

Documento de trabajo

Enlace de las series de paro 1976-2000 según la definición EPA-2002

Agosto 2005

Subdirección de Estadísticas del Mercado Laboral

Documento preparado por:

Javier Trejo Malfaz

Lourdes Ortega Núñez

Colaboración de:

Esperanza Gil Moraleda

Asesoramiento técnico de:

Ramiro López Paños

Florentina Álvarez Álvarez

Miguel Ángel García Martínez

Índice

Resumen	3
Introducción	3
Líneas de análisis	5
Procedimiento adoptado: Regresión con variable respuesta binaria. Modelo de regresión probit	7

1 Resumen

Según el *Acuerdo de la Comisión Delegada del Gobierno para Asuntos Económicos sobre mejoras en la transparencia en el ámbito de la información económica y estadística proporcionada por el Gobierno*, el Instituto Nacional de Estadística *procederá al enlace histórico de todas las series cuya base metodológica haya cambiado, con el fin de preservar la información histórica.*

En este documento se describe el procedimiento utilizado para el enlace de series de paro que salva la discontinuidad producida en dichas series por el cambio de definición ocurrido en el año 2002.

El enlace se extiende desde el tercer trimestre de 1976 hasta el cuarto trimestre de 2000 y se ha realizado para los principales agregados (paro por sexo y edad menor de 25 años ó 25 años y más) a nivel nacional y por comunidades autónomas.

En primer lugar se hace una introducción al problema de la ruptura de series de paro al que se aplica el enlace, en la que se detallan las causas y la naturaleza del cambio de definición. A continuación se enumeran los métodos explorados para salvar la discontinuidad en las series, con las ventajas e inconvenientes de cada uno. Finalmente, se explica más extensamente el procedimiento elegido y se muestran sus resultados.

2 Introducción

La Encuesta de Población Activa (EPA), además de ser una de las fuentes estadísticas más importantes sobre el mercado laboral en España, es la encuesta que proporciona datos armonizados sobre empleo y paro para la Unión Europea.

En este sentido, está regulada por varios reglamentos, fundamentalmente por el Nº 577/98 del Consejo, de 9 de marzo de 1998, y por varios de la Comisión que desarrollan el anterior. Los cambios que, a instancias de la Oficina Estadística de la Unión Europea (Eurostat), se introducen en la normativa estadística de la Unión han de aplicarse también en la EPA.

En 2002 se modificó la definición operativa de paro utilizada hasta entonces en la EPA, precisamente para adaptarse a la normativa europea.

En la EPA se considera paradas a las personas de 16 o más años que están sin trabajo, disponibles para trabajar y buscando activamente empleo. Son también parados quienes no tienen trabajo y no lo buscan porque ya han encontrado uno, siempre que estén disponibles para incorporarse a él. La encuesta utiliza la definición genérica de paro de la Organización Internacional del Trabajo.

La aplicación concreta de esta definición conlleva una serie de convenciones adicionales que es necesario tomar en cuenta, como por ejemplo, delimitar un 'plazo de disponibilidad' para trabajar (la persona debe estar disponible para traba-

jar en un plazo de dos semanas siguientes a la de referencia de la encuesta) o lo que se considera una 'búsqueda activa de empleo'.

En aras de una mayor armonización de las cifras de empleo y paro en la UE, en el Reglamento (CE) N° 1897/2000 de la Comisión, de 7 de septiembre de 2000 se establecen (además de otros requisitos relacionados con el orden de las preguntas, tratamiento de colectivos especiales, etc.) normas prácticas para la aplicación concreta de las condiciones para ser considerado parado en la Unión Europea. Estas condiciones difieren ligeramente de las que se venían considerando hasta el año 2002. En particular, según lo enunciado en el Anexo I punto 1 del Reglamento se establecen como únicos métodos de búsqueda de empleo los siguientes:

- Estar en contacto con una oficina pública de empleo con el fin de encontrar trabajo, cualquiera que sea la parte -la persona o la oficina- que haya tomado la iniciativa (la renovación de la inscripción por razones puramente administrativas no constituye un planteamiento activo).
- Estar en contacto con una oficina privada (oficina de empleo temporal, empresa especializada en contratación, ...) con el fin de encontrar trabajo.
- Enviar una candidatura directamente a los empleadores.
- Indagar a través de relaciones personales, por mediación de sindicatos, ...
- Anunciarse o responder a anuncios de periódicos.
- Estudiar las ofertas de empleo.
- Participar en una prueba, concurso o entrevista, en el marco de un procedimiento de contratación.
- Buscar terrenos, locales o material.
- Realizar gestiones para obtener permisos, licencias o recursos financieros.

En el caso de España, la condición de que el contacto con la oficina de empleo deba realizarse con el fin de encontrar trabajo, no considerándose como método activo la mera inscripción por razones administrativas, tiene un considerable impacto en las cifras y ha tenido consecuencias apreciables en la medida del nivel de paro dada por la EPA.

A lo largo del año 2001 se modificó el cuestionario y se recogió información en campo de forma que era posible calcular las cifras de paro tomando en cuenta la definición operativa del Reg. 1897/2000 y la anterior. Los datos de este año sirvieron como enlace entre la antigua y nueva definición.

Los cuestionarios anteriores a 2001 no recogían este matiz sobre si el contacto con la oficina de empleo se produjo con el fin de encontrar trabajo o por otras razones, por lo que no es posible determinar mediante las variables de la encuesta, para períodos anteriores a 2001, qué cifras de paro hubiera dado la EPA de utilizar la nueva definición. Sí es posible, por el contrario, obtener para 2001-2004 el dato de paro tanto con la antigua como con la nueva (y oficial) definición.

Las personas paradas según la antigua definición que no lo son según la nueva pasan a engrosar la categoría de inactivos, ya que no cumplen la condición ope-

rativa establecida de 'búsqueda activa' de empleo. El efecto del cambio de definición, por tanto, es una disminución del número de parados, de la tasa de paro y de la tasa de actividad.

3 Líneas de análisis

Para los años anteriores a 2001, se trata de asignar a los parados considerados históricamente como tales (de acuerdo a la definición de paro aplicada en su momento), la situación de parado o inactivo según el nuevo criterio.

Del análisis del comportamiento de ambas definiciones de desempleo en el periodo 2001-2004 y teniendo en cuenta la especificidad registrada en las distintas comunidades autónomas, se han considerado las mismas de forma independiente.

Respecto a otras variables explicativas que puedan influir en la determinación de las diferencias entre la nueva definición de paro y la anterior, se han considerado sólo las principales, como son el sexo y la distinción de mayor o menor de 25 años (límite usual para determinar el paro juvenil). No se podía descender a más detalle sin riesgo de que la escasez de muestra invalidara los resultados en alguno de los dominios.

Inicialmente, para los métodos ensayados se utilizó el año 1987 como punto de partida (es decir, asumiendo que en 1987 ambas definiciones no diferían) y desagregando por comunidad autónoma y sexo. La elección de esta fecha vino determinada por dos factores. Por un lado, su proximidad a 1985, año en que se publicó la Orden de 11 de marzo del Ministerio de Trabajo y Seguridad Social, que establece las definiciones a considerar en relación con las demandas de empleo en las oficinas públicas del INEM (actual Servicio Estatal de Empleo) y que, cabe suponer, influyó en lo concerniente al comportamiento de los individuos en relación con las oficinas públicas de empleo. Por otra parte, en 1987 se introdujo un importante cambio metodológico y de cuestionario en la encuesta, que permitió medir más profundamente el fenómeno del paro y que, en lo sustancial, ha permanecido inalterada hasta 2004.

Atendiendo las recomendaciones del Grupo de Trabajo sobre Estadísticas Coyunturales del Mercado Laboral (junio de 2005) se decidió ampliar el periodo de inicio hasta 1976¹ e incluir como factor discriminador el grupo de edad (menor de 25 años y 25 años y más). A continuación se enumeran los métodos ensayados en el estudio:

- Métodos de clasificación. Análisis discriminante con regresión logística binaria.

El Análisis Discriminante nos permite analizar si existen diferencias significativas entre grupos de individuos respecto a un conjunto de variables medidas sobre los mismos para, en el caso de que existan, explicar en qué sentido se dan y

¹ Salvo para el conjunto de Ceuta y Melilla cuya información sólo está disponible desde 1998.

proporcionar procedimientos de clasificación sistemática de nuevas observaciones de origen desconocido en uno de los grupos analizados.

En nuestro estudio se intenta clasificar a los individuos en dos grupos distintos según su situación laboral con ambas definiciones de paro (antigua y nueva). Estos dos grupos serán:

- ✓ Ser parado con la definición nueva de paro siéndolo también con la definición antigua.
- ✓ Ser inactivo con la definición nueva de paro siendo parado con la definición antigua.

Esta clasificación la obtenemos a partir de las encuestas de los años 2001 a 2004 donde, con la información disponible de cada individuo, se puede determinar su situación laboral tanto con la antigua como con la nueva definición de paro.

La aplicación de este método es imposible antes de 1999 ya que el cuestionario no contenía suficiente información para ello. En efecto, la variable que recoge la información sobre cuándo se produjo el último contacto con la oficina de empleo (variable CONTACT), en el caso de que el método de búsqueda utilizado sea la inscripción en la Oficina de Empleo de la Administración, no está disponible en el cuestionario EPA administrado en el periodo 1992-1998. Esta variable es fundamental para distinguir operativamente la antigua y la nueva definición de paro.

- Modelos Multivariantes o Econométricos.

Tratar de explicar el comportamiento del paro (nueva definición) en función de la evolución de variables que se consideran explicativas (variables causales del fenómeno) a través de modelos econométricos podría aportar resultados eficientes. La principal variable explicativa sería la caracterización del contacto con la oficina de empleo, como activo o no. Sin embargo, en estos modelos, cuando se desea realizar estimaciones, el desconocimiento de los valores de las variables explicativas determina la necesidad de utilizar predicciones para éstas.

En nuestro caso, no se ha optado por estos modelos ante la dificultad o imposibilidad de conocer los valores de las magnitudes que aparecerían como exógenas. Además, dado que el estudio se realiza por comunidades autónomas y debido a que la cuantía de la muestra de la EPA que proporciona los datos básicos es reducida en aquellas comunidades de menor población, el método nos llevaría, a causa de los errores de muestreo, a estimaciones '*poco significativas*' sobre todo en estas comunidades menos representadas.

- Modelos estocásticos de Series Temporales. Modelo de Box-Jenkins.

Cuando el objeto del estudio es exclusivamente la '*predicción*', como es nuestro caso, no siempre es necesario que se especifique un modelo causal en el que la variable explicada se exprese en función de un conjunto de variables explicativas, ya que en muchos casos se pueden obtener predicciones satisfactorias mediante modelos univariantes o de series de tiempo. Además, cuando se espera que no existan bruscas alteraciones respecto al comportamiento actual de la va-

riable, estos métodos pueden proporcionar buenas predicciones. Entre las técnicas más rigurosas para la predicción univariante se encuentran los modelos Box-Jenkins.

Para cada uno de los colectivos analizados se estudió el efecto del cambio de la definición de paro introducida en el año 2002 mediante modelos ARIMA combinados con modelos de intervención.

Una vez ajustado el modelo se recalculó la nueva serie de paro, estimada para cada sexo, incorporando el efecto de la nueva definición de paro en cada comunidad autónoma. El total nacional se obtiene por agregación de todas ellas.

Este método proporciona un buen ajuste del modelo con respecto a la serie de paro histórica. Sin embargo, las variables demográficas relacionadas con la situación laboral de la población se ven afectadas por factores externos muy difíciles de cuantificar como pueden ser los ciclos económicos, políticas de empleo o el fenómeno de la inmigración. Estos factores repercuten en la variabilidad del modelo, aumentando los errores de predicción y suavizando el efecto del cambio de definición de paro, esto es, disminuye de forma cuantitativa la estimación de la intervención. Todo ello repercute en la predicción proporcionando, en algunos trimestres, estimaciones para el número de parados con la definición nueva mayores que con la definición antigua, lo que resulta conceptualmente imposible.

- Regresión con variable respuesta binaria. Modelos de regresión probit.

Se pueden utilizar también modelos de regresión para predecir la respuesta binaria de un individuo, del que se conocen unas características medibles. Precisamente en el problema que nos ocupa la variable respuesta es binaria (la variable toma los valores 1 ó 0 para clasificar a los individuos como parados o no). Uno de los modelos de regresión más utilizados es el probit, que toma como referencia la distribución normal.

Tras evaluar las diferentes vías de actuación para la resolución del problema de la estimación de las series de paro retrospectivas con la nueva definición de paro, se decidió optar por este procedimiento, que se describe detalladamente en el punto siguiente.

4 Procedimiento adoptado: Regresión con variable respuesta binaria. Modelos de regresión probit.

➤ Información de partida disponible

Como datos para la estimación de las probabilidades de permanencia en la clasificación de parado según la antigua y nueva definición se han considerado los registros de los individuos parados según la definición antigua sobre la población de 16 y más años para el periodo comprendido entre el primer trimestre de 2001 y cuarto trimestre de 2004 (periodo del que disponemos información con ambas definiciones).

Una vez obtenidas estas probabilidades se han aplicado a la serie histórica de paro con el objeto de determinar una estimación del número de parados según la definición de paro implantada en el año 2002. Debido a que tanto la metodología y elaboración de la EPA como diferentes factores que influyen en el mercado laboral español y otros aspectos sociales y situaciones económicas han tenido cambios a lo largo del tiempo, se considerará que el efecto de esta nueva definición irá desapareciendo conforme nos vamos alejando de 2001.

➤ **Descripción del procedimiento**

Manteniendo una continuidad con los modelos anteriores, el estudio se ha realizado desagregando los datos por sexo y comunidad autónoma. De nuevo el total nacional se obtiene por agregación de todas ellas. Hay que advertir, no obstante, que los resultados de la estimación de la serie de parados obtenida por agregación de las comunidades autónomas y la obtenida aplicando el método directamente al total nacional son similares.

A su vez, este procedimiento permite realizar planteamientos diferentes que resumiremos a continuación.

El primero se basa en obtener las probabilidades de permanencia en la clasificación de parado según la antigua y nueva definición (ser parado con la definición nueva de paro siéndolo también con la definición antigua) para cada uno de los trimestres del período 2001-2004.

Seguidamente se modeliza matemáticamente la curva que describe la tendencia temporal de estas probabilidades, para cada grupo de colectivos considerados, por regresión respecto al tiempo. Las probabilidades de permanencia trimestrales anteriores al año 2000 se calculan por extrapolación a partir de estos modelos. La estimación del número de parados se realiza combinando estas probabilidades y las series históricas de paro.

Los modelos utilizados fueron, entre otros, de tipo lineal, polinómico, potencial, logarítmico o exponencial. También se estudiaron modelos con *variables retardadas* (pese a su nombre, referidas al valor de la probabilidad de transición de trimestres posteriores) similares a los empleados en el cálculo de las primeras proyecciones de tasa de actividad que publicó el INE en 1995. En todos los casos, la estimación de los parámetros de las ecuaciones se efectuó mediante el método de los mínimos cuadrados.

Dado que el periodo de tiempo para el que disponíamos de información era escaso (cuatro años) en comparación con el periodo a estimar se consideraron más adecuados otros procedimientos que permitieran elaborar estimaciones más robustas a largo plazo.

En un segundo enfoque se trató de obtener la probabilidad de permanencia en el estado de parado sin hacer un estudio temporal de las observaciones.

Una vez obtenida esta probabilidad mediante una regresión probit, se procedió a ajustarle una tendencia de modo que para la primera observación esta probabilidad fuera uno (finalmente el segundo trimestre de 1976), y en el último trimes-

tre de 2000 fuera el probit obtenido. Las tendencias analizadas fueron de tipo lineal y cuadrática.

El procedimiento de ajuste de una tendencia lineal consiste en modelizar la recta que pasa por los dos puntos de los que disponemos (correspondientes con el primer trimestre de la serie histórica, cuya probabilidad es uno, y el trimestre correspondiente al cuarto trimestre de 2000, cuya probabilidad es la obtenida mediante el modelo de regresión probit).

Para ajustar una tendencia cuadrática se precisaba un punto adicional que sólo se podía obtener en el periodo 2001 a 2004, resultando muy próximo al del final de la serie. Al ajustar una curva polinómica de grado dos los resultados obtenidos con la tendencia cuadrática no permitían un tratamiento general satisfactorio.

También se consideró tomar en cuenta las posibles variaciones estacionales y se analizó ajustar la tendencia en función del trimestre dentro del año al que pertenecen las observaciones obteniéndose, por tanto, una probabilidad diferente para cada uno de los cuatro trimestres. Sin embargo, al distinguir por trimestres, la cuantía de la muestra de la EPA que proporciona los datos básicos era reducida en aquellas comunidades de menor población, afectando a las estimaciones de los parámetros, no existiendo por otro lado diferencias importantes para las comunidades grandes respecto a lo obtenido prescindiendo de las variaciones estacionales. Por esta razón se decidió calcular un único probit común para los cuatro trimestres.

Todas las consideraciones expuestas anteriormente nos han llevado a optar finalmente por utilizar un modelo de regresión probit corregido mediante una tendencia lineal.

En esencia, el procedimiento seguido consta de las siguientes fases:

1. Cálculo de las probabilidades de transición de parado con la antigua definición a parado con la nueva definición mediante un modelo de regresión probit, considerando para ello como variable endógena una variable binaria que mide si un individuo es parado con ambas definiciones o no, y como variables exógenas el sexo, grupo de edad (menor de 25 ó 25 y más años) y la interacción entre ambas, ponderando por el factor de elevación.
2. Una vez obtenidas estas probabilidades de transición, se les aplica una tendencia lineal de forma que en el tercer trimestre del año 1976 haya desaparecido totalmente el efecto de este cambio de definición.
3. A partir de aquí, aplicamos las probabilidades ajustadas de forma lineal a la serie histórica de parados.

Los modelos obtenidos con este procedimiento, tanto para las comunidades autónomas como para el total nacional se presentan a continuación. Se incluye el modelo teórico del total nacional ya que, a pesar de que sus resultados se han obtenido por agregación de todas las comunidades autónomas de las que se compone, es interesante modelizarlo para poder comparar su comportamiento con el de las comunidades.

El modelo es del tipo

$$p = P(Y = 1/\text{sexo}, \text{edad}) = \Phi(\beta_0 + \beta_1 \cdot \text{sexo} + \beta_2 \cdot \text{edad} + \beta_{12} \cdot \text{int})$$

donde denotamos por

$$Y = \begin{cases} 1 & \text{si parado con ambas definiciones} \\ 0 & \text{si no} \end{cases}$$

$$\text{edad} = \begin{cases} 1 & \text{si} < 25 \text{ años} \\ 0 & \text{si} \geq 25 \text{ años} \end{cases}$$

$$\text{sexo} = \begin{cases} 1 & \text{si varón} \\ 0 & \text{si mujer} \end{cases}$$

$$\text{int} = \text{edad} \cdot \text{sexo} = \begin{cases} 1 & \text{si varón} < 25 \text{ años} \\ 0 & \text{en otro caso} \end{cases}$$

$$p = P(Y = 1/\text{sexo}, \text{edad})$$

$$\Phi(\cdot) = \text{Función de distribución de una } N(0,1)$$

Los modelos ajustados para el total nacional y las comunidades autónomas son los siguientes:

1. Total nacional:

$$p = \Phi(0.8814 + 0.1980 \cdot \text{sexo} + 0.2099 \cdot \text{edad} - 0.1439 \cdot \text{int})$$

2. Andalucía:

$$p = \Phi(0.9293 + 0.2796 \cdot \text{sexo} + 0.1557 \cdot \text{edad} - 0.2420 \cdot \text{int})$$

3. Aragón:

$$p = \Phi(0.4767 + 0.3464 \cdot \text{sexo} + 0.2413 \cdot \text{edad} - 0.1555 \cdot \text{int})$$

4. Asturias:

$$p = \Phi(0.5064 - 0.0214 \cdot \text{sexo} + 0.1348 \cdot \text{edad} - 0.0159 \cdot \text{int})$$

5. Baleares:

$$p = \Phi(1.2164 + 0.2549 \cdot \text{sexo} - 0.1422 \cdot \text{edad} - 0.1729 \cdot \text{int})$$

6. Canarias:

$$p = \Phi(0.9877 + 0.0042 \cdot \textit{sexo} + 0.2400 \cdot \textit{edad} - 0.0606 \cdot \textit{int})$$

7. Cantabria:

$$p = \Phi(0.8439 + 0.0614 \cdot \textit{sexo} + 0.1411 \cdot \textit{edad} - 0.1244 \cdot \textit{int})$$

8. Castilla – La Mancha:

$$p = \Phi(0.5847 + 0.1796 \cdot \textit{sexo} + 0.2603 \cdot \textit{edad} - 0.2214 \cdot \textit{int})$$

9. Castilla y León:

$$p = \Phi(0.8079 + 0.1838 \cdot \textit{sexo} + 0.2393 \cdot \textit{edad} - 0.1982 \cdot \textit{int})$$

10. Cataluña:

$$p = \Phi(1.9164 + 0.0299 \cdot \textit{sexo} + 0.0813 \cdot \textit{edad} + 0.1827 \cdot \textit{int})$$

11. Comunidad Valenciana:

$$p = \Phi(1.0515 + 0.3167 \cdot \textit{sexo} + 0.2786 \cdot \textit{edad} - 0.1830 \cdot \textit{int})$$

12. Extremadura:

$$p = \Phi(0.2660 + 0.3525 \cdot \textit{sexo} + 0.2351 \cdot \textit{edad} - 0.3022 \cdot \textit{int})$$

13. Galicia:

$$p = \Phi(0.9522 + 0.0935 \cdot \textit{sexo} + 0.2403 \cdot \textit{edad} - 0.2689 \cdot \textit{int})$$

14. Madrid:

$$p = \Phi(0.5429 + 0.0524 \cdot \textit{sexo} + 0.2332 \cdot \textit{edad} + 0.0060 \cdot \textit{int})$$

15. Murcia:

$$p = \Phi(1.1251 + 0.3664 \cdot \textit{sexo} + 0.1863 \cdot \textit{edad} - 0.2417 \cdot \textit{int})$$

16. Navarra:

$$p = \Phi(0.9591 + 0.1577 \cdot \textit{sexo} + 0.3882 \cdot \textit{edad} - 0.1288 \cdot \textit{int})$$

17. País Vasco:

$$p = \Phi(1.1264 + 0.1955 \cdot \textit{sexo} + 0.1961 \cdot \textit{edad} - 0.0252 \cdot \textit{int})$$

18. Rioja:

$$p = \Phi(0.4835 + 0.3708 \cdot \textit{sexo} + 0.1234 \cdot \textit{edad} - 0.2268 \cdot \textit{int})$$

19. Ceuta y Melilla:

$$p = \Phi(-0.5076 + 0.3515 \cdot \textit{sexo} + 0.0873 \cdot \textit{edad} - 0.3648 \cdot \textit{int})$$

Se presentan a continuación tres tablas. En la Tabla 1 aparecen las probabilidades marginales por sexo y menores de 25 años o de 25 años y más. En las dos siguientes se muestran las probabilidades de permanencia en el estado de paro considerando conjuntamente ambas variables, en la Tabla 2 sin introducir el efecto de la interacción y en la Tabla 3 introduciéndola.

Tabla 1.

	Varones	Mujeres	< 25 años	>= 25 años
Total nacional	0.86365	0.82242	0.86790	0.83087
Andalucía	0.88222	0.83241	0.86482	0.85108
Aragón	0.80148	0.69977	0.78858	0.72263
Asturias (Principado de)	0.69641	0.70304	0.73358	0.69068
Balears (Illes)	0.92807	0.89424	0.91921	0.90623
Canarias	0.85009	0.85028	0.88484	0.83880
Cantabria	0.81844	0.80751	0.82974	0.80735
Castilla y León	0.84197	0.80344	0.85108	0.80727
Castilla - La Mancha	0.78084	0.73857	0.79593	0.74042
Cataluña	0.97805	0.97346	0.98202	0.97313
Comunidad Valenciana	0.91822	0.86705	0.91745	0.87923
Extremadura	0.72706	0.62104	0.69948	0.65503
Galicia	0.85048	0.84007	0.86665	0.83784
Madrid (Comunidad de)	0.74496	0.72114	0.78965	0.71332
Murcia (Región de)	0.92983	0.88027	0.91328	0.89584
Navarra (Comunidad Foral de)	0.88126	0.84910	0.91312	0.84528
País Vasco	0.91375	0.87677	0.91985	0.88421
Rioja (La)	0.79714	0.69659	0.74579	0.73862
Ceuta y Melilla	0.40617	0.31503	0.33508	0.36117

Tabla 2.

	Varones		Mujeres	
	< 25 años	>= 25 años	< 25 años	>= 25 años
Total nacional	0.88613	0.85513	0.85116	0.81431
Andalucía	0.88974	0.87955	0.84239	0.82943
Aragón	0.83583	0.78752	0.74796	0.68739
Asturias (Principado de)	0.72923	0.68544	0.73741	0.69422
Balears (Illes)	0.93394	0.92535	0.90287	0.89144
Canarias	0.88396	0.83757	0.88571	0.83977
Cantabria	0.83415	0.81308	0.82527	0.80352
Castilla y Leon	0.81606	0.76667	0.78084	0.72658
Castilla - La Mancha	0.86911	0.83165	0.83740	0.79457
Cataluña	0.98354	0.97561	0.98043	0.97129
Comunidad Valenciana	0.93914	0.91007	0.89848	0.85690
Extremadura	0.75510	0.71922	0.65481	0.61341
Galicia	0.87090	0.84346	0.86311	0.83461
Madrid (Comunidad de)	0.79735	0.72446	0.78184	0.70622
Murcia (Región de)	0.93876	0.92584	0.89389	0.87462
Navarra (Comunidad Foral de)	0.92422	0.86431	0.90363	0.83358
País Vasco	0.93367	0.90634	0.90544	0.87036
Rioja (La)	0.80556	0.79479	0.70655	0.69316
Ceuta y Melilla	0.38684	0.41439	0.29736	0.32250

Tabla 3.

	Varones		Mujeres	
	< 25 años	>= 25 años	< 25 años	>= 25 años
Total nacional	0.87400	0.85980	0.86245	0.81096
Andalucía	0.86921	0.88665	0.86105	0.82364
Aragón	0.81831	0.79477	0.76362	0.68321
Asturias (Principado de)	0.72707	0.68618	0.73931	0.69373
Baleares (Illes)	0.92515	0.92939	0.91286	0.88808
Canarias	0.87927	0.83938	0.89022	0.83835
Cantabria	0.82173	0.81734	0.83768	0.80064
Castilla y Leon	0.84914	0.83932	0.85250	0.79042
Castilla - La Mancha	0.78903	0.77763	0.80093	0.72061
Cataluña	0.98646	0.97420	0.97713	0.97235
Comunidad Valenciana	0.92839	0.91438	0.90826	0.85350
Extremadura	0.70932	0.73187	0.69184	0.60486
Galicia	0.84543	0.85214	0.88346	0.82950
Madrid (Comunidad de)	0.79801	0.72419	0.78116	0.70640
Murcia (Región de)	0.92452	0.93209	0.90513	0.86973
Navarra (Comunidad Foral de)	0.91561	0.86796	0.91105	0.83125
País Vasco	0.93225	0.90689	0.90699	0.86999
Rioja (La)	0.77365	0.80352	0.72805	0.68564
Ceuta y Melilla	0.33231	0.43799	0.33715	0.30588

Del análisis de estos resultados se deduce que los varones tienen una mayor probabilidad de permanecer como parado, según la nueva definición, que las mujeres. También dicha probabilidad es mayor entre los jóvenes menores de 25 años que entre los de 25 años y más (Tabla 1). Sólo Asturias y Canarias respecto al comportamiento por sexo y el conjunto de Ceuta y Melilla en lo que se refiere a la edad no siguen esta pauta general.

Por otro lado, la interacción existente entre las variables sexo y grupo de edad (Tabla 3) pone de manifiesto un efecto oculto en las probabilidades conjuntas, consistente en una general disminución de la probabilidad de permanencia para los varones menores de 25 años (comparando con Tabla 2), salvo en Madrid y Cataluña donde esta probabilidad aumenta.