

«La calidad de la EPA en la estimación de transiciones en el mercado de trabajo»

En el presente artículo se analizan, por un lado, las implicaciones del desgaste muestral para el estudio de transiciones en el mercado de trabajo y, por otro, los errores de clasificación implícitos en las preguntas retrospectivas. En ambos casos el análisis se lleva a cabo con datos de la Encuesta de Población Activa. Respecto al primer objetivo, aunque se encuentra que la no respuesta no puede ser ignorada, hay poca evidencia de que el desgaste muestral cause, en general, problemas importantes de selección. Aún así, se detectan algunos problemas de selección en rangos extremos de edades. En lo que respecta al segundo objetivo, se encuentra que la información retrospectiva contiene errores sistemáticos. Sin embargo, tanto la modesta magnitud de los errores como su importancia decreciente sustentan el uso de la información retrospectiva para el estudio de las transiciones en el mercado de trabajo.

Artikulu honetan aztertzen da, alde batetik, lan merkatuan gertatzen diren iragaiteak ikertzeko orduan laginaren higadurak dituen ondorioak eta, beste aldetik, atzera begirako galderetan dauden sailkapen errakuntza inplizituak. Bi kasuetan, azterlana Biztanleria Aktiboaren gaineko Inkestaren datuekin egin da. Lehenengo helburuari dagokionez, erantzun eza kontuan hartu behar dela ondorioztatu bada ere, ez dago ebidentzia handirik esateko laginaren higadurak, oro har, arazo garrantzitsurik sortzen duenik hautaketan. Hala eta guztiz ere, hautaketa arazo zenbait aurkitu dira muturretako adin taldeetan. Bigarren helburuari dagokionez, ondorioa honako hau da: atzera begirako informazioan errakuntzak sistematikoki agertzen dira. Hala ere, errakuntza horien tamaina txikia da eta beraren eragina gero eta motelagoa eta, horren ondorioz, atzera begirako informazioa lan merkatuaren iragaiteak aztertzeke erabiltzea egokitzat jotzen da.

This article analyses the implications of sample deterioration in the study on the labour market transitions and, additionally, the implicit mistakes of classification in retrospective questions. In either case the analysis is carried out with data from the Labour Force Survey (EPA). With respect to the first objective, although it has become clear that the nonresponse cannot be ignored, there is little evidence of the fact that sample deterioration, as a rule, causes important selection problems. Even so, some selection problems in extreme age ranges have been detected. Reference to the second objective, it is clear that the retrospective Information contains systematical errors. However, both the small magnitude of such errors and their diminishing importance sustain the use of retrospective information for the study of transitions in the labour market.

1. Introducción
 2. Una descripción de la EPA
 3. Desgaste muestral
 4. Información retrospectiva vs. medición reiterativa
 5. Conclusiones
- Referencias bibliográficas

Palabras clave: Transiciones en el mercado laboral, dinámica de población activa, desgaste muestral, EPA.
Nº de clasificación JEL: J20, J21, J62

1. INTRODUCCIÓN

En este artículo examinamos algunos aspectos que surgen al analizar la dinámica de la población activa usando la Encuesta de Población Activa (EPA). Nuestro análisis se basa en los datos individuales disponibles desde 1987 hasta 1997.

Dos vías pueden ser usadas para estudiar las transiciones entre estados de actividad de la población en base a la EPA. La primera explota su naturaleza de panel rotativo para enlazar la información entre olas consecutivas¹. La otra explota, en cambio, el hecho de que la encuesta

correspondiente al segundo trimestre contiene información retrospectiva sobre la actividad en el año precedente. Cada una de estas dos vías tiene sus propias ventajas e inconvenientes, siendo uno de los objetivos del presente artículo clarificarlas.

Empezaremos describiendo los problemas de no respuesta después de haber entrado en la muestra o desgaste muestral y de reentrada en la muestra. El desgaste muestral puede tener consecuencias importantes ya que la muestra puede perder representatividad de la población en estudio². Distinguimos dos tipos de desgaste: monótono y no monótono. En el primer caso el individuo que deja la muestra no

* Agradecemos la financiación recibida de la Fundación BBV a través del proyecto «La reforma de la Seguridad Social en España».

¹ Bover, Arellano y Bentolila (1996), Alba-Ramírez (1997-1998) son ejemplos notables de estudios que explotan la naturaleza rotativa de la EPA.

² Villagarcía (1995) constituye un excelente ejemplo de las consecuencias de la no respuesta sobre los indicadores del mercado de trabajo.

vuelve a ser observado. En el segundo, los individuos vuelven a ser observados algunos periodos más adelante. Otro fenómeno de importancia es la entrada después de la primera entrevista de nuevos individuos. Nótese que las nuevas entradas y las reentradas (desgaste no monótono) pueden mitigar las consecuencias del desgaste muestral. En nuestro trabajo, presentamos un simple modelo que ayuda a entender la evolución de los patrones de respuesta en la EPA.

En la segunda parte de nuestro papel analizamos las consecuencias del desgaste muestral sobre la validez de las transiciones estimadas. En particular, planteamos un contraste no paramétrico en el que se comparan las probabilidades de transición de los que responden a todas las entrevistas y los que no responden, por cualquier causa, a alguna entrevista. Avanzándonos a lo que más adelante matizaremos, no encontramos evidencia de que la no respuesta tenga consecuencias importantes sobre las probabilidades de transición entre estados.

Además, estudiamos la importancia y los determinantes de los errores que se cometen al analizar transiciones en base a las preguntas retrospectivas de la encuesta. Afortunadamente, aunque detectamos algunos errores sistemáticos, su magnitud modesta y su decreciente importancia permiten afirmar la utilidad de las preguntas retrospectivas de la EPA.

El plan del resto del artículo es el siguiente. La sección 2 describe los datos. La sección 3 examina las cuestiones de desgaste muestral y los efectos del mismo sobre la bondad de las

estimaciones de probabilidades de transición. La sección 4 compara las estimaciones de la probabilidades de transición basadas en la información retrospectiva con las basadas en la medición reiterativa. Finalmente, la sección 5 concluye.

2. UNA DESCRIPCIÓN DE LA EPA

La EPA es una encuesta trimestral rotativa a cargo del Instituto Nacional de Estadística (INE). La muestra planeada consta de 64.000 hogares conteniendo aproximadamente a 150.000 individuos adultos. El esquema de rotación de la encuesta comporta que cada nuevo grupo de rotación permanece en la muestra seis periodos (un año y medio). Dado este esquema de rotación, si no hubiera desgaste, entonces 5/6 de la muestra en cada dos encuestas consecutivas podría ser enlazado y, además, en cada trimestre aquellos en los grupos de rotación 1-2 podrían ser enlazados con los grupos 5-6 en el mismo trimestre del año siguiente.

Aunque la encuesta se viene realizando desde 1964, sólo están disponibles los datos individuales desde 1976 y, específicamente, el enlace de datos de encuestas adyacentes sólo es posible desde el segundo trimestre de 1987. El cuestionario de 1976 fue modificado en 1987 (cuando se introdujeron las preguntas retrospectivas), que fue a su vez modificado en el primer trimestre de 1992 y, más recientemente, en el primer trimestre de 1999, aunque este último cambio queda fuera de nuestro horizonte muestral.

2.1. Características de la EPA

Los individuos que viven en hogares privados constituyen la población objeto de estudio de la EPA. Los hogares colectivos (hospitales, cuarteles, etc..) y las residencias secundarias están excluidas de la muestra. Las secciones censales son las unidades muestrales primarias, mientras que los hogares (de hecho, direcciones postales) son la unidades secundarias.

El INE distingue entre personas encuestables y no encuestables. Son personas encuestables todos los miembros adultos del hogar presentes o ausentes por menos de un trimestre. Otros miembros del hogar ausentes por más de un trimestre son considerados no encuestables, excepto en el caso de que regresen al hogar durante el periodo de permanencia del hogar en la muestra, en cuyo caso vuelve a ser considerados como encuestables. Los que se añaden al hogar durante el periodo muestral son añadidos a la lista de personas encuestables y aquellos que dejan el hogar desaparecen de dicha lista.

El cuestionario se somete, en general, a una única persona, que responde por todos los miembros del hogar. El informante puede cambiar entre entrevistas sucesivas. El hecho de que la EPA no es una encuesta estrictamente personal puede explicar las moderadas tasas de desgaste muestral y, más importante aún, puede reducir la importancia del desgaste endógeno para ciertos grupos de población (por ejemplo, individuos en edades intermedias o próximos al retiro) pero, desafortunadamente puede introducir

otros errores, especialmente en la respuestas retrospectivas. La naturaleza y el tamaño de estos errores de medida dependen del tamaño del hogar y de la relación entre el informante y el resto e los miembros del hogar.

Según el resultado del proceso de entrevista, el INE clasifica los hogares en una de las tres categorías siguientes:

- Entrevistado: Cuando se obtiene en entrevista de, al menos, uno de los miembros del hogar.
- Ausente: Si todas las personas encuestables del hogar están ausentes del hogar.
- Negativa: Si todos los encuestables rechazan colaborar en el proceso de la entrevista.

La EPA sigue la metodología propuesta por la Oficina Internacional del Trabajo (OIT) y sigue las recomendaciones de la oficina estadística de la Comunidad Europea (EUROSTAT)³. Más aún, la penúltima modificación del cuestionario de la EPA (1992) constituyó un paso al frente en el proceso de integración de las encuestas sobre el mercado de trabajo en la Comunidad Europea. Pocas diferencias perduran hoy en día, especialmente después de la más reciente modificación acaecida en el primer trimestre de 1999.

Como resultado, la EPA contiene información, durante la semana de referencia, relativa a la actividad laboral, educación y características familiares,

³ De hecho, el sustento español de la encuesta de la Comunidad Europea sobre el mercado de trabajo («European Community Labor Force Survey») a cargo de Eurostat.

aunque presenta algunas deficiencias importantes. Por ejemplo, no contiene información sobre salarios y otras fuentes de ingresos.

2.2. Chequeo, seguimiento y evaluación de la calidad de los datos

En esta sección resumizamos los procedimientos que sigue la oficina estadística al objeto de reducir la tasa de no respuesta debida a no contacto o negativas a responder.

En primera entrevista, los hogares ausentes o inaccesibles son revisitados durante seis semanas hasta que se obtiene una entrevista o una negativa firme. Los hogares aún inaccesibles después de dicho periodo son visitados de nuevo en segunda entrevista.

Por otra parte, los hogares que rehusan responder en primera entrevista son reemplazados aleatoriamente por otros en la misma sección censal. La negativa en entrevistas sucesivas es tratada de forma diferente, ya que los hogares no son, por regla general, reemplazados, aunque, al objeto de no alterar la representatividad de la muestra, las respuestas obtenidas en la entrevista inmediatamente anterior son repetidas. En el caso de no respuesta reiterada, las respuestas dejan de imputarse. Sin embargo, cuando la no respuesta afecta sólo a un conjunto limitado de preguntas, el INE utiliza un método de imputación automático llamado DIA (Detección y Imputación Automática)⁴.

⁴ Véase INE (1994) para una descripción detallada de la metodología DIA.

Anualmente el INE elabora una encuesta de evaluación de la calidad de los datos en la EPA (INE, 1995). Dicha encuesta, realizada desde 1970, tiene dos objetivos: a) controlar el trabajo de campo y, b) evaluar la calidad de los resultados.

2.3. Enganche de la EPA y la Estadística de flujos

Desde 1987 el INE publica resultados sobre flujos en el mercado de trabajo, explotando la naturaleza rotativa de la encuesta, y pone a disposición de los investigadores los ficheros de datos longitudinales, llamados Estadísticas de Flujos (EF). Sin embargo, dichos ficheros, aunque permiten seguir los individuos hasta seis periodos, están sujetos a severas limitaciones ya que contienen menos información que la EPA en sí⁵.

Afortunadamente, estas dificultades o limitaciones pueden ser solucionadas mediante el enganche de los hogares en la EPA con los identificadores individuales en la estadística de flujos (EF). Nuestro problema es muy diferente del descrito en, por ejemplo, Peracchi y Welch (1995), que enganchan información individual entre olas adyacentes del «Current Population Survey» de los Estados Unidos. Nosotros, en cambio, engarzamos dos versiones de sección cruzada del mismo conjunto de datos.

⁵ El identificador del hogar y la residencia son eliminados. La edad se resume en intervalos quinquenales y el sector y ocupación se recolocan en un menor número de categorías, lo que disminuye la utilidad de la información y dificulta su uso en el análisis de, especialmente, transiciones en el mercado de trabajo.

La variable clave del engarce es el identificador individual, un número de 18 dígitos, único para cada persona que consta, a su vez, de tres partes: El identificador de la sección censal, el identificador del hogar dentro de la sección censal y, finalmente, el identificador de la persona dentro del hogar. Para identificar dichas partes, usamos un proceso secuencial que consta de los siguientes pasos:

1. Ordenación de los registros en las dos fuentes en base a los valores de todas las variables comunes. Este primer paso engancha los registros únicos en ambas fuentes y ayuda a identificar la sección muestral a la que un individuo o un hogar pertenece.
2. Ordenación de los identificadores de hogar dentro de una sección muestral y de los individuos dentro de un hogar (ambos identificadores son parte del identificador total), lo que engancha por completo ambas fuentes.

Con este procedimiento, conseguimos enganchar el ciento por ciento de los registros en la EF con sus correspondientes en la EPA. En base a dicho enganche, añadimos el identificador a la EPA, con lo que obtenemos la versión longitudinal de la EPA. Consecuentemente, a partir de aquí deberá entenderse que estamos utilizando esta última fuente de datos.

2.4. Dinámica de la muestra

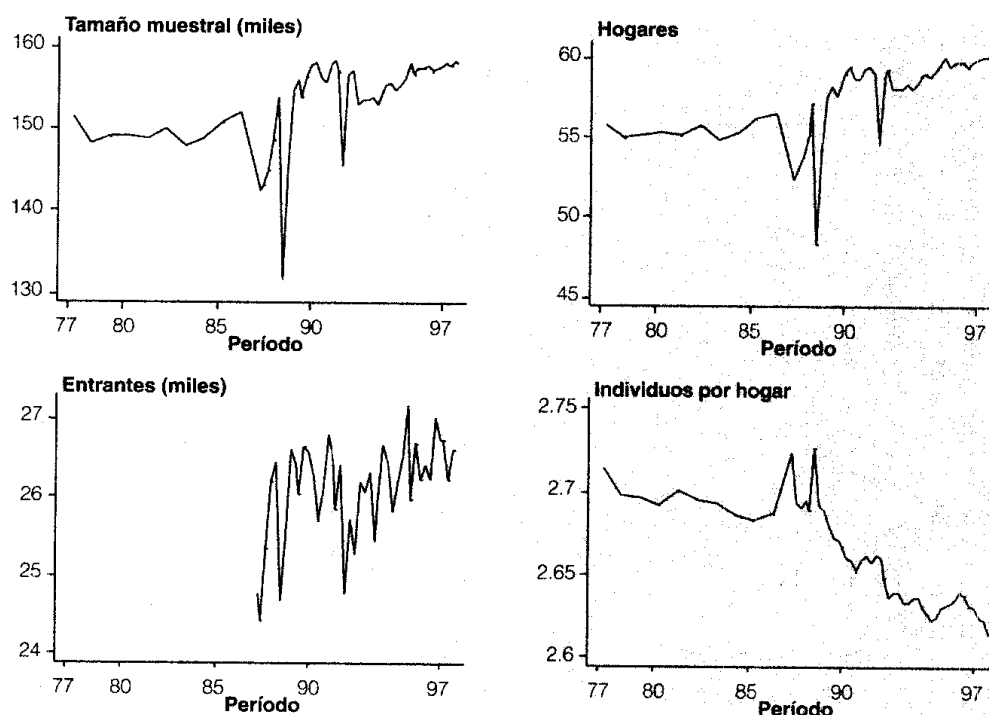
Tal y como ya hemos comentado brevemente, la EPA considera 3024 secciones muestrales y, en promedio, 20

hogares en cada una de ellas. Consecuentemente, el tamaño muestral teórico es de aproximadamente 64.000 hogares. El Gráfico n.º 1 describe la evolución de la muestra realizada durante el periodo 1977-1997. Para el periodo 1977-1987 únicamente presentamos los datos del segundo trimestre. Desde 1987, presentamos datos de todos los trimestres.

Nótese que la muestra realizada, tanto en términos de individuos como de hogares entrevistados, se redujo drásticamente en el segundo trimestre de 1987 como consecuencia de la introducción de un nuevo cuestionario, que causó un fuerte aumento en la tasa de no respuesta. El impacto de este cambio se acentúa en el tercer trimestre de 1988, primer periodo en el que todos los entrevistados han realizado su primera entrevista después del cambio de cuestionario. Por otra parte, los cambios introducidos en el cuestionario en el primer trimestre de 1992 también redujeron transitoriamente, aunque en menor medida, el tamaño de la muestra realizada. Nótese que en años recientes el INE ha realizado un gran esfuerzo para incrementar el tamaño muestral en base a aumentar el número de entrantes.

Un fenómeno de especial interés que también presentamos en el Gráfico n.º 1 es la caída en el número de personas entrevistadas por hogar, fiel reflejo de los recientes cambios demográficos. Nótese que el pico en el número de personas entrevistadas por hogar se observa justo cuando la muestra realizada es menor, reflejando el hecho de que los hogares más grandes son más fácilmente observables.

Gráfico n.º 1. Tamaño muestral de la EPA en el periodo 1977-1997



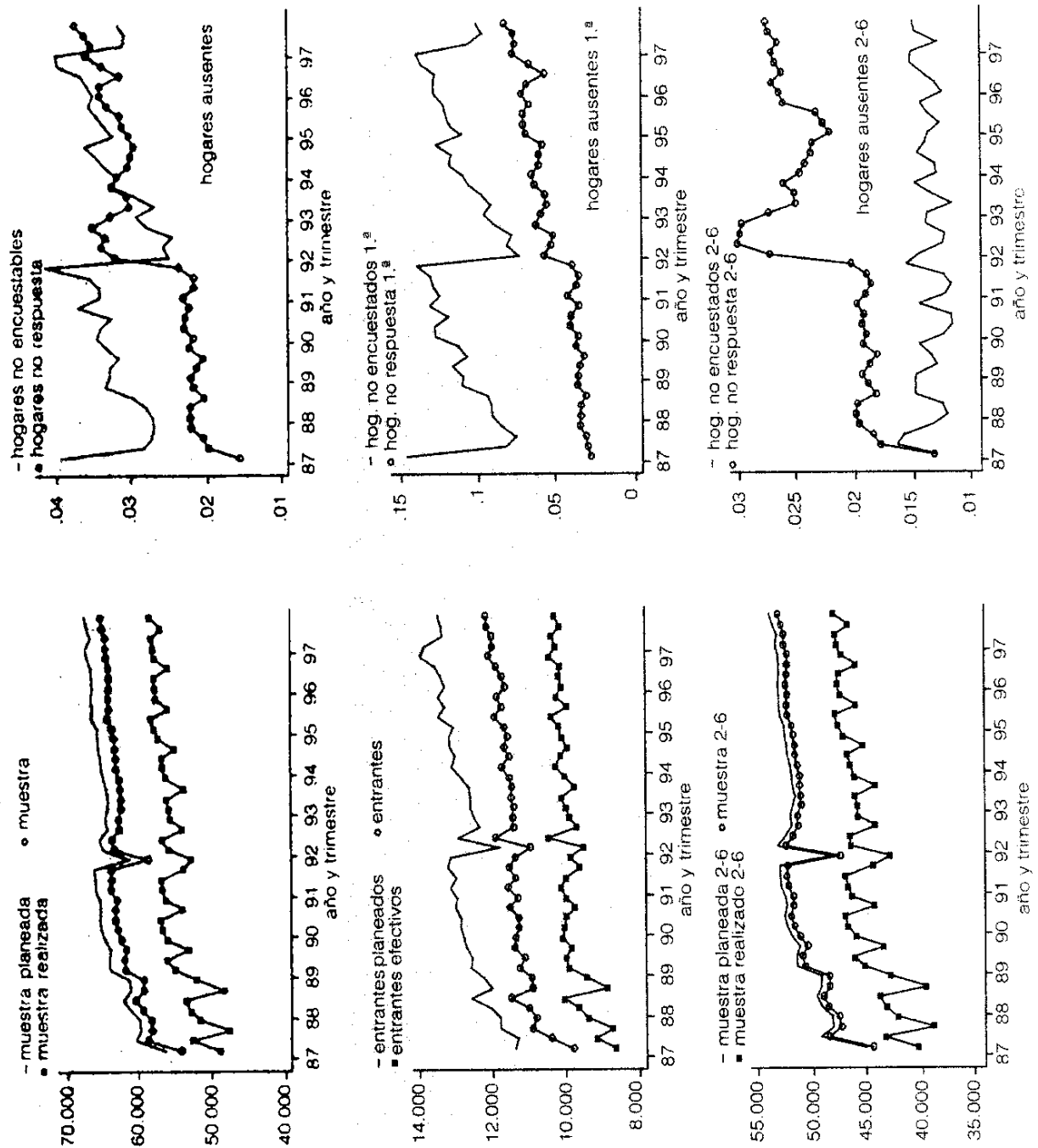
El Gráfico n.º 2 nos permite ahondar en la dinámica de la muestra. En él analizamos el comportamiento de las tasas de respuesta durante el periodo 1987-1997 (véase también la sección 2.1), distinguiéndose entre respuestas a la primera y otras entrevistas (segunda a sexta).

La primera columna del gráfico compara el tamaño de la muestra potencial y la realizada. Encontramos tres características remarcables. Primero, mientras que la muestra potencial se incrementa un 10 por ciento en el periodo, la muestra realizada aumenta en menor medida. Segundo, los datos presentan dos discontinuidades importantes: una

primera en el tercer trimestre de 1988 y otra en el primero de 1992 debido a las reducciones, en ambos casos, del número de hogares entrevistables y finalmente entrevistados. Tercero, la muestra realizada presenta fuertes fluctuaciones estacionales.

La columna central del gráfico muestra la tendencia temporal tanto de la fracción de hogares no encuestables como de las negativas a responder. Nótese el importante número de hogares no encuestables en primera entrevista (cerca del 10 por ciento) y la configuración, reflejo de la estacionalidad, en dientes de sierra. Las caídas súbitas se localizan en los periodos de cambio de cuestionario

Gráfico n.º 2. Muestra potencial y realizada y causa de la no respuesta: 1987-1997



(1987,1992 y, en menor medida, 1997). Después de la primera entrevista, la fracción de hogares no encuestables es mucho menor y, más aún, no presenta ninguna tendencia especial, excepto la componente estacional ya apuntada (más acentuada en el cuarto trimestre).

La tasa de negativa muestra una clara tendencia creciente. Empieza alrededor del 2 por ciento en 1987 y alcanza el 3,7 por ciento en 1997. Como era predecible, la fracción de negativas es mucho más importante en primera entrevista que en sucesivas entrevistas. Nótese el incremento temporal en el número de negativas en segunda y sucesivas entrevistas después del cambio de cuestionario en 1992, que se contrapone al modesto incremento de negativas observado en primera entrevista en la misma referencia temporal. De hecho, el continuo incremento en la fracción de negativas puede ser atribuido, en parte, al cambio en la composición según el nivel educativo de la población, ya que los individuos relativamente más educados rehúsan colaborar más frecuentemente que los individuos relativamente menos educados. Sin duda, el diferente valor del tiempo para los dos grupos mencionados ayuda a explicar dicha diferencia.

Finalmente, la última columna del gráfico muestra las pautas de la fracción de hogares ausentes. Dicha fracción presenta una fuerte estacionalidad en torno a una tendencia decreciente, que refleja los esfuerzos del INE tendentes a reducir el número de hogares no contactados o ausentes, mientras que el pico estacional del tercer trimestre es reflejo del periodo vacacional.

En suma, el INE ha intentado mitigar el continuo incremento en la fracción de hogares que rehúsan colaborar en la encuesta mediante las dos medidas siguientes: (a) Incrementando el tamaño de la muestra potencial o teórica; (b) Aumentando el esfuerzo dedicado a disminuir la fracción de contacto fallidos en dos dimensiones, número y estacionalidad. Estos esfuerzos, tal y como mostraremos a continuación, han tenido consecuencias importantes en las tasas de desgaste muestral observadas durante el periodo.

3. DESGASTE MUESTRAL

3.1. Desgaste en muestras longitudinales

El desgaste muestral es un tipo especial de no respuesta que aparece en muestras rotativas o longitudinales, y puede ser definido como la incapacidad para obtener información de una unidad muestral en cualquier ola después de haber sido seleccionada para participar en la muestra.

Resulta importante distinguir entre desgaste exógeno y endógeno. Mientras que el primero sólo crea problemas de eficiencia pero no sesga los resultados, el desgaste endógeno, es decir aquel sistemáticamente relacionado con la variable de interés, puede sesgar considerablemente los resultados (Little y Rubin, 1987).

En las encuestas a hogares, el desgaste está típicamente relacionado con transiciones importantes en la vida de una persona: ir a la universidad, encontrar un nuevo trabajo, matrimonio o divorcio, retiro

Cuadro n.º 1. La dinámica de la respuesta: patrones y persistencia

Pautas más comunes de respuesta					
Pauta	Porcentaje	Pauta	Porcentaje	Pauta	Porcentaje
++++++	64.44	+----+-	2.13	-+ +--+	0.49
+++++-	3.46	-+++++	2.04	-+----+	0.47
+-----	2.89	+---++-	1.90	++--++	0.46
++++-++	2.84	+--+++-	1.88	+++-++	0.44
++-++++	2.70	++++--	0.74	--++++	0.42
++++-+	2.55	-+++-+	0.73	+---+++	0.39
+--++++	2.42	-----+	0.50	Otros	6.11
Persistencia en la respuesta					
Primera respuesta en	Aún presente después de (trimestres)				
	1	2	3	4	5
1.ª entrevista	85.7	81.1	77.0	73.5	69.7
2.ª entrevista	79.0	62.4	45.0	40.6	—
3.ª entrevista	70.7	50.1	36.0	—	—
4.ª entrevista	52.5	16.7	—	—	—
5.ª entrevista	13.6	—	—	—	—

Nota: + indica respuesta; - indica no respuesta

y muerte. A menudo, estos eventos constituyen el objeto de estudio, lo que ocasiona que el desgaste muestral no pueda ser ignorado, ya que ignorarlo puede invalidar los resultados obtenidos, incluso cuando las tasas de desgaste son modestas.

En el caso de individuos, es importante distinguir entre desgaste monótono y no monótono, dependiendo si una persona que deja la muestra vuelve a la misma en algún momento posterior. En los casos en que el desgaste y la entrada (nueva entrada o reentrada) se compensan y los que entran son similares a los que se van, la muestra continua siendo representativa

de la población en estudio, incluso en el caso de que las tasas de desgaste y entrada sean altas⁶.

3.2. El desgaste muestral en la EPA

El Cuadro n.º 1 estudia los patrones de participación en la muestra de los individuos de más de 18 años, observados por primera vez entre

⁶ Una lección del Panel Study of Income Dynamics (PSID) es que el desgaste y la entrada o reentrada pueden contribuir a hacer la muestra más representativa de la población en estudio (ver Becketti era/., 1998).

el segundo trimestre de 1987 y el cuarto de 1996. Dichas pautas combinan cambios en el tamaño de los grupos de rotación iniciales, desgaste, entrada y reentrada.

El panel superior muestra los veinte (de 64) patrones de respuesta más frecuentes. Participación completa (aquellos que responden a todas las entrevistas) es, con mucho, el caso más común, representando casi dos tercios de los individuos considerados. Los dos casos siguientes, en términos cuantitativos, son el de los que no responden a la última entrevista (3,5 por ciento) y los que sólo responden a la primera entrevista (2,9 por ciento). Los cuatro casos siguientes responden a casos en los que los individuos fallan una sola entrevista (por ejemplo, 2,8 por ciento fallan la cuarta, 2,7 por ciento la tercera, 2,6 por ciento la quinta y 2,4 la segunda). También son importantes los casos de aquellos que entran después de la primera entrevista y permanecen después (2,0 y 0,4 por ciento) y los que sólo responden una vez al año (2,1 por ciento responden a la primera y la quinta entrevista). Otros patrones de respuesta son mucho menos frecuentes.

El panel inferior muestra la fracción de aquellos que respondiendo por primera vez en la entrevista t ($t = 1 \dots 5$), todavía responde después de q trimestres. Nótese que los que responden a la primera se comportan muy diferente del resto. De hecho, la probabilidad de responder a la entrevista $t+1$ condicional a haber empezado en t decrece con t .

En el Cuadro n.º 2 analizamos las diferencias en las pautas de respuesta en

la EPA entre grupos de participantes según diversas características individuales. Consideramos cinco tipos de participantes: participantes en todos los periodos (64,4 por ciento de la muestra), participantes con un patrón monótono de desgaste (7,4 por ciento), participantes iniciales que fallan una sola entrevista (10,5 por ciento), entrantes que no vuelven a salir (3,1 por ciento y, finalmente, otros participantes (14,6 por ciento). Detectamos importantes diferencias por edad y nivel de estudios. En particular, la fracción de los que responden a todas las entrevistas entre los que tienen estudios universitarios es mucho más pequeña que entre los que no tienen estudios. En cambio, las diferencias según el status laboral y sexo son mínimas. De hecho, las últimas son inexistentes si no consideramos la distorsión que introduce la diferencia de mortalidad entre ambos sexos.

Al objeto de ilustrar las consecuencias de la no respuesta en el análisis de los resultados en el mercado de trabajo, comparamos las tasas de participación en primera entrevista para dos grupos de participantes: completos (presentes en todas las entrevistas) y el resto. El Gráfico n.º 3 presenta los resultados de dicho ejercicio por sexo, grupo de edad (18-29, 30-39, 40-54 y 55-65), nivel de estudios y periodo de la primera entrevista.

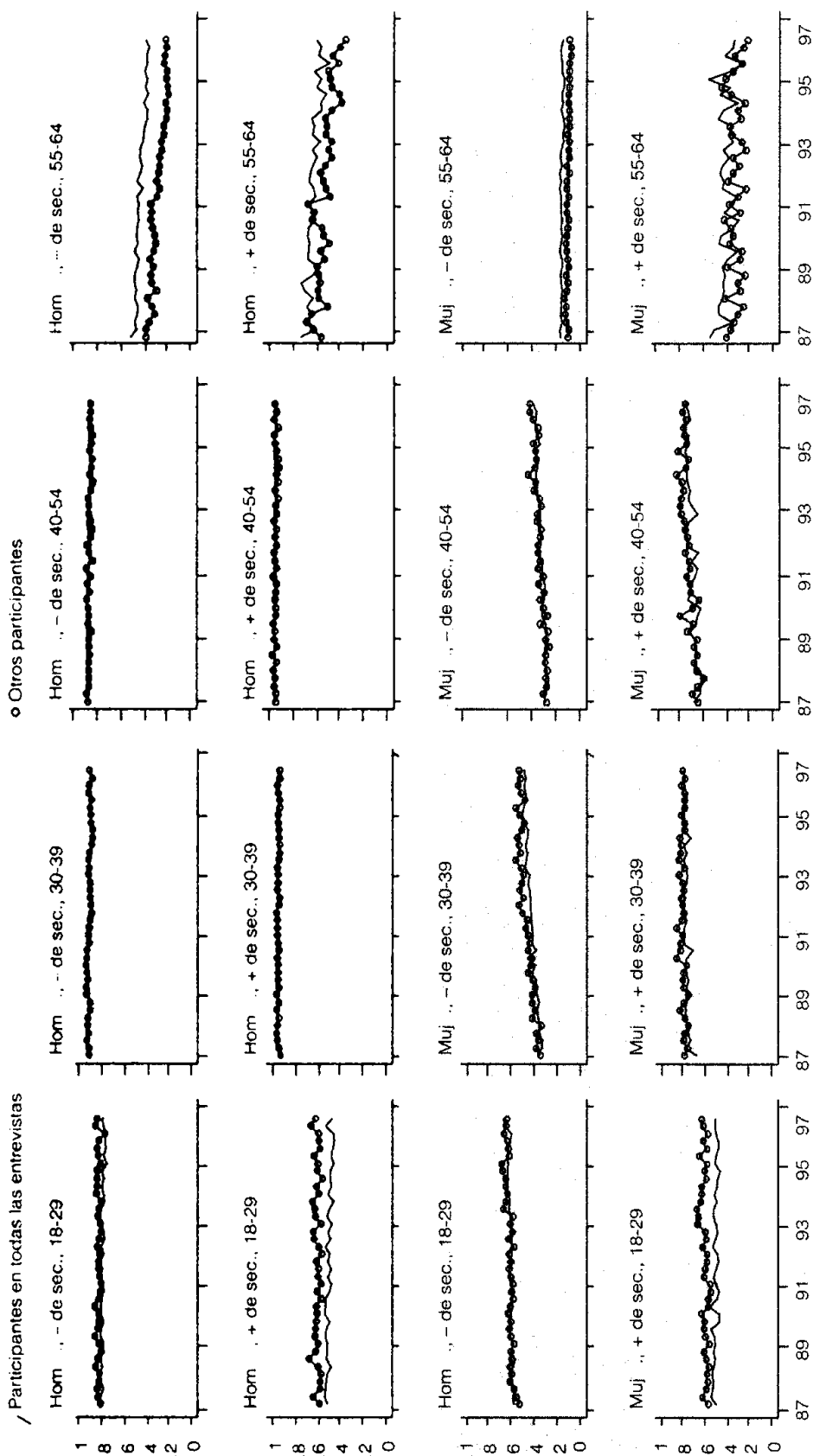
Si bien no detectamos diferencias para el grupo central de edades (30-54), si las hay para otros grupos de edad. En particular, encontramos diferencias significativas en las tasas de participación para los jóvenes de 18-30 con estudios

Cuadro n.º 2. Pautas de respuesta según características individuales

	Participantes completos	Desgaste monótono	Fallo en una entrevista	Entrada monótona	Otros casos	Total
Según el sexo						
Hombres	65.01	7.29	7.55	2.99	17.16	100.00
Mujeres	63.9	7.44	7.77	3.19	17.68	100.00
Según el sexo (edad 18-69)						
Hombres	63.86	7.80	7.47	3.00	17.87	100.00
Mujeres	63.24	7.87	7.72	3.09	18.08	100.00
Según el nivel de educación						
Sin estudios	67.31	6.95	6.92	2.68	16.14	100.00
Est. primarios	65.84	7.04	7.70	2.97	16.46	100.00
Est. secund.	62.48	7.76	7.67	3.45	18.64	100.00
Est. univ.	52.36	9.75	9.19	4.24	24.46	100.00
Según grupo de edad en 1.ª entrevista						
18-24	66.73	7.18	6.63	2.86	16.60	100.00
25-29	54.70	9.86	7.37	3.60	24.47	100.00
30-34	57.83	8.38	8.61	3.92	21.26	100.00
35-39	62.54	7.42	8.72	3.64	17.68	100.00
40-44	67.33	6.64	8.03	3.16	14.84	100.00
45-49	70.89	6.00	7.42	2.69	13.00	100.00
50-54	71.45	5.75	7.35	2.38	13.07	100.00
55-59	70.53	5.92	7.78	2.37	13.40	100.00
60-64	67.83	6.38	8.42	2.64	14.73	100.00
65-69	65.20	6.84	8.57	2.91	16.49	100.00
70-74	63.54	7.44	8.37	3.02	17.63	100.00
75-79	61.13	8.08	7.34	3.40	20.05	100.00
80 +	53.96	10.65	5.54	4.00	25.84	100.00
Según relación con la actividad en 1.ª entrevista						
Empleados a TC	63.41	7.58	7.69	3.23	18.09	100.00
Empleados a TP	63.62	7.57	7.49	3.28	18.43	100.00
Desempleados	62.76	7.81	6.89	3.25	19.30	100.00
Inactivos	65.59	7.11	7.81	2.96	16.53	100.00
Total	64.44	7.37	7.66	3.10	17.43	100.00

Nota: 1.080.019 individuos 18 +. Periodo muestral: 87.II a 96.IV

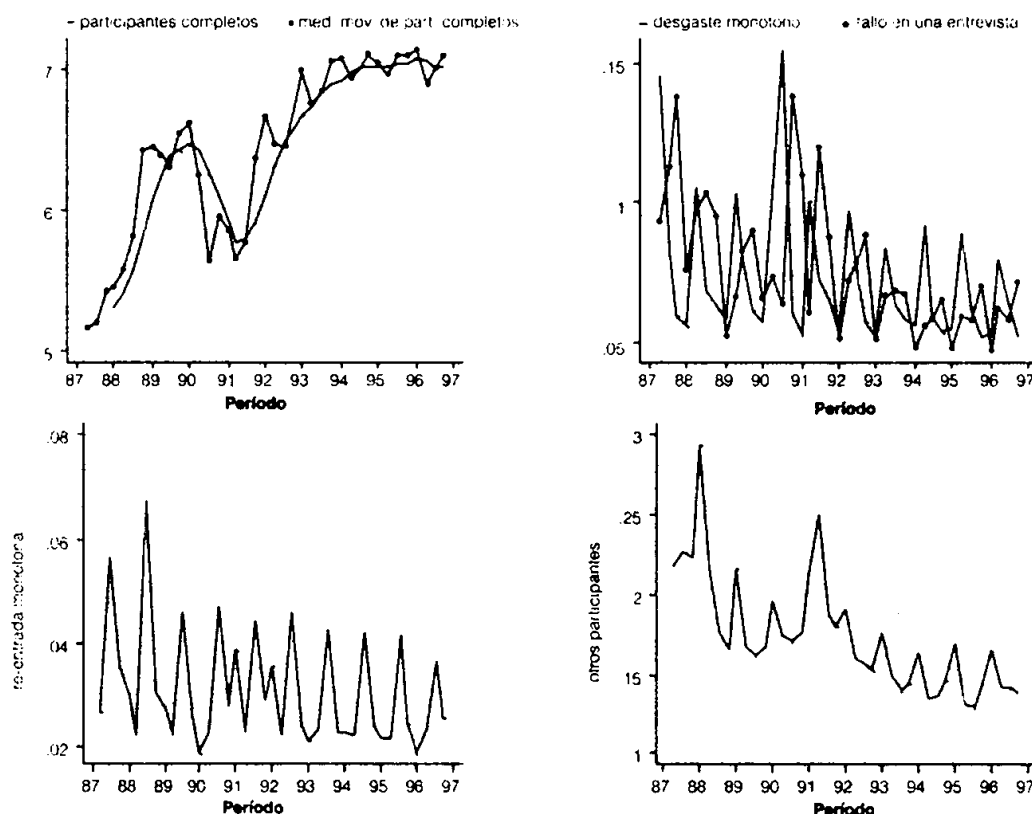
Gráfico n.º 3. Una comparación entre las tasas de participación en primera entrevista para participantes en todas las entrevistas y otros participantes



secundarios o universitarios y los mayores de 55 sin estudios o estudios primarios. En consecuencia, la EPA tiende a subestimar la participación de los hombres jóvenes y educados. Para los mayores menos educados, la evidencia indica justo lo contrario, ya que la participación de los individuos de respuesta completa es más alta que la correspondiente a otros tipos de participantes en la muestra. Nótese que esta evidencia tiene consecuencias importantes en, por ejemplo, las tasas de salida del mercado de trabajo de trabajadores en edad avanzada.

El panel superior derecho del Gráfico n.º 4 ilustra el efecto del año y el trimestre de la primera entrevista sobre la fracción de participantes completos. Al objeto de resaltar la tendencia presentamos también una versión suavizada de la serie. El gráfico muestra un incremento en la fracción de participantes completos desde algo más del 50 por ciento en 1987 hasta cerca del 70 por ciento en 1997. También revela la presencia de diferencias sistemáticas dependiendo del trimestre de la primera entrevista. Particularmente interesante es el caso de aquellos a los que se

Gráfico n.º 4. Tipos de participantes según año y trimestre de la primera entrevista



entrevista por primera vez en el segundo trimestre, que tienen menos probabilidad de participar en todas las entrevistas debido a que son entrevistados dos veces en el tercer trimestre, en periodo de vacaciones.

3.3. Desgaste y reentrada individual

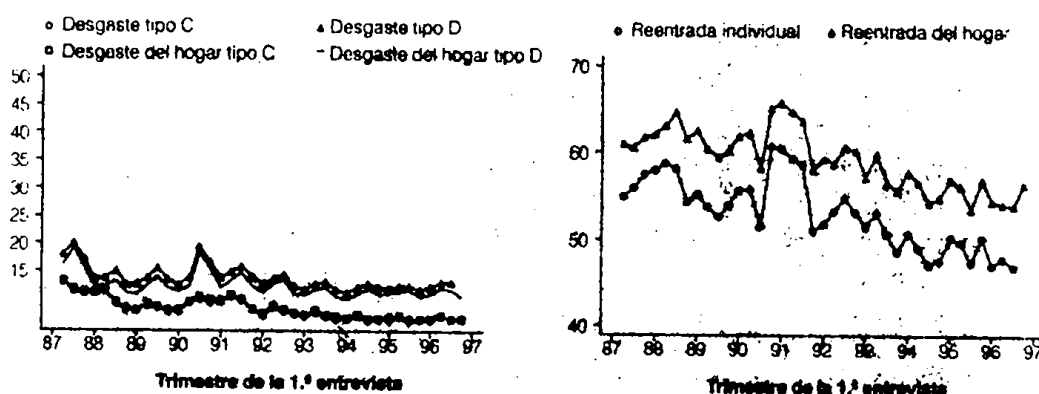
Podemos plantear diferentes medidas de desgaste y reentrada individual o del hogar, por ejemplo:

1. Fallo en la sexta entrevista (A).
2. Fallo en cualquiera de las seis entrevistas (B),
3. Fallo o desgaste entre entrevistas sucesivas (C),
4. Fallo o desgaste después de cuatro entrevistas (desgaste sobre un año) (D).

La primera medida no es muy informativa mientras que la segunda es justo el complemento de la probabilidad de ser participante a tiempo completo. De hecho, las más interesantes desde cualquier punto de vista son las definiciones de desgaste sobre el trimestre y el año (tipos C y D, respectivamente) y de reentrada condicional a no respuesta en el trimestre anterior.

El panel izquierdo del Gráfico n.º 5 presenta las tasas de desgaste individuales y del hogar según el periodo de la primera entrevista. Independientemente de la definición considerada, las tasas de desgaste individual han decrecido en el tiempo. Durante el periodo, la fracción de los que no responden después de un trimestre ha caído desde el 11-12 hasta el 6 por ciento, mientras que la fracción de los que no responden después de un año

Gráfico n.º 5. Desgaste y reentrada en la EPA según el trimestre de la primera entrevista



(tipo D) ha caído desde el 18-20 hasta el 13 por ciento. Las tasas de desgaste de los hogares son sólo algo menores. Para ambas, se observa una estacionalidad moderada.

El panel derecho del gráfico muestra las tasas de reentrada de los individuos y hogares en la entrevista t , condicional a falta de respuesta en la entrevista $t - 1$. Las tasas de reentrada son muy importantes, por encima del 50 por ciento, aunque decrecen en el periodo considerado, en paralelo al decaimiento de las tasas de desgaste.

En los Cuadros n.^{os} 3 y 4 analizamos las diferencias en las tasas de desgaste tipo-C y las de reentrada según diversas características individuales (sexo, nivel de estudios, relación con el cabeza de familia, grupo de edad y región de residencia). El núcleo de la familia (el cabeza de familia y su cónyuge y sus hijos mayores) es menos probable que deje de responder que otros miembros del hogar (abuelos, otros parientes y otras personas que conviven en el hogar). En general, los más educados es más probable que no respondan que los menos educados. También los adultos que viven solos que los que conviven con otros adultos. De hecho, las tasas de desgaste decrecen suavemente con el tamaño del hogar, por dos razones: porque las familias de mayor tamaño son menos móviles y porque el método de obtención de la información aumenta las tasas de respuesta de las familias de mayor tamaño.

Por otra parte, la probabilidad de reentrada es mayor para los hombres

cabeza de familia y sus cónyuges, y es menor para las hijas con menor educación y sus cónyuges. Nótese que aunque las tasas de desgaste y reentrada tienden a estar positivamente correlacionadas (por ejemplo en las regiones donde se observa un más alto desgaste también se observa una mayor tasa de reentrada), hay importantes excepciones. Por ejemplo, los hijos, otros parientes y los adultos que viven solos están sujetos a un mayor desgaste y una menor reentrada.

3.4. Análisis del desgaste y la reentrada

El Cuadro n.º 5 presenta las tasas de desgaste condicional e incondicional según el número de entrevista. Tal y como se muestra en el panel superior, las tasas de desgaste incondicional aumentan con el número de entrevista desde 7,6 por ciento en la primera entrevista hasta el 17 por ciento en la última. Por otra parte, las tasas de desgaste condicionales a no respuesta en el periodo anterior (mostradas en el panel inferior) empiezan en 45,8 por ciento y se incrementan ligeramente en entrevistas posteriores. Por contra, las tasas de desgaste condicionales a respuesta en el periodo anterior empiezan en 9,3 por ciento en segunda entrevista y decrecen ligeramente hasta el 7,4 por ciento en la última entrevista. Condicionar en las dos entrevistas anteriores no parece alterar decisivamente los resultados. En consecuencia, una cadena de Markov de primer orden parece suficiente para explicar las pautas de participación en la EPA.

Cuadro n.º 3. **Desgaste y reentrada individual según características individuales en la EPA**

			Nivel de estudios								Total
			-De primarios		Secundarios		Universitarios		Todos		
			Hom.	Muj.	Hom.	Muj.	Hom.	Muj.	Hom.	Muj.	
Según la relación con el cabeza de familia											
Cabeza fam.	C		7,4	10,1	10,4	14,7	12,2	16,5	8,2	10,8	8,7
	R		58,0	53,4	57,4	48,7	59,4	52,9	58,1	53,0	56,9
Cónyuge	C		9,5	7,0	13,3	10,5	14,6	12,0	11,1	7,6	7,7
	R		45,0	61,5	46,0	59,3	47,4	60,7	45,7	61,2	60,9
Hijos	C		6,8	7,4	6,7	6,6	8,2	8,1	6,9	7,2	7,1
	R		42,6	39,1	50,8	50,6	51,2	51,0	46,7	45,7	46,2
Cónyuges de hijos	C		7,2	7,8	8,2	9,3	10,0	10,4	7,5	8,1	7,7
	R		33,6	28,5	33,9	37,4	25,7	34,7	33,1	30,3	32,1
Otros	C		10,8	10,9	11,5	12,8	15,3	16,4	11,1	11,3	11,2
	R		33,7	36,5	31,7	35,7	36,3	37,8	33,6	36,5	35,5
Según el grupo de edad											
18-24	C		7,6	9,4	6,5	6,9	7,2	7,5	7,1	8,0	7,5
	R		42,9	37,1	53,8	52,4	60,9	56,7	48,7	45,8	47,2
25-29	C		10,0	10,5	10,1	10,8	10,8	11,6	10,2	10,8	10,5
	R		41,0	46,0	43,9	46,0	43,4	43,4	42,3	45,3	43,8
30-34	C		9,0	8,6	10,6	10,2	14,0	13,2	10,1	9,8	9,9
	R		50,4	55,5	51,7	53,8	50,2	52,1	50,7	54,3	52,5
35-39	C		7,9	7,6	9,7	9,5	12,5	12,4	8,9	8,5	8,7
	R		57,1	59,4	56,3	59,6	57,8	59,9	57,1	59,6	58,3
40-44	C		6,9	6,4	8,8	8,9	12,2	11,3	7,8	7,1	7,5
	R		60,0	61,6	60,8	59,1	60,8	63,2	60,3	61,5	60,9
45-49	C		6,1	5,8	7,5	8,5	10,2	10,2	6,7	6,2	6,4
	R		58,8	60,6	61,6	62,0	63,6	64,9	60,0	61,2	60,6
50-54	C		5,8	5,9	7,4	9,2	9,6	10,0	6,2	6,2	6,2
	R		57,9	60,8	59,4	55,6	63,5	62,4	58,8	60,6	59,7
55-59	C		6,0	6,5	9,0	10,0	9,3	10,5	6,4	6,7	6,6
	R		59,5	61,4	58,0	61,7	63,2	62,2	59,7	61,4	60,6
60-64	C		6,7	7,4	8,4	11,2	9,7	11,1	7,0	7,6	7,3
	R		60,0	61,7	61,8	63,3	61,1	64,1	60,2	61,8	61,1
65-69	C		7,5	8,3	10,3	11,9	11,1	12,4	7,8	8,4	8,1
	R		59,9	61,6	57,6	61,4	66,4	61,4	60,3	61,6	61,0
70 +	C		8,6	9,6	11,1	11,1	12,2	13,0	8,9	9,7	9,3
	R		48,9	47,8	55,3	55,3	57,8	59,9	49,7	48,2	48,8
Total	C		7,5	7,9	8,4	8,6	10,9	10,9	8,0	8,3	8,1
	R		52,9	54,1	53,1	53,1	56,0	54,8	53,4	54,0	53,7

Nota: 1.080.019 individuos 18 +. Período muestral: 87.II a 96.IV.

C: Desgaste de tipo C.

R: Reentrada.

Cuadro n.º 4. **Desgaste y reentrada individual según tamaño del hogar y región de residencia en la EPA**

			Nivel de estudios								
			-De primarios		Secundarios		Universitarios		Todos		
			Hom.	Muj.	Hom.	Muj.	Hom.	Muj.	Hom.	Muj.	
Tamaño del hogar (16 +)											
Un adulto	C	14,9	13,4	23,5	19,0	22,5	19,8	16,6	14,1	14,8	
	R	45,8	52,4	37,2	46,4	45,4	51,6	44,6	51,9	49,7	
Dos	C	9,4	9,3	11,7	11,8	14,0	14,0	10,2	10,0	10,1	
	R	57,5	58,5	56,0	56,4	57,7	57,5	57,3	58,1	57,7	
Tres	C	6,8	7,0	8,1	8,3	10,3	10,3	7,3	7,4	7,4	
	R	57,5	58,5	56,1	56,4	57,7	57,5	57,3	58,1	57,7	
Cuatro	C	6,0	6,2	6,9	6,9	8,4	8,5	6,4	6,5	6,5	
	R	51,0	51,7	53,6	53,3	56,9	54,8	52,3	52,4	52,4	
Cinco	C	5,8	6,2	6,0	6,2	7,6	7,6	6,1	6,4	6,2	
	R	44,3	43,5	48,6	48,5	54,5	49,9	46,6	45,4	46,0	
Región de residencia											
Andalucía	C	7,0	7,4	7,5	7,8	10,7	11,0	7,4	7,7	7,6	
	R	52,4	53,1	50,7	52,3	56,4	53,7	52,7	53,1	52,9	
Aragón	C	7,5	8,2	7,9	7,6	9,8	9,7	7,8	8,2	8,0	
	R	53,2	53,8	50,1	51,9	56,5	53,5	53,1	53,5	53,3	
Asturias	C	7,1	7,4	7,4	7,8	9,4	10,1	7,4	7,7	7,5	
	R	58,1	59,6	60,2	59,6	63,7	57,1	59,2	59,3	59,2	
Balears	C	7,0	7,2	7,9	7,9	10,3	10,5	7,4	7,5	7,5	
	R	49,6	52,6	53,0	52,3	55,4	55,4	51,0	52,8	51,9	
Canarias	C	10,4	10,2	1,9	11,8	14,3	13,5	11,0	10,8	10,9	
	R	56,5	58,6	58,1	57,5	56,5	56,1	56,8	58,1	57,5	
Cantabria	C	7,1	7,4	7,0	7,9	9,8	9,7	7,4	7,7	7,5	
	R	50,9	53,1	48,8	54,6	54,1	49,9	50,9	53,0	52,0	
Castilla-L	C	7,1	7,8	8,3	8,3	11,0	11,4	7,6	8,1	7,8	
	R	53,8	53,2	55,0	52,6	57,3	55,4	54,4	53,4	53,8	
Castilla-M	C	7,5	8,1	7,9	8,1	9,6	10,3	7,8	8,3	8,1	
	R	53,6	54,9	53,0	50,6	56,3	55,2	53,8	54,3	54,1	
Cataluña	C	8,0	8,1	9,1	9,5	11,8	11,6	8,6	8,7	8,6	
	R	51,5	53,4	52,5	53,1	52,7	52,7	51,9	53,2	52,6	
Com. Valenciana	C	7,4	7,9	8,5	8,2	10,1	10,2	7,8	8,1	7,9	
	R	56,7	58,4	60,0	58,0	63,5	61,3	58,1	58,7	58,4	
Extremadura	C	8,1	8,8	9,5	9,5	13,1	13,1	8,7	9,2	9,0	
	R	51,3	51,8	49,8	51,8	50,4	52,0	51,0	51,9	51,4	
Galicia	C	4,7	4,9	6,5	6,6	9,1	8,9	5,4	5,5	5,4	
	R	45,9	48,2	45,5	45,3	55,1	51,6	47,0	48,2	47,7	
Madrid	C	9,4	10,0	9,6	10,2	13,0	12,4	10,0	10,3	10,2	
	R	55,2	56,9	56,5	57,0	55,7	57,5	55,6	57,0	56,4	
Murcia	C	6,1	6,4	7,1	7,9	9,2	10,4	6,5	6,9	6,7	
	R	53,5	54,2	55,4	54,7	58,5	59,8	54,5	54,9	54,7	
Navarra	C	8,3	9,0	8,5	9,0	11,2	11,6	8,7	9,3	9,0	
	R	52,6	54,0	52,7	51,6	54,6	54,8	52,9	53,7	53,3	
País Vasco	C	9,3	9,7	8,8	9,0	10,9	10,9	9,4	9,7	9,5	
	R	48,8	50,4	49,3	48,8	54,2	52,8	49,7	50,5	50,1	
La Rioja	C	5,0	5,8	5,5	5,7	7,2	7,1	5,3	5,9	5,6	
	R	53,4	55,0	60,5	52,6	58,7	57,1	55,5	54,9	55,2	
Ceuta y Melilla	C	14,0	14,5	17,1	15,5	20,7	20,5	15,3	15,1	15,2	
	R	52,2	51,0	47,9	53,7	48,3	47,0	50,6	51,0	50,8	

Nota: 1.080.018 individuos 18 +. Período muestral: 87.II a 96.IV.

Cuadro n.º 5. Tasas de desgaste condicional e incondicional
(+ = respuesta; - = no respuesta)

	Trimestre de la entrevista					
	1	2	3	4	5	6
Tasas de no respuesta incondicionales	7,6	12,0	13,5	14,7	16,1	17,0
Tasas de desgaste condicional a:						
R _{t-1} = -	-	45,8	51,6	58,9	65,0	67,3
R _{t-1} = +	-	9,3	8,3	7,8	7,7	7,4
R _{t-1} = -, R _{t-2} = -	-	-	60,9	74,5	81,3	82,3
R _{t-1} = -, R _{t-2} = +	-	-	47,9	45,6	45,8	45,3
R _{t-1} = +, R _{t-2} = -	-	-	18,5	17,5	17,4	17,8
R _{t-1} = +, R _{t-2} = +	-	-	7,8	7,1	7,0	6,7

Muestra: Individuos 18+, 1.^a entrevista entre 87.II y 96.I

Consideremos, en consecuencia, los siguientes modelos de desgaste individual (A^j ; $j = C; D$) y reentrada (R):

$$P(A^j = 1) = \phi(X_k^o + Z_k^{\otimes j}); j = C; D$$

$$P(R = 1) = \phi(X_k^o + Z_k^{\otimes r})$$

donde i denota al individuo, X es un vector de características personales y del hogar, $^o j$ y $^o r$ son los correspondientes vectores de parámetros, Z es una función lineal de controles del número, trimestre y el año de la entrevista y $^{\otimes j}$ y $^{\otimes r}$ sus correspondientes parámetros. Finalmente, ϕ es la función de distribución logística.

Las estimaciones que presentaremos se basan en dos muestras alternativas. La primera está formada por individuos que responden a la primera entrevista, mientras que la segunda muestra no está

restringida (esto es, considera también los entrantes en 2 y sucesivas entrevistas). En ambos casos, tomamos una muestra aleatoria del 10 por ciento del total de la muestra inicial (que combina más de 40 trimestres consecutivos), con las que estimamos modelos de la probabilidad de desgaste tipos C y D, así como de reentrada. Aunque muchos coeficientes son estadísticamente significativos, en gran medida debido al gran tamaño muestral, los modelos en conjunto tienen una capacidad explicativa muy limitada (el pseudo R^2 sólo supera ligeramente el 5 por ciento en los modelos de desgaste y el 8 por ciento en los modelos de reentrada).

En vez de presentar las estimaciones de los coeficientes y estadísticos t, en el Cuadro n.º 6 mostramos las probabilidades estimadas para individuos tipo, obtenidas a partir de un caso base. El caso base escogido

Cuadro n.º 6. Probabilidades estimadas de desgaste muestra y reentrada

	Desgaste tipo C		Tipo D	Reentrada	
	Muestra 1	Muestra 2	Muestra 1	Muestra 1	Muestra 2
Caso base	19,3	17,5	25,2	56,4	60,1
Entrevista 3	17,5	15,6	—	—	—
Entrevista 4	16,9	14,7	—	48,1	51,3
Entrevista 5	16,5	13,2	—	41,3	43,7
Entrevista 6	15,9	13,7	23,2	35,0	35,2
2.º trimestre	19,3*	17,4*	25,8*	49,0	52,6
3.º trimestre	31,6	29,2	29,9	46,0	49,7
4.º trimestre	24,6	23,0	28,0	65,2	68,0
Características personales					
Mujeres	19,3*	17,4*	26,2	55,9*	59,4*
Solteros	16,4	16,3	23,8	56,2*	58,5*
Edad 25	23,4	21,6	32,5	50,4	54,1
Edad 45	16,6	15,2	19,7	61,4	63,8
Edad 55	14,4	13,5	17,3	62,6	65,6
Edad 65	14,8	13,8	18,2	64,1	66,6
Edad 75	15,1	14,4	20,3	60,8	62,1
Edad 82	17,3	16,1	28,2	49,0	44,4
Sin estudios o primarios	16,6	14,3	23,0	56,3*	59,1*
Estudios universitarios	22,9	20,3	29,9	57,4*	58,4*
Cónyuge del cabez fam.	18,3	17,2*	22,7	60,8	62,4
Hijo del cabeza fam.	18,1	15,6	25,2*	49,6	51,5
Otros miembros hogar	29,0	24,4	37,7	40,3	43,3
Características del hogar					
Número de empleados en el hogar	19,2*	17,5*	25,9	55,8*	59,8*
Sin hijos 0-5	21,6	20,4	27,5	57,5*	60,9*
Hijos 6-10	17,3	15,6	22,8	57,1*	60,4*
Hijos 11-15	17,4	15,8	22,8	57,3*	60,6*
Tamaño del hogar 1	29,6	27,6	36,7	50,9	55,9
Tamaño del hogar 3	14,5	13,0	21,1	53,8	59,1*
Tamaño del hogar 3	12,8	11,3	18,6	52,8	55,9
Tamaño del hogar 5	11,9	10,5	19,0	47,1	51,8

Caso base: entrevista 2, hombre, edad 35, est. secundario, cabeza fam, casado, con al menos un hijo 0-5 viviendo en Madrid con otro adulto (tamaño del hogar 2), empleado a tiempo completo y con experiencia; tomando el resto de variables la respuesta negativa.

Otras variables: periodo de la primera entrevista (39), edad, edad², edad³.

Criterio selección muestra: 1/10 de la muestra total (87-II a 97-IV).

Muestra 1: todos los individuos. Muestra 2: individuos que responden a la primera entrevista.

Nota: el asterisco denota diferencias respecto al caso base no significativas (a nivel de significación del 5%).

Cuadro n.º 6. Probabilidades estimadas de desgaste muestra y reentrada (continuación)

	Desgaste tipo C		Tipo D	Reentrada	
	Muestra 1	Muestra 2	Muestra 1	Muestra 1	Muestra 2
Status laboral y actitudes ante el mercado de trabajo					
Desempleado	19,2	17,6*	25,3*	56,1	62,4*
Inactivo	19,0	17,6*	23,9*	56,8	62,9*
Contrato temporal	17,9	16,9	25,1*	53,7	57,5
Funcionario	20,4	18,4	26,3*	53,2	62,1
Sin empleo a TC	18,7*	17,3*	25,4*	53,2	61,6
buscando empleo	19,9*	17,3*	27,7	54,6*	57,0*
Sin experiencia previa	19,4*	17,9*	24,9*	58,0	61,7
Ayuda familiar	16,7	14,9	20,3	59,1*	63,3*
Estudiando	17,5	15,4	24,5*	65,0	68,5
Disponibilidad a cambiar residencia	20,4	19,5	28,5	53,0	58,0*
Trabajo doméstico	17,1	16,1	19,7	59,9	59,3*
Sur	16,9	15,8	23,4	57,5*	60,5*
Centro	15,5	14,6	21,0	56,6*	61,1*
Este	15,4	14,5	21,7	54,5	56,9
Norte	14,7	13,8	20,7	52,2	56,6
Observaciones	449595	427164	185470	50367	45415
Wald WO: todos los coef. = 0	10075.2	9467.7	4867.7	679.5	4709.2
Wald W1: var. fict. ent. y trim.	2696.8	2642.8	164.4	2068.2	2149.2
Wald W2: var. fict. periodo 1.ª ent.	2018.1	1853.3	577.1	469.4	425.2
Wald W3: conjuntamente W1 y W2	4881.1	4527.8	886.3	2517.7	2568.6
Pseudo R ²	.040	.041	.033	.070	.078

Caso base: entrevista 2, hombre, edad 35, est. secundario, cabeza fam, casado, con al menos un hijo 0-5 viviendo en Madrid con otro adulto (tamaño del hogar 2), empleado a tiempo completo y con experiencia; tomando el resto de variables la respuesta negativa.

Otras variables: periodo de la primera entrevista (39), edad, edad², edad³.

Criterio selección muestra: 1/10 de la muestra total (87-II a 97-IV).

Muestra 1: todos los individuos. Muestra 2: individuos que responden a la primera entrevista.

Nota: el asterisco denota diferencias respecto al caso base no significativas (a nivel de significación del 5%).

es un hombre de 30-39, cabeza de familia y casado, con al menos un hijo entre 0-5, observado en segunda entrevista, con estudios secundarios, empleado a tiempo completo con contrato indefinido, teniendo experiencia previa.

En primer lugar, nótese que la probabilidad de desgaste para el individuo base es menor que para el resto de individuos considerados y su probabilidad de reentrada es mayor en la muestra restringida a los participantes en todas las encuestas. Segundo, en todos los modelos las probabilidades de desgaste y reentrada decrecen con el número de entrevista. Tercero, el desgaste es notoriamente mayor en tercera entrevista que en otras. Alternativamente, la reentrada es significativamente más importante en el cuarto trimestre (esto es, cuando las personas no localizables en el tercero por causa de las vacaciones son reentrevistadas). En todos los casos la hipótesis nula de que el número entrevista o el trimestre de la entrevista no tiene efecto sobre la probabilidad de desgaste o reentrada resulta rechazada.

Muchas características individuales tienen notoria influencia en las probabilidades de desgaste muestral. Por ejemplo, los solteros tienen menos probabilidad de dejar la muestra que los casados. Según la edad, el mínimo desgaste se observa a los 50 y la máxima reentrada a los 60 años de edad. Además, ambas probabilidades crecen con el nivel de estudios. Por otra parte, los cabeza de familia y sus cónyuges tienen menor probabilidad de dejar la muestra que otros miembros del hogar. Finalmente, la probabilidad de volver a la muestra, después de no haber

respondido al menos una vez, es menor para los hijos que para el resto del núcleo familiar.

Respecto al efecto de otras características, reseñamos que el número de empleados en el hogar apenas tiene efecto sobre la propensión a dejar la muestra. Alternativamente, la presencia de hijos menores de 16, reduce significativamente la probabilidad de desgaste y, por contra, no afecta a las probabilidades de reentrada. Mientras que la propensión al desgaste se reduce con el tamaño del hogar, la de reentrada alcanza un máximo en hogares de tres adultos.

También hemos considerado un gran número de controles de las características del empleo o las actitudes frente al mercado de trabajo. La situación ante la actividad y las variables que describen las características del empleo, aunque son significativas en la mayoría de los casos, tienen un efecto muy limitado sobre ambas probabilidades. Alternativamente, estar estudiando o ayudando en el negocio familiar reducen notoriamente el desgaste de tipo C y, en menor medida, el de tipo D. La predisposición a cambiar de residencia incrementa la probabilidad de desgaste y disminuye la probabilidad de reentrada. Finalmente, los individuos que llevan a cabo tareas domésticas tienen menor predisposición a dejar la muestra, especialmente sobre el periodo de un año.

Finalmente, cabe destacar que se detectan diferencias muy notorias según la región de residencia. Tanto el desgaste tipo C como el D son significativamente menores en el Norte, Centro y Este que en el Sur y Madrid. A cambio, éstas

últimas regiones muestran unas tasas de reentrada significativamente mayores.

3.5. Consecuencias de la no respuesta en la probabilidades de transición

En esta sección evaluamos contrastes de la hipótesis de que el desgaste muestral no causa sesgos importantes en las estimaciones de transiciones basadas en los enlaces trimestrales de la información de la EPA. Con dicho objeto, comparamos las tasas de transición de los participantes completos (aquellos que responden a las seis entrevistas) y otros participantes (aquellos que responden menos de seis entrevistas) en un modelo con cuatro estados posibles: ocupado a tiempo completo (35 horas o más), ocupado a tiempo parcial (35 horas), desempleado e inactivo. Diferencias estadísticas significativas entre las dos estimaciones constituyen evidencia contra la hipótesis nula de que el desgaste no afecta a los resultados. Veamos, a continuación, una breve descripción de la metodología empleada.

Sea $Y_{ij} = 1$ si un individuo se mueve desde el estado i en el periodo t al estado j en el periodo $t + 1$, y sea D un indicador binario sobre si la persona es un participante completo ($D = 1$) o no. El desgaste ocasiona sesgo si las siguientes probabilidades de transición entre los estados i y j entre t y $t + 1$ (λ_{ij}) son diferentes:

$$\lambda_{ij}^{(0)}(x) = E(Y_{ij} | X = x; D = 0) \neq E(Y_{ij} | X = x; D = 1) = \lambda_{ij}^{(1)}$$

donde X es un conjunto de condicionantes. Cuando X es un vector

que sólo toma un número finito de valores discretos podemos particionar la muestra según dicho conjunto de valores y llevar a cabo el contraste no paramétricamente. En cuyo caso basta con analizar la diferencia entre las transiciones observadas $I_{ij}^{(D)}$ para los dos grupos de observaciones ($D = 0; 1$). Los valores críticos del contraste están basados en el hecho de que la diferencia $I_{ij}^{(0)} - I_{ij}^{(1)}$ dividida por su error estándar estimado se distribuye, siempre que haya un número suficiente de observaciones en cada valor del vector X , como una t de Student.

Los Gráficos n.ºs 6 y 7 muestran los valores del estadístico t para celdillas caracterizadas por el sexo, el grupo de edad y la situación laboral en el primer periodo. En una amplia mayoría de casos, el estadístico cae dentro de la región de aceptación $(-2,2)$ de la hipótesis nula, lo que indica que el desgaste y/o la reentrada no causa importantes problemas de selección.

Sin embargo la evidencia general oculta algunos casos particulares de interés. Así, cuando concentramos nuestra atención en rangos de edades extremos y transiciones particulares detectamos que el contraste del tránsito desde la inactividad al empleo se sitúa en la región de rechazo para los jóvenes de 25-30. Asimismo, detectamos diferencias significativas en las transiciones de empleo a inactividad (retiro) para los individuos de más de 50 años. En consecuencia, la evidencia muestra que debemos ser especialmente cuidadosos en el análisis de las transiciones en el mercado de trabajo de los jóvenes y las personas acercándose a la edad de retiro legal.

Gráfico n.º 6. Estadístico t de la comparación de las probabilidades de transición en la muestra balanceada y no balanceadas. Hombres

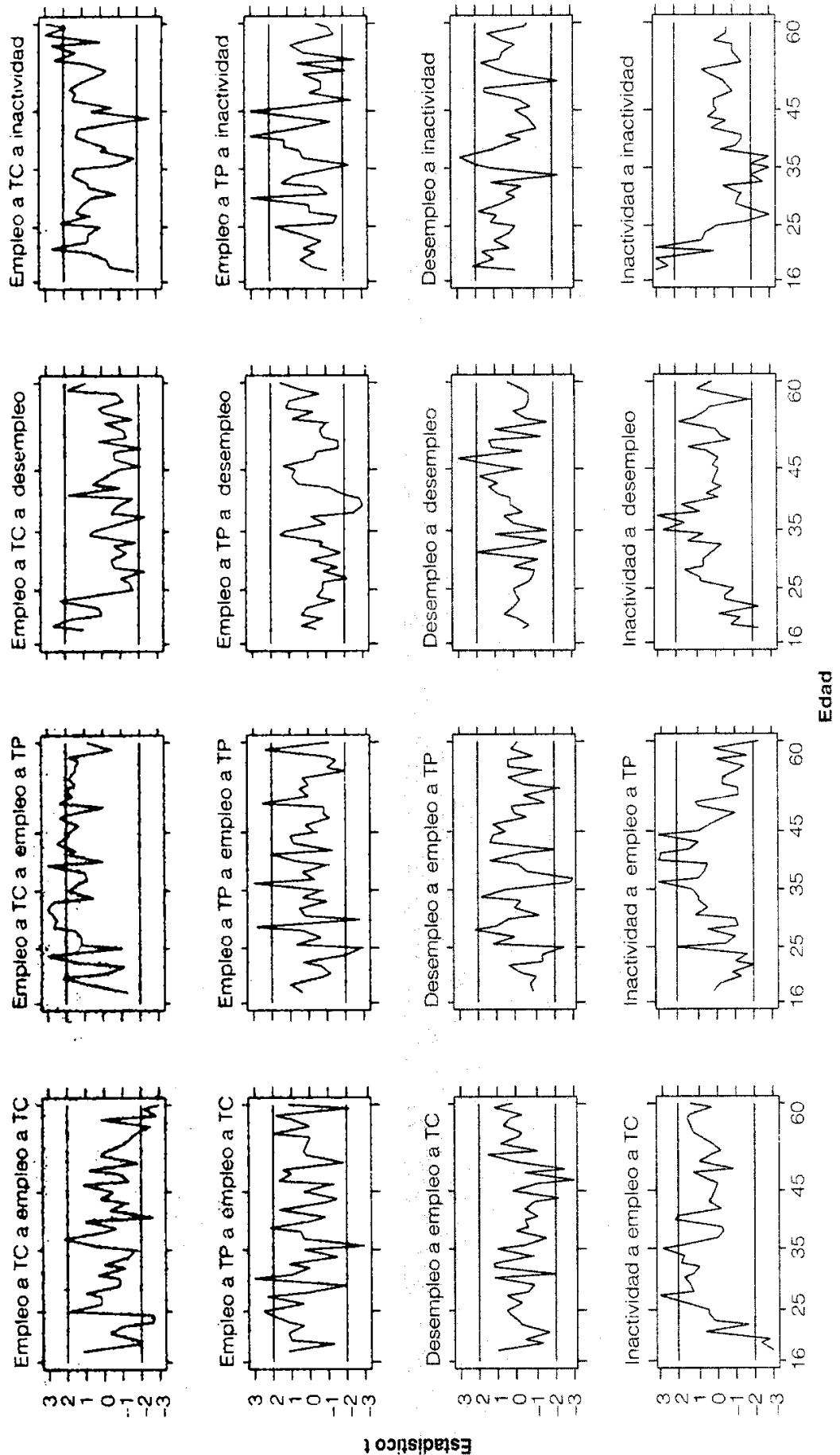
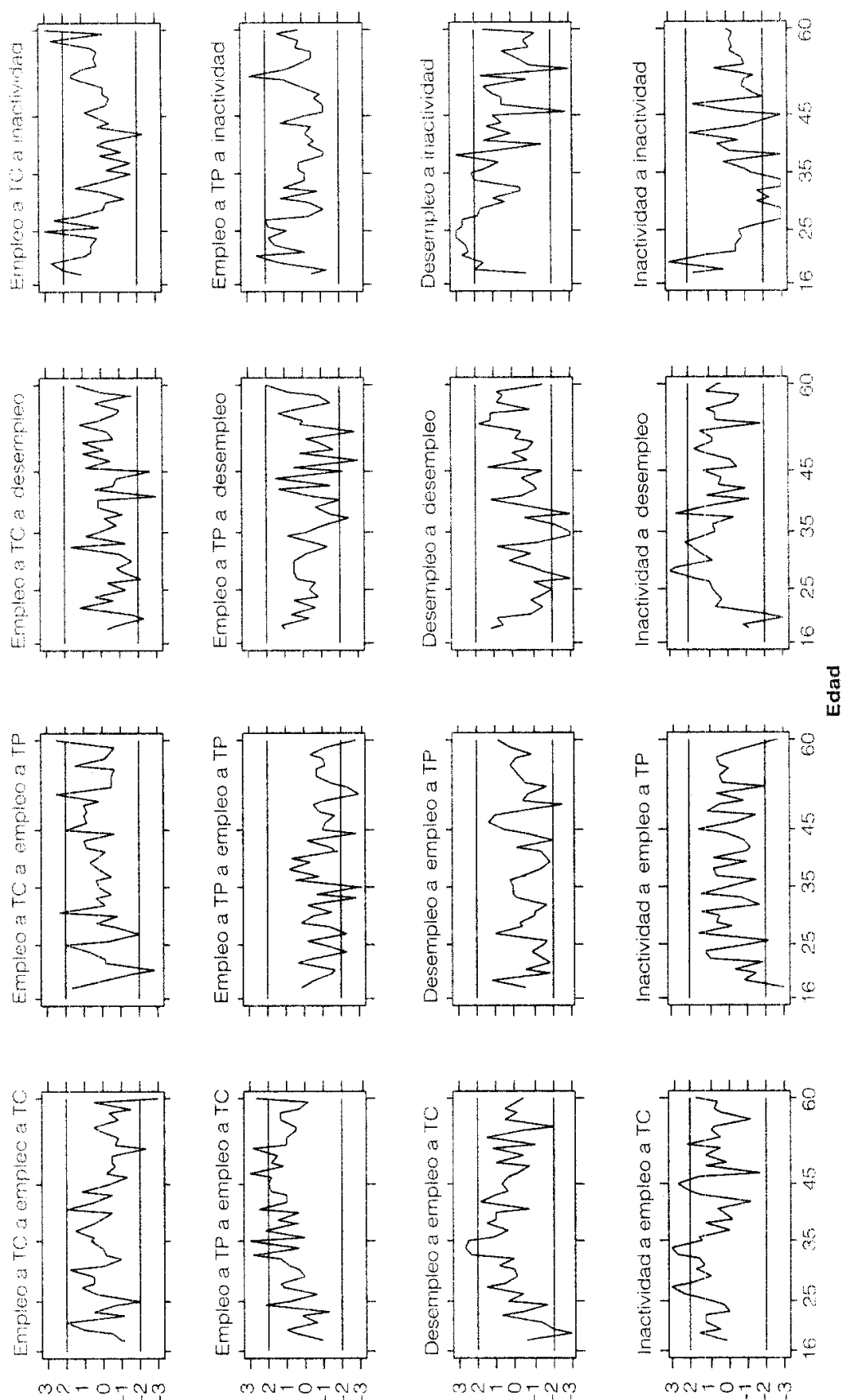


Gráfico n.º 7. Estadístico t de la comparación de las probabilidades de transición en la muestra balanceada y no balanceada. Mujeres



4. INFORMACIÓN RETROSPECTIVA VS. MEDICIÓN REITERATIVA

En esta sección evaluamos la bondad de la preguntas retrospectivas del segundo trimestre de la EPA al objeto de estudiar transiciones entre estados laborales en el mercado de trabajo. En nuestro análisis consideramos que existe un error de clasificación o «recall error» cuando la respuesta a la pregunta retrospectiva no coincide con la respuesta a la entrevista $t - 4$.

En el Cuadro n.º 7 presentamos evidencia sobre el error de clasificación entre la pregunta retrospectiva (PR) y la respuesta a la pregunta de la entrevista corriente (PC) según diversas características del individuo y el hogar. Según la relación con el cabeza de familia, el error de clasificación es mayor para los hijos y sus cónyuges. Según la edad, la mala clasificación derivada de la pregunta retrospectiva es más frecuente para los jóvenes y los individuos de edad avanzada. Alternativamente, no se observan diferencias sistemáticas según el sexo y el nivel de educación. Pero sobre todo, destacamos las importantes diferencias entre desempleados y ocupados o inactivos (el error de clasificación para los primeros es cuatro veces mayor que para los segundos). Además, el error de clasificación aumenta con el número de adultos en el hogar, desde 3,8 por ciento para hogares con sólo un adulto hasta 9,0 en hogares con más de cinco adultos. Finalmente, también se observan importantes diferencias entre regiones. El error de clasificación en Andalucía, Canarias o Ceuta y Melilla dobla al de Asturias y La Rioja. Sin duda, las diferencias en el nivel

de desempleo en las regiones mencionadas explican gran parte de las diferencias observadas. Así pues, la información retrospectiva tiende a sistemáticamente sobrestimar la permanencia en el mismo estado y subestima las transiciones a otros estados. Es decir, la pregunta retrospectiva tiende a subestimar la movilidad en el mercado de trabajo.

Como complemento al análisis descriptivo, presentamos en el Cuadro n.º 8 un análisis de la probabilidad de que la RR no coincida con la RC, esto es, de la probabilidad de observar error de clasificación. En nuestro análisis consideramos tres grupos de población (grupo de edad 18-34, grupo 35-69 y adultos que viven solos) y dos modelos en cada caso: el modelo A, que considera la información sobre transiciones en el pasado inmediato y el modelo B, que no considera dicha información. Nótese que en todos los casos el poder explicativo, medido por el pseudo R^2 , de los modelos tipo A es sustancialmente mayor.

Entre la evidencia obtenida destacamos los siguientes puntos. Primero, cuanto mayor es el número de cambios de estados mayor es la probabilidad de observar un error de clasificación, siendo la transición más influyente la del $t - 4$ al $t - 3$. La situación laboral un año atrás también afecta a la probabilidad de error. Por ejemplo, para un individuo de 20-34 la probabilidad base es de un 5,0, 11,9 y 6,2 por ciento, según su situación sea la de empleado, desempleado o inactivo, respectivamente. Consecuentemente, la información sobre estados y transiciones es crucial a la hora de explicar la probabilidad de observar un error.

Cuadro n.º 7. **Fracción de no coincidencia entre la respuesta retrospectiva y la respuesta a la encuesta**

	Nivel de estudios								Total
	-De primarios		Secundarios		Universitarios		Todos		
	Hom,	Muj.	Hom.	Muj.	Hom.	Muj.	Hom.	Muj.	
Según relación con el cabeza de familia									
Cabeza familia	4,9	4,7	3,3	7,1	2,0	3,1	4,4	4,7	4,5
Cónyuge	7,6	5,8	4,3	9,2	2,1	5,6	6,0	6,1	6,1
Hijos	11,8	12,0	9,8	10,2	9,8	11,3	10,8	11,1	10,9
Cónyuges de hijos	8,9	7,5	6,6	14,9	3,0	7,6	8,3	8,5	8,4
Otros	7,7	5,3	9,5	8,7	11,2	9,4	8,2	6,2	7,1
Por grupos de edad									
18-24									
25-29	14,4	14,6	10,1	10,0	7,5	11,2	11,9	12,0	11,9
30-34	9,3	13,0	8,9	13,6	11,7	13,4	9,6	13,3	11,4
35-39	7,4	10,1	4,7	10,0	4,9	7,8	6,4	9,7	8,0
40-44	5,7	8,4	3,2	8,0	2,1	4,7	4,7	7,8	6,3
45-49	5,0	7,4	2,7	6,1	1,4	3,8	4,2	6,9	5,6
50-54	4,8	5,8	3,5	4,8	1,6	2,1	4,3	5,5	4,9
55-59	5,0	4,9	3,2	4,3	2,3	3,6	4,6	4,8	4,7
60-64	6,0	3,6	3,8	4,3	2,7	2,9	5,6	3,6	4,6
65 +	5,4	2,5	5,0	2,4	3,4	2,1	5,2	2,5	3,8
	2,4	1,4	2,6	1,4	2,5	1,6	2,5	1,4	1,9
Según el status laboral un año atrás									
Empleado	4,3	6,2	3,6	5,5	2,0	3,3	3,9	5,5	4,4
Desempleado	19,6	23,2	26,5	26,1	28,0	23,8	21,5	24,1	22,8
Inactivo	6,8	4,6	7,1	7,0	7,3	9,4	6,9	5,2	5,7
Según el número de adultos del hogar									
Uno	4,8	3,1	4,0	6,1	2,5	2,4	4,5	3,3	3,8
Dos	5,3	6,7	3,7	10,2	2,3	6,2	4,7	7,2	6,0
Tres	6,2	5,7	7,6	8,5	5,3	8,3	6,4	6,3	6,4
Cuatro	7,2	6,9	8,5	9,6	6,7	9,4	7,5	7,7	7,6
Cinco +	9,4	8,0	9,3	10,4	7,6	10,6	9,1	8,8	9,0
Según la región de residencia									
Andalucía	10,5	9,5	10,2	12,3	6,8	9,6	10,1	9,9	10,0
Aragón	3,5	4,4	5,2	9,1	4,6	6,8	4,0	5,5	4,7
Asturias	4,3	3,9	5,3	6,1	4,5	5,9	4,5	4,5	4,5
Baleares	7,3	7,9	8,3	11,1	5,4	9,1	7,3	8,5	7,9
Canarias	10,9	10,7	11,6	13,2	7,3	11,6	10,7	11,2	11,0
Cantabria	5,8	4,2	6,9	10,6	5,0	9,5	6,0	6,1	6,0
Castilla-L	5,5	5,1	5,8	8,8	5,0	8,0	5,5	5,9	5,7
Castilla-M	5,0	5,4	6,5	9,7	5,3	8,9	5,3	6,6	6,0
Cataluña	5,1	5,6	6,6	8,6	3,3	6,9	5,3	6,3	5,8
Com, Val,	5,4	6,6	6,7	9,3	4,8	8,0	5,6	7,2	6,4
Extremadura	10,3	9,1	7,5	12,1	5,9	8,1	9,5	9,4	9,5
Galicia	5,3	5,2	6,1	7,8	5,2	8,0	5,4	5,9	5,7
Madrid	4,0	2,6	4,9	5,6	3,5	5,6	4,1	3,7	3,9
Murcia	7,8	8,6	8,0	14,8	5,8	10,0	7,7	9,6	8,7
Navarra	5,9	7,0	6,9	8,7	4,1	8,3	5,9	7,6	6,7
País Vasco	4,6	5,0	6,8	9,3	4,6	6,9	5,2	6,2	5,7
La Rioja	2,9	2,3	2,6	6,0	1,9	4,1	2,7	3,2	3,0
Ceuta y Melilla	11,1	11,7	8,6	12,8	8,1	19,4	10,2	12,6	11,4
Total	6,7	6,6	7,2	9,7	5,0	8,1	6,6	7,3	7,0

Muestra: edad, 18, 1.ª entrevista entre 87.II y 96.IV.

Cuadro n.º 8. Modelos de error de clasificación o recall error

	Edad 20-34				Edad 35-69				Hogares de un miembro			
	Modelo A		Modelo B		Modelo A		Modelo B		Modelo A		Modelo B	
	Coef.	T-st	Coef.	T-st	Coef.	T-st	Coef.	T-st	Coef.	T-st	Coef.	T-st
Constante	-2,648	20,6	-1,984	16,7	-3,224	35,4	-2,351	29,2	-2,924	8,9	-2,140	7,5
Constante después de 1992	0,351	8,4	0,337	8,8	0,436	9,2	0,327	7,9	0,626	3,0	0,416	2,2
Tendencia (88-II = 1)	-0,068	9,4	-0,037	10,6	-0,92	11,0	-0,087	11,8	-0,089	2,6	-0,051	1,6
Características personales												
Hom.	-0,106	4,5	-0,013	,6	0,051	1,1	0,047	1,1	0,371	3,6	0,452	4,9
Edad 25-29	0,006	,2	-0,091	3,5	—	—	—	—	-0,196	,6	-0,219	,8
Edad 30-34	-0,062	4,5	-0,352	10,6	—	—	—	—	0,018	,1	-0,344	1,3
Edad 35-39	—	—	—	—	—	—	—	—	-0,070	,2	-0,440	1,8
Edad 40-44	—	—	—	—	-0,018	,4	-0,057	1,5	-0,374	1,3	-0,693	2,7
Edad 45-49	—	—	—	—	-0,077	1,6	-0,168	4,0	-0,142	,5	-0,556	2,2
Edad 50-54	—	—	—	—	-0,095	1,8	-0,225	5,0	-0,238	,9	-0,834	3,2
Edad 55-59	—	—	—	—	-0,095	1,8	-0,267	5,7	-0,285	1,0	-0,739	2,8
Edad 60-64	—	—	—	—	-0,123	2,2	-0,373	7,5	-0,866	2,8	-1,424	5,3
Edad 65-69	—	—	—	—	-0,654	9,9	-0,996	16,1	-1,410	4,7	-2,133	8,0
Educación primaria	0,013	,2	0,036	,6	0,036	1,1	-0,121	4,4	0,302	2,4	0,307	2,7
Educación secundaria	-0,068	1,0	-0,089	1,5	0,016	,3	-0,269	5,3	0,294	1,5	0,240	1,4
Educación universitaria	0,041	,5	0,124	1,9	-0,064	,9	-0,466	7,6	0,083	,4	-0,014	,1
Cónyuge de cabeza fam.	0,397	7,3	0,487	9,6	-0,038	,8	0,010	,2	—	—	—	—
Hijo del cabeza fam.	0,242	3,9	0,345	6,0	0,106	1,9	0,126	2,5	—	—	—	—
Cónyuge del hijo	0,355	3,7	0,426	4,7	0,063	,6	0,181	1,9	—	—	—	—
Nieto, cónyuge de nieto	0,450	4,6	0,536	5,9	-0,802	1,0	-0,589	,9	—	—	—	—
Padre del cc. o su cónyuge	—	—	—	—	0,018	,1	0,038	,7	—	—	—	—
Otros	0,303	3,0	0,316	3,4	0,032	,3	0,011	,1	—	—	—	—
Características del hogar												
Número de empleados	0,007	,5	0,022	1,8	0,056	3,3	0,060	3,9	—	—	—	—
Tamaño del hogar 2	-0,168	1,6	-0,229	2,4	0,082	1,3	0,025	,4	—	—	—	—
Tamaño del hogar 3	-0,092	,9	-0,131	1,39	0,167	2,7	0,176	3,2	—	—	—	—
Tamaño del hogar 4	-0,066	,6	-0,089	,9	0,130	2,0	0,196	3,3	—	—	—	—
Tamaño del hogar 5+	-0,010	,1	-0,040	,4	0,179	2,5	0,271	4,2	—	—	—	—
Número de hijos 0-15	0,027	2,1	0,045	3,9	0,006	,4	0,043	3,3	0,255	3,2	0,297	4,2
Status laboral retrospectivo y transiciones entre estados												
Empleados en t-4	-0,294	10,3	-0,276	10,0	-0,667	20,5	-0,412	13,3	-1,008	8,1	-0,853	7,2
Desempleados en t-4	0,651	24,0	1,097	43,8	0,524	13,8	1,667	53,1	-0,014	,1	1,027	8,8
Transición de t-1 a t	0,597	22,9	—	—	1,158	32,1	—	—	0,883	6,4	—	—
Transición de t-2 a t-1	0,487	18,7	—	—	0,813	22,7	—	—	0,807	6,0	—	—
Transición de t-3 a t-2	0,338	13,6	—	—	0,522	15,3	—	—	0,795	6,1	—	—
Transición de t-4 a t-3	1,948	87,3	—	—	2,814	97,6	—	—	2,412	21,5	—	—
Estudiando	-0,093	3,5	-0,195	7,8	—	—	—	—	-0,130	,8	0,054	,4
Observaciones	116390		116390		189611		189611		15166		15166	
Test de Wald ($\chi^2 = 0$)	161178,3		5613,2		21170,7		6863,4		1520,3		744,6	
Pseudo R ²	0,1982		0,0388		0,2801		0,0908		0,2813		0,1378	
Prob. (caso base 88-II)(%)	6,2		11,4		3,5		8,0		4,7		10,0	
Prob (base case 92-II)(%)	7,1		12,7		4,1		8,5		6,6		12,7	

Categorías omitidas: 2.º trimestre de 1988, mujer, edad 18-24, sin estudios, cabeza familia, tamaño hogar 1, inactiva en t-4, sin transiciones entre t-4y t (sólo en el modelo A), el resto de variables toman respuesta negativa o valor cero.

Por otra parte, en nuestro análisis encontramos evidencia de que la probabilidad de error decae en el tiempo. Por ejemplo, para el caso base en el grupo de jóvenes (modelo A), mientras que la probabilidad de error en el 2.º trimestre de 1988 era de un 6,2 por ciento, en el 2.º trimestre de 1991 pasó a un 5,45. Aunque en el 2.º trimestre de 1992, como consecuencia del cambio de constante ligado al cambio de cuestionario, dicha probabilidad vuelve a subir a 7,1 por ciento, posteriormente vuelve a bajar hasta el 5,2 por ciento en el 2.º trimestre de 1997.

Frecuentemente, se ha argumentado que la amplitud del cuestionario resulta crucial a la hora de determinar las tasas de respuesta y la probabilidad de observar un error de clasificación. Nosotros mismos hemos aportado suficiente evidencia de lo primero y, tal como mostramos a continuación, de lo segundo⁷. En nuestro análisis detectamos que, por ejemplo, para los individuos de 35-69 la probabilidad del caso base aumenta desde 2,1 al 2,6 por ciento, y desde el 2,8 al 4,4 por ciento en el caso de hogares con un único adulto.

También hemos encontrado otros rasgos de interés. Primero, mientras que para los jóvenes el error de clasificación alcanza el mínimo en hogares de dos adultos, para el grupo de los adultos (35-69) el mínimo se alcanza en hogares de tres miembros. Segundo, no se detectan diferencias importantes según el

nivel educativo para los jóvenes pero sí para los adultos y los hogares de un sólo miembro. Tercero, según la relación con el cabeza de familia, el error es relativamente más importante para las esposas jóvenes, los hijos y sus cónyuges. Cuarto, los jóvenes aún estudiando tienen una menor probabilidad de error. Finalmente, no se detectan diferencias importantes según el sexo, excepto para hogares de un sólo adulto.

5. CONCLUSIONES

En el presente artículo hemos analizado las implicaciones del desgaste muestral para el estudio de transiciones en el mercado de trabajo usando información de la EPA. Si nuestro objetivo es predecir la evolución futura de la participación debemos estar razonablemente seguros de que la fuente es, en un sentido estadístico, correcta. A continuación exponemos nuestras principales conclusiones.

Primero, la no respuesta no debe ser ignorada. A pesar de los esfuerzos del INE para compensar el continuo incremento en la fracción de hogares que rehusan participar, la no respuesta es probable que sea cada vez más importante debido sobre todo al deslizamiento de la composición de la población hacia grupos más propensos a no responder (como lo son los individuos de más nivel educativo).

Segundo, mostramos que como resultado de las diferencias en la no respuesta o en el desgaste según la educación y el grupo de edad, la EPA tiende a subestimar el empleo y la participación de los jóvenes y a sobrestimar el empleo y la

⁷ De hecho, el mismo INE reconoce que el número de preguntas, esto es, la longitud del cuestionario reduce las tasas de respuesta y aumenta los errores de clasificación.

participación de los mayores con menor nivel educativo.

Tercero, las pautas de desgaste y reentrada en la muestra pueden ser representadas por una cadena de Markov de primer orden, en la que las probabilidades de ambos eventos dependen de características observables del individuo y otras variables, entre las que destacamos el tiempo, que aproxima el esfuerzo del INE para controlar la dinámica muestral.

Cuarto, en consonancia con lo encontrado para otros estudios del mismo ámbito, encontramos poca evidencia de

el desgaste muestral cause, en general, problemas importantes de selección. Aún así, detectamos algunos problemas de selección en rangos extremos de edades.

Finalmente, encontramos que la información retrospectiva contiene errores sistemáticos. En particular, la información retrospectiva subestima la movilidad entre estados en el mercado de trabajo. Sin embargo, tanto la modesta magnitud de los errores como su importancia decreciente, sustentan el uso de la información retrospectiva para el estudio de las transiciones en el mercado de trabajo.

REFERENCIAS BIBLIOGRÁFICAS

- BOVER, O., ARELLANO, M. y BENTOLILA, S. (1996): «Unemployment duration, benefit duration and the business cycle», *Economic Studies*, 56, Bank of Spain.
- ALBA-RAMÍREZ, A. (1998): «Reemployment probabilities of young workers in Spain», *Investigaciones Económicas*, 22(2).
- (1997): «Labor force participation and transition of older workers in Spain», Universidad Carlos III de Madrid, WP 97-39.
- BECKETTI, S., W. GOULD, L. LILLARD, y F. WELCH (1988): «The Panel Study of Income Dynamics after fourteen years: An evaluation», *Journal of Labor Economics*, 6, 472-492.
- HOROWITZ, J.L. y MANSKI, C.F. (1995): «Censoring of outcomes and regressors due to survey nonresponse: Identification and estimation using weights and imputations», *Journal of Econometrics*, 84, 37-58.
- INE (1991): *Encuesta de Población Activa: Descripción de la Encuesta, Definiciones e instrucciones para la cumplimentación del cuestionario*, Madrid.
- (1993): *Incidencia en los Trabajos de Campo en la Encuesta de Población Activa 1992*, Madrid.
- (1994): *Encuesta de Población Activa: Tratamiento de la Información*, Madrid.
- (1995): *Evaluación de la Calidad de los Datos de la Encuesta de Población Activa*: 1995, Madrid.
- (1996): *Estadística de Flujos: Metodología*, Madrid.
- LITTLE, R.J.A., y D.B. RUBIN (1987): *Statistical Analysis with Missing Data*, Wiley, New York.
- PERACCHI, F. y WELCH, F. (1994): «Trends in Labor Force Transitions of Older Men and Women», *Journal of Labor Economics*, 12(2): 210-242.
- (1995): «How representative are matched Cross-sections? Evidence from the Current Population Survey», *Journal of Econometrics*, 68, 153-180.
- PERACCHI, F. (1998): «The European Community Household Panel Survey: A review», mimeo.
- VILLAGARCÍA, T. (1995): «Existe un sesgo de inactividad en la encuesta de población activa?», *Documento de Trabajo* 95-04, Universidad Carlos III de Madrid.