.001	2.711	0.0003		
con experiencia	0.102	1.839	0.038		
con subsidio	0.056	1.008	0.021		
tasa de paro provincial	-0.016	-6.291	-0.006		
con trabajo en t=2	0.514	7.294	0.193	0.352	5.967
con trabajo en t=3	0.289	6.087	0.109		
con trabajo en t=4	0.356	7.622	0.134		
con trabajo en t=5	1.488	31.696	0.52		
constante	-0.378	-2.51			
- ln función de verosimilitud			3514.01		
- ln f. de verosimilitud restringida			5339.46		
% de predicciones correctas			81.05		
Tamaño muestral			7772		

Tabla 10.— Estimación MV del modelo (4)-(5). MUJERES: Alternativa cuántas.

VARIABLE	Coeficiente	Estad. t	Efecto Marginal	Varianza	
				Coeficiente	Estad. t
Edad en la encuesta inicial	-0.006	-1.088	-0.002		
Estudios terminados					
ninguno	-0.31	-0.169	-0.008		
primarios	0.007	0.089	0.002		
EGB		(referencia)			
BUP / COU	0.084	0.855	0.023	-0.233	-3.469
FP-I	0.221	2.475	0.062		
FP-II o III	0.013	0.134	0.003		
univ. 3 cursos	0.213	1.804	0.06		
univ. > 3 cursos	0.387	3.18	0.111		
Formación complementaria	-0.074	-0.513	-0.02	0.213	2.196
nº de encuestas anteriores en las que estaba cursando estudios no reglados					
una	0.129	1.239	0.036		
dos	0.48	3.263	0.139		
tres	0.298	1.593	0.085		
cuatro	0.607	2.749	0.179		
cinco	1.242	5.038	0.368		
Situación e historial laboral					
años en paro	-0.038	-2.3	-0.012		
(años en paro) ²	0.0005	1.446	0.0002		
con experiencia	0.164	2.563	0.044		
con subsidio	-0.156	-1.568	-0.041		
tasa de paro provincial	-0.01	-2.607	-0.003	0.006	2.402
Método de búsqueda					
pasivo(s)		(referencia)			
medio	0.114	1.152	0.031		
activo	0.089	1.622	0.024		
con trabajo en t=2	1.045	7.487	0.307	0.369	5.126
con trabajo en t=3	0.217	2.694	0.06		
con trabajo en t=4	0.564	6.349	0.16	0.134	2.36
con trabajo en t=5	2.356	11.922	0.635		
constante	-1.295	-7.337			
- ln función de verosimilitud			2566.28		
- ln f. de verosimilitud restringida			4504.91		
% de predicciones correctas			86.6		
Tamaño muestral			7174		

Tabla 11.— Estimación MV del modelo (4)-(5). MUJERES: Alternativa cuándo.

VARIABLE	Coeficiente	Estad. t	Efecto Marginal	Varianza	
				Coeficiente	Estad. t
Edad en la encuesta inicial	-0.007	-1.097	-0.002		
Estudios terminados					
ninguno	-0.027	-0.15	-0.007		
primarios	0.008	0.102	0.002		
EGB		(referencia)			
BUP / COU	0.075	0.761	0.021	-0.229	-3.342
FP-I	0.219	2.454	0.061		
FP-II o III	0.011	0.117	0.003		
univ. 3 cursos	0.196	1.674	0.055		
univ. > 3 cursos	0.384	3.143	0.11		
Formación complementaria	-0.091	-0.638	-0.024	0.214	2.203
n^o de encuesta en la que cursó estudios no reglados					
primera	0.296	2.28	0.084		
segunda	0.165	1.282	0.046		
tercera	0.268	1.984	0.076		
cuarta	0.16	1.161	0.045		
quinta	-0.008	-0.052	-0.002		
Situación e historial laboral					
años en paro	-0.039	-2.393	-0.013		
(años en paro) ²	0.001	1.526	0.0002		
con experiencia	0.164	2.557	0.044		
con subsidio	-0.151	-1.523	-0.04		
tasa de paro provincial	-0.01	-2.615	-0.003	0.006	2.38
Método de búsqueda					
pasivo(s)		(referencia)			
medio	0.106	1.071	0.029		
activo	0.085	1.562	0.023		
con trabajo en t=2	1.046	7.465	0.308	0.371	5.135
con trabajo en t=3	0.219	2.708	0.061		
con trabajo en t=4	0.56	6.308	0.159	0.135	2.362
con trabajo en t=5	2.348	11.819	0.633		
constante	-1.286	-7.288			
- ln función de verosimilitud			2567.82		
- ln f. de verosimilitud restringida			4504.91		
% de predicciones correctas			86.53		
Tamaño muestral			7174		