

**SERIES ENLAZADAS DE LOS PRINCIPALES AGREGADOS
NACIONALES DE LA EPA, 1964-2009
(REGDAT_EPA_NAC_V10)**

*Angel de la Fuente**

D-2012-07

Octubre 2012

*Instituto de Análisis Económico (CSIC)

Este trabajo forma parte de un proyecto de investigación cofinanciado por BBVA Research y la Fundación Rafael del Pino. Agradezco también la financiación del Ministerio de Economía y Competitividad a través del proyecto ECO2011-28348 y los comentarios y sugerencias de Luis González Calbet, Alfredo Cristóbal y dos evaluadores anónimos de Estadística Española.

Dirección para correspondencia: Angel.delaFuente@uab.es

En la página web de los Documentos de Trabajo de la Dirección General de Presupuestos figura, además de este documento, tres anexos en Excel con las series de datos:

<http://www.sepg.pap.minhap.gob.es/sitios/sepg/es-ES/Presupuestos/Documentacion/Paginas/Documentacion.aspx>

Los Documentos de Trabajo de la Dirección General de Presupuestos no representan opiniones oficiales del Ministerio de Hacienda y Administraciones Públicas. Los análisis, opiniones y conclusiones aquí expuestos son los del autor, con lo que no tiene que coincidir, necesariamente la citada Dirección. Ésta considera, sin embargo, interesante la difusión del trabajo para que los comentarios y críticas que suscite contribuyan a mejorar su calidad.

Resumen

En el presente trabajo se recopilan y extienden diversas series históricas de los principales agregados nacionales de la Encuesta de Población Activa (EPA) y se construyen nuevas series anuales homogéneas de las mismas variables para el período 1964-2009 corrigiendo algunas de las rupturas que persisten en las series históricas más recientes del INE.

Palabras clave: EPA, series históricas, empleo.

Código JEL: J21

1. Introducción

En el presente trabajo se recopilan y extienden diversas series históricas de los principales agregados nacionales de la Encuesta de Población Activa (EPA) y se construyen nuevas series anuales homogéneas de las mismas variables para el período 1964-2009. Estas nuevas series extienden hacia atrás y enlazan entre sí las series históricas más recientes del INE y corrigen algunas de las rupturas que persisten en las mismas. Aunque las correcciones realizadas no cambian cualitativamente la senda de los agregados de interés, sí eliminan parte del ruido contenido en las series existentes, lo que a su vez debería redundar en una mayor precisión y fiabilidad de los resultados de análisis estadísticos en los que los principales indicadores del mercado de trabajo aparecen como variables dependientes o independientes.

El trabajo está organizado como sigue. En la sección 2 se describe brevemente el diseño de la EPA y su metodología de estimación y se repasan los principales cambios metodológicos que pueden haber generado discontinuidades en las series de la encuesta a lo largo de su historia. Entre estos cabe destacar ciertos cambios de criterio en la clasificación de algunos colectivos durante los primeros años de la encuesta, la introducción de nuevos cuestionarios y de ciertas mejoras técnicas en la estimación y las sucesivas renovaciones de la muestra. En la sección 3 se describen las series de las que partiré en este trabajo, esto es, las sucesivas series históricas que el INE ha construido para el período posterior a 1976 y los intentos de reconstrucción de series homogéneas para el período 1964-76 que encontramos en la literatura. Una de estas series, la construida por Mas et al (2002) es la que se utiliza en la sección 4 como referencia básica para extender hacia atrás hasta el año 1964 cada una de las series históricas del INE. En las secciones 5 y 6 se aborda una revisión de las más recientes series históricas del INE con el fin de corregir ciertas discontinuidades presentes en las mismas. Seguidamente, en la sección 7 se recopilan algunas variables auxiliares que resultan de utilidad para poder pasar de las series de ocupados de la EPA a las series de empleo y horas trabajadas de Contabilidad Nacional. La sección 8 concluye con una descripción de los archivos de datos que acompañan al trabajo. También se incluyen dos anexos con detalles técnicos y un listado cronológico detallado de los distintos cambios metodológicos y otras incidencias que ha experimentado la EPA desde su aparición en 1964.

2. La Encuesta de Población Activa: diseño y metodología de estimación

La Encuesta de Población Activa es una de las principales fuentes de información sobre el mercado de trabajo español. En ella se ofrece información detallada sobre la población en edad de trabajar desglosada por su relación con la actividad económica (en activos e inactivos, con los primeros clasificados a su vez en ocupados y parados) y sobre un buen puñado de características de estos colectivos (incluyendo su distribución por sexo, edad, nivel educativo y nacionalidad) así como sobre su actividad laboral (duración de la jornada de

trabajo, tipo de contrato, etc.) o su búsqueda de empleo (duración de la misma y procedimientos utilizados).

La EPA se ha realizado desde 1964 con frecuencia (generalmente, aunque no siempre) trimestral¹ y proporciona información desagregada territorialmente (por regiones y/o provincias) desde 1972.² La encuesta se dirige a la población en edad de trabajar que reside en viviendas familiares principales, excluyendo por tanto los llamados hogares colectivos (cuarteles, residencias y prisiones, entre otros)³ y las viviendas secundarias o de temporada. En la actualidad, el tamaño inicial de la muestra es de unas 65.000 familias, de las que aproximadamente 60.000 llegan a ser entrevistadas, lo que supone unas 180.000 personas. La muestra se renueva por completo cada año y medio procediendo de forma escalonada de manera que cada hogar es encuestado durante 6 trimestres y cada trimestre se renueva una sexta parte de la muestra. Las encuestas se realizan a un ritmo uniforme a lo largo de las 13 semanas de cada trimestre, de forma que los resultados se refieren a "una semana típica" del trimestre.⁴

2.1. Diseño muestral

A efectos estadísticos y electorales, España se divide en provincias, municipios, distritos municipales y algo más de 30.000 secciones censales. Cada sección ha de incluir entre sus residentes entre 500 y 2.000 potenciales votantes. La división del territorio nacional en secciones o *seccionado* se actualiza al comienzo de cada año, coincidiendo con la revisión del censo electoral, y tras cada censo o padrón. Con motivo de cada actualización, se procede a subdividir o a fusionar aquellas secciones que hayan dejado de cumplir los límites de población establecidos.

Las secciones censales juegan un papel importante en las encuestas a hogares que llevan a cabo el INE y otros organismos. En la mayor parte de ellas se utiliza un esquema de muestreo en dos etapas en el que primero se eligen una serie de secciones censales, que se mantienen fijas durante un período relativamente largo de tiempo, y después se toma una muestra aleatoria de hogares o de individuos dentro de cada sección. Esto es mucho más barato que tomar la muestra trabajando directamente con el conjunto de la población española porque permite concentrar las entrevistas en determinadas áreas geográficas, reduciendo así los costes y tiempos de desplazamiento de los entrevistadores, y porque contribuye también a reducir los costes derivados del mantenimiento del *marco* de la encuesta, esto es, el listado de

¹ Desde 1969 hasta mediados de los setenta, la encuesta se realizó semestralmente. Durante 1966-68 y 1971-72, además, algunas de las encuestas previstas no llegaron a realizarse.

² Ceuta y Melilla se incorporan a la muestra en 1988.

³ En principio, las personas ingresadas en hospitales o prisiones se contabilizan en sus residencias familiares de origen, al igual que los militares que residen temporalmente en cuarteles (incluyendo los que están prestando el servicio militar) o los estudiantes internos en residencias. Sin embargo, los religiosos que viven en residencias colectivas y otras personas que carecen de residencia familiar están excluidos de la muestra.

⁴ Los detalles técnicos del diseño de la encuesta se resumen en diversas publicaciones metodológicas del INE, algunas de las cuales están disponibles en la sección dedicada a la EPA en la página web del Instituto (véase http://www.ine.es/inebmenu/mnu_mercalab.htm). Esta sección se basa en particular en INE (2008 y 2009) así como en García España (1969), Porras (1999) y Álvarez (2000).

unidades del que se selecciona la muestra. Una vez elegidas las secciones censales, por ejemplo, sólo resulta necesario mantener actualizado el listado de viviendas familiares correspondiente a aquellas secciones que han sido incluidas en la muestra, que sólo suponen en torno a un 11% del total.

En el caso de la EPA, en la primera etapa del proceso de selección de la muestra se utiliza un procedimiento *de estratificación* que permite asegurar que se obtendrán resultados de una precisión aceptable no sólo para la población española en su conjunto sino también para determinados subconjuntos de la misma y puede ayudar a aumentar la precisión de las estimaciones derivadas de la encuesta. Con este fin, las secciones censales de cada provincia se agrupan en una serie de *estratos* básicamente de acuerdo con el tamaño del municipio al que pertenecen (medido por su población) y, dentro de cada estrato, en subestratos en base a la condición socioeconómica de sus habitantes. Para realizar ambas clasificaciones se utilizan datos detallados del último censo disponible y de la Agencia Tributaria sobre la población de los distintos municipios y sobre las características de sus residentes (incluyendo su nivel educativo, tipo de ocupación, tasas de actividad y paro, nacionalidad y nivel de renta). Puesto que el proceso de estratificación exige una información muy detallada, el mismo sólo se puede revisar con cada nuevo censo.

Una vez realizada la estratificación y determinado el número de secciones censales a las que resulta posible encuestar con los recursos disponibles (unas 3.600 en la actualidad), éstas se reparten entre estratos y subestratos en un proceso conocido como *afijación*. En primer lugar, las secciones se reparten entre provincias teniendo en cuenta la población de éstas pero primando a las más pequeñas de forma que en ellas también se puedan obtener resultados de una precisión aceptable. La asignación de cada provincia se reparte después entre estratos. En este caso se prima a los municipios más grandes por motivos de eficiencia, pues en este estrato se observan los menores costes por entrevista y la mayor variación en términos de las características objeto de estudio, lo que ayuda a mejorar la precisión de las estimaciones. Finalmente, dentro de cada estrato el reparto entre subestratos es estrictamente proporcional al tamaño de los mismos, medido por el número de viviendas familiares, lo que asegura que la probabilidad de selección es igual para todos los hogares de un determinado estrato (incluyendo a la provincia en la definición de estrato).

Una vez fijadas las secciones censales, se elige aleatoriamente una muestra de hogares dentro de cada una de ellas. Actualmente, se entrevistan 18 hogares por sección, excepto en las provincias más pobladas (Madrid, Barcelona, Sevilla, Valencia y Zaragoza) en las que el número de entrevistas por sección es de 22. Una complicación importante que surge en este punto es la *no-respuesta* de algunas de las unidades incluidas en la muestra original pues en ocasiones resulta imposible contactar con los integrantes de un hogar determinado o estos indican que no pueden o no desean colaborar con la encuesta. Cuando el problema afecta a un hogar que ha respondido a la encuesta en el trimestre anterior, a éste se le imputan sus propias respuestas en la última entrevista. Cuando no se localiza a ningún miembro de la familia en el primer trimestre en el que ésta entra en la muestra, la vivienda se mantiene en la muestra a la espera de conseguir realizar la entrevista en el futuro. Si esto no se consigue, la

vivienda se elimina de la muestra sin reemplazarla por otra. Finalmente, cuando la negativa a colaborar se produce durante el primer contacto, la vivienda se substituye por otra tomada de una lista de suplentes elegidas al azar.

La renovación del seccionado

Las secciones censales que conforman la muestra de la EPA tienden a permanecer aproximadamente fijas durante un periodo prolongado de tiempo. En circunstancias normales, de un año a otro sólo son necesarios pequeños ajustes para reemplazar a aquellas secciones en las que ya se han visitado todas las viviendas encuestables y para acomodar la posible fusión o subdivisión de secciones incluidas en la muestra.⁵ Cada cierto tiempo se lleva a cabo una revisión más profunda del seccionado en la que salen de la muestra algunas secciones que han perdido peso en la población y se incorporan otras que lo han ganado. Sin embargo, dado que la introducción de nuevas secciones tiene un coste significativo, incluso en estos casos se procede generalmente de una forma que permite mantener al mayor número posible de secciones dentro de la muestra.⁶ La principal excepción en este sentido ha sido la actualización del seccionado realizada durante 1995 y la primera mitad de 1996 en la que, abandonando el procedimiento habitual, se produjo una renovación de más del 80% de la muestra de secciones, lo que refleja la convicción del INE de que la muestra existente estaba ya muy lejos de ser una buena representación del conjunto de la población.

En principio, las revisiones del seccionado se realizan cada vez que se dispone de los resultados detallados de un nuevo censo, lo que permite revisar la estratificación, modificar la afijación de secciones si resulta necesario y recalcular la probabilidad de selección de cada una de las secciones, que se obtiene como el cociente entre el número de viviendas familiares de la misma y el número total de tales viviendas en el estrato en el que ésta está incluida. Una vez realizados tales cálculos se procede a adaptar la muestra de secciones a la nueva distribución territorial de la población trabajando de la forma siguiente. Aquellas secciones incluidas en la muestra cuya probabilidad de selección ha aumentado (esto es, aquellas que han ganado peso dentro de su estrato) se mantienen en la muestra, mientras que las que lo han perdido salen de la muestra con una probabilidad que depende de la reducción observada de su peso dentro del estrato correspondiente. Las secciones que salen de la muestra se substituyen por otras del mismo estrato que son elegidas aleatoriamente entre

⁵ Por ejemplo, si una sección incluida en la muestra se subdivide en varias, se elige aleatoriamente una de las subdivisiones para continuar en la muestra con probabilidades proporcionales al tamaño de las mismas. Si una sección incluida en la muestra se fusiona con otra que no lo está, la sección resultante continúa en la muestra. En ambos casos, el número de entrevistas que se realiza en la sección se ajusta al alza o a la baja de forma que la probabilidad de selección de las viviendas de la nueva sección siga siendo la misma que en el pasado (que a su vez es común para todo el estrato).

⁶ Lo dicho sobre la permanencia del seccionado es cierto desde 1971 en adelante. En ese año, la EPA se integra en el marco más amplio de la Encuesta General de Población (EGP) que proporciona un diseño muestral común para las encuestas del INE dirigidas a los hogares. Con anterioridad a esta fecha, el seccionado se renovaba parcialmente cada trimestre o semestre, aunque siempre utilizando la división en secciones del censo de 1960 y, posteriormente, la del padrón de 1965. Por lo tanto, tales renovaciones no servían para mitigar el problema del envejecimiento de la muestra del que hablaremos más adelante, pues ésta seguía eligiéndose con probabilidades que reflejaban la distribución espacial de la población en el momento de un censo o padrón determinado.

aquellas no incluidas en la muestra original que han ganado peso con la revisión.⁷ En algunas ocasiones, el cambio de la muestra (salida de viejas secciones y entrada de las nuevas) se ha realizado de una forma gradual durante un período de 6 trimestres.

Un procedimiento similar al que acabo de describir se utiliza también para realizar revisiones parciales del seccionado (aunque sin modificar la estratificación) durante el período comprendido entre dos censos utilizando datos del padrón. Las grandes revisiones basadas en datos censales se han llevado a cabo en 1971, 1985, 1995 y 2005 (a partir de los censos de 1970, 1981, 1991 y 2001) respectivamente, mientras que revisiones menores basadas en datos del padrón se han hecho en 1968, 1978, 1989 y 2001 (a partir de los padrones de 1965, 1975 y 1986 y el padrón continuo de 1998) y, en años recientes, se realizan aproximadamente cada dos años.

2.2. Estimación de los totales poblacionales

Una vez realizadas las entrevistas y tabulados sus resultados, los datos de la encuesta han de *elevarse a la población* con el fin de estimar el número total de ocupados, parados e inactivos en España en su conjunto y en cada una de sus provincias, así como la distribución poblacional de las características de cada uno de estos grupos. Para ello, resulta necesario sumar sobre los hogares de la muestra, multiplicando los resultados obtenidos para cada uno de ellos por un *factor de elevación* que nos dice a cuántos hogares de la población está representando este elemento de la muestra.

Por ejemplo, sea X el número total de ocupados en la población. Esta magnitud se puede estimar como la suma ponderada del número de ocupados que habitan en cada vivienda i de la muestra M , donde la ponderación w_i se obtiene como la inversa de la probabilidad π_i de que la vivienda i haya sido incluida en la muestra dado el diseño de la encuesta:

$$(1) \hat{X} = \sum_{i \in M} w_i X_i \quad \text{con } w_i = \frac{1}{\pi_i}$$

En el caso de la EPA, el diseño muestral implica que el factor de elevación ha de ser el mismo para todos los hogares que pertenecen al mismo estrato, M_e . Podemos por tanto sumar los hogares dentro de cada estrato antes de aplicar el factor de elevación, para sumar a continuación los distintos estratos, convenientemente ponderados. De esta forma se obtiene

$$(2) \hat{X} = \sum_{i \in M} \frac{1}{\pi_i} X_i = \sum_e \frac{1}{\pi_e} \left(\sum_{i \in M_e} X_i \right)$$

donde π_e es la probabilidad de selección de una vivienda dentro del estrato e . Finalmente, el estimador tradicionalmente utilizado por el INE aproxima el inverso de π_e mediante el cociente entre la población total estimada (residente en viviendas familiares) del estrato e (en el día 15 del mes central del trimestre de acuerdo con las estimaciones de población que

⁷ Para más detalles, véase INE (2009), pp. 24-6.

realiza el propio INE, \tilde{P}_e) y el número de personas de la muestra que pertenecen al mismo estrato, p_e , lo que nos deja con la siguiente expresión:⁸

$$(3) \hat{X}_{EPA} = \sum_e \frac{\tilde{P}_e}{P_e} \left(\sum_{i \in M_e} X_i \right)$$

Obsérvese que si aplicamos este estimador a la población total, P , en vez de al número de ocupados, recuperaremos la proyección de población total del INE:

$$\hat{P}_{EPA} = \sum_e \frac{\tilde{P}_e}{P_e} \left(\sum_{i \in M_e} p_i \right) = \sum_e \frac{\tilde{P}_e}{P_e} p_e = \sum_e \tilde{P}_e = \tilde{P}$$

Sin embargo, no hay ninguna garantía de que el estimador reproduzca exactamente la estructura real de la población en términos de su desglose por sexo y edad entre otras variables.

Algunos problemas

Un problema bien documentado de la EPA es que, al menos hasta años recientes, la encuesta ha tendido a infravalorar el número de activos y ocupados en la población, como revela una comparación cuidadosa con otras fuentes estadísticas. La visión de consenso en la literatura es que tal fenómeno se debe fundamentalmente a dos factores: el sesgo de no respuesta y el gradual envejecimiento del marco de la encuesta y, en particular, de su seccionado.⁹ Ambos fenómenos se traducen en una infravaloración de los grupos de edad con mayores niveles de actividad y el segundo de ellos genera también una tendencia a subestimar la población que reside en las zonas más dinámicas del territorio, lo que tiene un efecto adicional sobre la estimación.

La no respuesta tiende a sesgar los resultados de la estimación porque es prácticamente seguro que los hogares que se niegan a colaborar con la encuesta desde el principio (en torno a un 6 o 7% en años recientes) o cuyos miembros no están nunca localizables no se distribuyen aleatoriamente sino que tienden a concentrarse en determinados segmentos de la población que tienen tasas de actividad y ocupación claramente superiores a la media. Esto hace que su ausencia de la muestra finalmente entrevistada tienda a sesgar la composición de la misma, convirtiéndola en una representación inexacta de la población. En particular, los hogares integrados por personas de mayor edad, y que por lo tanto cuentan con más miembros inactivos que son más fáciles de entrevistar por pasar más tiempo en casa, tienden a estar sobrerrepresentados en la muestra final de la EPA mientras que aquellos segmentos de la población con una mayor tasa de ocupación (en particular, los que tienen ente 25 y 49 años de edad) están infrarrepresentados. Esto es fácil de comprobar: cuando la ecuación (3) se utiliza para estimar la estructura por edades de la población total a partir de la muestra de la EPA, el resultado se aleja significativamente de la estructura real de la población española.

⁸ Véase INE (2009) y Porras (1999, pp. 551-52).

⁹ Véanse entre otros Toharia (1998), Alvarez (2000) y Pérez Infante (2000 y 2006).

El envejecimiento de la muestra plantea problemas de una naturaleza similar a los generados por la no-respuesta. Puesto que las secciones censales encuestadas no varían durante períodos relativamente largos, es muy probable que la muestra termine no siendo una buena fotografía de la población si la localización de ésta varía de forma significativa de un censo a otro. Así, las áreas que pierden peso en la población total, que suelen ser las más envejecidas, estarán cada vez más sobrerrepresentadas en la encuesta, mientras que sucederá lo contrario con las zonas más dinámicas, en las que es probable además que los niveles de actividad sean mayores, incluso tras controlar por la estructura de la población por sexo y edad.

El método de calibrado o reponderación

A lo largo del tiempo, el INE ha ido tomando diversas medidas para intentar reducir los sesgos derivados de los problemas descritos en el apartado anterior. Con el fin de reducir la tasa de no respuesta se han ido introduciendo modificaciones en el trabajo de campo tales como la flexibilización de los horarios de los entrevistadores y la utilización de entrevistas telefónicas. Por otra parte, y con el objetivo de paliar el problema que supone el envejecimiento de la muestra de secciones, el INE ha aumentado en años recientes la frecuencia de las revisiones del seccionado utilizando datos del padrón continuo y del censo electoral (Álvarez (2000), p. 221).

A estas medidas hay que añadir la introducción en 2002 de una importante mejora técnica en la metodología de estimación que permite mitigar los efectos de los dos problemas descritos en el apartado anterior mediante un ajuste de los factores de elevación. El ajuste busca corregir *ex-post* la infra o sobre-representación de ciertos segmentos de la población que, como hemos visto, parece ser la principal causa inmediata de los sesgos detectados en la encuesta. Este procedimiento, conocido como *reponderación o calibrado de los factores de elevación*, supone sustituir el estimador tradicional dado en (3) por otro estimador de la forma

$$(4) \hat{X} = \sum_{i \in M} w_i^* X_i$$

donde los factores de elevación originales de los hogares, w_i , han sido sustituidos por otros, w_i^* , que se alejan de los anteriores sólo el mínimo imprescindible para conseguir que las predicciones de la encuesta reproduzcan exactamente ciertas características de la población que se conocen por alguna fuente externa considerada fiable. Formalmente, el nuevo conjunto de factores de elevación se obtiene resolviendo un problema de la forma¹⁰

$$(5) \min_{\mathbf{w}^*} D(\mathbf{w}, \mathbf{w}^*) \text{ s. a. } \sum_{i \in M} w_i^* A_i^k = \tilde{A}^k \text{ para todo } k = 1, \dots, K$$

donde \mathbf{w} y \mathbf{w}^* son, respectivamente, los vectores de factores de elevación originales y reponderados, $D()$ un indicador de la distancia entre ambos y $\{A^k; k = 1, \dots, K\}$ un conjunto de

¹⁰ Para más detalles sobre el procedimiento de calibrado que utiliza el INE y sus efectos, véanse Álvarez (2000), INE (2009) y <http://www.ine.es/epa02/repondera.htm>. Lundström y Särndal (2001) ofrecen una introducción accesible y detallada a los problemas de estimación derivados de la falta de respuesta en encuestas y a las posibles ventajas en este contexto de los estimadores reponderados.

variables auxiliares investigadas en la encuesta para cuyos totales poblacionales se dispone de una estimación externa fiable, \tilde{A}^k .

En la práctica, las estimaciones externas, \tilde{A}^k , provienen de las proyecciones de población del propio INE. Las variables auxiliares que se han utilizado para calibrar los resultados de la EPA son la distribución por sexo y segmentos de edad de la población (residente en viviendas familiares) de las comunidades autónomas, el total provincial de la misma variable y, desde 2005, el desglose de la población de cada región entre españoles y extranjeros. Puesto que todas estas variables están altamente correlacionadas con la actividad laboral, la modificación de los factores de elevación para corregir la infra o sobreponderación de determinados colectivos tiene efectos significativos sobre las estimaciones de los principales agregados de la encuesta.

2.3. Posibles fuentes de discontinuidades en las series

La discusión precedente nos permite identificar una serie de factores que tienden a generar rupturas en las series de la EPA así como posibles formas de corregir algunas de ellas. Examinando la ecuación (3) resulta evidente que cualquier revisión de las proyecciones de población del INE generará una discontinuidad en las series de la EPA, pero también que dicha ruptura resulta muy fácil de corregir por el procedimiento de reestimar las series retroactivamente utilizando las nuevas proyecciones de población para recalcular los factores de elevación. También hemos visto que las actualizaciones del seccionado tienden a generar discontinuidades. Si se dispone de una estimación del tamaño de las mismas, en principio lo lógico sería repartir cada una de ellas sobre el período transcurrido desde la anterior actualización. Por otra parte, parece plausible que la reponderación de los factores de elevación pueda suavizar las discontinuidades de este tipo (al corregir los efectos del envejecimiento del seccionado sobre la estructura demográfica de la muestra), pero no parece probable que esta técnica pueda eliminar el problema por completo dado que ésta no es la única vía a través de la cual el problema citado puede distorsionar las estimaciones de empleo y actividad.

A estos factores hay que añadir los muchos otros cambios metodológicos que ha experimentado la EPA en sus casi cincuenta años de historia y que en principio también pueden dar lugar a discontinuidades en las series. Entre tales cambios se encuentran las numerosas modificaciones que se han ido introduciendo en los cuestionarios de la encuesta, buscando siempre una mayor claridad y precisión de las preguntas, los sucesivos cambios en las prácticas y procedimientos de trabajo de campo y algunas variaciones en las definiciones de ciertas variables de la encuesta, en los esquemas de clasificación (de estudios, de ocupaciones, de actividades económicas...) y en los criterios de estratificación utilizados en la misma. En el Anexo 2 se ofrece un listado cronológico de tales cambios metodológicos.

Durante los primeros años de la encuesta se observa una cierta inestabilidad en los criterios utilizados para calcular algunos de sus grandes agregados. Tales cambios de definición han afectado, entre otras cosas, a la clasificación como ocupados, parados o inactivos de las

personas que se encuentran realizando el servicio militar (o el servicio social sustitutorio), los temporeros sin trabajo entre campaña y campaña, los trabajadores agrícolas andaluces y extremeños acogidos al subsidio agrario y sus antecesores (lo que popularmente se conoce como PER), las llamadas ayudas familiares¹¹ y los activos marginales, esto es, aquellas personas que, autoclasificándose como inactivos (jubilados, estudiantes, amas de casa...), han trabajado o han buscado trabajo durante la semana anterior a aquella en la que se realiza la encuesta. También se ha modificado la definición de la población en edad de trabajar como resultado de la elevación de la edad legal mínima para comenzar a trabajar desde los 14 a los 16 años que se produce en 1980 con la aprobación del Estatuto de los Trabajadores. Los criterios utilizados para clasificar a algunos de estos grupos en las sucesivas metodologías de la EPA se recogen en el Cuadro 1.

Cuadro 1: Cambios de criterio en la clasificación de algunos colectivos en la EPA

	<i>Varones que prestan el servicio militar no se investigan</i>	<i>Activos marginales inactivos</i>	<i>Temporeros sin trabajo inactivos</i>	<i>Ayudas familiares y trabajadores ocasionales como en 1965</i>
<i>EPA 1964</i>				
<i>EPA 1965 1965T1 a 1971S2</i>	se clasifican según su situación en el momento de la llamada a filas	inactivos	inactivos	En principio ocupados, pero se incluyen entre los activos marginales si se autoclasifican como inactivos (estudiantes, amas de casa, jubilados...)
<i>EPA 1972 1972 a 1976T2</i>	sin cambios	inactivos	grupo aparte dentro de activos, no ocupados	Ocupados si trabajan en la semana de referencia y además trabajan habitualmente 6 o más meses al año y al menos 1/3 de la jornada habitual
<i>EPA 1976 1976T3 a 1987T1</i>	población contada aparte	ocupados si trabajan más de 1/3 de jornada e inactivos en caso contrario	inactivos	Ocupados si han trabajado al menos 1/3 de la jornada normal durante los 3 meses anteriores. Si no, inactivos o parados, dependiendo de si buscan o no empleo
<i>EPA 1987 1987T2 a 2001</i>	población contada aparte	ocupados si trabajan más de 1h; se elimina la distinción entre ocupados estrictos y marginales	inactivos	Se aplica el criterio general: ocupados si han trabajado al menos 1 h. en la semana de referencia

- Fuentes: Véase el Anexo 1.

Dos discontinuidades importantes en la encuesta son las que se producen en 1976 y 1987 cuando se introducen profundas modificaciones en el cuestionario en combinación con otros cambios metodológicos de distinta índole. En el primer caso se buscaba ampliar la información disponible sobre el paro tras su rápido aumento en los años precedentes y en el segundo adaptar la encuesta a los criterios de EUROSTAT para permitir comparaciones

¹¹ Este colectivo está integrado por aquellas personas que trabajan sin remuneración explícita en el negocio o empresa de un familiar con el que conviven.

homogéneas con otros países comunitarios así como recabar información sobre nuevas formas de contratación autorizadas tras la reforma del Estatuto de los Trabajadores. En años más recientes destaca la revisión del cuestionario que se produce en 1999¹² y la introducción en 2002 del procedimiento de calibrado y de una nueva definición operativa de paro que endurece las condiciones exigidas para concluir que una persona está buscando trabajo activamente y puede, por lo tanto, considerarse parado en lugar de inactivo.¹³ Finalmente, en 2005 se introducen cambios en el cuestionario de la EPA que han permitido medir con mayor precisión la duración de la jornada laboral así como la existencia de ciertos empleos de corta duración o de jornada muy reducida que con frecuencia no se detectaban en versiones anteriores de la encuesta. Al mismo tiempo se añade la nacionalidad a las variables auxiliares utilizadas en el proceso de calibrado y se introducen ciertas mejoras en los procedimientos de recogida de datos, incluyendo la utilización de entrevistas telefónicas asistidas por ordenador.¹⁴

3. El punto de partida: principales series disponibles

En esta sección se describen las dos fuentes básicas de las que partiré para elaborar las series revisadas de los grandes agregados de la EPA: las sucesivas series históricas de tales variables que ha elaborado el propio INE para el período que comienza en 1976 y las series homogeneizadas que algunos investigadores han construido para períodos anteriores.

3.1. Series históricas de la EPA elaboradas por el INE

El INE ha elaborado diversas series históricas en principio homogéneas de la EPA que, en el caso de los principales agregados de la encuesta, comienzan en el tercer trimestre (T3) de 1976. Estas series retrospectivas incorporan dos tipos de mejoras en relación con los resultados de la encuesta publicados originalmente. En primer lugar, están basadas en estimaciones actualizadas de la población (elaboradas generalmente a partir de los resultados detallados de cada nuevo censo o padrón). Y en segundo lugar, tales series generalmente incorporan aquellas novedades metodológicas que pueden ser aplicadas a posteriori a partir de los cuestionarios originales correspondientes a períodos anteriores, tales como la reponderación o ciertos cambios en la definición de algunos agregados.

Actualmente (en octubre de 2012) la página web del INE (2012^a) proporciona tres series históricas diferentes de población en edad de trabajar (entendida durante todo el período como aquella de 16 años de edad o más) clasificada por su relación con la actividad

¹² Véase por ejemplo Pérez Infante (2000).

¹³ Véase <http://www.ine.es/epa02/definiiparo.htm>. El cambio de criterio afecta, en particular, a aquellas personas no ocupadas cuya única forma de búsqueda de trabajo es a través de las oficinas públicas de empleo. Para este colectivo, ya no basta con mantener vigente su inscripción como parado a efectos administrativos, sino que se exige algún tipo de contacto activo con tales oficinas para buscar empleo. Sobre éste y otros cambios introducidos en la EPA en 2002, véase la web del INE (<http://www.ine.es/epa02/meto2002.htm>) así como Banco de España (2002) y Pérez Infante (2006) entre otros.

¹⁴ Sobre las novedades que introduce la EPA 2005 véanse entre otros INE (2005) y Albacete y Laborda (2005).

económica (en activos, ocupados, parados, inactivos y población contada aparte¹⁵). La más antigua, a la que denominaré “primera serie histórica del INE” o “serie H1”,¹⁶ cubre el período 1977- 2001. En la segunda (a la que llamaré serie H2¹⁷) el período muestral se extiende hasta 2004 y en la tercera y última por el momento (serie H3¹⁸) se llega hasta el momento actual, revisándose en ambos casos a posteriori diversos tramos de la serie anterior. En particular, H2 modifica H1 en su integridad, mientras que H3 difiere de H2 sólo de 1996 en adelante (tanto a nivel nacional como regional).

Las sucesivas revisiones de las series han servido, en primer lugar, para actualizar las estimaciones de población total en edad de trabajar en base a la información más reciente disponible en cada momento. Así, la serie H2 incorpora desde el año 1996 en adelante nuevas proyecciones de población, elaboradas en el tercer trimestre de 2001, que modifican al alza la estimaciones previas del INE para intentar recoger el fuerte influjo de inmigrantes que se produjo durante los últimos años del siglo XX. Una vez se dispuso de los resultados del censo de (noviembre de) 2001, resultó evidente que tal influjo había sido muy superior al estimado inicialmente, lo que hizo necesaria una nueva revisión al alza de las proyecciones de población de 1996 en adelante que se incorpora a la serie H3. Los efectos de las sucesivas revisiones de la población se aprecian con claridad en el Gráfico 1.a. Con la serie H2 se introduce también de forma retroactiva la reponderación de los factores de elevación desde 1976T3, lo que tiene un efecto apreciable sobre algunos agregados de la EPA durante todo el período muestral, tal como se observa en el Gráfico 1.b para el caso de la población ocupada. En la serie H3 el calibrado también se aplica retrospectivamente desde 1996 (que es cuando comienza a ser significativo el influjo de población extranjera) hasta 2004, incorporándose el peso de los extranjeros en la población de cada comunidad autónoma a las variables auxiliares que se utilizan en la reponderación.

¹⁵ En esta categoría se incluyen las personas que están realizando el servicio militar o el servicio social sustitutorio. Este colectivo no se integra ni en la población activa ni en la inactiva y se contabiliza en su región de origen o residencia habitual con independencia de donde pueda estar destinado. En algunos casos esta variable no se ofrece explícitamente pero se puede obtener por diferencia a partir de los demás agregados.

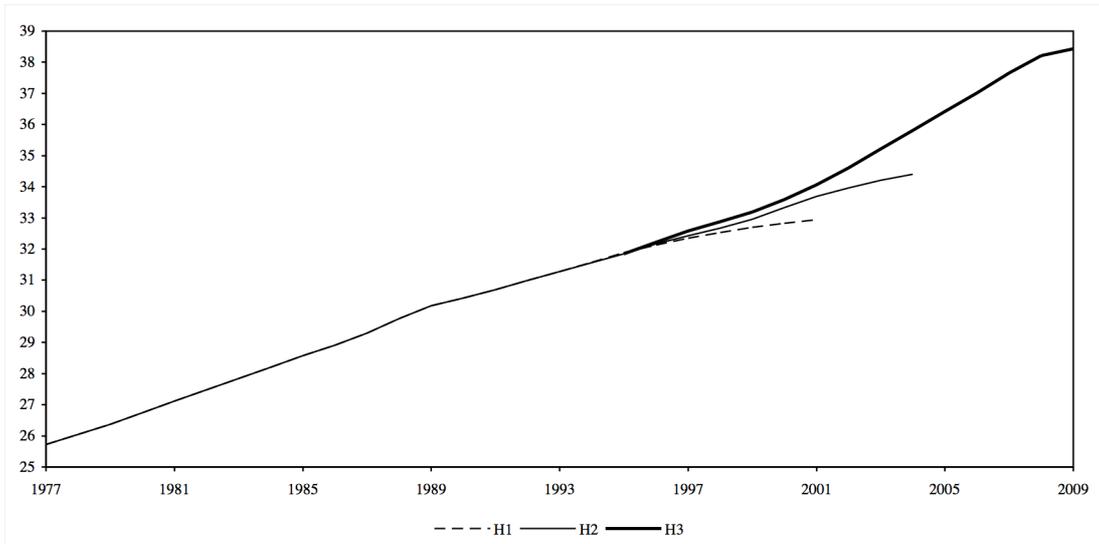
¹⁶ En su página web, el INE se refiere a H1 como la “serie publicada”. Esto no quiere decir, sin embargo, que se trate de la serie originalmente publicada trimestre a trimestre sin ninguna modificación. Desde 1976T3 a 1987T1, H1 coincide con la “serie revisada” publicada por el INE en 1990. Esta última serie se construye aplicando retroactivamente la nueva metodología de la EPA 1987 a los cuestionarios individuales del período citado, lo que corrige los posibles cambios de definición pero no los efectos de las modificaciones en el cuestionario. La serie está disponible en <http://www.ine.es/epa02/repercusion.htm> junto con la serie reponderada que se obtiene a partir de la misma aplicando el procedimiento de calibración descrito más arriba. Esta última serie coincide casi exactamente con H2, presentando sólo muy pequeñas discrepancias con la misma en algunos trimestres (de no más de 300 personas en el caso de los ocupados).

¹⁷ La serie H2 (<http://www.ine.es/jaxi/menu.do?type=pcaxis&path=/t22/e308/&file=inebase>) incluye la serie elaborada con la metodología EPA 2002 y, para años anteriores al comienzo de esta serie, una revisión a posteriori de la serie homogénea precedente. En particular, los datos para 1976-2000 coinciden con las “series históricas revisadas” publicadas por el INE en 2003. En 2001 existen diferencias entre las dos fuentes que parecen deberse a que la serie de la EPA 2002 que aparece en la web del INE incorpora la nueva definición operativa de parado en ese año, mientras que la serie histórica revisada conserva la antigua definición para preservar su homogeneidad.

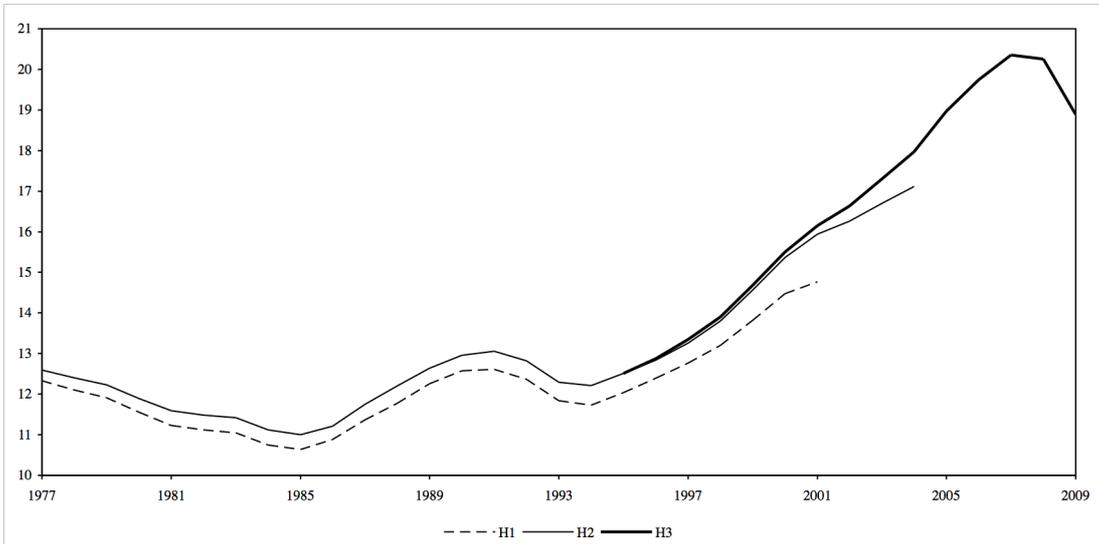
¹⁸ Véase

http://www.ine.es/jaxi/menu.do?type=pcaxis&path=/t22/e308_mnu&file=inebase&N=&L=0.

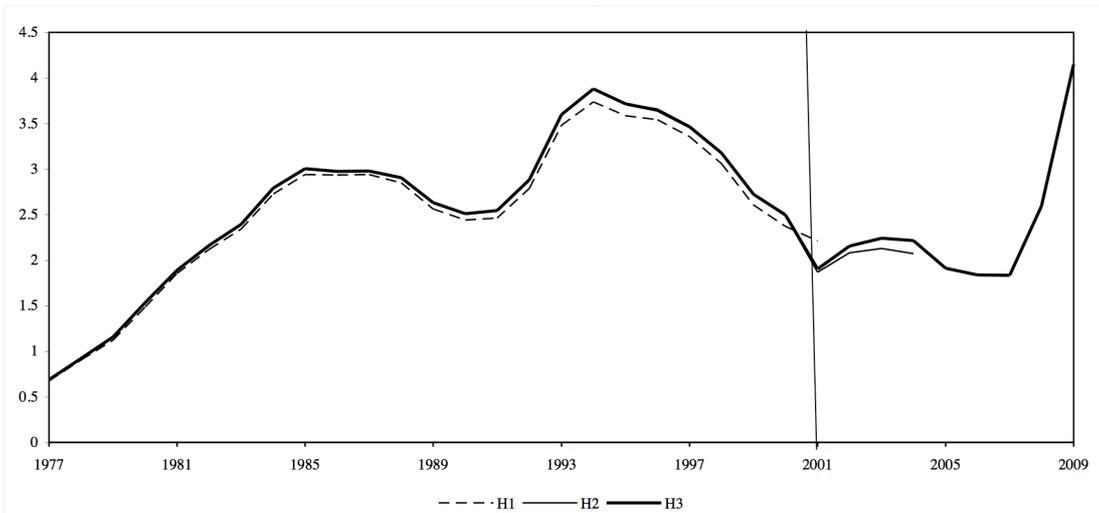
Gráfico 1: Comparación de las sucesivas series históricas del INE, España en su conjunto
a. Población de 16 o más años de edad en viviendas familiares



b: Población ocupada



c: Población parada



- *Notas:* Millones de personas; datos anuales, contruidos como la media no ponderada de los cuatro trimestres. Gráfico c: H2 y H3 presentan una discontinuidad en 2001, que es cuando comienza a aplicarse la nueva definición operativa de desempleo.

Lo que el INE generalmente no ha hecho en sus sucesivas revisiones de las series históricas de la EPA es eliminar aquellas discontinuidades de las mismas que tienen su origen en modificaciones del cuestionario o en otros cambios metodológicos que no se pueden aplicar retroactivamente a años anteriores. En algunos casos, sin embargo, el propio Instituto ha estimado los efectos de tales cambios en el momento de su introducción, lo que en principio permite enlazar los sucesivos segmentos de la serie de una forma que elimine las discontinuidades.

El Cuadro 2 contiene un listado de los principales cambios metodológicos que ha experimentado la EPA desde 1976. Siempre que tal información está disponible, el cuadro incluye la estimación que ha hecho el propio INE de la afluencia de empleo, actividad y paro derivada de cada cambio e indica en su última columna si la ruptura de la serie provocada por el mismo ha sido o no corregida en la versión actual de la serie histórica del Instituto. En relación con esta cuestión existe una importante área gris sobre la que volveré más adelante: no está claro hasta qué punto el proceso de calibrado corrige el sesgo derivado del envejecimiento del seccionado.

Cuadro 2: Posibles discontinuidades de las series históricas del INE, 1976-2009

<i>fecha</i>	<i>cambio metodológico</i>	<i>efecto estimado INE</i>	<i>¿corregido en serie histórica?</i>
1978T2	Actualización del seccionado		¿?
1980T2	Cambio en la edad legal para trabajar	A: -161; O: -76; P: -85	si
	Cambios de definición		si
1984T1	Cambio tratamiento beneficiarios PER	O: -120; P: +120	si
1985T2	Actualización del seccionado		??
1986T2	Nuevas proyecciones de población		si
1987T2	Cambios de definición	A: +29,4; O: +54,7; P: -25,4	si
	Cambios en el cuestionario	A: +38,7; O: +86,2; P: -47,4	no
1988T2	Inclusión de Ceuta y Melilla	A: +48,8; O: +31,8; P: +17	no
1989T1	Actualización del seccionado		¿?
1990T4	Nuevas proyecciones de población		si
1995T1	Actualización seccionado (hasta 1996T2)		¿?
1996T1	Nuevas proyecciones de población		si
1999T1	Cambios en el cuestionario		no
2000T1	Actualización del seccionado	A: +86; O: +75; P: +8,5	¿?
2002T2	Nueva definición operativa de paro	P: -480,1*	no
	Reponderación factores elevación	A: +953; O: +857; P: +96	si
	Nuevas proyecciones de población		si
2005T1	Actualización seccionado (hasta 2006T2)		¿?
	Revisión proyecciones de población		si
	Nacionalidad se incluye en calibrado		si
	Cambios en el cuestionario y en la recogida de información	A: +54,2; O: +132; P: -77,8	no

- Fuentes: Pérez Infante (2006), EPA, INE (1995 y 2002) y Anexo 2.

- Nota: En *efecto estimado*, A indica el número de activos, O el de ocupados y P el de parados. El número que aparece detrás de cada letra indica el número de miles de personas aflorados en cada grupo en el trimestre en el que se introduce el cambio metodológico.

(*) Diferencia entre el número medio anual de parados en 2001 de acuerdo con la nueva definición y con la antigua definición.

Como se observa en el cuadro, hay una serie importante de modificaciones metodológicas de la EPA cuyos efectos no han sido corregidos en las series históricas del INE. Una ruptura muy clara de la serie es la que se produce como consecuencia de la adopción en 2001-02 de la nueva definición operativa de desempleo. El número estimado de parados se reduce bruscamente con este cambio de definición como resultado del endurecimiento de los criterios utilizados para distinguir entre parados e inactivos dependiendo de si se produce o no una búsqueda activa de empleo. El INE ha estimado el número de parados utilizando ambas definiciones durante los cuatro trimestres de 2001 y ha construido una serie “homogénea” de parados de acuerdo con la nueva definición de desempleo introducida en 2001 (Trejo y Ortega, 2005) aunque ha optado por no incorporarla formalmente a la serie histórica propiamente dicha. La serie homogénea de paro del INE (de acuerdo con la nueva definición operativa del concepto) se obtiene, básicamente, por interpolación entre la nueva serie en 2001 y la serie antigua de paro en 1976, trabajando con datos desagregados por sexo, grupo de edad y comunidad autónoma.

En la misma línea, la incorporación de Ceuta y Melilla a la muestra en 1988 y las modificaciones introducidas en el cuestionario de la EPA en 1987, 1999 y 2005 también generan rupturas en las series históricas del INE que el Instituto hasta el momento no ha corregido. Para 1987 y 2005, el INE ha estimado el impacto de los cambios en el cuestionario por el procedimiento de realizar una encuesta auxiliar a una muestra más pequeña de lo habitual utilizando el antiguo cuestionario en el mismo trimestre en el que comienza a utilizarse el nuevo cuestionario (la llamada *EPA testigo* en el caso de 2005). En ambos casos, sin embargo, existe la sospecha de que la encuesta auxiliar podría estar subestimando el impacto del nuevo cuestionario.

3.2. Estimaciones para los años anteriores a 1977

Como hemos visto, las series históricas elaboradas por el INE sólo comienzan en el tercer trimestre de 1976. Sin embargo, algunos autores han construido series homogeneizadas de la EPA que se remontan hasta el comienzo de la encuesta en 1964. A los efectos que nos interesan aquí, los estudios relevantes son dos: el elaborado en 1979 por el Grupo de Trabajo para el estudio de los problemas de empleo organizado por el Ministerio de Economía (GTE, 1979) y el realizado por un equipo del IVIE (Mas et al, 2002) como parte de un proyecto cuyo objetivo central era la estimación de los niveles de escolarización de la población en edad de trabajar y de sus componentes por su relación con la actividad económica. Existen otros intentos de construir series históricas homogénea de las principales variables de la EPA, pero todos los que he encontrado hasta el momento se apoyan en algunos de los dos citados para el período que aquí nos interesa.¹⁹

En el estudio del GTE (1979) se reconstruyen series de los principales agregados de la EPA para España en su conjunto (excluyendo Ceuta y Melilla) durante el período 1964-78. Con el fin de eliminar las rupturas de las series, el GTE enlaza las tasas de actividad y ocupación

¹⁹ Véase por ejemplo García Perea (1991) y García Perea y Gómez (1994).

obtenidas en la encuesta original para distintos subgrupos de la población y aplica tales tasas a estimaciones mejoradas a posteriori de la población total en edad de trabajar y de su composición por edades.

En primer lugar, el GTE construye una serie de población de derecho de catorce o más años de edad desagregada por sexo y por tramos de edad, esencialmente por interpolación entre los censos de 1960 y 1970 y los padrones de 1965 y 1975, corrigiendo también en algunos casos la estructura por edades de la población que parece estar distorsionada por la atracción de los encuestados hacia los números redondos. También se elabora una estimación del número de varones que prestan en cada trimestre el servicio militar a partir del *Anuario Estadístico Militar*. Seguidamente, se toman las tasas de actividad y ocupación de cada subconjunto de la población de acuerdo con la encuesta original y se calculan sus variaciones interanuales, comparando cada período (trimestre o semestre) con el correspondiente del año anterior. Las tasas de variación que corresponden a períodos en los que se producen cambios metodológicos se eliminan, substituyéndose por otras "suavizadas" que se obtienen interpolando linealmente entre las tasas correspondientes a los períodos inmediatamente anterior y posterior al intervalo eliminado. Utilizando la serie de tasas de crecimiento así corregida, las series de tasas de actividad y ocupación se reconstruyen para cada grupo, trabajando hacia atrás y partiendo de los valores observados al final del período, que se dan por buenos del tercer trimestre de 1976 en adelante. Finalmente, estas tasas se aplican a la población total de cada grupo para obtener estimaciones de la población activa y ocupada desagregada por sexo y por tramos de edad, estimándose los parados por diferencia.

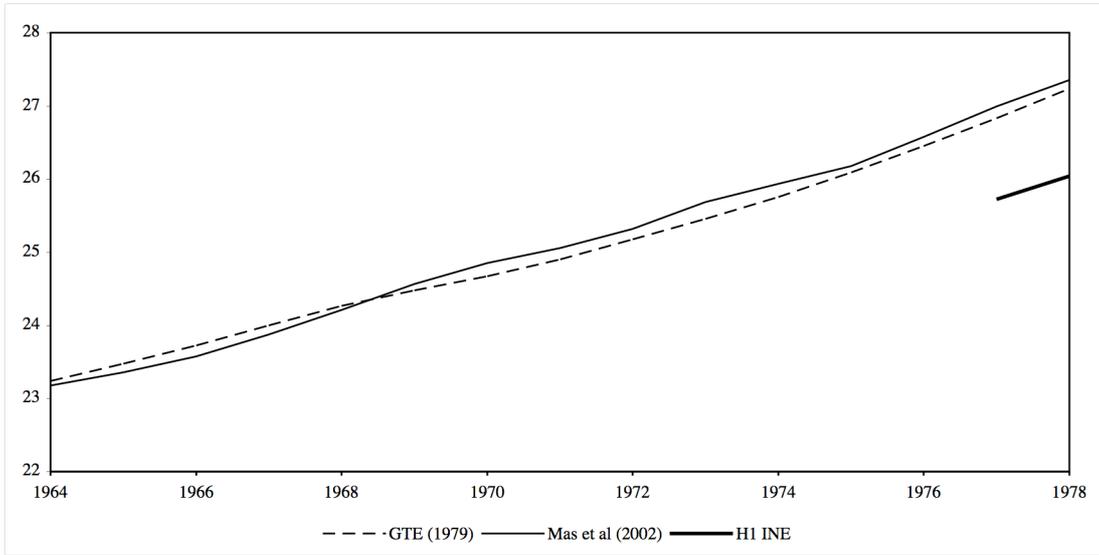
La otra fuente que también cubre el período 1964-76 es el estudio realizado por Mas et al (2002).²⁰ A diferencia del GTE, Mas et al no utilizan métodos indirectos para suavizar las discontinuidades de las series sino que intentan corregirlas directamente, estimando la evolución de los diversos colectivos cuya clasificación ha variado de una metodología de la encuesta a la siguiente. La población de referencia es la de 14 o más años de edad, pero sólo hasta 1980. Por lo tanto, las series que se ofrecen en este trabajo presentan un corte en esa fecha y durante el período que nos interesa no son directamente comparables con las series históricas del INE porque se refieren a un grupo distinto de población.

El Gráfico 2 compara las series de Mas et al (2002) con las de GTE (1979), mostrando también como referencia el primer tramo de la más antigua de las series históricas del INE (la serie H1) que debería en principio ser consistente con los datos de Mas et al si éstos se refiriesen a la población de dieciséis o más años de edad (16+). GTE y Mas et al (2002) coinciden aproximadamente en sus estimaciones de la población en edad de trabajar, pero presentan

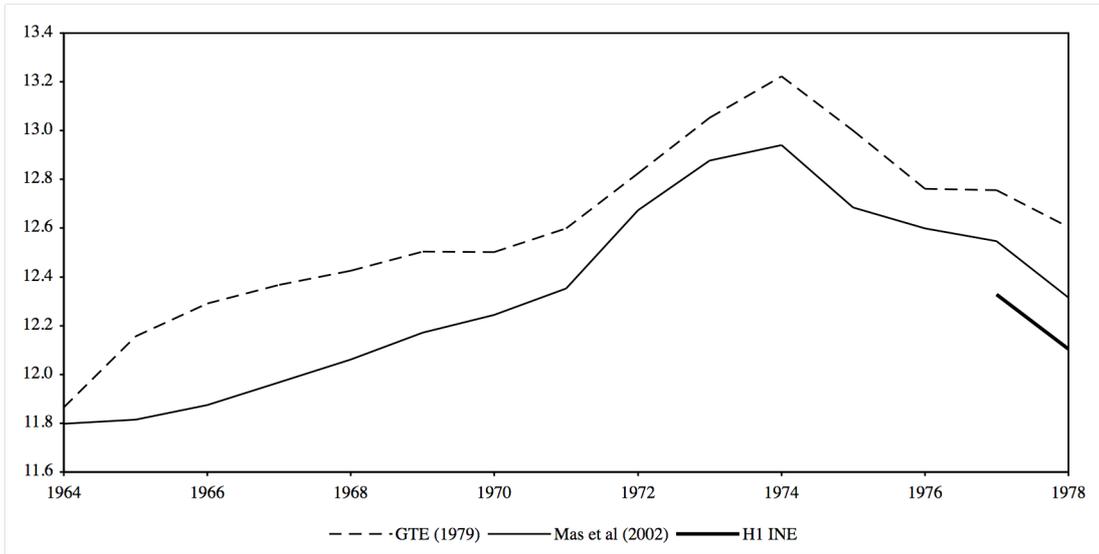
²⁰ Este trabajo forma parte de una serie de monografías del IVIE en las que se parte de la EPA para elaborar indicadores del nivel de estudios para distintos grupos de población. Aunque el objetivo central de estos trabajos es el de cuantificar la dotación de capital humano de España y sus regiones (y provincias), un subproducto interesante es la elaboración de series enlazadas de los principales agregados de la EPA que se remontan hasta 1964. Puesto que mi interés se centra principalmente en los primeros años de la muestra, he elegido trabajar con el primero de tales estudios. En versiones posteriores del estudio, este tramo de la serie no se revisa directamente, pero sí se altera al enlazarlo con series posteriores para eliminar discontinuidades debidas a cambios metodológicos (véase por ejemplo Serrano y Soler, 2008).

Gráfico 2: Comparación de diversas series, 1964-78

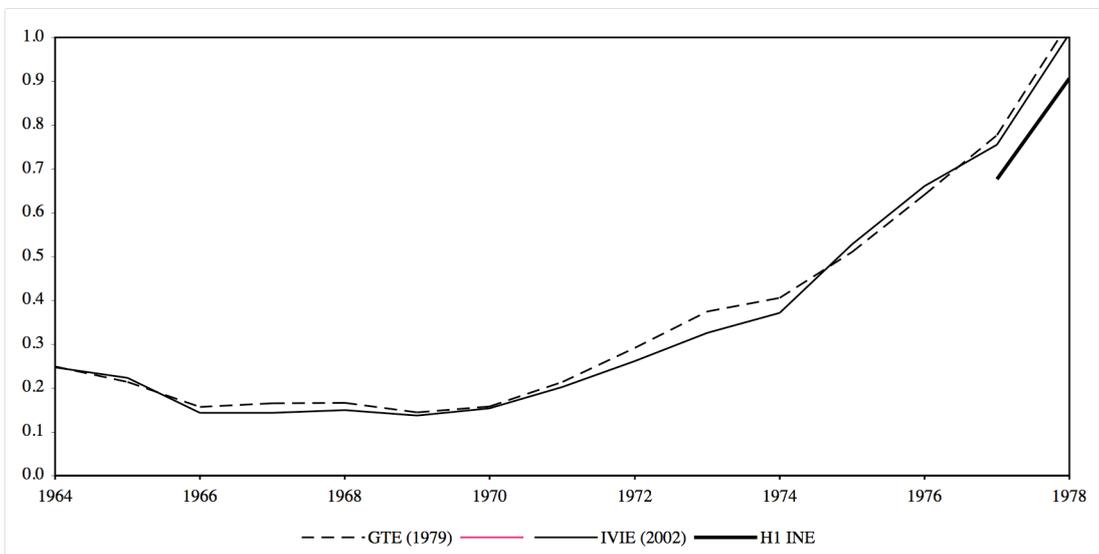
a. Población en edad de trabajar



b. Ocupados



c. Parados



diferencias muy significativas en lo que respecta a la población activa y ocupada, con la serie del GTE situándose siempre por encima de la de Mas et al, con una diferencia media del 2,2% en el caso de los activos y del 2,1% en el de los ocupados. Las diferencias de nivel entre ambas series se deben en parte a que cada estudio intenta homogeneizar las series tomando como referencia los criterios de la EPA vigente en el momento de su elaboración (la EPA 1976 en el caso del GTE y la EPA 1987 en el de Mas et al), que difieren en el tratamiento de ciertos colectivos. Una segunda diferencia importante es que en GTE se corrige la estructura por edades y sexo de la muestra de la EPA de una forma que debería tener un efecto similar al de la reponderación de los factores de elevación que el INE ha introducido a posteriori en 2002 y que las series de Mas et al todavía no incorporan.²¹ Finalmente, los perfiles temporales de ambas series son muy distintos, especialmente durante la primera mitad del período considerado. En buena parte al menos, esto se debe a que en los dos estudios se han utilizado procedimientos distintos para eliminar las discontinuidades de las series originales de la encuesta. La fuerte divergencia que se observa entre ambas series al pasar de 1964 a 1965 parece deberse al menos en parte a que el GTE no tiene en cuenta el hecho de que en la EPA de 1964 no se investigaba a los varones que prestaban el servicio militar, mientras que Mas et al sí corrigen por este hecho.

Dadas las notables diferencias existentes entre las dos series, ¿de cuál de ellas deberíamos partir para extender hacia atrás las series históricas del INE? He optado por la serie de Mas et al (2002) por dos razones. La primera es que el método directo que eligen estos autores para corregir las rupturas en las series de la EPA original resulta en principio preferible al procedimiento indirecto utilizado por el GTE con el mismo objetivo, basado como hemos visto en el suavizado las tasas de variación de las magnitudes relevantes. La segunda es que la serie de Mas et al es en principio consistente con la primera serie histórica del INE (excepto por el tramo de edad con el que se trabaja) mientras que la serie del GTE no es consistente con ninguna de las series históricas del INE por dos motivos ya comentados. Primero, porque se construye en principio con los criterios de la EPA 1976. Y segundo, porque el GTE introduce una corrección para recoger la estructura real por edades de la población, lo que aleja sus resultados de la primera serie histórica del INE, en la no se realiza ninguna corrección de este tipo. Puesto que por otra parte la corrección se lleva a cabo utilizando un método distinto del procedimiento de calibrado adoptado en la segunda serie histórica del INE, la serie del GTE tampoco es directamente comparable con ésta última.

4. Extensión de las series históricas del INE al período 1964-76

En esta sección y en las dos siguientes se describe la construcción de series enlazadas de los principales agregados nacionales de la EPA para el período 1964-2009. Estas series se construyen en dos etapas. En primer lugar, las tres series históricas del INE se extienden

²¹ Recuérdese que las tasas estimadas de actividad y ocupación para distintos segmentos de la población se aplican a la estructura demográfica real, conocida a posteriori, en vez de a la derivada de las respuestas a la encuesta, que como sabemos está sesgada a favor de los colectivos menos activos.

hacia atrás desde 1977 hasta 1964. Seguidamente, la más reciente de las series históricas así extendidas se corrige para eliminar algunas discontinuidades que persisten en la misma. El *output* de la primera etapa del proceso no es, por tanto, una única serie para 1964-76, sino dos series diferentes que nos permiten extender hacia atrás, por un lado, la primera serie histórica del INE, H1, y, por otro, las series H2 y H3 (que coinciden entre sí para los años anteriores a 1996). Aunque la primera de estas series puede considerarse como un paso intermedio en la construcción de la segunda, merece la pena incluirla en la base de datos porque puede ser una referencia útil a la hora de revisar las series de empleo de la Contabilidad Nacional.

4.1. Extensión hacia atrás de la primera serie histórica del INE

La primera tarea que abordaremos consiste en extender hacia atrás la primera de las series históricas del INE (la serie H1) tomando como referencia para el período 1964-76 la serie de Mas et al (2002). Sabemos que ambas series deberían coincidir al menos aproximadamente de 1977 en adelante si se refiriesen al mismo tramo de edad pues en ambas se utilizan criterios similares de clasificación (consistentes con la metodología de la EPA 1987 en lo que respecta al cálculo de los principales agregados de la encuesta) y que en ninguna de ellas se ha introducido todavía la reponderación de los factores de elevación. Así pues, la única diferencia que en principio debería existir entre ambas series en este período se debe a que la del INE se refiere a la población 16+ mientras que la de Mas et al lo hace a la población 14+ durante el período que aquí nos interesa.

Para corregir esta diferencia procederé en dos etapas. En primer lugar, introduciré una corrección a las series de Mas et al utilizando la información disponible sobre la población de 14 y 15 años y sobre sus niveles de actividad y ocupación. De esta forma, se obtiene una serie corregida preliminar que, en una segunda etapa, se enlaza por retropolación²² con la serie H1 del INE en 1977 para llegar a la “serie H1 extendida.” Este procedimiento en dos etapas parece preferible al enlace directo de las series de población 14+ de Mas et al con las series de población 16+ del INE (que parece haber sido la metodología elegida en versiones posteriores del estudio del IVIE) porque este segundo procedimiento supondría implícitamente que el peso del colectivo de 14 y 15 años en la población total, activa y ocupada se mantiene constante entre 1964 y 1976, lo que podría no ser el caso. La primera etapa del procedimiento utilizado aquí permite incorporar la información disponible sobre la población de 14 y 15 años para obtener un perfil temporal de la serie en principio más ajustado a la realidad, mientras que en la segunda etapa se introduce una última corrección de nivel a partir de la estimación de las variables de interés que realiza el propio INE para 1977.

a) Población en edad de trabajar

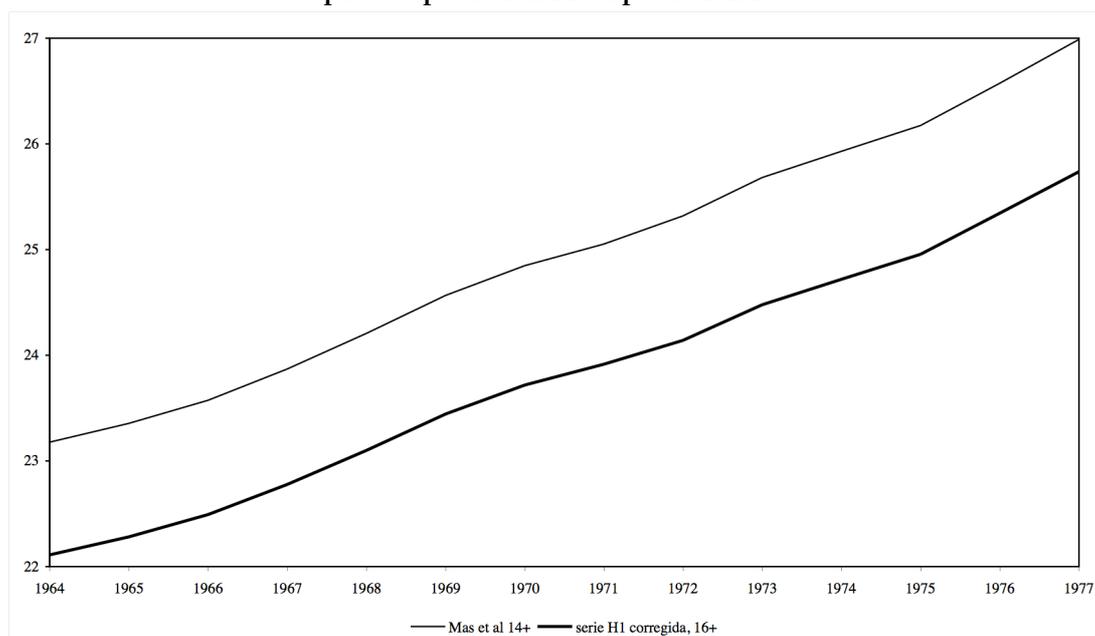
En primer lugar se procede a corregir la serie de población total en edad de trabajar de Mas et al para pasar de la población 14+ a la población 16+ utilizando datos de población desglosada

²² Esto es, la serie más reciente se extiende hacia atrás desde su primera observación utilizando las tasas de crecimiento de la serie más antigua.

por edades simples tomada del censo de 1960 y de las estimaciones intercensales de población del INE (2011d), que están disponibles desde 1971 en adelante.

El peso de la población de 14 y 15 años de edad (*w1415/14+*) en la población total de 14 o más años de edad (población 14+ de aquí en adelante) se obtiene de las fuentes citadas para (31 de diciembre de) 1960 y para (1 de julio de) los años de 1971 en adelante, interpolándose linealmente entre ambas fechas para completar la serie.²³ Esta serie de pesos se aplica seguidamente a la población 14+ de Mas et al para estimar la población total de 14 y 15 años por un lado y los mayores de 15 años (población 16+) por otro. A continuación la serie preliminar de población 16+ obtenida por este procedimiento se enlaza por retroprolación con la serie homóloga del INE que comienza en 1977 para obtener la serie definitiva de población 16+. Esto es, la serie H1 de población 16+ del INE se extiende hacia atrás hasta 1964 utilizando las tasas de crecimiento de la serie preliminar de población 16+ obtenida a partir de la serie de población 14+ de Mas et al. Finalmente, la serie preliminar de población de 14 y 15 años se ajusta de forma que sea consistente con la serie definitiva de población 16+ y con la población total 14+ de Mas et al. El resultado de la corrección final a la serie de población en edad de trabajar se muestra en el Gráfico 3.

Gráfico 3: Corrección de la serie de población en edad de trabajar de Mas et al (2010) paso de población 14+ a población 16+



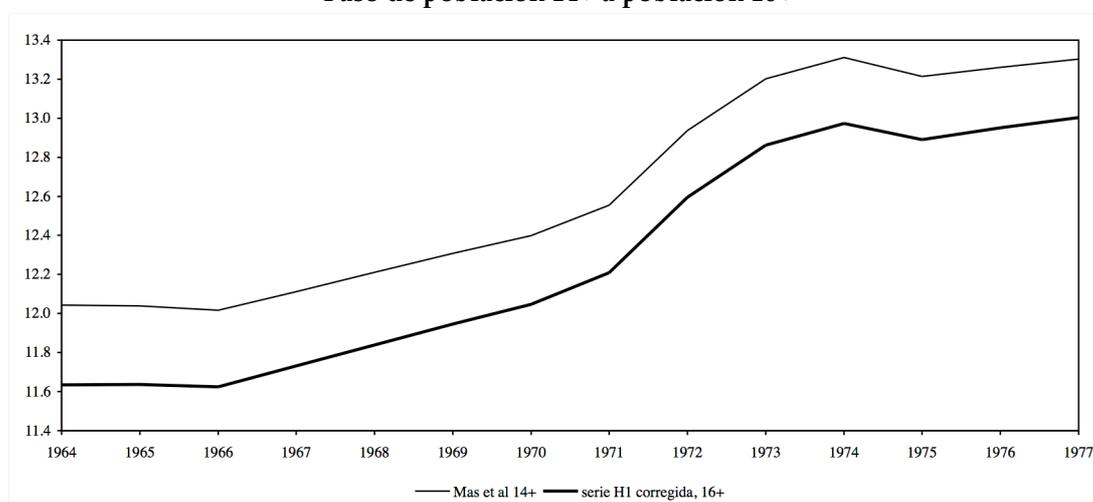
b) Población activa

Una vez completado el paso anterior, disponemos de una estimación de la población total de 14 y 15 años de edad (*POB1415*). Suponiendo que este colectivo es demasiado joven para haber comenzado el servicio militar, queremos estimar su desglose entre ocupados, parados e

²³ Un posible problema con los datos censales de población por edades es la comprobada tendencia de los censados a concentrarse en "edades redondas". En el presente caso, sin embargo, parece que la probable sobrevaloración de la población de 15 años se compensa aproximadamente con la infravaloración de la población de 14 años, lo que hace innecesario introducir una corrección. Véase la sección 1 del Anexo 1.

inactivos. Utilizando datos censales, podemos calcular la tasa de actividad de este grupo de edad, pero sólo en 1970 puesto que el censo de 1960 ofrece una desagregación de la población activa por grupos de edad que no permite aislar al colectivo deseado y el censo de 1981 no es una buena referencia puesto que se realiza después de la elevación a 16 años de la edad legal para trabajar.²⁴ Dado esto, una posible forma de proceder sería suponer que la tasa de actividad de la población de 14 y 15 años de edad se mantiene constante entre 1964 y 1977 al nivel observado en 1970 (bien en términos absolutos o bien en términos de su ratio con la tasa de actividad del conjunto de la población). Sin embargo, este supuesto no parece demasiado plausible dado que estamos hablando de unos años en los que la tasa de escolarización secundaria se elevó rápidamente, lo que presumiblemente se tradujo en un descenso de la tasa de participación en el mercado laboral del grupo de interés.

**Gráfico 4: Corrección de la serie de población activa de Mas et al (2010)
Paso de población 14+ a población 16+**



En consecuencia, he optado por suponer que la tasa de actividad de la población de 14 y 15 años de edad evoluciona entre 1964 y 1977 de la misma forma que la fracción no escolarizada de la población de 14 a 17 años. Para estimar esta última variable, he utilizado la información sobre el número de estudiantes de secundaria que proporciona el *Anuario Estadístico de España* (para más detalles, véase la sección 2 del Anexo 1). Para cada año entre 1964 y 1977, la fracción no escolarizada de la población se expresa como un índice, normalizando a 1 su valor en 1970. Este índice se multiplica por el valor observado de la tasa de actividad del colectivo de interés de acuerdo con el censo de 1970 (que es del 31,53%) para estimar la tasa de actividad de la población de 14 y 15 años en cada año entre 1964 y 1970. Aplicando esta tasa a la población total de 14 y 15 años se obtiene una estimación preliminar del número total de activos de esa edad y restando éstos del total de activos de 14 o más años según Mas et al, se obtiene una serie preliminar de población activa de 16 o más años. Finalmente, esta serie se enlaza por retropolación con la serie H1 de activos 16+ del INE que comienza en

²⁴ De acuerdo con el censo de 1981, la tasa de actividad de la población de 14 y 15 años sería del 16,4% y su tasa de paro del 70%. Estas cifras, sin embargo, no son una buena referencia para el período anterior a la aprobación del Estatuto de los Trabajadores (en 1980) pues parece muy probable que el cambio que esta ley introdujo en la edad legal para trabajar haya generado una discontinuidad importante en la senda de estas variables.

1977²⁵ (ajustando de la forma necesaria la serie de activos de 14 y 15 años, que también se conserva de cara a futuros posibles ajustes de las cifras de empleo de Contabilidad Nacional). El resultado final del ejercicio se compara con la serie original de Mas et al en el Gráfico 4.

c) Población ocupada y parada

El censo de 1970 ofrece un desglose por edades menos detallado para la población parada que para la población activa. Dentro del tramo de edades que nos interesa se distingue únicamente entre los menores de 15 años y los que tienen entre 15 y 19 años de edad. Para aproximar la tasa de paro de la población de 14 y 15 años en 1970, he utilizado una media ponderada de las tasas de paro de estos dos grupos, asignando un peso algo mayor (0,60) al grupo de 14 años o menos que a la población 15-19 por entender que la media de este segundo grupo seguramente estará dominada por su segmento de mayor edad, que probablemente presente un comportamiento muy distinto que el de las personas de 15 años que todavía están en edad de cursar estudios secundarios. Trabajando con datos del censo he estimado también la tasa media de paro de la población 15+ y calculado el diferencial entre esta última variable y mi estimación de la tasa de paro de la población de 14 y 15 años en 1970, que resulta ser de 8,7 puntos porcentuales.

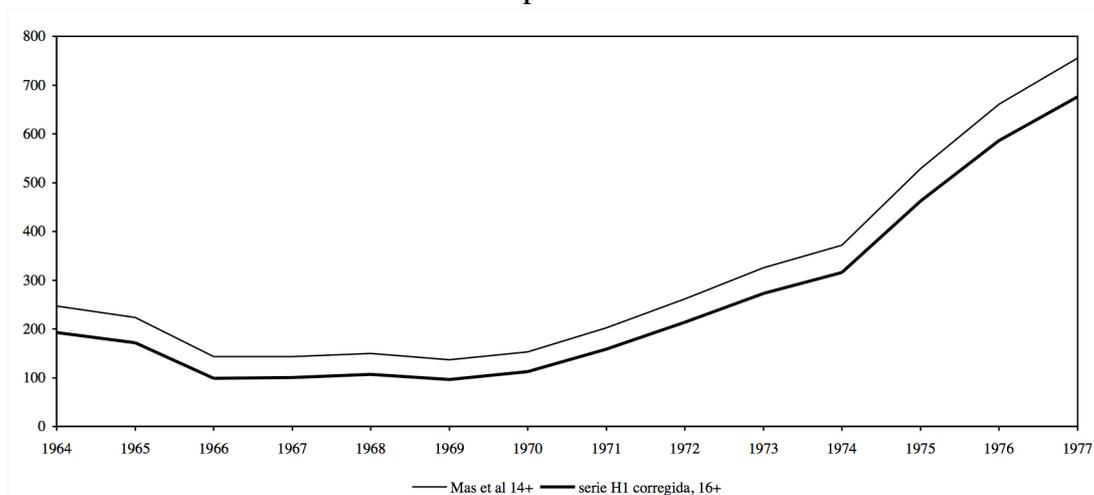
Para estimar la tasa de paro de la población de 14 y 15 años durante el resto del período de interés, he supuesto que el diferencial de paro entre la población de 14 y 15 años y el conjunto de la población 14+ se mantiene constante en el tiempo. Esto nos permite aproximar la tasa de paro de la población 14-15 como la suma del diferencial calculado arriba y la tasa de paro de la población 14+ de acuerdo con Mas et al. Aplicando esta tasa a la población activa estimada de 14 y 15 años (calculada en el apartado anterior) se obtiene una estimación del número de parados de 14 y 15 años que se substraen del total de parados 14+ para llegar a una serie preliminar del número total de parados 16+. Seguidamente, esta serie preliminar se enlaza por retropolación con la serie *publicada* de parados del INE que comienza en 1977 para obtener la serie definitiva de parados 16+ entre 1964 y 1976. Restando este colectivo de los activos de 16+ se llega finalmente a una estimación de los ocupados 16+ y, por diferencia, de los ocupados de 14 y 15 años de edad.²⁶ El Gráfico 5 compara las series corregidas de parados y ocupados (para la población 16+) con las series originales de Mas et al (para la población 14+).

²⁵ La discrepancia entre las dos series en el punto de enlace es mínima: el número estimado de activos mayores de 15 años es de 12,993 millones en la serie preliminar y de 13,004 millones en la serie publicada del INE con la que se enlaza la anterior.

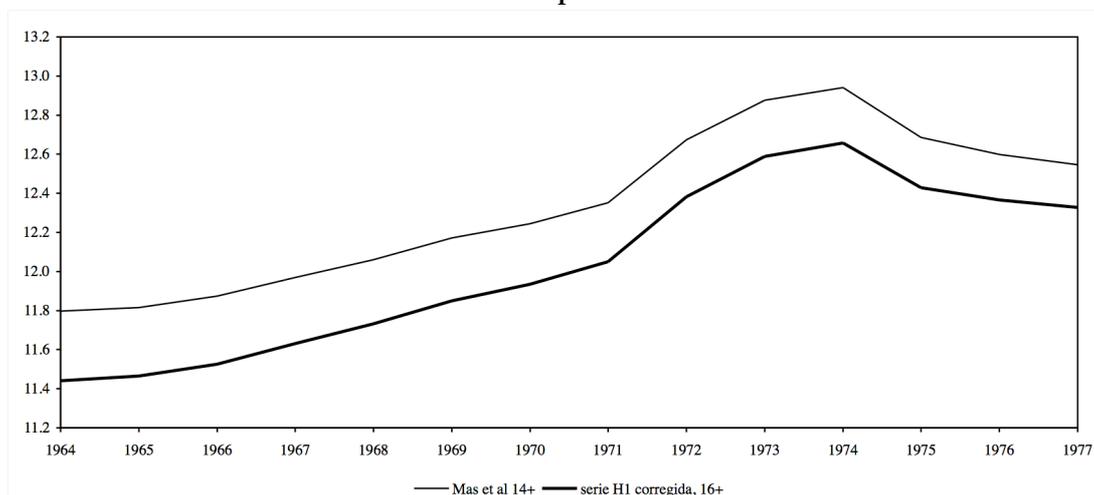
²⁶ Por su interés a la hora de comparar los datos de ocupación de la EPA con los de Contabilidad Nacional, la serie de ocupados de 14 y 15 años de edad se extiende hacia delante hasta 1984. Desde 1977 hasta 1981 la serie se prolonga por interpolación entre el último dato estimado por el procedimiento indicado en el texto y el dato censal de 1981 (de 66.592 efectivos). Finalmente, he supuesto que los ocupados de este grupo de edad convergen a cero cuatro años después de la elevación de la edad legal para trabajar a los 16 años, con decrementos iguales entre 1981 y 1984.

**Gráfico 5: Corrección de las series de Mas et al (2010),
paso de población 14+ a población 16+**

a. parados



b. ocupados



d) Población contada aparte y población inactiva

Como hemos visto, el GTE ha construido una estimación del número de varones llamados a filas para prestar el servicio militar durante cada año del período de interés. Esta serie se utiliza para extender desde 1977 hacia atrás la serie publicada de población contada aparte del INE. El enlace se realiza por retropolación. Combinando esta serie con los resultados descritos más arriba, se obtiene por diferencia la serie de población inactiva 16+.

4.2. Determinantes del tamaño de la corrección por reponderación

Tras los cálculos descritos en el apartado anterior, disponemos de una extensión hacia atrás (hasta 1964) de la primera serie histórica del INE, H1. Como hemos visto, el INE ha elaborado dos series históricas más recientes que coinciden entre sí durante el período 1977-96 y cuya única diferencia con H1 durante el mismo período parece provenir de la aplicación retroactiva del procedimiento de reponderación de los factores de elevación descrito en una sección anterior. Para prolongar hacia atrás tales series hasta 1964, por lo tanto, necesitamos

aproximar el efecto que la reponderación habría tenido sobre las series construidas en el apartado anterior para el período 1964-76. Con este fin, en este apartado investigaré los factores que influyen sobre el tamaño de la corrección por reponderación, definida como la diferencia porcentual entre las series H2 y H1 del INE, utilizando datos anuales del período 1977-2000. El ejercicio es también relevante de cara a la corrección de ciertas discontinuidades de las series históricas del INE pues podría darnos alguna indicación sobre el grado en el que la reponderación mitiga las discontinuidades en las series de la EPA que tienen su origen en cambios en el seccionado de la encuesta.

Como hemos visto, la reponderación sirve para imponer *a posteriori* ciertos rasgos de la estructura real de la población (fundamentalmente su desglose por sexo y tramos de edad) que la muestra original no estaría recogiendo correctamente debido a la no respuesta selectiva así como, posiblemente, al envejecimiento del marco de la encuesta. Puesto que el sexo y la edad están altamente correlacionados con la actividad económica, esta corrección serviría a su vez para mitigar los sesgos generados por los dos problemas citados. Dado esto, parece razonable suponer que el tamaño de la corrección resultante de la reponderación debería estar ligado a la magnitud de los problemas que pueden hacer que la muestra de hogares entrevistados no sea una buena representación de la población, esto es, al nivel de no respuesta y a la “distancia” existente entre la distribución espacial de la población en la fecha del censo que se utiliza para definir el seccionado y en la fecha de realización de la encuesta. También parece razonable suponer que el impacto de ambos problemas tenderá a ser mayor cuanto más envejecida esté la población.

El Cuadro 3 recoge los resultados de una serie de regresiones que intentan contrastar estas hipótesis trabajando con datos anuales a nivel nacional correspondientes al período 1977-2000, para el que se dispone a la vez de datos reponderados y no reponderados.²⁷ La variable dependiente es la corrección porcentual de las series originales de ocupados y parados que se introduce como resultado de la reponderación $((\text{serie reponderada} - \text{serie H1}) / \text{serie H1})$. Las variables explicativas son i) el porcentaje de hogares que se niegan a colaborar con la encuesta (*negativas*), sin considerar las ausencias, que en principio son menos problemáticas, ii) una variable que intenta aproximar el grado de envejecimiento del marco de la encuesta y iii) el peso de los mayores de 64 años en la población de 16 o más años de edad (*w65+*). Mi hipótesis de partida es que los coeficientes que miden el impacto de las dos primeras variables serán una función creciente del valor de la tercera, lo que se recoge mediante la introducción del producto de esta última variable con cada una de las otras dos en la ecuación a estimar.

El grado de envejecimiento del marco se aproxima mediante un indicador de la “distancia” que existe entre la distribución espacial de la población observada en el momento en el que se realiza la encuesta y la existente en el momento del censo (o padrón) que sirve como base para fijar el seccionado de la misma. Para aproximar esta variable se construye en primer

²⁷ La serie no reponderada es H1, mientras que la reponderada es muy similar a H2 como ya se ha indicado en la nota al pie número 16. Ambas series están disponibles en <http://www.ine.es/epa02/repercusion.htm>.

lugar un indicador de la distancia existente entre las distribuciones de la población observadas en cada par de censos sucesivos que viene a medir aproximadamente la fracción de la población que cambia de sitio entre un censo y otro. Seguidamente, la distancia acumulada entre dos censos sucesivos se distribuye sobre el período intercensal utilizando el perfil temporal del indicador anual de intensidad migratoria construido en de la Fuente (2010), lo que a su vez permite calcular la distancia entre cada año de interés y el censo o padrón correspondiente. (Para más detalles véase la sección 3 del Anexo 1).

En el Cuadro 3 se utilizan tres indicadores alternativos de distancia. El primero (*disttot*) mide la distancia “total” entre las distribuciones relevantes que se construye utilizando únicamente datos censales, el segundo (*distintraprov1*) se construye con los mismos datos pero utilizando sólo las variaciones en la localización de la población dentro de cada provincia y el tercero (*distintraprov2*) es, como el segundo, un indicador de distancia intraprovincial pero difiere de éste en que en su construcción se incorporan datos de los padrones y no sólo del censo, lo que permite tener en cuenta las renovaciones menores del seccionado que se suelen realizar durante el período intercensal sin modificar la estratificación (que se ignoran en los dos primeros indicadores de distancia).

Los tres indicadores que he utilizado son con seguridad medidas muy ruidosas de lo que se pretende medir (la intensidad de los cambios experimentados por la distribución espacial de la población entre dos puntos del tiempo) porque se basan en información muy agregada. La población de cada provincia se desagrega únicamente en tres partes: la residente en la capital, la que vive en el mayor del resto de los municipios (por población), y la que tiene su domicilio en el resto de los municipios de la provincia. Ninguno de los indicadores, por tanto, tiene en cuenta los cambios de localización que se producen dentro de los municipios y todos ellos recogen de forma muy imperfecta la movilidad entre municipios dentro de cada provincia. En principio, el indicador de movilidad intraprovincial debería ser el más relevante dado que las variaciones de las poblaciones totales de cada provincia no suponen un problema a efectos de las estimaciones de la EPA (pues tales poblaciones se recogen directamente y se utilizan para construir los factores de elevación). Sin embargo, es muy probable que la movilidad no observada a nivel de unidades menores, que es en principio la que más debería preocuparnos por sus efectos sobre la estimación, esté correlacionada con la movilidad interprovincial. Por este motivo, he experimentado con ambos indicadores.

Los resultados recogidos en el Cuadro 3 son consistentes con las hipótesis de partida y sugieren que como cabría esperar la reponderación de los factores de elevación corrige parcialmente los sesgos derivados de la no respuesta y del envejecimiento del seccionado. Los efectos del calibrado son mayores cuanto más elevado es el grado de no respuesta y mayor es la distancia entre la realización de la encuesta y la definición de su seccionado, medida en términos del cambio experimentado por la distribución espacial de la población entre ambos momentos. Además, el impacto de ambos factores aumenta con el envejecimiento de la población, tanto para los ocupados como para los parados. El porcentaje de negativas a participar en la encuesta tiene el signo esperado y es significativo cuando se incluye en la ecuación como único indicador del grado de no respuesta (ecuaciones [1], [4] y [7]). Sin

Cuadro 3: Determinantes de la corrección porcentual al alza de cada serie como resultado de la reponderación

	<i>serie de ocupados</i>								
	[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]	[9]
<i>constante</i>	0.0127 (2.11)	0.0408 (5.59)		0.0125 (2.08)	0.0406 (5.63)		0.0162 (2.55)	0.0312 (2.82)	
<i>negativas</i>	0.87 (5.19)	-5.31 (3.90)		0.87 (5.28)	-5.20 (3.89)		0.78 (5.22)	-4.08 (2.33)	
<i>negativas*w65+</i>		27.08 (4.41)	6.00 (16.67)		26.49 (4.38)	5.93 (16.23)		23.27 (3.04)	5.28 (10.19)
<i>distintraprov1</i>	0.07 (1.01)	-1.52 (2.33)							
<i>distintraprov1*w65+</i>		10.36 (2.49)	1.71 (6.74)						
<i>disttot</i>				0.07 (1.06)	-1.29 (2.37)				
<i>disttot*w65+</i>					8.78 (2.54)	1.49 (6.79)			
<i>distintraprov2</i>							0.04 (0.29)	0.83 (0.79)	
<i>distintraprov2*w65+</i>								-4.26 (0.68)	2.59 (5.57)
<i>Root MSE</i>	0.0055	0.0039	0.00602	0.00549	0.00387	0.00599	0.00562	0.00462	0.00679
	<i>serie de parados</i>								
	[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]	[9]
<i>constante</i>	-0.0027 (0.58)	0.0114 (1.55)		-0.0028 (0.60)	0.0112 (1.54)		0.0130 (2.36)	0.0151 (1.26)	
<i>negativas</i>	1.11 (8.42)	-1.83 (1.34)		1.10 (8.53)	-1.74 (1.29)		0.80 (6.18)	-0.31 (0.17)	
<i>negativas*w65+</i>		12.49 (2.03)	5.22 (22.01)		12.00 (1.97)	5.17 (21.49)		5.69 (0.68)	5.17 (11.54)
<i>distintraprov1</i>	0.15 (2.72)	-1.01 (1.54)							
<i>distintraprov1*w65+</i>		7.50 (1.80)	1.16 (6.95)						
<i>disttot</i>				0.13 (2.76)	-0.84 (1.52)				
<i>disttot*w65+</i>					6.25 (1.79)	1.01 (7.00)			
<i>distintraprov2</i>							-0.11 (1.03)	0.37 (0.33)	
<i>distintraprov2*w65+</i>								-2.79 (0.41)	1.28 (3.20)
<i>Root MSE</i>	0.00431	0.00392	0.00397	0.00429	0.00391	0.00395	0.00489	0.00502	0.00586

Notas:

- Estadísticos *t* entre paréntesis debajo de cada coeficiente estimado.
- Las ecuaciones se estiman con datos agregados anuales del período 1977-2000.

embargo, la introducción de la interacción entre esta variable y el grado de envejecimiento generalmente permite mejorar los resultados y parece indicar con claridad que el efecto de la no respuesta aumenta con el envejecimiento de la población. La introducción del efecto de

interacción es todavía más importante en el caso de la variable de distancia. Cuando trabajamos con la serie de ocupados, la distancia por sí sola no resulta significativa, y es sólo cuando se introduce su producto con el peso de la población 65+ que los efectos de esta variable resultan aparentes (ecuaciones [2] y [3] o [5] y [6]). Por otra parte, el Cuadro 3 no permite discriminar entre los dos indicadores de distancia basados sólo en datos censales, que generan resultados muy similares (compárese la ecuación [2] con la [5] o la [3] con la [6]). Sin embargo, el indicador de distancia que incorpora datos padronales y tiene en cuenta las renovaciones menores del seccionado (*distintraprov2*) no resulta significativo en la mayor parte de las especificaciones y suele presentar además el signo “incorrecto” (véanse las ecuaciones [7] y [8]). Este resultado sugiere que las renovaciones menores del seccionado realizadas entre censos podrían no ser demasiado efectivas para combatir el envejecimiento del marco.

Finalmente, los resultados sugieren que el procedimiento de calibrado puede corregir en parte el sesgo derivado del envejecimiento del seccionado, aunque resulta difícil ser mucho más preciso. Para el caso de 1995-96, las estimaciones de Toharia (1998) implican una afloración de 378.000 ocupados como resultado de la renovación del seccionado, lo que supone un 3,05% del empleo que se habría observado en 1997 sin la renovación de la muestra. Para intentar ver hasta qué punto la recalibración corrige esta discontinuidad podemos calcular el salto en el índice de distancia que se produce como resultado de la renovación (esto es la diferencia entre el índice real de distancia tras el cambio del seccionado en 1997 y el que se habría observado sin él en el mismo año) y utilizar las ecuaciones estimadas en el Cuadro 3 para calcular el salto inducido en la corrección por calibrado.

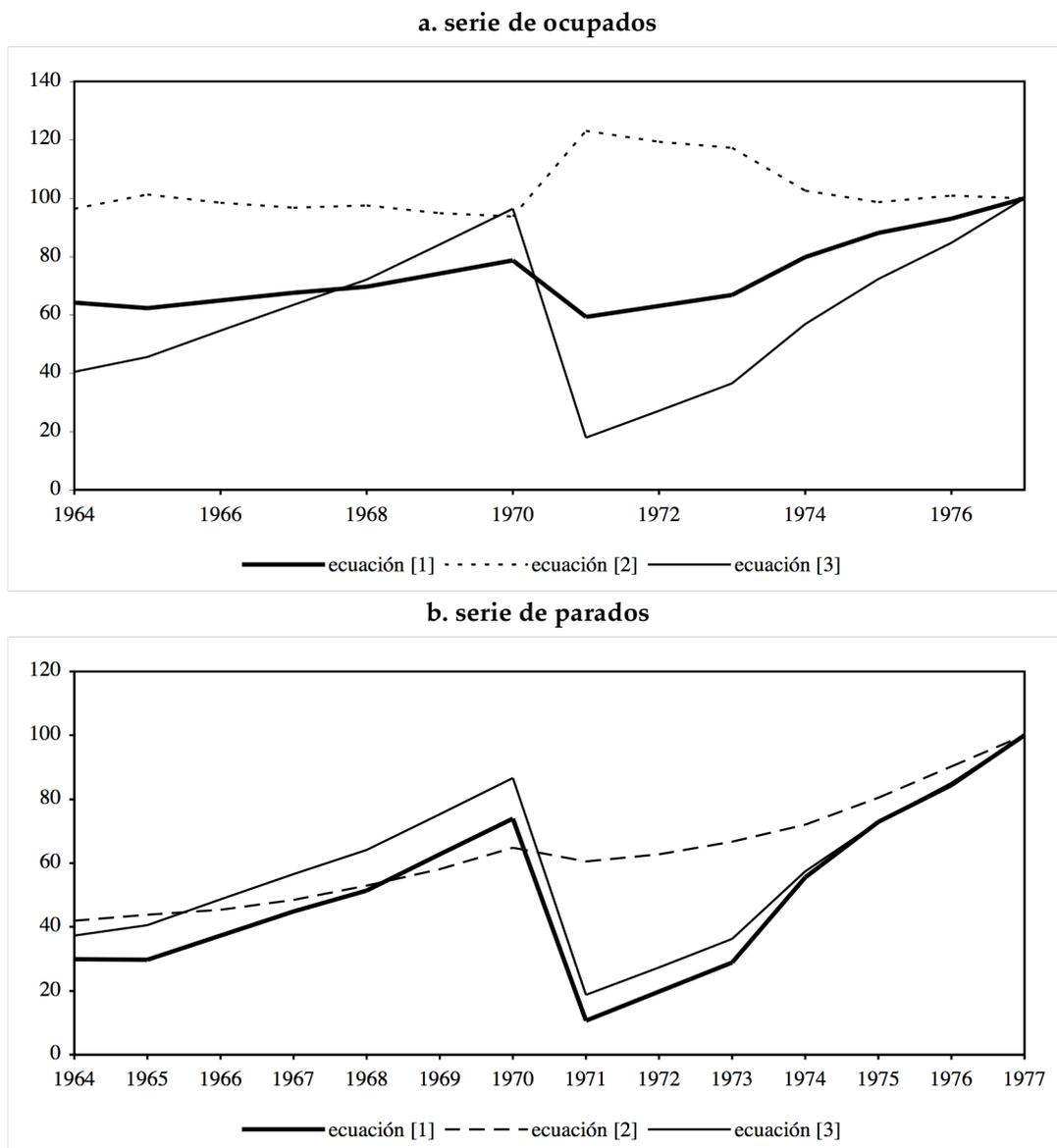
El resultado del cálculo es el 0,96% con los coeficientes estimados en la ecuación [3] y el 0,21% con la ecuación [1], lo que implicaría que el calibrado corrige entre un 7% y un 32% de la discontinuidad inducida por la renovación del seccionado de 1995-96. Por otra parte, la presumible mala calidad del indicador de distancia hace muy probable que la estimación subestime la sensibilidad del calibrado al envejecimiento de la muestra. La cifras que acabamos de calcular, por tanto, han de tratarse más bien como una cota inferior, y dejan abierta la posibilidad de que la corrección real sea mayor, aunque en principio uno no esperaría que la reponderación pudiese eliminar totalmente el sesgo derivado del envejecimiento del marco por la razón ya indicada más arriba.

En cualquier caso, hay que observar que el resultado no es extrapolable a otras actualizaciones del seccionado. La diferencia entre la predicción del modelo y la estimación directa del tamaño de la ruptura refleja seguramente lo bien o mal que el indicador de distancia que he construido recoge el envejecimiento real del seccionado. En 1995 la diferencia es probablemente mayor que en otros casos, dado que mi indicador sugiere un envejecimiento moderado del marco, mientras que el INE ha optado por renovar la muestra casi por completo.

4.3. Extensión hacia atrás de la segunda y tercera series históricas del INE (H2 y H3)

Puesto que la corrección resultante de la reponderación parece presentar un patrón predecible en función de variables conocidas, los resultados del apartado anterior nos permiten corregir las series construidas en el apartado 4.1 para el período 1964-76 de forma que sean aproximadamente consistentes con las series históricas más recientes del INE para 1977 y años posteriores, en las que ya se ha utilizado el procedimiento de reponderación.

Gráfico 6: Corrección generada por diversas ecuaciones estimadas (1977 = 100)

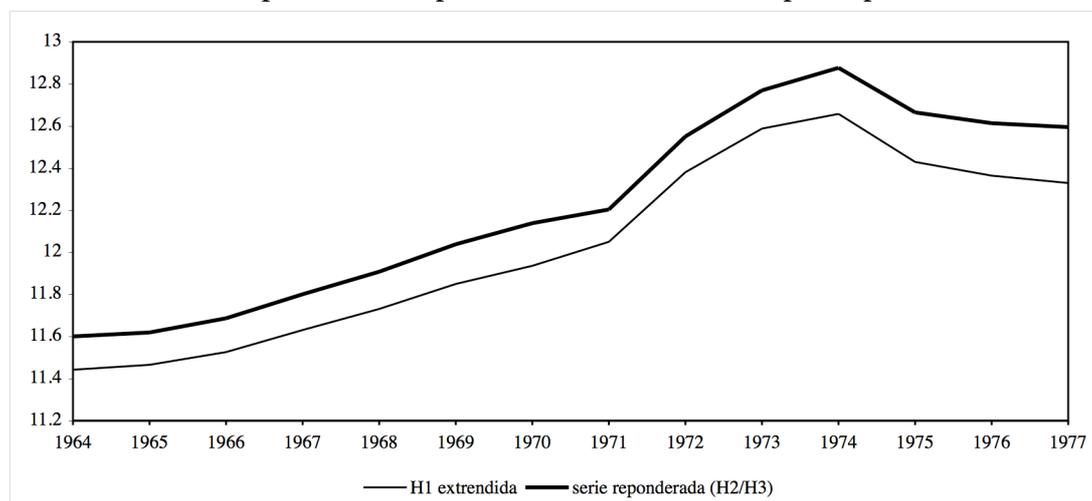


Utilizando las ecuaciones estimadas en el Cuadro 3 y el valor real de los regresores en cada año del período de interés, se construyen sendas series de predicciones del modelo para el período 1964-77 que se convierten en índices normalizando el valor de 1977 a 1. El Gráfico 6 muestra los resultados obtenidos con las ecuaciones [1] – [3], estimadas con el indicador de

distancia intraprovincial que en principio es el que mejor debería recoger el fenómeno causante del envejecimiento del seccionado. En el caso de la serie de ocupados, he descartado la ecuación [2] por generar una senda poco razonable de corrección durante el período 1964-77.²⁸ De las dos ecuaciones restantes, he optado por la ecuación [1] por ser la que genera un perfil más suave. Ésta es también la ecuación elegida para corregir la serie de parados, aunque en este caso no hay mucha diferencia entre los perfiles de corrección que generan las ecuaciones [1] y [3].

Finalmente, el índice generado por la ecuación [1] se multiplica por el valor observado de la corrección por reponderación en 1977 para estimar el valor de tal corrección en cada año entre 1964 y 1976. Aplicando tales correcciones a la serie H1 de ocupados y parados construidas en la sección anterior, obtenemos nuevas series “reponderadas aproximadamente” que enlazan con la segunda y tercera series históricas del INE (H2 y H3) en 1977. El Gráfico 7 muestra la serie de ocupados obtenida de esta forma y la compara con la serie análoga no reponderada.

Gráfico 7: Reponderación aproximada de la serie de ocupados para 1964-77



Finalmente, la población contada aparte de las series históricas H2 y H3 se extiende hacia atrás enlazándola por retroprolación con la serie (“H1 extendida”) de población contada aparte obtenida en el apartado anterior para 1964-76.

5. Estimación del tamaño de las discontinuidades en las series históricas extendidas del INE

En esta sección y en la siguiente se construyen series homogéneas revisadas de los principales agregados de la EPA en las que se intenta corregir las principales discontinuidades que persisten en las series históricas extendidas construidas en la sección anterior. Las nuevas

²⁸ Con los valores de los coeficientes estimados en [2], el coeficiente que multiplica a la variable de distancia (esto es, la suma del coeficiente estimado de esta variable más el producto del coeficiente del término de interacción por el peso de la población 65+) es negativo durante el período 1964-76 y esto hace que la corrección aproximada por reponderación presente un perfil poco plausible. En particular, el escalón que se produce en 1971 con la renovación del seccionado sería al alza en vez de a la baja tal como se observa en el Gráfico 6a.

series se construyen en dos etapas. En esta sección se aproxima el tamaño de las discontinuidades relevantes, mientras que en la siguiente se procede a la eliminación de las mismas mediante el enlace de los distintos segmentos de las series originales.

El Cuadro 4 identifica los cambios metodológicos que podrían generar discontinuidades en las series históricas extendidas y muestra, en su caso, mi estimación del tamaño de las mismas, indicando tanto el número de efectivos aflorados con cada cambio metodológico como el salto porcentual (de hecho logarítmico) que los mismos inducen en la serie en el punto de corte. Los detalles de tales estimaciones se presentan en el resto de esta sección, comenzando con un ejemplo relativamente complejo que utilizaré para exponer en detalle la metodología utilizada en la mayor parte de los casos.

Cuadro 4: Estimación del tamaño de algunas discontinuidades de la serie histórica extendida del INE (serie 2005 extendida)

<i>fecha</i>	<i>cambio metodológico</i>	<i>efecto estimado ocupados miles/dif log</i>	<i>efecto estimado parados</i>
1968T1	actualización menor del seccionado	0	
1971S2	actualización mayor del seccionado	0	
1972S2	introducción nuevo cuestionario	+204,6/1,64%	
1978T2	Actualización menor del seccionado	0	
1985T1	Nueva estratificación	+216,9/2,00%	
1985T2	Actualización mayor del seccionado	0	
1987T2	Cambios cuestionario	+176,8/1,52%	
1988T2	Inclusión de Ceuta y Melilla (en 1989)*	+35,6/0,28%	+15,0/0,57%
1989T1	Actualización menor del seccionado		
1995T1	Actualización mayor del seccionado (hasta 96T2)	+277,4/2,19%	
1999T1	Cambios cuestionario	+113,1/0,79%	
2000T1	Actualización menor del seccionado	0	
2002T2	Nueva definición operativa de paro		-480,1/-22,48%**
2005T1	Actualización mayor del seccionado Cambios cuestionario y recogida de información	0 +203,5/1,11%	

* Ocupados y parados en Ceuta y Melilla en 1989, media anual.

** = Diferencia entre las medias anuales correspondientes al año 2001 con la nueva definición de paro (serie 2005) y con la antigua definición.

Conviene distinguir entre las discontinuidades generadas por las actualizaciones del seccionado y los efectos de otros cambios metodológicos. A los efectos que nos interesan en esta sección, el primer tipo de discontinuidades plantean un problema que no existe en otros casos debido a que, por lo que hemos visto en secciones anteriores, sus efectos se ven corregidos parcialmente por la práctica del recalibrado, aunque en un grado que resulta imposible de precisar. Una consecuencia de este hecho es que las estimaciones existentes de su impacto a menudo no pueden utilizarse directamente puesto que han sido construidas trabajando con las series aún sin recalibrar. En tales casos, por lo tanto, resulta necesario examinar la serie recalibrada para intentar determinar el tamaño de la discontinuidad que pudiera persistir en la misma. En el resto de los casos, por el contrario, las estimaciones

existentes pueden en principio ser utilizadas directamente, si bien también pueden existir motivos para modificarlas.

Con las renovaciones del seccionado, la incidencia más frecuente son los cambios de cuestionario. Un objetivo común de las sucesivas mejoras que se han ido introduciendo en el cuestionario de la encuesta ha sido el de mejorar su capacidad para detectar a trabajadores no asalariados en situaciones de empleo ocasional o marginal que generalmente se habrían clasificado como inactivos en las primeras versiones de la EPA. Así pues, el efecto esperado sobre las cifras de paro de tales cambios metodológicos es en general muy limitado. Esta es una de las razones por las que en muchos casos he supuesto que los cambios metodológicos no generan discontinuidades en las series de parados. La otra es que mientras que en el caso de las series de ocupados existen ciertas referencias externas que resultan de utilidad a la hora de estimar el tamaño de las posibles discontinuidades, en el caso de los parados existe menos información externa.

Dos referencias muy útiles en un ejercicio como el que sigue son las que proporcionan la serie de afiliaciones a la Seguridad Social (MESS, 2012a)²⁹ y la serie de ocupados de la Contabilidad Nacional Trimestral de España (CNTR, INE, 2012c). Las discrepancias entre estas series y la serie de ocupados de la EPA pueden alertarnos sobre posibles anomalías en esta última, que es la fuente que ha experimentado con mayor frecuencia cambios metodológicos susceptibles de generar rupturas.

La serie de afiliaciones y la serie de ocupados de la EPA presentan diferencias significativas de nivel pero un perfil temporal generalmente similar (véase el Gráfico 8). Las diferencias de nivel entre ambas series se deben fundamentalmente a que la EPA recoge en principio una serie de colectivos que con frecuencia no cotizan a la Seguridad Social, tales como las ayudas familiares, los funcionarios incluidos en el régimen de clases pasivas, ciertos trabajadores del servicio doméstico y los ocupados en la "economía sumergida." El comportamiento de algunos de estos grupos (y en particular el de las ayudas familiares, que descienden acusadamente durante todo el período y el de los funcionarios integrados en el régimen de clases pasivas, que se mantienen aproximadamente constantes) tiende a hacer que la tasa de crecimiento interanual de las afiliaciones se sitúe generalmente algo por encima de la de la ocupación de la EPA. La inversión de este patrón puede ser una útil señal de alarma.

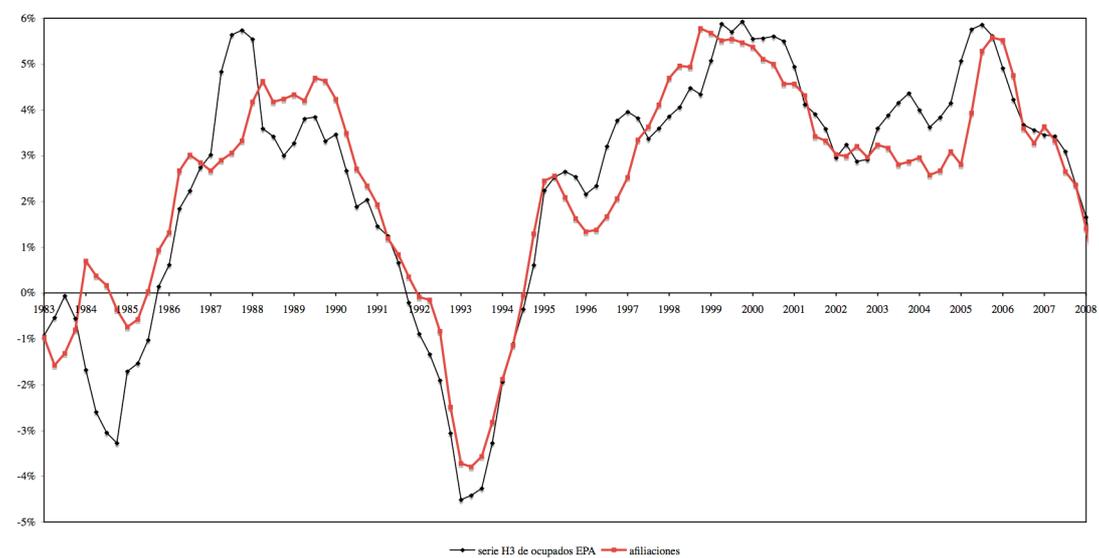
Algo similar sucede con las series de ocupados de la Contabilidad Trimestral (CNTR). La EPA es la fuente primaria pero no única de las estimaciones de empleo de esta estadística de síntesis en la que ha de asegurarse una cierta consistencia entre las estimaciones de producción y ocupación. Por lo tanto, cualquier desviación significativa de las tasas de crecimiento del empleo de la CNTR en relación con la EPA es un indicio claro de que el perfil

²⁹ La serie tiene en principio periodicidad mensual y recoge el número de afiliados de acuerdo con los ficheros de la Seguridad Social en el último día de cada mes. En esta categoría se incluyen los trabajadores en alta laboral y situaciones asimiladas, tales como la baja temporal por enfermedad, que generan obligación de cotizar. Una misma persona puede ser contabilizada varias veces si tiene varios puestos de trabajo. Para más detalles véase la sección de fuentes y notas de MESS (2012a).

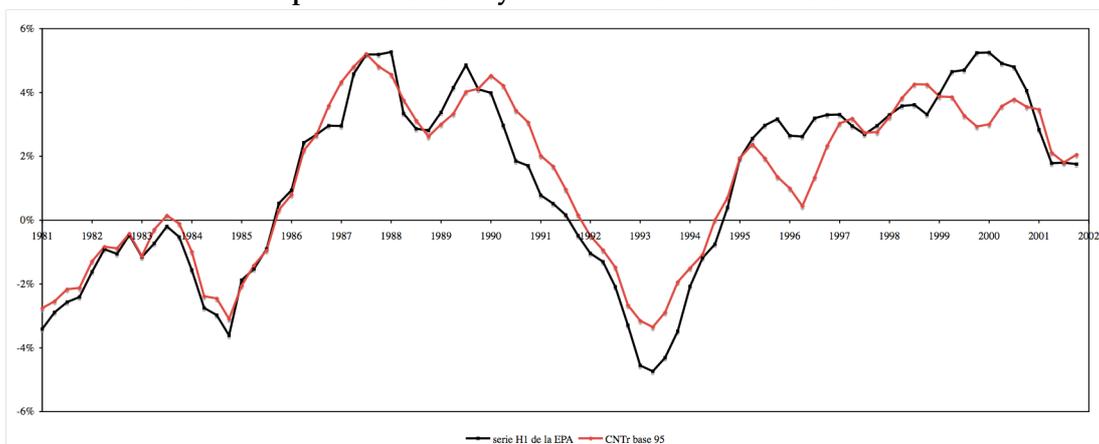
de la serie original de la EPA podría estar distorsionado por algún cambio metodológico cuyos efectos se han corregido de alguna forma en la CNTR.

Gráfico 8: Tasas de crecimiento interanual,

a. serie H3 de ocupados de la EPA y serie de afiliaciones a la Seguridad Social



b. serie H1 de ocupados de la EPA y serie de Contabilidad Trimestral base 1995



Finalmente, una tercera referencia que resulta útil especialmente en los primeros años de la EPA, cuando no se dispone de las dos series comentadas, es la serie "suavizada" de variaciones interanuales que utilizó en su día el GTE para depurar los primeros años de las series de la EPA. Como el lector recordará, el procedimiento utilizado por el GTE consiste en eliminar las tasas de variación en períodos afectados por cambios metodológicos para sustituirlas por otras construidas por interpolación entre las primeras tasas "no contaminadas" a ambos lados del corte. Cuando no se dispone de otra referencia mejor, este procedimiento puede utilizarse para aproximar el tamaño de la ruptura generada por un cambio metodológico como la diferencia entre la variación interanual observada con los datos de la encuesta en el momento del cambio y la tasa interanual que se obtiene por interpolación entre observaciones no contaminadas por el cambio. Un refinamiento de este procedimiento, que también emplearé más adelante, consiste en utilizar el perfil temporal de las series de

afiliación a la Seguridad Social o de la CNTR para estimar la tasa de crecimiento que la serie habría presentado en ausencia del cambio metodológico.

5.1. La actualización del seccionado de 1995-96

Durante 1995 y la primera mitad de 1996 se llevó a cabo una actualización del seccionado de la EPA que generó una importante ruptura de las series del INE. Comenzaré analizando el impacto de esta incidencia como forma de ilustrar la metodología que utilizaré repetidamente a lo largo de esta sección en un caso relativamente complejo en el que el cambio metodológico no se introduce de una vez sino que toma la forma de una renovación gradual de la muestra de secciones a lo largo de seis trimestres.

Existen diversos estudios en los que se han realizado estimaciones del tamaño de la discontinuidad en la serie original de la EPA que ha sido generada por esta incidencia. La primera de tales estimaciones (*INE1* en el Cuadro 5) se puede construir a partir de los datos que proporciona el INE en un estudio específico sobre el tema (INE, 1996a) realizado en el tercer trimestre de 1995, una vez completada la renovación de la primera mitad de la muestra. El estudio compara la mitad “nueva” de la muestra con la mitad “vieja” y encuentra que la primera de ellas es más joven que la segunda, con un peso mayor de los grupos de edad con tasas más elevadas de ocupación y participación. Aplicando las diferencias entre las tasas de ocupación y paro correspondientes a las distintas submuestras a la población total en edad de trabajar, se obtienen las estimaciones de afloración de empleo y paro que se muestran en la primera fila del Cuadro 5. El INE, sin embargo, advierte que el cálculo que acabo de describir podría ser engañoso pues existen indicios de que cada una de las dos mitades de la muestra, tomada por sí sola, no sería representativa del conjunto de la población, lo que hace aconsejable recurrir a algún otro procedimiento si se desea afinar los cálculos.

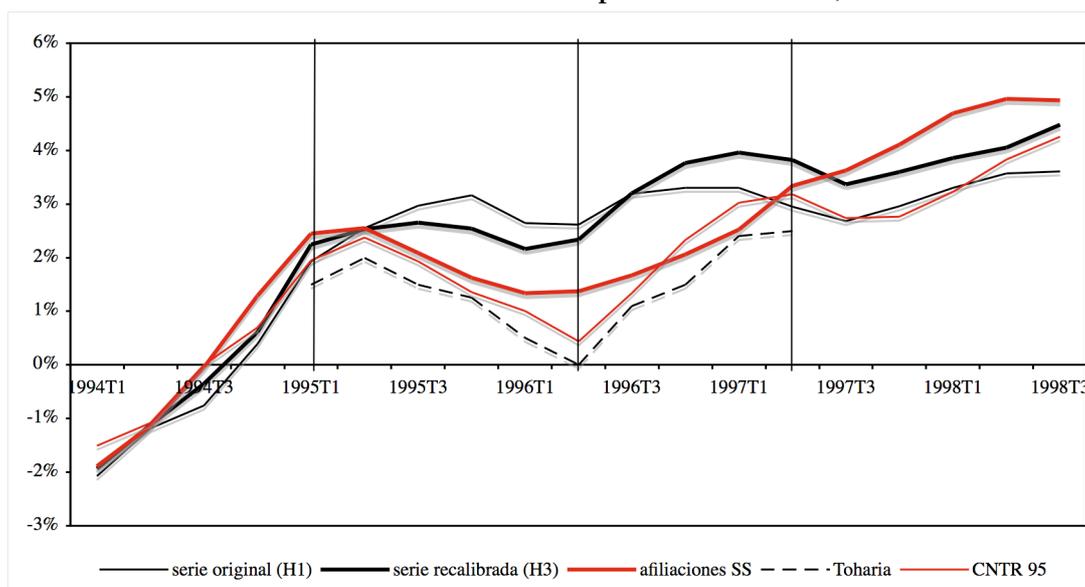
Cuadro 5: Afloramiento de ocupados y parados como resultado de la actualización del seccionado de 1995-96. Diversas estimaciones miles de personas

	<i>ocupados</i>	<i>parados</i>
<i>INE 1</i>	454,4	61,0
<i>INE 2</i>	225	-46,9
<i>Perez Infante</i>	299,5	
<i>BBVA</i>	425-500	0
<i>Toharia</i>	378	0

El propio INE utiliza un procedimiento alternativo para obtener una segunda aproximación de la discontinuidad (*INE2* en el Cuadro 5) que tampoco pretende ser demasiado ajustada. En una nota interna no publicada (INE, 1996b) se utilizan los datos detallados del Censo de 1991, desagregados a nivel de sección, para calcular la variación en la ocupación y el paro que se habría producido en ese año al cambiar la antigua muestra de secciones por la nueva.

Además de los cálculos del INE, disponemos de (al menos) tres estimaciones alternativas de analistas privados que también se resumen en el Cuadro 5. Pérez Infante (1998, citado en Pérez Infante 2006, p. 108) estima la afloración de empleo como la diferencia entre las tasas de crecimiento de esta magnitud derivadas de la EPA y de la CNE. El Servicio de Estudios del BBV (Izquierdo, 1998) ajustó una serie de modelos univariantes a los datos de empleo y paro desagregados en cuatros grandes sectores incluyendo también una “tendencia truncada” durante el período 1995-T1 a 1996-T2 para recoger el efecto del cambio en el seccionado y, en algunas versiones, un indicador “externo” de la evolución del empleo (la serie de afiliaciones). Trabajando con tales modelos se obtiene la estimación que se recoge en la cuarta fila del Cuadro 5. Finalmente, la última estimación proviene de Toharia (1998), quien aproxima el crecimiento que habría tenido la ocupación en ausencia de la renovación del seccionado trabajando con la parte de la muestra que se mantiene constante en cada par de trimestres consecutivos. Las estimaciones de Toharia implican una afloración de 378.000 ocupados³⁰ como resultado de la renovación del seccionado, lo que supone un 3,05% del empleo que se habría observado en 1997 de acuerdo con la serie publicada del INE de no haberse actualizado la muestra.

Gráfico 9: Variación interanual de los ocupados cerca de 1995, varias fuentes



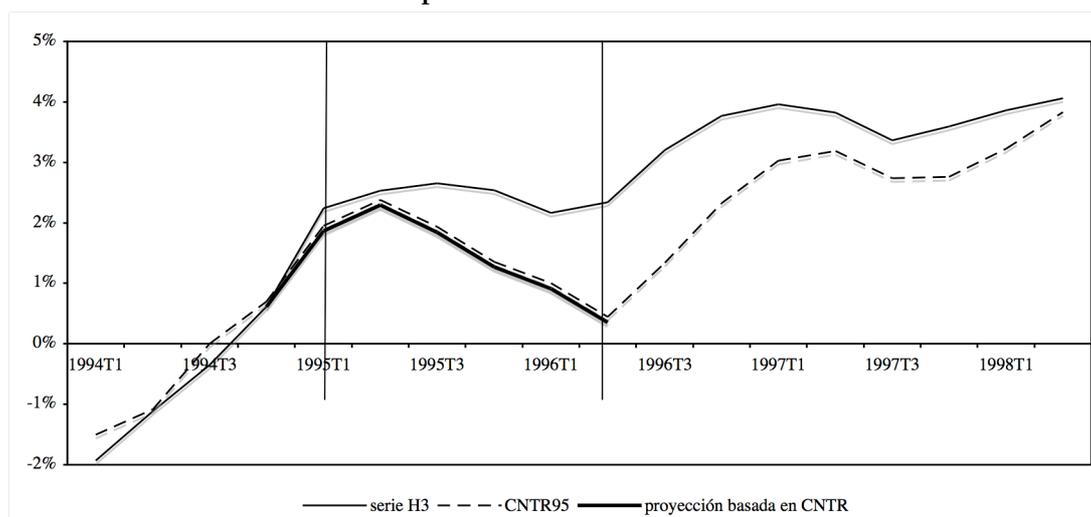
El Gráfico 9 muestra la evolución trimestre a trimestre de la tasa de variación interanual de cuatro series distintas de ocupación (las series históricas H1 y H3 de la EPA, la serie de ocupados de la CNTR base 1995 y la serie de afiliaciones a la Seguridad Social) junto con la estimación de Toharia de la misma magnitud. La ventana marcada por las dos primeras líneas verticales que aparecen en el gráfico corresponde al período durante el cual se realiza la actualización del seccionado, con la consiguiente distorsión en la tasa de variación

³⁰ Trabajando con la parte de la muestra que no cambia en cada trimestre, este autor estima que el crecimiento del empleo, neto del efecto de la renovación de la muestra, habría sido del 1,7% en 1995, el 1,2% en 1996 y el 2,6% en 1997 en vez del 2,7%, 2,9% y 3% de la serie original del INE (Toharia 1998, p. 10). Aplicando las tasas de crecimiento estimadas por Toharia al empleo de 1994 se obtiene una cifra de empleo “homogéneo” para 1997 que se compara con la cifra real observada para estimar el tamaño de la discontinuidad inducida por la renovación del seccionado.

interanual del empleo. Puesto que estamos trabajando con variaciones interanuales, tal distorsión afecta también al período comprendido entre la segunda y la tercera líneas verticales, pues la tasa de variación también se calculará en este período por diferencia entre un trimestre afectado por la renovación de la muestra y otro que no lo está.

La comparación entre las distintas series que se muestran en el gráfico revela que la renovación del seccionado tuvo el efecto de “camuflar” la clara desaceleración del crecimiento del empleo que se aprecia en las series de Toharia, CNTR y afiliaciones durante 1995 y 1996. Por otra parte, la introducción del recalibrado en la serie H3 parece mitigar en alguna medida el problema que presenta la serie original, pero no llega ni mucho menos eliminarlo del todo. Resulta por tanto, necesario, introducir una corrección en la serie. A la vista del Gráfico 11, he optado por utilizar el perfil temporal de la tasa de variación interanual de la serie de ocupados de la CNTR base 1995, que ocupa una posición intermedia entre las series de tasas de crecimiento de Toharia y de afiliaciones, para estimar el tamaño de la discontinuidad que la renovación del seccionado ha inducido en la serie H3 del INE.

Gráfico 10: Variaciones interanuales de la serie H3 de ocupados del INE y de la serie de CNTR95 y proyección de la variación interanual de los ocupados cerca de 1995 en base a la serie de ocupados de la Contabilidad Trimestral base 1995



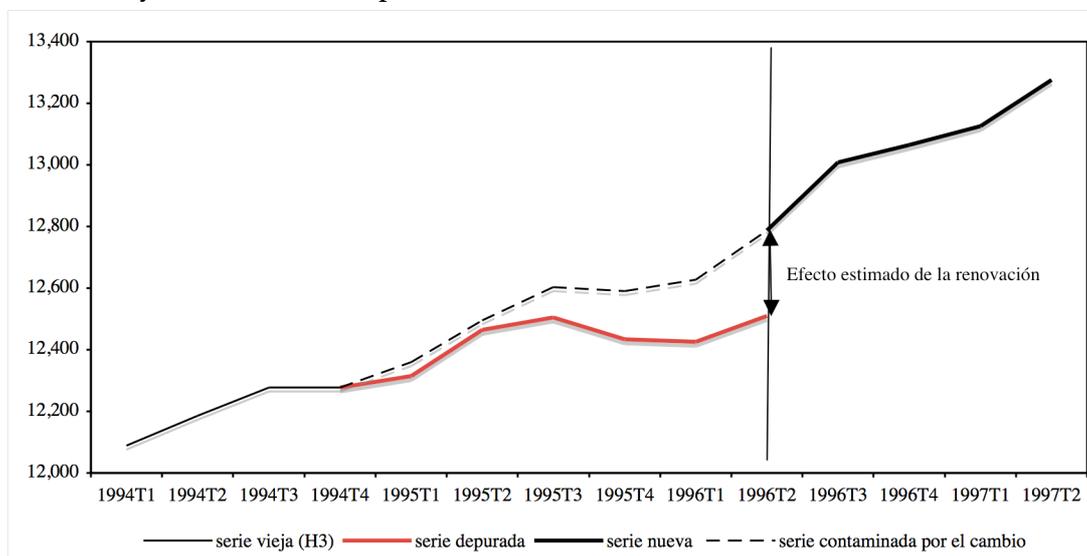
Tal como se ilustra en el Gráfico 10, la idea es utilizar la serie de la CNTR como referencia para estimar la tasa de crecimiento que la serie de empleo de la EPA habría registrado durante los trimestres en los que se produce la renovación de la muestra de secciones en ausencia de este cambio metodológico.³¹ Aplicando estas tasas de variación hipotéticas a la última observación de la serie (“vieja”) en niveles no contaminada por el cambio metodológico (1994-T4), se construye una “serie vieja depurada” de los efectos de tal cambio. Finalmente, el impacto de la renovación se estima como la diferencia entre la primera observación de la “serie nueva” (esto es la primera observación de la serie de empleo una vez

³¹ En particular, sea $tcia_t^S$ la tasa de crecimiento interanual de la serie S en el trimestre t . La proyección de la tasa de crecimiento de la serie de ocupados de la EPA para t entre 1995-T1 y 1996-T2 se obtiene sumando a la tasa observada de crecimiento de esta serie en 1994-T4 la variación observada de la tasa de crecimiento de la serie de ocupados de la CNTR entre t y 1994-T4, esto es:

$$\widehat{tcia}_t^{epa} = tcia_{94-T4}^{epa} + (tcia_t^{cntr} - tcia_{94-T4}^{cntr})$$

completada la renovación de la muestra) y la serie vieja depurada en el mismo trimestre (1996-T2). El Gráfico 11 muestra los niveles de las series relevantes y el tamaño estimado de la ruptura debida a la renovación del seccionado.

Gráfico 11: Series de empleo en niveles y estimación del impacto de la renovación del seccionado de 1995-96



El cálculo que acabo de describir me lleva a cifrar la afluencia de ocupados resultante de la renovación del seccionado en 277.400 efectivos, lo que supone una corrección al alza del 2,19% en la serie H3 de ocupados del INE en el segundo trimestre de 2006. Si el cálculo se realiza trabajando con la serie original (H1) en vez de con la serie recalibrada (H3), el impacto estimado es algo mayor, ascendiendo a 302.900 efectivos, o un 2,48% del empleo, lo que se sitúa dentro del rango de las estimaciones recogidas en el Cuadro 5.

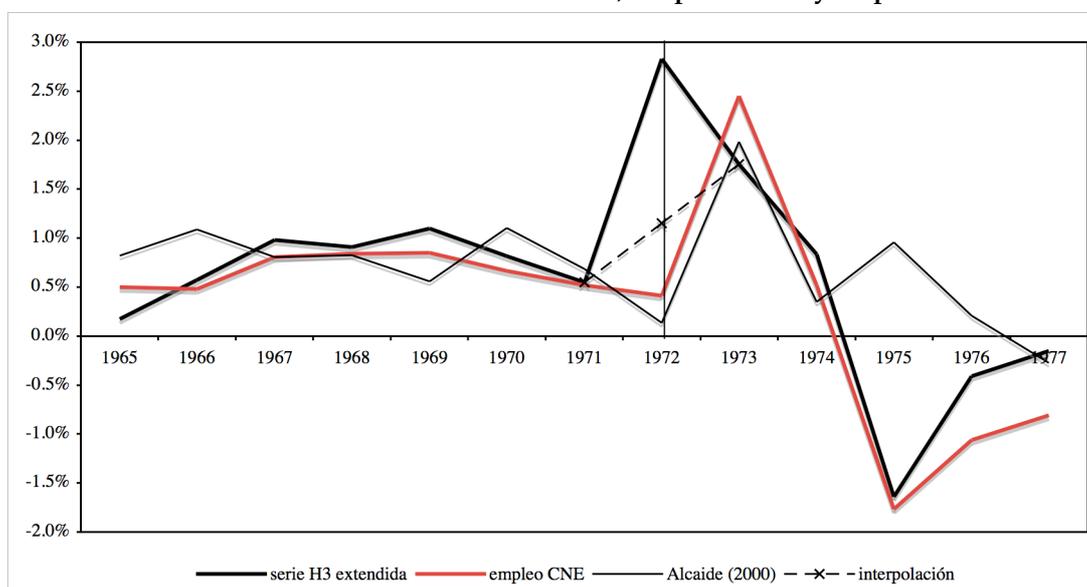
5.2. De 1964 a 1976

Durante el primer tramo de la EPA se producen tres incidencias que podrían haber generado discontinuidades aún no corregidas en las series de la encuesta: las actualizaciones del seccionado de 1968 y 1971 y la introducción de un nuevo cuestionario en 1972. El Gráfico 12 muestra las tasas de variación interanual año a año de la serie H3 extendida de ocupados construida en la sección 4 junto con el crecimiento anual del empleo (ocupados) de la CNE, tomado de Maluquer y Llonch (2005), y el de la serie de ocupados construida por Alcaide (2000b).³² El perfil de las series sugiere que la única posible discontinuidad importante en las fechas indicadas sería la ligada a la introducción del nuevo cuestionario en 1972, que podría explicar el fuerte repunte de la ocupación de la EPA en un año que, de acuerdo con las otras dos fuentes, estaría caracterizado por una ralentización del crecimiento del empleo. Por otra

³² Maluquer y Llonch (2005) extienden hacia atrás la serie de empleo de la CNE base 1986, enlazándola con bases anteriores de la propia CNE. En estos años todavía no se dispone de las afiliaciones a la Seguridad Social ni de la Contabilidad Nacional Trimestral. Alcaide (2000) construye una serie de ocupados que actualiza sus trabajos previos en el BBVA.

parte, el vigoroso crecimiento del PIB real que la CNE estima para el mismo año (en torno al 8%) plantea ciertas dudas sobre el perfil de estas dos últimas series de empleo.

Gráfico 12: Tasas de variación interanual, ocupados EPA y empleo CNE



Dado esto, he optado por una opción intermedia entre la EPA y las dos alternativas a la misma y he estimado la tasa de crecimiento del empleo en 1972 en ausencia del cambio de cuestionario por interpolación entre las observaciones de 1971 y 1973 de la propia serie de tasas de crecimiento de la ocupación EPA, tal como se muestra en el Gráfico 12. Comparando el nivel de ocupación que se habría alcanzado en 1972 bajo este supuesto hipotético con el dato real, se obtiene una estimación del tamaño del salto en la serie, que sería de 204.600 efectivos o un 1,64% del empleo.

5.3. De 1977 a 1986

Entre 1977 y 1986 se producen únicamente dos incidencias que podrían generar discontinuidades en las series de la EPA. La primera de ellas es la actualización menor del seccionado que se produce en 1978-T2. El Gráfico 13, donde se compara la tasa de crecimiento de la ocupación EPA con la de su homóloga en la CNE86, no ofrece indicios claros de la existencia de una discontinuidad. Aunque hay diferencias entre las dos series, su perfil temporal es similar entre 1977 y 1979.

La segunda incidencia digna de mención es la renovación del seccionado que se produce en 1985 junto con un cambio en los criterios de estratificación. Sobre la fecha exacta de estos cambios existe una cierta incertidumbre pues la misma publicación del INE (2002) ofrece dos fechas diferentes para el cambio en la estratificación: enero de 1985 (p. 44) y el segundo trimestre del mismo año (p. 78). Aunque no he conseguido una confirmación oficial de la fecha relevante, el examen de las series apunta hacia el primer trimestre, que es cuando se detecta un corte claro en la evolución del empleo.

Gráfico 13: Tasa de crecimiento interanual, 1978-84

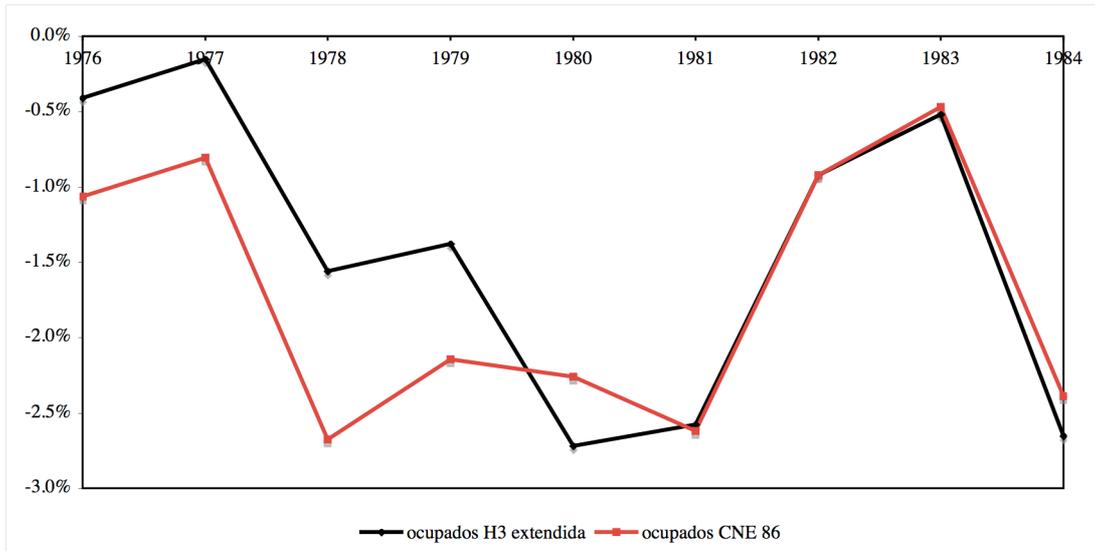
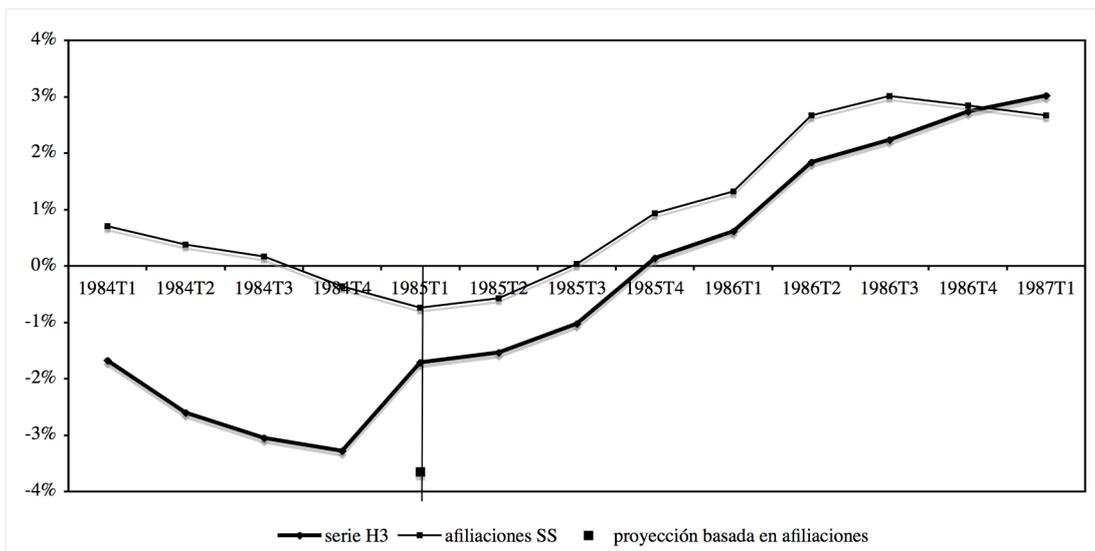


Gráfico 14: Evolución del empleo en torno a 1985T2, EPA vs. afiliación a la Seguridad Social tasas interanuales de variación trimestre a trimestre

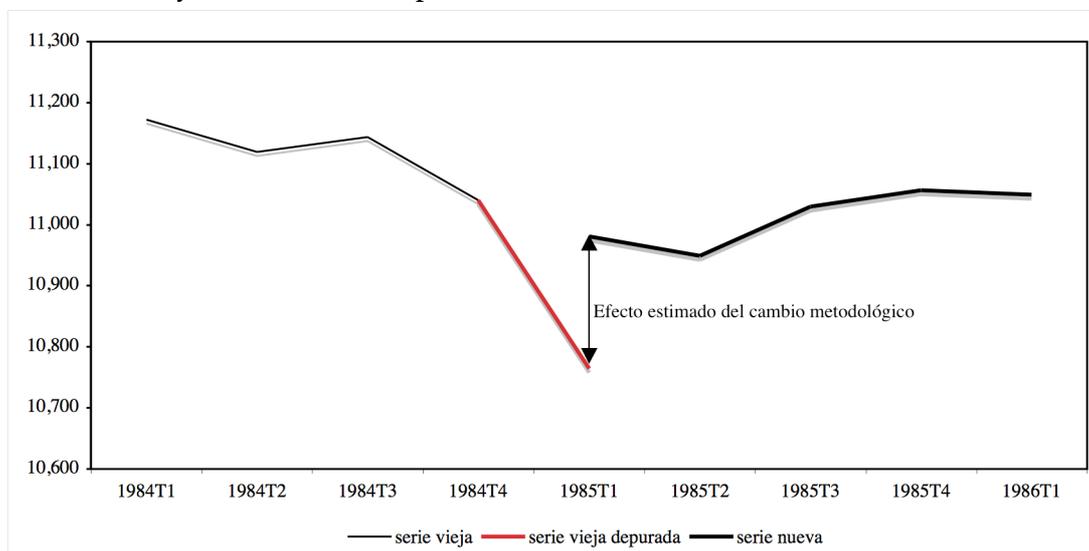


El Gráfico 14 compara las tasas de variación interanual de la serie H3 de ocupación del INE con las de la serie de afiliación a la Seguridad Social. El Gráfico revela un repentino y sospechoso estrechamiento del diferencial entre ambas series en el primer trimestre de 1985: mientras que la destrucción de empleo continúa acelerándose durante este trimestre de acuerdo con los datos de afiliación, la EPA muestra una fuerte mejoría de esta variable, con una caída de punto y medio en la tasa de destrucción de empleo.

El tamaño de la ruptura detectada en la serie se aproxima utilizando el perfil temporal de la serie de afiliaciones para estimar la tasa de crecimiento del empleo que se habría observado en ausencia del cambio metodológico siguiendo el procedimiento expuesto en mayor detalle en el apartado 5.1. La tasa homogénea de crecimiento interanual así estimada se aplica seguidamente al nivel de la serie de ocupación en 1984-T1 para estimar el nivel de esta variable que se habría observado en 1985-T1 en ausencia del cambio metodológico (véase el

Gráfico 15). Finalmente, este dato se compara con el dato real de ocupación en la misma fecha para estimar el tamaño de la discontinuidad. El resultado es una afloración de 216.900 ocupados (o un 2,00%, si tomamos la diferencia logarítmica entre ambas cifras).

Gráfico 15: Series de empleo en niveles y estimación del impacto de la nueva estratificación de 1985T1



5.4. Introducción de la EPA 1987

En 1987 se introducen cambios sustanciales en la metodología de la EPA que incluyen modificaciones en la definición de ciertas variables y mejoras en el cuestionario. El INE ha estimado separadamente los efectos de ambos cambios metodológicos, aplicando las nuevas definiciones a los antiguos cuestionarios y utilizando los resultados de una encuesta paralela a 10.000 familias. Los efectos estimados del cambio de las definiciones se han incorporado a las series históricas de la EPA elaboradas por el INE, pero no así la estimación de los efectos del cambio de cuestionario, que el propio INE recomienda utilizar con precaución y que han sido cuestionados por diversos autores.

Cuadro 6: Afloramiento de ocupados y parados como resultado de los cambios en el cuestionario de la EPA introducidos en 1987, estimaciones alternativas, miles de efectivos

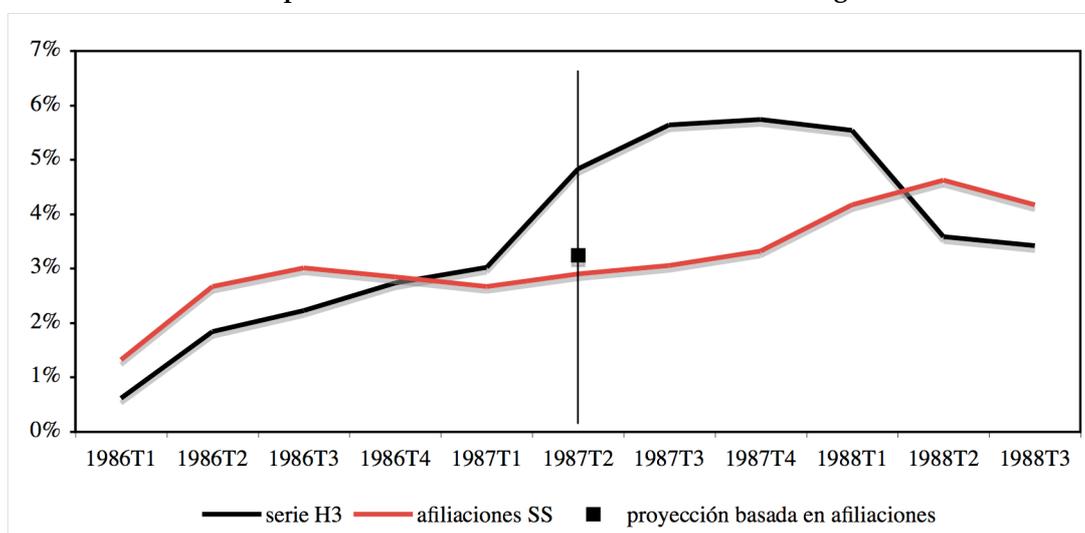
	<i>ocupados</i>	<i>parados</i>
<i>INE</i>	86,2	-47,4
<i>DGPC</i>	200,9	7,8

- Fuente: MTSS (1987). En esta publicación se compara el incremento observado entre 1986-T4 y 1987-T4 en el número de activos, ocupados y parados de acuerdo con cuatro fuentes: 1) las series originales de la EPA sin corrección alguna, 2) las series revisadas por el INE tras aplicar las nuevas definiciones a los viejos cuestionarios, 3) las series revisadas por la DG de Empleo del MTSS utilizando los resultados de la encuesta paralela del INE y 4) las series revisadas por la DGPC utilizando un modelo ARIMA. Las estimaciones del tamaño de la discontinuidad inducida por el cambio de cuestionario que se muestran en el Cuadro se obtienen como la diferencia entre 3) y 2) en el caso de la estimación del INE y como la diferencia entre 4) y 2) en el de la estimación de la DGPC.

El Cuadro 6 muestra dos estimaciones alternativas de la ruptura generada por el cambio de cuestionario. La primera es la basada en la encuesta paralela del INE, mientras que la

segunda ha sido elaborada por la Dirección General de Previsión y Coyuntura (DGPC, 1988) del Ministerio de Economía y Hacienda. Tras observar que los resultados de la encuesta paralela del INE no parecen excesivamente fiables,³³ la DGPC elabora una estimación alternativa del tamaño de la discontinuidad en las series de la EPA utilizando un modelo ARIMA para proyectar la evolución de tales series en ausencia del cambio de cuestionario. El análisis se realiza desagregando la serie de ocupados en asalariados y no asalariados. Como cabría esperar dada la naturaleza de los cambios en el cuestionario, la serie de asalariados no parece haber sido afectada por el cambio metodológico, mientras que la de no asalariados presenta un crecimiento mucho mayor de lo esperado. Los resultados de la DGPC apuntan hacia una afloración de empleo muy superior a la detectada por la encuesta del INE y un efecto prácticamente nulo sobre los parados.

**Gráfico 16: Tasa de variación interanual del empleo cerca de 1987T2
serie H3 de ocupados del INE vs. serie de afiliaciones a la Seguridad Social**



A la misma conclusión se llega también si utilizamos la serie de afiliaciones a la Seguridad Social para estimar lo que habría sido el crecimiento de la ocupación en ausencia del cambio de cuestionario. El Gráfico 16 compara la tasa de crecimiento interanual de la serie H3 de ocupados de la EPA con la de la serie de afiliación. En el momento del cambio metodológico, las dos series se separan, mostrando la primera de ellas una “joroba” durante cuatro trimestres consecutivos que parece un indicio claro de una significativa discontinuidad al alza. Si utilizamos el perfil de la serie de afiliaciones para estimar la tasa de crecimiento de la serie de ocupados en ausencia del cambio metodológico, la afloración de empleo estimada

³³ Véase DGPC (1988), p. 1187. De acuerdo con esta fuente, existen dos problemas fundamentales. El primero es que la comparación entre los resultados de la encuesta paralela y los de la encuesta oficial debería de haberse realizado utilizando la misma muestra de 10.000 familias, lo que no ha sido posible por no haberse publicado el segundo dato. El segundo problema es que la encuesta paralela presenta una elevada tasa de no respuesta muy concentrada, además, temporal y espacialmente. Presumiblemente por eso, la encuesta genera resultados atípicos, incluyendo entre otras cosas un brusco cambio en la composición por sexo y edad de la población ocupada. Fernández, Muro y Toharia (1991) comparten la desconfianza de la DGPC hacia los resultados de la encuesta paralela y señalan también que, como indica el propio INE, la encuesta paralela se realizó en una semana distinta de la encuesta oficial y por otro entrevistador, lo que también pudo distorsionar sus resultados.

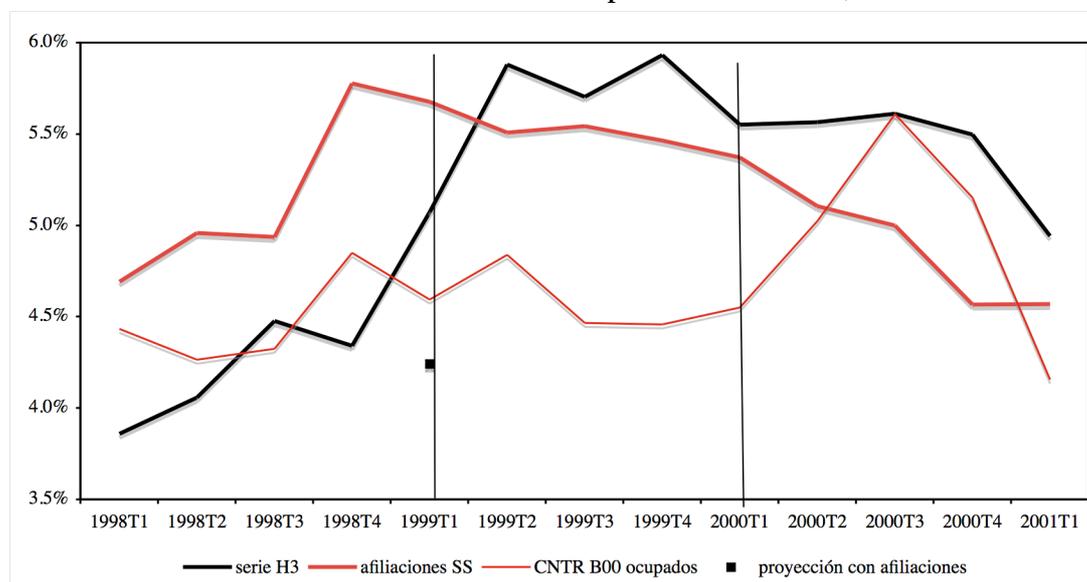
sería de 176.800 ocupados, una cifra muy cercana a la obtenida por la DGPC, lo que representaría una corrección al alza de la serie de ocupados del 1,52%.

En el primer trimestre de 1989 se produce una actualización menor del seccionado. La comparación de las tasas de crecimiento de las series de ocupados de la EPA y de las series de referencia que venimos manejando no revela indicios claros de una discontinuidad significativa.

5.5. Cambio de cuestionario en 1999

En el primer trimestre de 1999 vuelven a introducirse cambios importantes en el formulario de la EPA. Aunque el INE parece considerar que el nuevo cuestionario no ha generado una ruptura en las series de la encuesta, algunos analistas difieren al considerar que el incremento del empleo que muestra la EPA en el trimestre citado resulta sospechosamente elevado en comparación con otras fuentes (véase Pérez Infante, 2006, p. 109 y siguientes y Argentaria, 1999). El Gráfico 17 parece confirmar esta apreciación: de acuerdo con la EPA, el crecimiento de la ocupación se acelera significativamente en el primer trimestre de 1999 mientras que tanto la serie de afiliaciones como la de Contabilidad Trimestral muestran una ralentización del crecimiento del empleo. Utilizando el perfil de la serie de afiliaciones para estimar la tasa de crecimiento de la ocupación en ausencia del cambio de cuestionario, mi estimación es que el cambio de cuestionario genera la afloración de 113.100 ocupados, o una revisión al alza de la serie del 0,79%.

Gráfico 17: Variación interanual de los ocupados cerca de 1999, varias fuentes



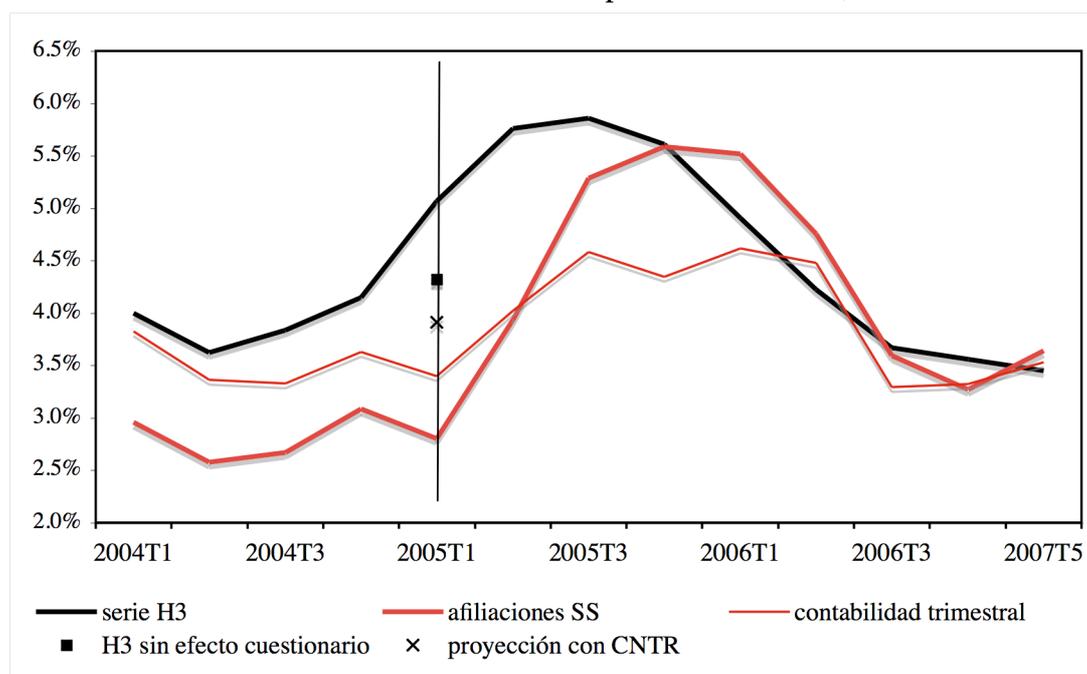
En el primer trimestre de 2000 se produce una nueva renovación menor del seccionado. El Gráfico 17 es difícil de interpretar pero no ofrece evidencia clara de que esta incidencia haya generado una discontinuidad significativa. Por un lado, la tasa de crecimiento de la ocupación se sitúa por encima de la de la serie de afiliaciones, lo que no es habitual. Por otra parte, sin embargo, el diferencial entre ambas series se estrecha en el trimestre de interés, en

vez de ensancharse como sería esperable si se hubiese producido una afloración significativa de empleo con la renovación. Finalmente, la serie de CNTR muestra un comportamiento muy distinto del de las otras dos durante todo 2000. Puesto que además en este año se enlaza con la base 95, es posible que la tasa de crecimiento anual esté un tanto distorsionada.

5.6. Cambios metodológicos en 2005

La siguiente ronda de cambios en la metodología y la muestra de la EPA se produce en el primer trimestre del año 2005. Por un lado, se introduce un nuevo cuestionario así como cambios en el procedimiento de recogida telefónica de datos. Por otro, se inicia también una nueva actualización del seccionado que se realizará de forma escalonada a lo largo de seis trimestres. El INE ha estimado el impacto del cambio de cuestionario utilizando una encuesta paralela, la llamada EPA testigo, pero no parece haber considerado necesario estimar la posible discontinuidad generada por la actualización del seccionado. Por otra parte, Perez Infante (2006, pp. 121 y siguientes) sostiene que es muy probable que la estimación del INE basada en la EPA testigo infravalore los efectos del cambio de cuestionario.

Gráfico 18: Variación interanual de los ocupados cerca de 2005, varias fuentes



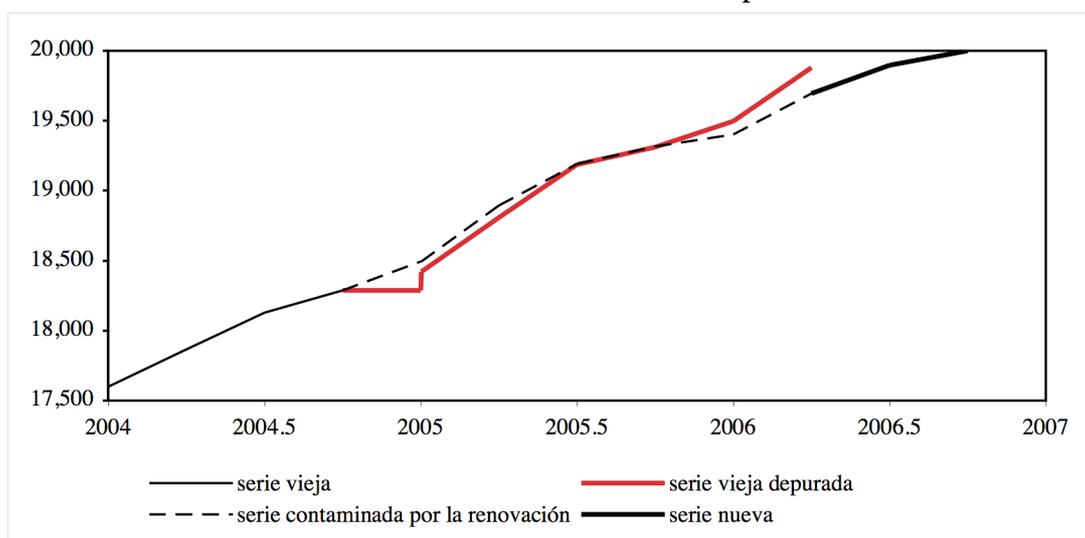
El examen de los datos de la EPA y su comparación con otras series de referencia sugiere que, en efecto, se produce una discontinuidad en 2005 y que ésta es probablemente algo mayor de lo que estima el INE en base a la EPA testigo. El Gráfico 18 resume los datos relevantes. Obsérvese que en 2005-T1 el crecimiento de la serie de ocupados de la EPA registra una significativa aceleración (la tasa de crecimiento interanual aumenta en casi un punto porcentual en relación al trimestre anterior), mientras que tanto la serie de afiliaciones a la Seguridad Social como la de CNTR muestran una cierta ralentización.³⁴

³⁴ Por otra parte, la serie de afiliaciones está distorsionada durante el segundo y tercer trimestres de 2005 por el proceso de regularización de inmigrantes que se produjo en esos meses. Según cálculos de

El gráfico también sugiere que la estimación que hace el INE del efecto del cambio de cuestionario (+132.000 ocupados) no parece suficiente para explicar la aparente aceleración del crecimiento del empleo. Corrigiendo a la baja la tasa de crecimiento del empleo en 2005-T1 por este factor (*H3 sin efecto cuestionario* en el gráfico), la cifra resultante sigue situándose por encima de la tasa que se obtiene proyectando el crecimiento de la EPA con el perfil temporal de la CNTR base 2000 (*proyección con CNTR* en el gráfico).

La diferencia entre ambas tasas de crecimiento puede deberse (además de, por supuesto, al error muestral de la propia encuesta) a dos cosas. Una es el efecto de la renovación del seccionado y otra la posible infravaloración del tamaño de la discontinuidad generada por el cambio de cuestionario. Si se tratase del primer factor citado, el tamaño de la discontinuidad debería aumentar con el paso del tiempo según se va renovando la muestra de secciones. Sin embargo, a juzgar por el Gráfico 19, éste no parece haber sido el caso. Para aproximar la evolución de la serie de ocupados de la EPA en ausencia de la renovación del seccionado, tras incorporar en 2005-T1 el salto en la ocupación que el INE atribuye al cambio de cuestionario, la serie se prolonga hacia delante en el gráfico utilizando el perfil temporal de la serie de la CNTR. En contra de lo que cabría esperar si tal renovación hubiese generado una afloración significativa de empleo, la diferencia entre la serie original y la serie depurada se reduce con el paso del tiempo y termina siendo negativa en 2006-T2, que es cuando se completa la renovación de la muestra de secciones. Lo mismo sería cierto, con un salto a la baja mucho mayor, si se tomase como referencia la serie de afiliaciones a la Seguridad Social. Dado que esto parece poco plausible, lo más probable es que tal renovación no haya tenido un efecto significativo sobre la senda de la serie de ocupados.

Gráfico 19: Proyección de la serie de ocupados de la EPA en ausencia de la renovación del seccionado con el perfil de la CNTR00



los ministerios de trabajo y de economía, se estima un incremento de las afiliaciones por este motivo de entre 450 y 550 miles. Véase por ejemplo: http://serviciosweb.meh.es/APPS/DGPE/TEXTOS/sie/historicos/sie_2006/sie_ene06.pdf

Si esto es cierto, para que las cifras cuadren tenemos que suponer que la encuesta testigo del INE infravalora los efectos del cambio de cuestionario. Utilizando el procedimiento habitual de proyectar el crecimiento del empleo en 2005-T1 en ausencia de cambios metodológicos utilizando el perfil temporal de la CNTR, la estimación de la discontinuidad en la serie de ocupados se eleva a 203.500 efectivos, lo que supone un ajuste al alza del 1,11%.

6. Construcción de las series enlazadas de ocupados y parados

Una vez estimada la magnitud de las discontinuidades generadas en la serie H3 de la EPA por distintos cambios metodológicos, se procede a construir la serie corregida final por el procedimiento de enlazar los distintos segmentos de las series de ocupados y parados separados por tales incidencias.

En este punto conviene introducir algo de notación para poder describir con precisión los procedimientos de enlace que utilizaré en esta sección. Supongamos que en el momento T_n se produce la n -ésima incidencia que genera una discontinuidad en algún agregado de la EPA al permitirnos medir mejor alguno de sus componentes. Llamemos x_t , a la serie original (no corregida) del agregado de interés medida en logaritmos, y_t^n al logaritmo de la serie corregida por la incidencia n que queremos construir para $0 \leq t \leq T_n$ y d_t^n a la diferencia logarítmica entre ambas series. Interpretaremos d_t^n como el error de medición contenido en la serie original antes de la introducción del cambio metodológico de interés e intentaremos adivinar su evolución en los años anteriores a T_n a partir del único valor conocido de esta magnitud -- el correspondiente al punto de enlace, $d_{T_n}^n$ -- que ha sido estimado en la sección anterior. Así pues, estimaremos y_t^n como

$$(1) \hat{y}_t^n = x_t + \hat{d}_t^n \text{ para } 0 \leq t \leq T_n$$

utilizando varios procedimientos alternativos sobre los que volveré enseguida para estimar la senda de \hat{d}_t^n entre el comienzo de la serie ($t = 0$) y el punto de enlace, T_n .

Una vez construida la serie corregida por la incidencia n trabajando en logaritmos, se recupera el nivel de la misma en cada período, $\hat{Y}_t^n = \exp(\hat{y}_t^n)$, y se calcula la corrección a la serie original en niveles, X_t , que viene dada por $\hat{D}_t^n = \hat{Y}_t^n - X_t$. Finalmente, se suman las correcciones correspondientes a cada una de las incidencias que estamos considerando durante cada año dado,

$$(2) \hat{D}_t = \sum_n \hat{D}_t^n,$$

y la serie final, corregida por todas las incidencias y expresada en niveles, se obtiene como

$$(3) \hat{Y}_t = X_t + \hat{D}_t$$

El procedimiento elegido para estimar d_t^n depende de la naturaleza de la incidencia cuyos efectos estamos tratando de corregir. Cuando se trata de una renovación del seccionado, supondré que la discontinuidad observada en el momento del enlace, d_{T_n} , se ha generado gradualmente durante el período transcurrido desde la anterior renovación (mayor) del seccionado, digamos en T_{n-k} , y que su crecimiento durante cada período se produce a un ritmo proporcional a la intensidad migratoria registrada en el mismo. En este caso, por tanto, el enlace se construye por interpolación tomando como referencia el indicador de intensidad migratoria, IM , construido en de la Fuente (2010) como el número de desplazamientos migratorios (interregionales o internacionales) observados durante el año por cada mil habitantes. Formalmente, definimos CIM_t como la fracción de los movimientos migratorios totales registrados entre ambas renovaciones del seccionado que se ha acumulado en el año $t \in [T_{n-k}, T_n]$,

$$(4) \quad CIM_t = \frac{\sum_{s=T_{n-k}}^t IM_s}{\sum_{s=T_{n-k}}^{T_n} IM_s}$$

y estimamos d_t^n como

$$(5) \quad \hat{d}_t^n = CIM_t * d_{T_n} \quad \text{para } t \in [T_{n-k}, T_n]$$

La otra incidencia más común es el cambio de cuestionario. En este caso supondremos que el error de medición en la serie original que aflora con la introducción del nuevo cuestionario ya estaba presente en alguna medida desde el comienzo de la serie, pero se va reduciendo según vamos hacia atrás en el tiempo de acuerdo con el perfil temporal generado por el procedimiento *mixto* de enlace propuesto en de la Fuente (2009b)³⁵ parameterizado por la vida media, H , del error de medición aflorado con el cambio metodológico. Fijando el valor de H en 10 años, tenemos

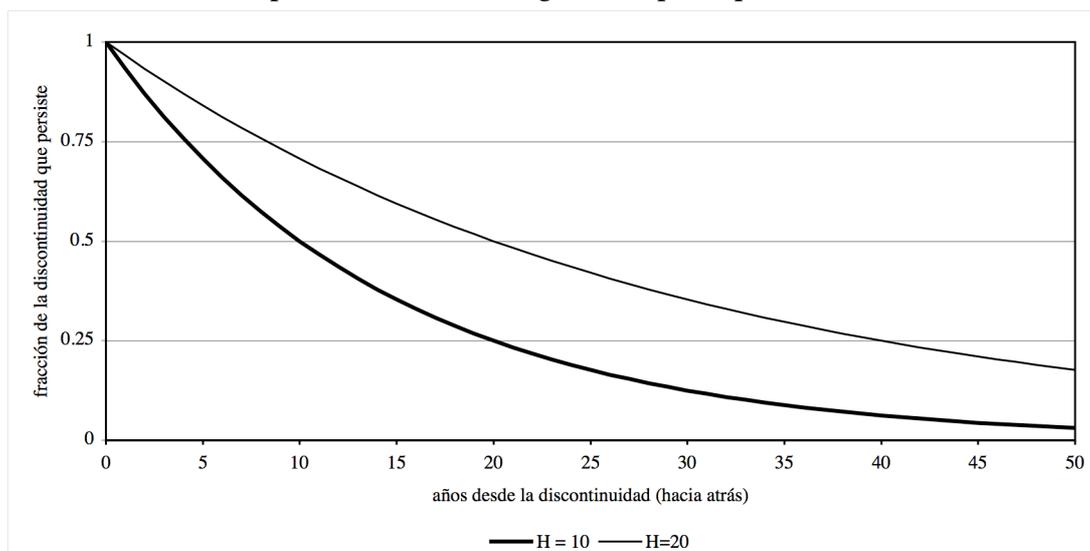
$$(6) \quad \hat{d}_t^n = d_{T_n} \left(\frac{1}{2} \right)^{\frac{T_n-t}{H}} \quad \text{para } t \in [0, T_n]$$

donde 0 corresponde en la práctica al año 1964 en el que comienzan las series de la EPA.

³⁵ El nombre del método hace referencia al hecho de que éste reparte la discrepancia aflorada en el momento del enlace a lo largo del tiempo de una forma que afecta tanto a las tasas de crecimiento de la serie más antigua como a su valor en el año base. La fórmula de enlace se obtiene a partir del supuesto de que el "error de medición" de la serie más antigua es proporcional a un componente "mal medido" del producto o del empleo total, Z , cuya tasa bruta de crecimiento (Z_{t+1}/Z_t) es un múltiplo constante de la tasa bruta de crecimiento del resto del agregado correspondiente. Bajo estos supuestos, d_t^n se puede

aproximar por $\hat{d}_t^n = d_{T_n} \rho^{\frac{T_n-t}{H}}$ donde el parámetro ρ mide la fracción del error de medición aflorado en T_n que ya estaba presente en la serie en $t = 0$. En el presente contexto, resulta útil reparameterizar esta expresión en términos de la vida media de dicho error, yendo hacia atrás. Para una discusión más detallada de este procedimiento de enlace, véase de la Fuente (2009b).

Gráfico 20: Senda temporal de la corrección generada por el procedimiento mixto de enlace



El Gráfico 20 ilustra el perfil de ajuste gradual generado por la ecuación (6) con dos valores alternativos de H . Con el valor elegido para H de 10 años, una afloración del 10% del empleo en el momento T_n implica una revisión al alza de toda la serie precedente con un perfil decreciente: la corrección de la serie “antigua” sería de 5 puntos porcentuales en $T_n - 10$, 2,5 puntos en $T_n - 20$ y 0,625 puntos en $T_n - 40$.

Las correcciones a la serie original se calculan generalmente trabajando con series de frecuencia anual. La única excepción tiene lugar cuando la discontinuidad se produce durante un trimestre distinto del primero y afecta, por lo tanto, a sólo una parte del año. En este caso se trabaja con datos trimestrales durante el año (o años) en el que se produce la discontinuidad. La corrección logarítmica estimada para cada trimestre se aplica a continuación a la serie depurada de los efectos del cambio metodológico para obtener la serie trimestral corregida, primero en logaritmos y luego en niveles. Finalmente, se calcula la media de la serie trimestral corregida en niveles para el año del enlace y se compara con la serie original para obtener la corrección correspondiente a ese año.

Incorporación de Ceuta y Melilla a la muestra de la EPA

Finalmente, hay dos incidencias que reciben un tratamiento *ad-hoc*. La primera es la incorporación a la muestra de la EPA de Ceuta y Melilla, que se produce en el segundo trimestre de 1988.³⁶ En este caso, he optado por estimar el número de activos, ocupados y parados de las dos plazas africanas a partir de su población mayor de 15 años (16+) y del supuesto de que los ratios entre las tasas de actividad y ocupación de Ceuta y Melilla y sus homólogas para el resto de España se mantienen constantes durante el período 1964-88 a los valores observados en 1989, que es el primer año completo en el que estas ciudades se incluyen en la muestra. La población 16+ de Ceuta y Melilla se aproxima proyectando hacia

³⁶ MTSS (1988b) ofrece algunas cifras sobre el impacto de la incorporación de las dos ciudades autónomas sobre las principales series de la EPA.

atrás el valor que ofrece la EPA de este agregado en 1989 utilizando la tasa de crecimiento de la misma variable de acuerdo con las estimaciones intercensales de población del INE (2010d).

Homogeneización de la definición de paro

Como hemos visto en la sección 3.1, las series históricas de parados (y por lo tanto de activos e inactivos) del INE presentan una discontinuidad en 2001 debida a la introducción en ese año de una definición operativa de paro más exigente que la anterior. La nueva definición endurece las condiciones necesarias para considerar que una persona está buscando trabajo activamente cuando su búsqueda se produce a través de una oficina pública de empleo. Con el cambio de criterio, la mera inscripción en tal oficina deja de considerarse un método activo de búsqueda de trabajo, exigiéndose además que haya existido un contacto con la misma durante las cuatro semanas previas a la de referencia.

El INE ha construido una serie que salva la discontinuidad generada por este cambio metodológico por interpolación, de forma que el efecto del cambio se reduce linealmente según vamos hacia atrás en el tiempo hasta desaparecer en 1976.³⁷ Dado que el cambio de criterio afecta sólo a aquellos parados que buscan trabajo a través de una oficina pública de empleo, parece más lógico utilizar el tamaño de este colectivo para construir el enlace. Aunque la EPA no nos dice cuántos parados *sólo* buscaron trabajo activamente a través de una oficina de empleo público, sí contiene información sobre el número de parados que se han inscrito en tales oficinas. Dividiendo esta cifra por el número total de parados se obtiene el indicador (*wPOPE*) cuyo perfil temporal he utilizado para construir por interpolación la serie corregida de parados que se muestra en el Gráfico 21.^{38,39} El logaritmo de la serie corregida viene dado por $\hat{y}_t^n = x_t + \hat{d}_t^n$ con

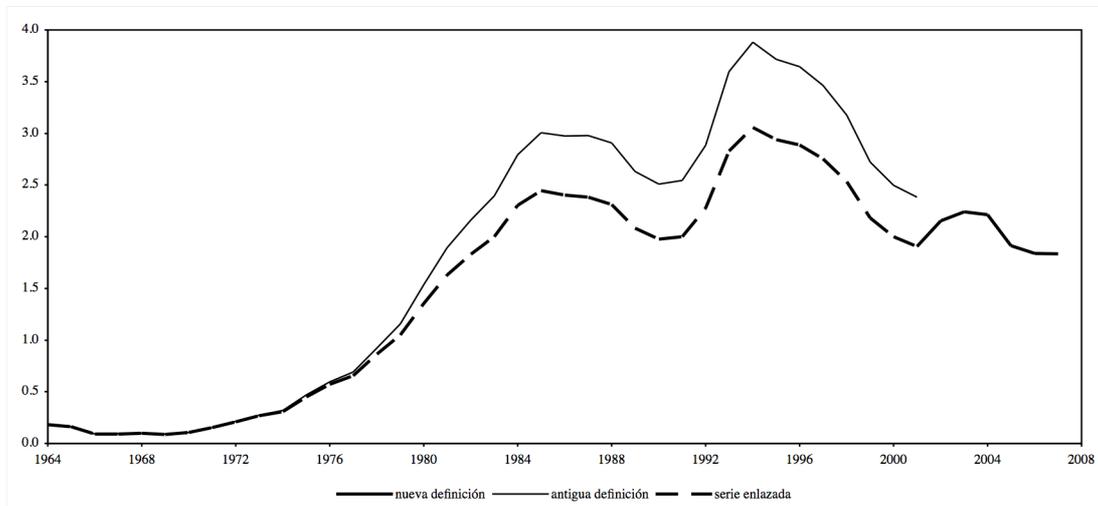
³⁷ El procedimiento de enlace se describe en detalle en Trejo y Ortega (2005) y se basa en la estimación de un modelo *probit* en el que la variable a explicar es la probabilidad de que un parado con la definición antigua siga siéndolo con la nueva definición. Las variables explicativas son el sexo, el tramo de edad, la interacción entre ambos y la comunidad autónoma de residencia. El modelo se estima con datos individuales de los años 2001 a 2004, para los que el cuestionario utilizado permite calcular el número de parados de acuerdo con ambas definiciones. Para extender la serie hacia atrás, se supone que la probabilidad estimada es 1 (esto es, que no hay cambios al aplicar la nueva definición) en 1976T2 y se interpola linealmente desde entonces hasta 2000T4, asignando a esta última fecha la probabilidad estimada con el *probit*.

³⁸ Para años anteriores a 1987 se usan datos tomados de las publicaciones de la EPA en papel. La serie de *wPOPE* se interpola para cubrir ciertos huecos en la misma durante los primeros años setenta. Véase la sección 4 del Anexo 1. Por otra parte, en 2001T1 se introducen ciertos cambios en el cuestionario ligados al cambio en la definición operativa de parado que generan una discontinuidad apreciable en la variable *wPOPE*. Para eliminar esta discontinuidad, he depurado la serie trimestral de *wPOPE* durante 2001 de los efectos del cambio por el procedimiento habitual de prolongar la serie precedente utilizando las tasas de variación interanual que se obtienen por interpolación entre observaciones no contaminadas a ambos lados del cambio. La media de las observaciones trimestrales de 2001 así depuradas nos da el valor de *wPOPE* en dicho año que se utiliza en la ecuación (5).

³⁹ En Banco de España (2002) se construye una serie enlazada de desempleo de acuerdo con la nueva definición desde 1987 hasta 2002 por un procedimiento similar al que aquí se utiliza. Aparentemente, en la construcción de esta serie se utiliza información sobre el número de parados que *sólo* buscan trabajo a través de una oficina de empleo. De acuerdo con el INE, esta variable no está disponible (aunque podría construirse a partir de una tabulación de los cuestionarios originales, que es quizás lo que ha hecho el Banco de España), o al menos no lo está para el período más largo que se considera en el presente trabajo.

$$(5) \hat{d}_t^n = d_{2001} * \frac{wPOPE_t}{wPOPE_{2001}} \text{ para } 1964 \leq t \leq 2000$$

Gráfico 21: Parados con la antigua y nueva definición y serie enlazada millones de personas

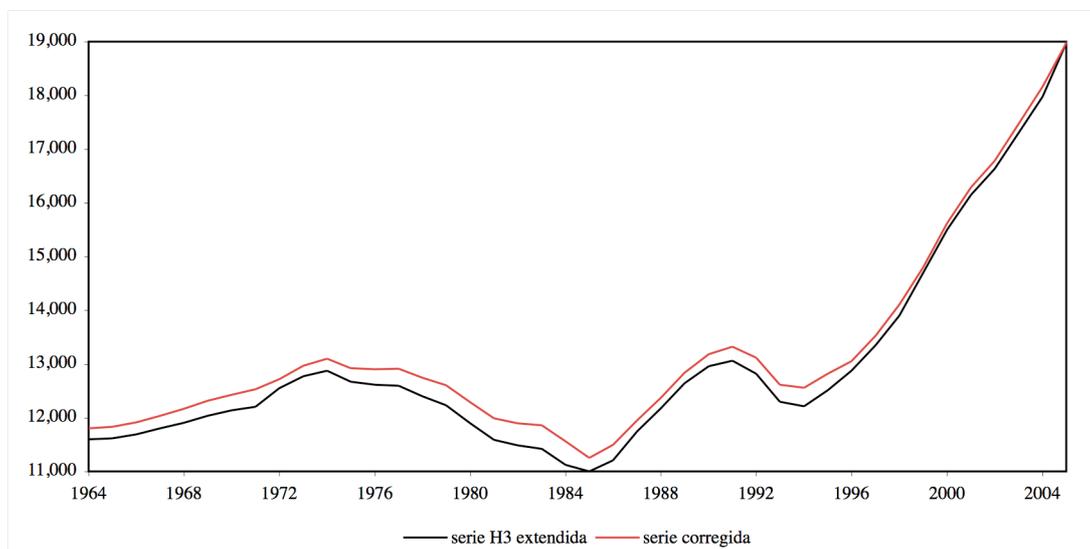


La serie enlazada

El panel *a* del Gráfico 22 compara la serie corregida de ocupados construida en esta sección con la serie H3 extendida descrita en la sección 3. Las diversas correcciones que he introducido elevan el número estimado de ocupados en varios cientos de miles durante la mayor parte del período muestral, alcanzando un máximo de casi 450.000 efectivos adicionales a mediados de los años ochenta, pero no cambian significativamente el perfil de la serie. Cuando se comparan las tasas de crecimiento de ambas series (véase el panel *b* del Gráfico 22), las diferencias son mínimas excepto en los años correspondientes a algunas de las incidencias analizadas, en los que la serie corregida presenta un perfil más suave.

Gráfico 22: Serie corregida de ocupados vs. serie H3 extendida

a. niveles



b. tasas de crecimiento interanual

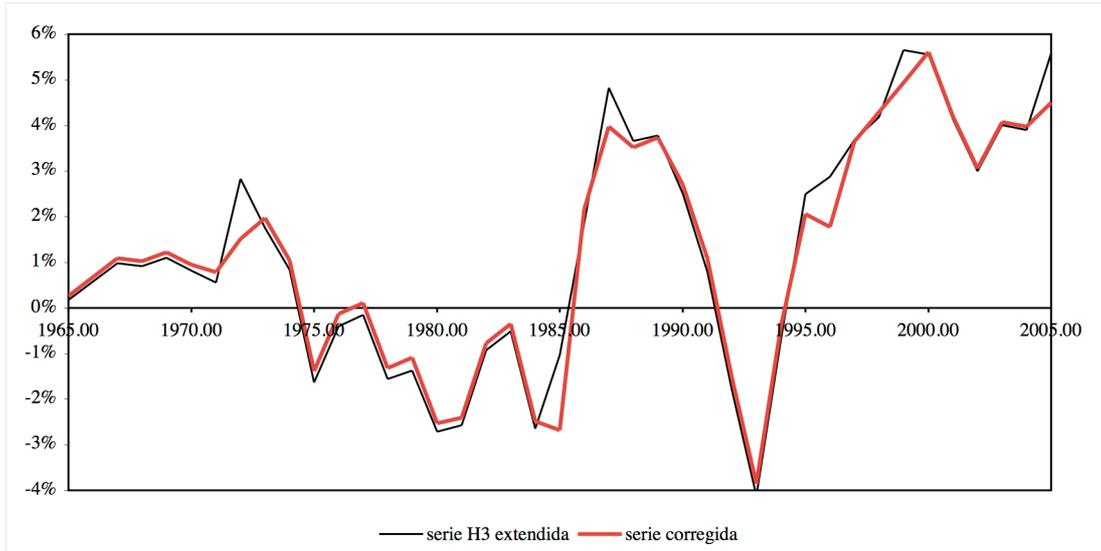
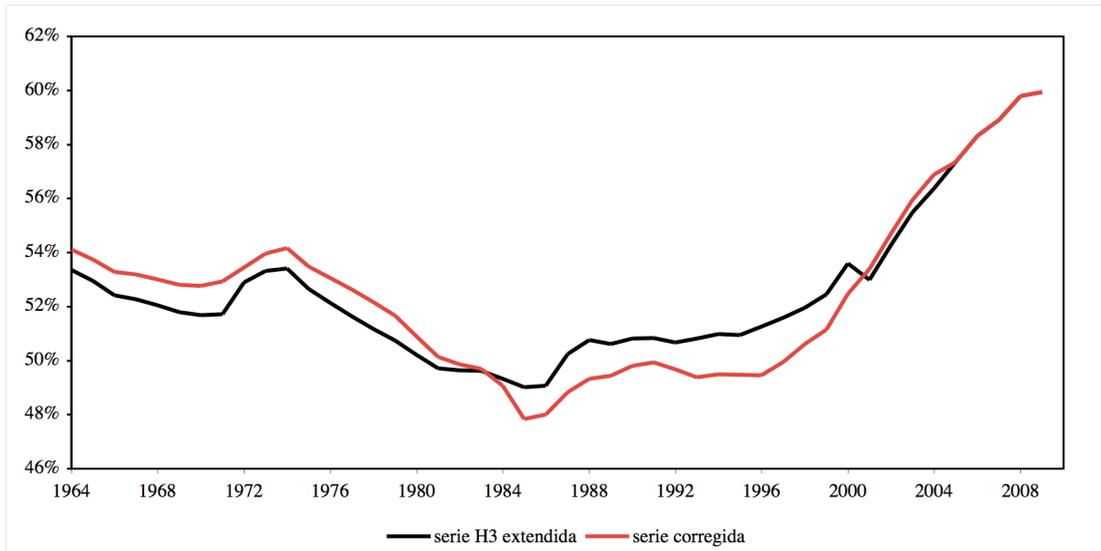
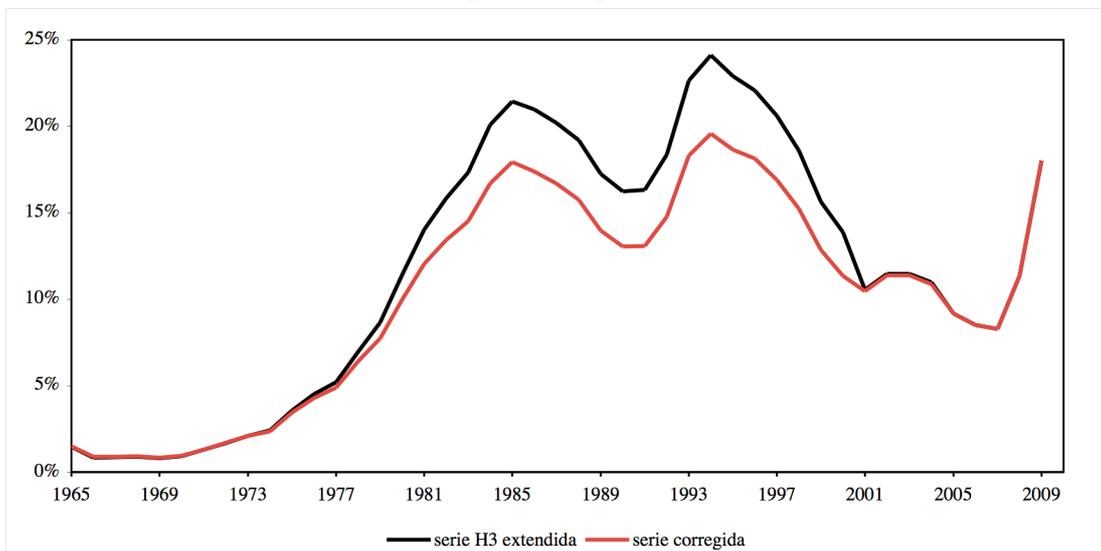


Gráfico 23: Serie corregida de ocupados vs. serie H3 extendida
a. Tasa de actividad de la población 16+



b. Tasa de paro de la población 16+



El Gráfico 23 muestra la evolución de la tasa de actividad y de la tasa de desempleo con ambas series durante el período muestral. La homogeneización de la definición de desempleo tiene un efecto muy significativo sobre la tasa de paro estimada en los años centrales de la muestra. El efecto de esta corrección también se aprecia en la senda de la tasa de actividad en estos años, que desciende por pasar a considerarse como inactivos varios cientos de miles de personas que en la EPA original se contabilizaban como parados y por lo tanto activos.

7. Puestos de trabajo y horas trabajadas

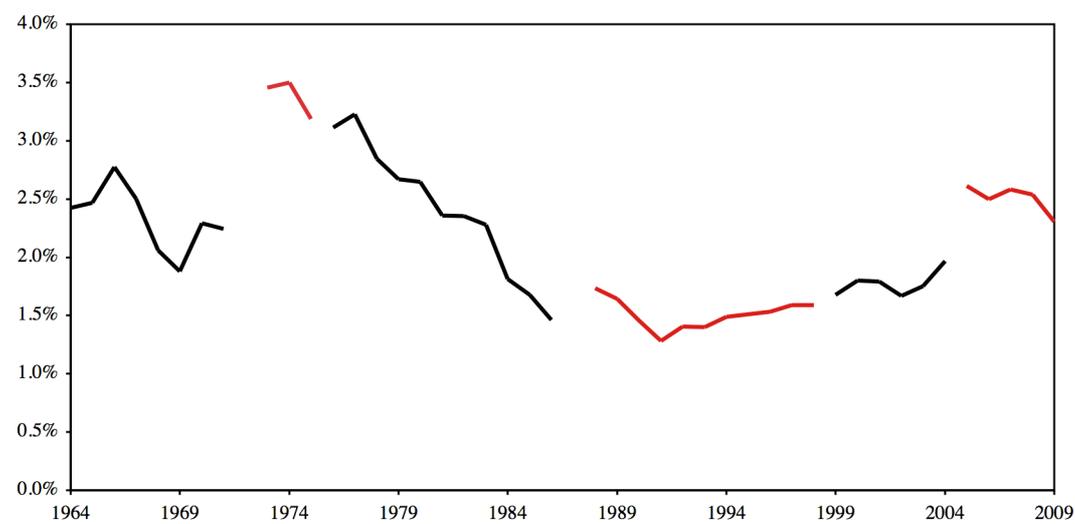
La página web del INE también proporciona información sobre algunas variables que pueden resultar útiles para construir series de puestos de trabajo y horas trabajadas. Se trata del número de ocupados que declaran un segundo empleo y del número medio de horas semanales trabajadas por los ocupados. La información disponible sobre estos agregados es, sin embargo, bastante más limitada que la referida a las variables analizadas en las secciones anteriores. A nivel nacional, la página web del INE ofrece una serie de la EPA 2002 (consistente con H2) que comienza en el segundo trimestre de 1987 y otra de la EPA 2005 (consistente con H3) que se desvía de la anterior de 1996 en adelante. Ambas presentan posibles rupturas en los primeros trimestres de 1999 y 2005 debido a los cambios en el cuestionario introducidos en esas fechas. Recurriendo a las publicaciones en papel de la EPA (INE, varios años), las dos series pueden extenderse hacia atrás hasta 1964, aunque con ciertas lagunas (véase la sección 4 del Anexo 1) y posibles rupturas en 1972-S2, 1973-T3 y 1987-T2. Sin embargo, ninguna de estas series está disponible a nivel regional en la página web del INE. En el resto de esta sección se construyen series homogéneas de duración media de la jornada laboral y del porcentaje de los ocupados que declaran un segundo empleo mediante el enlace de los distintos segmentos de las series originales.

7.1. Ocupados que declaran una segunda actividad

El Gráfico 24 muestra los distintos segmentos de la serie anual del porcentaje de ocupados que declara un segundo empleo (*%SEGEMP*) de acuerdo con la web del INE y con las publicaciones de la EPA en papel.⁴⁰ Como ya se ha visto, durante el primer tramo de la encuesta existen lagunas en ciertos trimestres o semestres en los que la EPA no se realizó y la variable de interés presenta oscilaciones importantes y no siempre plausibles. En períodos posteriores, la serie presenta un perfil generalmente más suave pero también claros indicios de rupturas en tres años en los que se produce un cambio de cuestionario (1972, 1987 y 2005).

⁴⁰ La serie anual se obtiene promediando las observaciones trimestrales o semestrales correspondientes. Antes de hacerlo, se rellenan por interpolación lineal las lagunas que presenta la serie original durante los años 1966 a 1971 (pero no en 1972 pues en este caso no se dispone de observaciones no contaminadas por posibles cambios metodológicos a ambos lados de la observación que falta en 1972-S1). En 1964, se descarta la observación de 1964-T2 por ser casi dos veces superior a las que le siguen. La observación anual de ese año se construye aplicando a la correspondiente a 1965 la razón entre la media de los trimestres III y IV en 1964 y 1965.

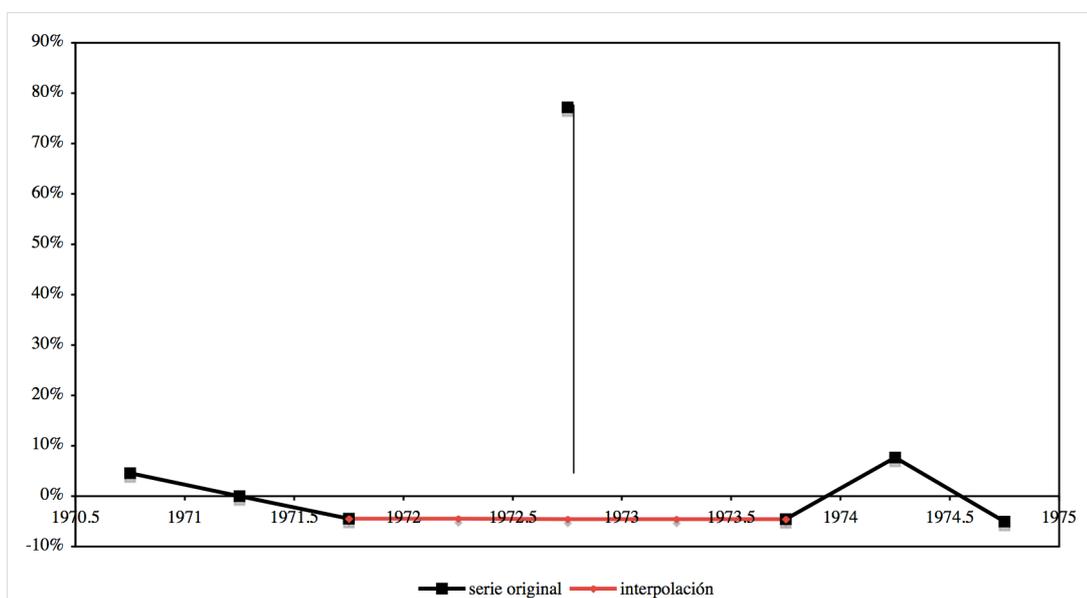
Gráfico 24: % de los ocupados que declara un segundo empleo (%SEGEMP), series anuales



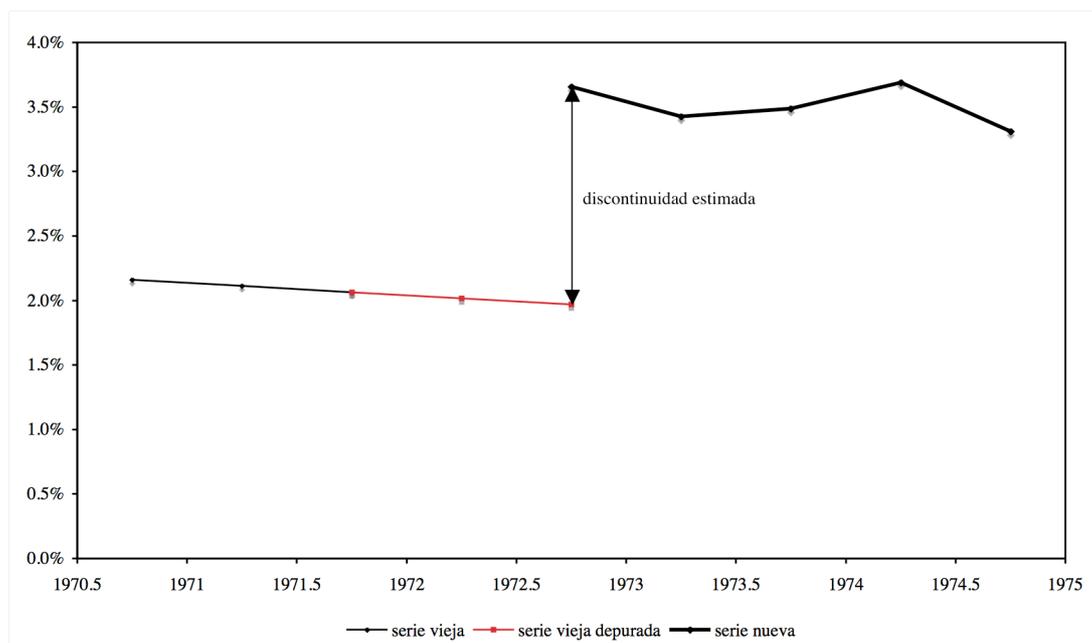
El primer paso para construir una serie anual enlazada consiste en estimar el tamaño de las sucesivas rupturas de la misma, lo que haré procediendo desde atrás hacia delante. La primera de las discontinuidades en la serie es la debida a la introducción de un nuevo cuestionario en 1972-S2. Para estimar la evolución de %SEGEMP en ausencia del cambio metodológico, he tomado como referencia la tasa de crecimiento interanual que se obtiene interpolando entre las tasas correspondientes al primer y al último semestres disponibles no afectados por el cambio, que son 1971-S2 y 1973-S2. Las tasas de crecimiento así estimadas para 1972-S1 (cuando no se realizó la EPA) y 1972-S2 (cuando se introduce el nuevo cuestionario) se utilizan para estimar el nivel de %SEGEMP que se habría observado en estos dos semestres, en el segundo caso en ausencia del cambio metodológico y, seguidamente, el tamaño del salto generado por tal cambio, tal como se ilustra en el Gráfico 25.

Gráfico 25: Estimación de la discontinuidad de la serie de %SEGEMP en 1972-S2

a. Tasas de variación interanual



b. Series en niveles



Cuadro 7: magnitud estimada de las discontinuidades en la serie del porcentaje de ocupados que declara un segundo empleo

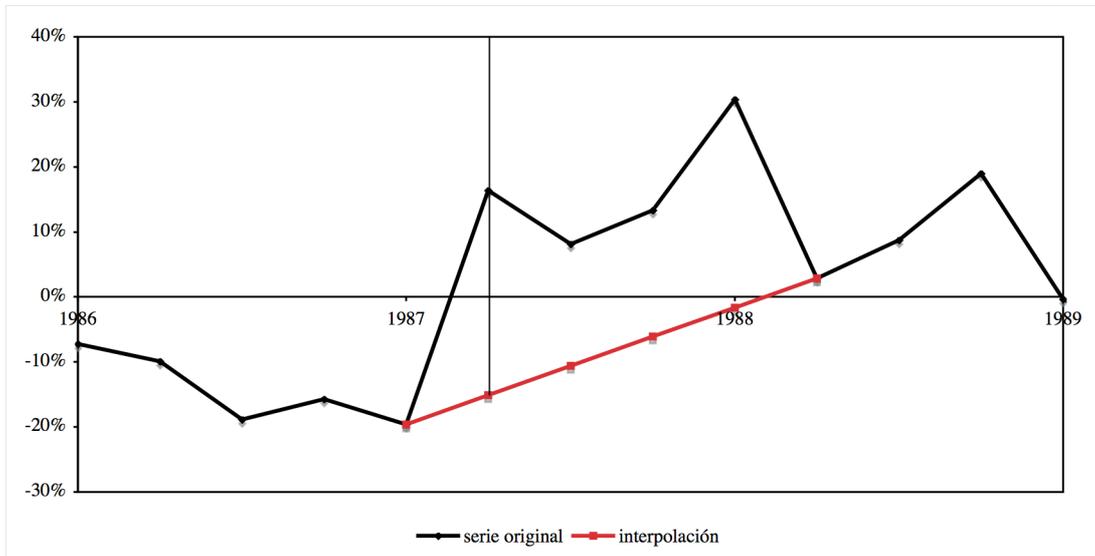
	<i>magnitud absoluta: puntos porcentuales</i>	<i>magnitud relativa: diferencia logarítmica</i>
1972-S2	1.69 pp	61.8%
1987-T2	0.48 pp	31.6%
2005-T1	0.65 pp	27.5%

El Cuadro 7 muestra el tamaño estimado de las tres rupturas de la serie, tanto en términos absolutos (puntos porcentuales de diferencia entre la serie depurada y sin depurar en cada punto de ruptura) como relativos (diferencia logarítmica entre ellas). La estimación se realiza utilizando el mismo procedimiento que acabo de exponer para el caso de 1972. El Gráfico 26 muestra el comportamiento de las tasas interanuales de variación de %SEGEMP en el entorno de los puntos de ruptura. En ambos años, la serie de variaciones interanuales presenta la joroba típica de una discontinuidad al alza de la correspondiente serie en niveles. En cada caso, he utilizado el primer punto de la serie interpolada de tasas de crecimiento que se aparece en rojo en el gráfico como referencia para estimar la evolución que habría tenido la serie en niveles en ausencia del cambio metodológico. Comparando esta hipotética observación depurada en el punto de ruptura con la que ya recoge los efectos del cambio metodológico, se obtiene una estimación del tamaño de la discontinuidad en la serie.

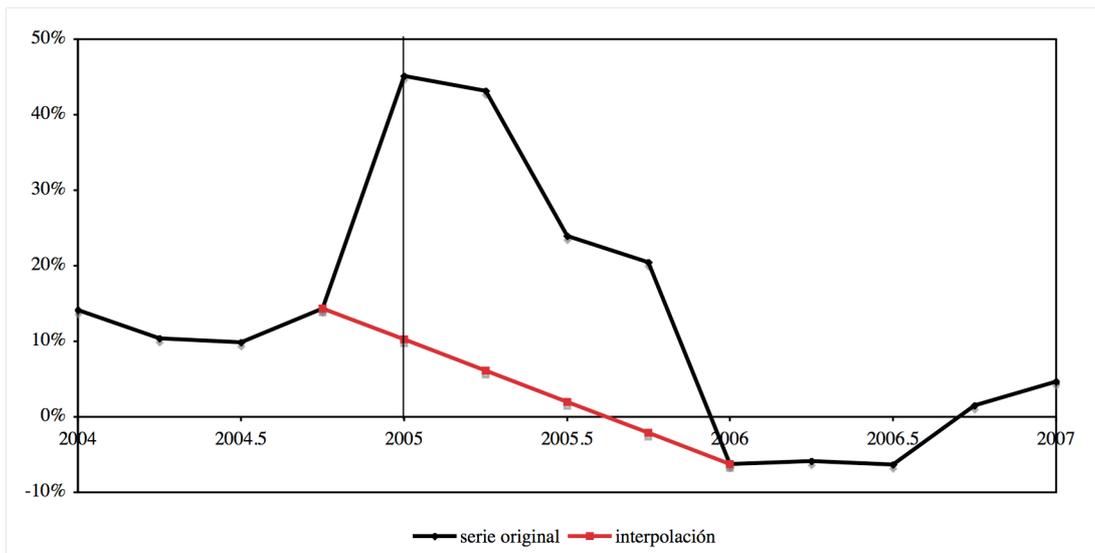
Finalmente, se construye una serie corregida por el procedimiento de ir enlazando, desde atrás hacia delante y de forma iterativa, los sucesivos segmentos de la serie por el procedimiento mixto utilizando una vida media de 10 años. Esto es, en primer lugar, la serie 1964-1972S1 se enlaza con la serie 1972S2-1987T1 para obtener una primera serie corregida 1964-1987T1. Esta serie se enlaza seguidamente con la serie 1987T2-2004T4 para obtener una segunda serie corregida 1964-2004T4. Finalmente, esta serie se enlaza con la serie 2005T1-2010

**Gráfico 26: Porcentaje de ocupados que declara un segundo empleo
tasas interanuales de variación trimestre a trimestre**

a. cerca de 1987-T2



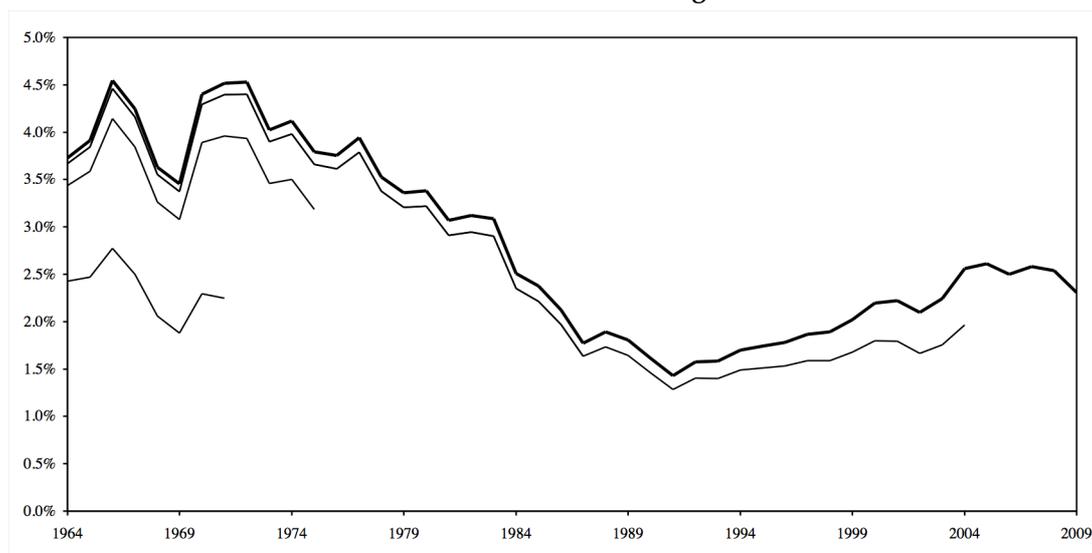
b. cerca de 2005-T1



para obtener la serie final corregida 1964-2010. Como en la sección anterior, los cálculos necesarios se realizan directamente con las series anuales excepto en aquellos casos en los que la ruptura no se produce en el primer trimestre o semestre del año, en cuyo caso se utilizan también las series trimestrales durante el año de la ruptura. El Gráfico 27 ilustra el procedimiento y muestra la serie final corregida, que aparece en trazo grueso.

La revisión eleva muy significativamente el porcentaje estimado de ocupados que declara un segundo empleo durante los primeros años de la muestra. La serie revisada muestra un perfil temporal decreciente desde mediados o finales de los años 60 hasta finales de los ochenta, cuando se produce una suave recuperación que se interrumpe con el inicio de la crisis actual.

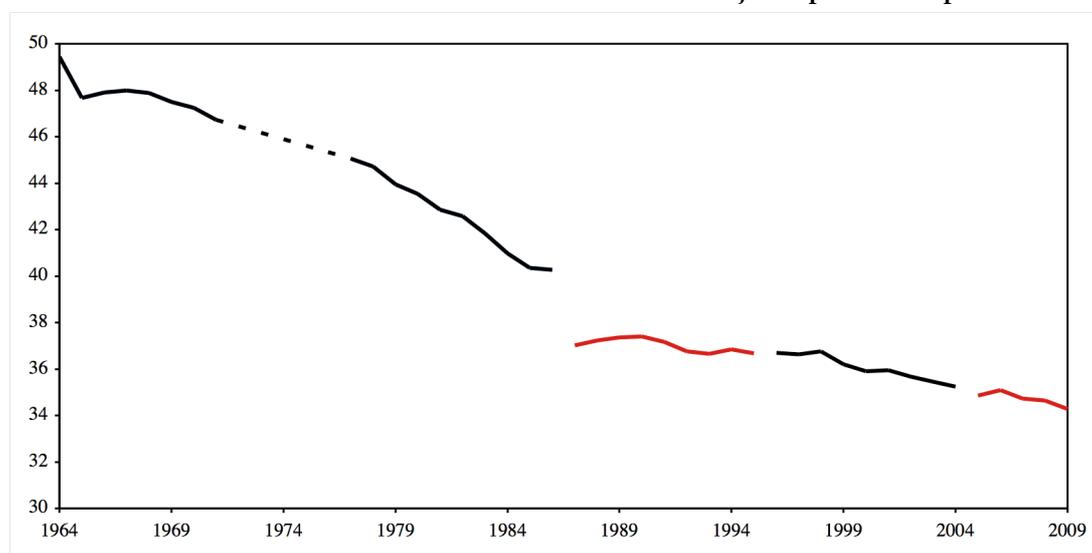
Gráfico 27: Construcción de la serie corregida de %SEGEMP



7.2. Duración media de la jornada laboral

El Gráfico 28 resume la información que proporciona la EPA sobre el número medio de horas semanales trabajadas por todos los ocupados, hayan o no trabajado en la semana de referencia, (*HMSEM*) dibujada con frecuencia anual.⁴¹

Gráfico 28: Número medio de horas semanales trabajadas por los ocupados

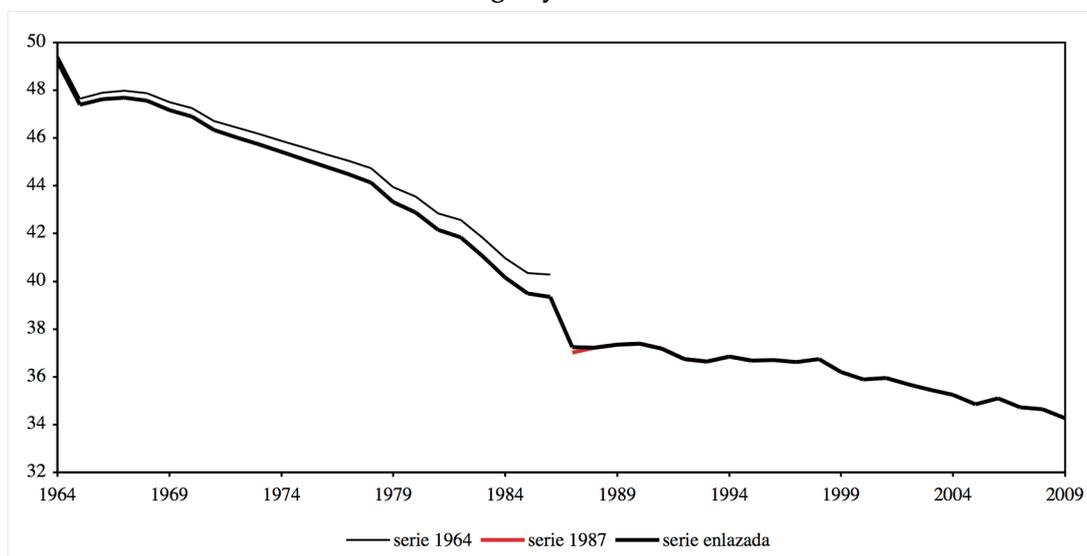


La serie de la EPA en papel presenta una laguna entre 1972 y 1976 que he rellenado por interpolación lineal dado que a primera vista no parece probable que los cambios metodológicos que se producen en dichos años hayan generado una ruptura significativa de la serie. Lo mismo es cierto en 1996 y 2005 pero no en 1987, cuando el cambio de cuestionario sí parece haber generado una apreciable ruptura a la baja. Para corregir esa ruptura, he

⁴¹ Sobre la evolución de la duración de la jornada media de trabajo véase entre otros Pérez Infante y Fina (1999).

enlazando la serie por el procedimiento habitual utilizando una vida media de 10 años. La serie corregida se muestra junto con la original en el Gráfico 29 (y sigue mostrando un apreciable escalón a la baja en 1987 porque el tercer trimestre de ese año muestra un número de horas atípicamente reducido pero que ya no puede ser atribuido al cambio metodológico).

Gráfico 29: Número medio de horas semanales trabajadas por los ocupados
Serie antigua y serie enlazada



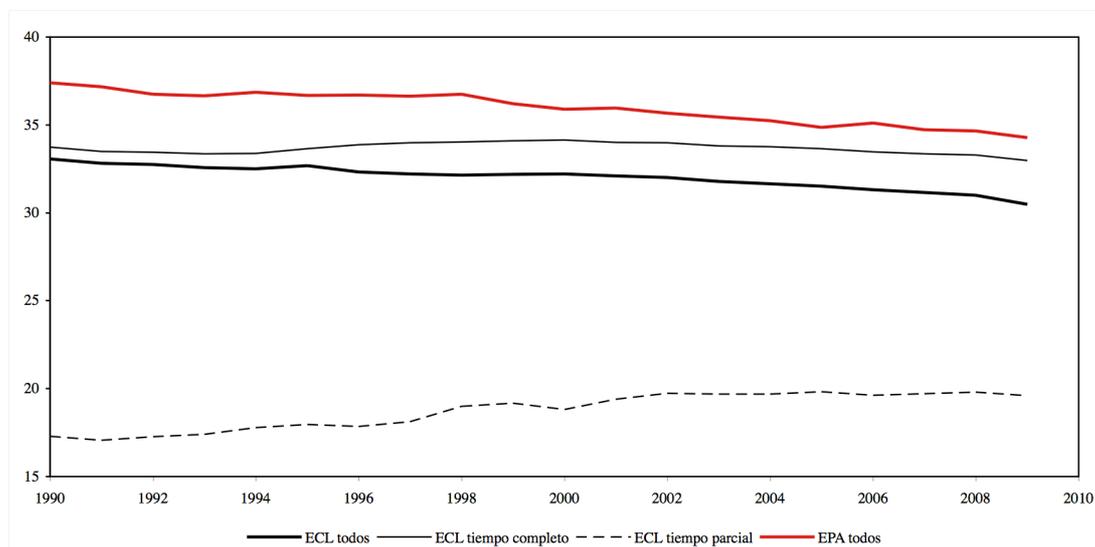
Una fuente adicional de datos sobre la duración media de la jornada laboral es la Encuesta de Coyuntura Laboral (ECL). La ECL (MESS, 2012b) es una encuesta trimestral dirigida a las empresas en la que se recoge información sobre sus efectivos laborales y vacantes, la movilidad del personal, su tiempo de trabajo y otros aspectos del mercado laboral. La encuesta se implantó en el segundo trimestre de 1990 y sufrió importantes cambios metodológicos en el primer trimestre de 1997 cuando, entre otras cosas, comenzaron a investigarse los establecimientos de menos de 6 trabajadores, que habían estado excluidos de la encuesta hasta el momento. La ECL contabiliza los puestos de trabajo correspondientes a asalariados afiliados al régimen general de la Seguridad Social y al régimen especial de la minería del carbón que trabajan en la industria, construcción o servicios, excluyendo de este último sector la administración pública, la defensa y las organizaciones religiosas. Su muestra no incluye a Ceuta y Melilla.

La variable de la ECL que nos interesa es la jornada efectiva media por trabajador, incluyendo las horas extra realizadas y excluyendo las horas pactadas pero no realizadas por cualquier motivo (incluyendo bajas por enfermedad o maternidad y huelgas). Esta magnitud se calcula para el conjunto de los trabajadores asalariados así como separadamente para los trabajadores a jornada completa y a jornada parcial.

El Gráfico 30 muestra las distintas series de la ECL junto con la serie de jornada media de la EPA. La duración media de la jornada laboral de los ocupados muestra un perfil temporal muy similar en ambas fuentes, si bien el nivel de la serie de la ECL es más bajo – lo que parece lógico dado que en realidad esta fuente viene a medir las horas medias trabajadas por puesto de trabajo mientras que la EPA nos da las horas medias por ocupado, incluyendo

aquellos ocupados que tienen más de un puesto de trabajo. Por lo tanto, si necesitamos una serie larga de horas medias por puesto de trabajo, no parece excesivamente arriesgado extender hacia atrás la serie de la ECL con la tasa de crecimiento de la serie de la EPA.

**Gráfico 30: Número medio de horas semanales trabajadas por los ocupados
ECL vs EPA**



Por otra parte, las jornadas medias de los trabajadores a tiempo completo y a tiempo parcial muestran perfiles diferentes. Mientras que la primera de estas variables se mantiene aproximadamente constante, la segunda muestra una clara aunque suave tendencia al alza.

8. Archivos de datos

El trabajo va acompañado de dos archivos de Excel. El primero (*NacDat_EPA_v10_64_09.xls*) contiene las series descritas y/o construidas en este trabajo, que constituyen la versión 1.0 de *NacDat-EPA*. El segundo archivo (*Dat_orig_nac_EPA_10.xls*) recoge el resto de los datos de otras fuentes que se han utilizado en el presente trabajo. Los Cuadros 8 y 9 describen el contenido de los archivos citados. Se incluye también un tercer archivo (*C&P_provincias.xls*) que contiene la población total de cada provincia y la parte de la misma que reside en la capital y en el segundo municipio de la misma de acuerdo con los distintos censos y padrones elaborados desde 1960 hasta 2001.

Cuadro 8: Contenido del archivo de datos *NacDat_EPA_v10_64_09.xls*

hoja 1: Extensión de las series históricas del INE hasta 1964

1. Primera serie histórica o serie H1 extendida, 1964-2001.

- *Variables disponibles:* población de 16 años de edad o más, desglosada en activos, ocupados, parados, inactivos y población contada aparte. Población de 14 y 15 años desglosada en ocupados, parados e inactivos para 1964-1980.

2. Segunda serie histórica del INE o serie H2 extendida, 1964-2004

- *Variables disponibles:* población de 16 años de edad o más, desglosada en activos, ocupados, parados, inactivos y población contada aparte.

3. Tercera serie histórica del INE o serie H3 extendida, 1964-2009

- *Variables disponibles:* población de 16 años de edad o más, desglosada en activos, ocupados, parados, inactivos y población contada aparte.

hoja 2: Serie final enlazada, 1964-2009

1. Serie final enlazada obtenida a partir de la serie H3 del INE

- *Variables disponibles:* población de 16 años de edad o más, desglosada en activos, ocupados, parados, inactivos y población contada aparte.

2. Correcciones a la serie H3 del INE

3. Otras series de interés

- Porcentaje de los ocupados que declara un segundo empleo.
- Horas medias semanales trabajadas por todos los ocupados.

Cuadro 9: Contenido del archivo de datos *Dat_orig_nac_EPA_10.xls*

hoja 1: Series históricas del INE. Datos trimestrales

1. Primera serie histórica o serie H1, 1976TIII a 2001TIV

- *Variables disponibles:* población de 16 años de edad o más, desglosada en activos, ocupados, parados, inactivos y población contada aparte.

- *Fuente:* INE. Encuesta de Población Activa, Metodología EPA 2002. Novedades metodológicas. Repercusión de los cambios.

<http://www.ine.es/epa02/repercusion.htm>

- *Nota:* A la página de la EPA 2002 se puede acceder desde la página dedicada a la EPA 2005 (último apartado: otra información disponible en INEbase).

2. Segunda serie histórica del INE o serie H2

- *Variables disponibles:* población de 16 años de edad o más, desglosada en activos, ocupados, parados, inactivos y población contada aparte.

- *Fuente:* INE. Encuesta de Población Activa, Metodología EPA 2002. Resultados

<http://www.ine.es/jaxi/menu.do?type=pcaxis&path=/t22/e308/&file=inebase>

Cuadro 9 – continuación

3. Tercera serie histórica del INE o serie H3

- *Variables disponibles:* población de 16 años de edad o más, desglosada en activos, ocupados, parados, inactivos, población contada aparte, ocupados que declaran un segundo empleo y horas medias semanales trabajadas por ocupado (por todos los ocupados, hayan o no trabajado).

- *Fuente:* INE. EPA 2005, Encuesta de Población Activa, Metodología 2005. Resultados http://www.ine.es/jaxi/menu.do?type=pcaxis&path=/t22/e308_mnu&file=inebase&N=&L=0

hoja 2: Series históricas del INE. Datos anuales.

- Mismo contenido que la hoja 1 pero con frecuencia anual (promedios de las observaciones trimestrales de cada año).

hoja 3: Series del Grupo de Trabajo sobre problemas de empleo del Ministerio de Economía (GTE, 1979)

- *Variables disponibles:* población de 14 años de edad o más, desglosada en activos, ocupados, parados, inactivos y población que presta el servicio militar.

- *Fuente:* Grupo de trabajo sobre problemas de empleo (GTE, 1979). Población, actividad y ocupación en España (reconstrucción de las series históricas 1960-78). Madrid, Ministerio de Economía.

hoja 4: Series de García Perea y Gómez (1994)

- *Variables disponibles:* activos, ocupados y parados.

- *Fuente:* García-Perea, P. y R. Gómez (1994). "Elaboración de series históricas de empleo a partir de la Encuesta de Población Activa (1964-1992)." Documento de Trabajo, núm. 9409. Banco de España, Madrid.

hoja 5: Series de Mas et al (2002)

- *Variables disponibles:* población de 14 años o más, desglosada en activos, ocupados y parados.

- *Fuente:* Mas, M., F. Pérez, E. Uriel, L. Serrano y A. Soler (2002). Capital humano, series 1964-2001. Bancaja.

hoja 6: varios

- Población por edades del censo de 1960.

- Estudiantes de secundaria en ciertos años, tomados del Anuario Estadístico de España.

- % de hogares que se niega a colaborar con la EPA. Fuente: EPA, INE.

- distancia estimada entre la población de cada año y la población en el año del censo que se utiliza para fijar el seccionado.

- Series anuales de población 16+, activos y ocupados, 1950-78 de Alcaide (2000b).

hoja 7: otras series trimestrales

- Afiliaciones a la Seguridad Social: trabajadores afiliados en situación de alta laboral en el último día del mes, promedio para cada trimestre.

- *Fuente:* Ministerio de Empleo y Seguridad Social (MESS), <http://www.empleo.gob.es/series/>

- Contabilidad trimestral, bases 1995 y 2000

- variables disponibles: ocupados, puestos de trabajo, puestos de trabajo equivalentes a tiempo completo y horas trabajadas

- *Fuente:* INE

hoja 8: Encuesta de Coyuntura Laboral

- Jornada media efectiva por trabajador (horas anuales), total y por tipo de jornada.

- *Fuente:* MESS, <http://www.empleo.gob.es/series/>

Cuadro 9 – continuación

hoja 9: Datos de la EPA en papel, 1964-1987

- Con frecuencia trimestral o semestral: ocupados, ocupados que declaran una segunda actividad, número medio de horas trabajadas en la semana por todos los ocupados, parados y parados que utilizan oficinas públicas de colocación.

- Con frecuencia anual: % de ocupados que declaran una segunda actividad, número medio de horas trabajadas en la semana por todos los ocupados y % de parados que utilizan oficinas públicas de empleo. (véase la sección 4 del Anexo 1).

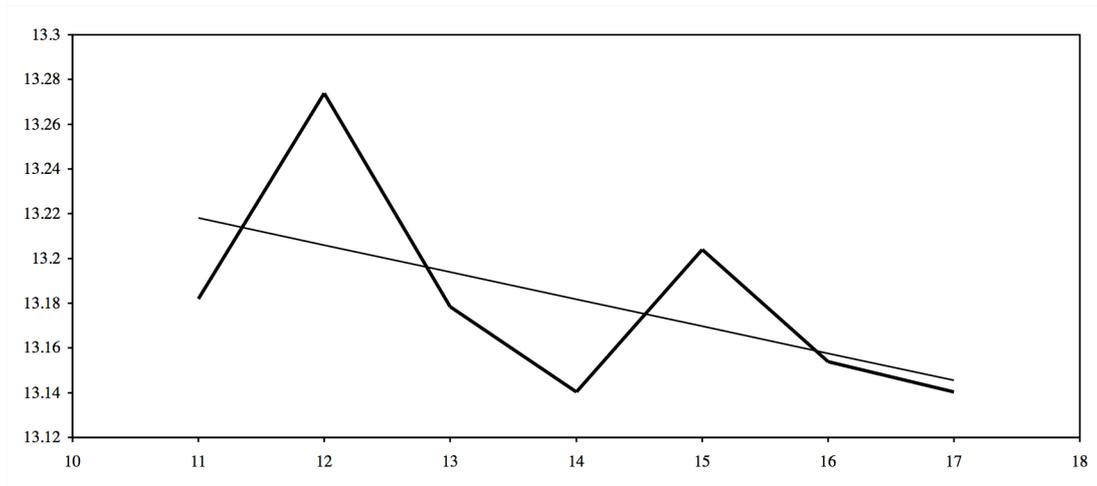
- Fuente: INE.

ANEXO 1

1. La distribución por edades de la población en el censo de 1960

El Gráfico A.1 ilustra el fenómeno ya comentado en el texto de la “atracción” de los censados hacia ciertas “edades redondas.” Mientras que uno esperaría que el número de efectivos variase de forma monótona y “suave” con la edad, el gráfico revela la existencia de altibajos significativos que sugieren una sobrevaloración de la población de 12 y 15 años de edad, acompañada de la infravaloración del número de personas con edades cercanas a las citadas. En el gráfico se muestra también el resultado del ajuste de una relación lineal entre el logaritmo del tamaño de la población con una edad determinada y la propia edad. Si tomamos la línea ajustada como una estimación del número real de personas de cada edad, vemos que las desviaciones con relación a la recta de regresión que se producen a los 14 y 15 años de edad son aproximadamente del mismo tamaño, lo que sugiere que, el peso conjunto de la población de 14 y 15 años en el total que se obtiene del censo sería aproximadamente correcto.

Gráfico A.1: logaritmo de la población nacional por edades simples vs. edad, censo de 1960



2. Estimación de la tasa de escolarización de la población de 14 a 17 años en 1964-77

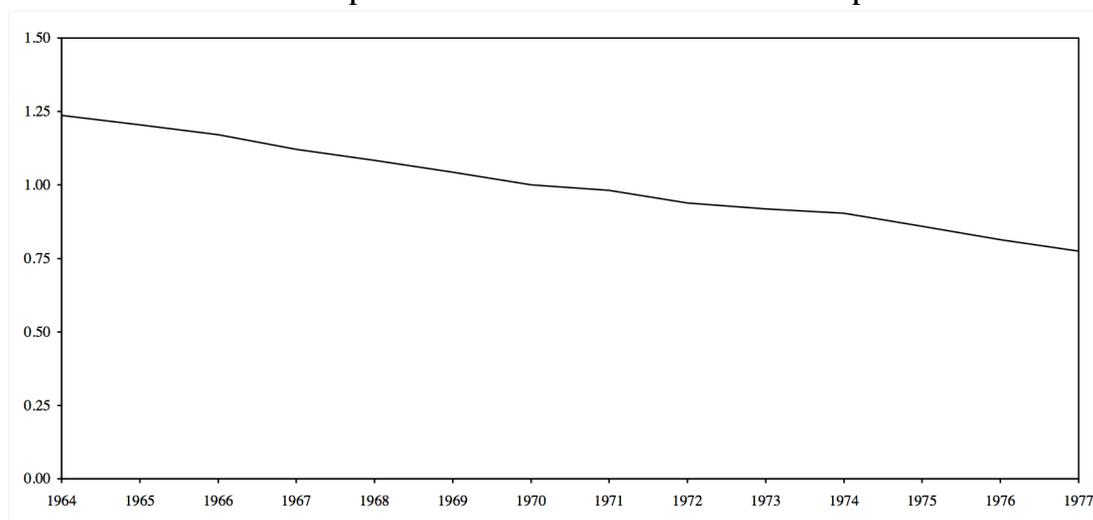
La serie de estudiantes de secundaria tomada del *Anuario Estadístico de España* presenta una discontinuidad durante estos años como resultado de una reforma educativa. En el plan antiguo, el bachillerato consta de seis años (cuatro de bachillerato elemental y dos de bachillerato superior) a lo que hay que añadir el curso preuniversitario (PREU). Tras la reforma de Villar Palasí en 1970, el bachillerato elemental desaparece, integrándose en la EGB, y el bachillerato superior se alarga un año, pasando a llamarse BUP, mientras que el curso preuniversitario pasa a llamarse curso de orientación universitaria o COU. Puesto que el Anuario no desglosa los estudiantes de bachillerato por años o incluso por ciclos, no resulta posible identificar con precisión a los alumnos que tendrían que tener catorce o quince años

en condiciones normales y resulta necesario construir una aproximación necesariamente burda pero seguramente suficiente para nuestros propósitos.

La serie de estudiantes de bachillerato bajo el plan antiguo (incluyendo PREU) se extiende hasta el curso 1973-74. Durante los últimos tres años de esta serie, sin embargo, el número de estudiantes de bachillerato se reduce debido a que sólo quedan en el plan antiguo los estudiantes de los cursos superiores (puesto que los más jóvenes ya comienzan a cursar el segundo ciclo de EGB en vez de el bachillerato elemental). Para corregir este efecto, la cifra de estudiantes de bachillerato y FP que ofrece el Anuario en estos años se multiplica, por $7/6$, $7/5$ y $7/4$ respectivamente.

Comenzando en 1974-75 disponemos de una serie de estudiantes de BUP (y COU) que se adecua mejor al segmento de población cuya tasa de escolarización queremos aproximar. Esta serie se extiende hacia atrás tomando como referencia el perfil temporal de la serie ajustada de estudiantes de secundaria con el plan antiguo. Para pasar del primer curso del plan nuevo al último del plan antiguo (1973-74), he supuesto que la tasa de crecimiento del número de estudiantes entre estos dos años es la misma que observamos entre 1974-75 y el curso siguiente. Para seguir hacia atrás desde 1973-74 se utiliza la tasa de crecimiento de la serie de estudiantes del plan antiguo. Seguidamente se suman los estudiantes de FP a los de bachillerato y COU y el número medio de estudiantes en cada año natural se aproxima como una media ponderada de los matriculados durante el curso que termina en junio del año relevante (con un peso de $8/12$) y en el que comienza en septiembre (con un peso de $4/12$).

Gráfico A.2: Perfil temporal de la fracción no escolarizada de la población 14-17



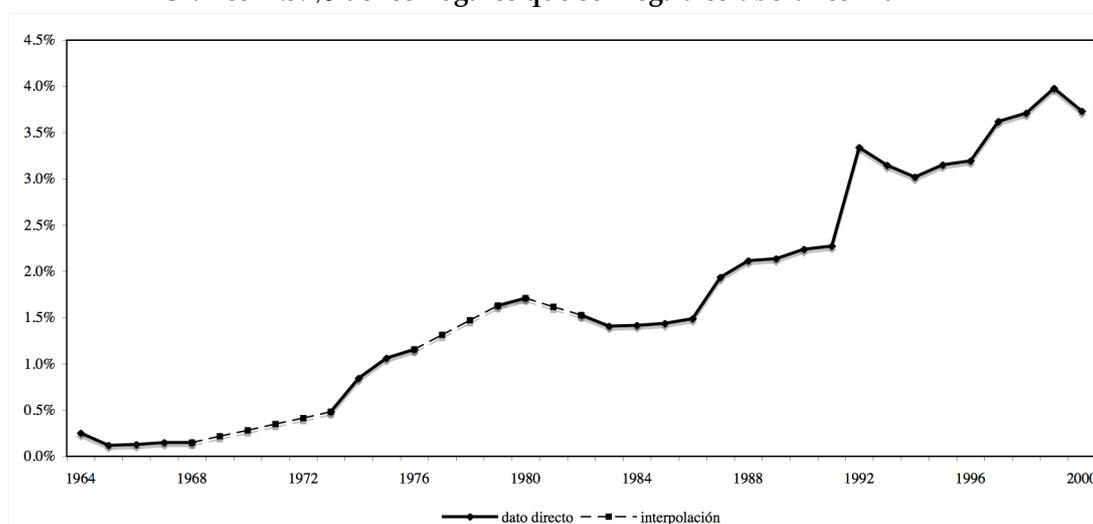
Por otra parte, he estimado la población total de 14 a 17 años. De 1971 en adelante el dato se toma directamente de la estimaciones intercensales de población del INE. Para 1964-71 se procede como en el caso de la estimación de la población de 14 y 15 años, esto es, interpolando linealmente entre el censo de 1960 y el primer dato anual del INE en 1971. Dividiendo el número estimado de estudiantes por la población total de 14 a 17 años se obtiene finalmente la estimación de la tasa de escolarización que se utiliza para aproximar el perfil temporal de la tasa de actividad de la población de 14 y 15 años tal como se indica en el

texto. El Gráfico A.2 muestra el perfil temporal de la fracción no escolarizada de este segmento de la población (normalizada de forma que 1970 = 1).

3. Datos para las regresiones en las que se exploran los determinantes del impacto de la reponderación

El indicador de no-respuesta utilizado en el texto es el porcentaje de los hogares que se niega a colaborar con la EPA. Para los primeros años de la encuesta, el dato se toma de la publicación anual de resultados de la misma. En años posteriores se utilizan las evaluaciones de calidad de la EPA y los informes sobre incidencias en los trabajos de campo (INE, varios años, a y b). Finalmente, se interpola para los años en los que no se dispone de datos directos, tal como se ilustra en el Gráfico A.3.

Gráfico A.3: % de los hogares que se niega a colaborar con la EPA



Indicadores de distancia entre cada EPA y el censo o padrón utilizado para definir su seccionado

Para cada censo y padrón entre 1960 y 2001 he recogido datos de población de derecho por provincias desagregada en tres tramos integrados, respectivamente, por los residentes en la capital, en el municipio de mayor población tras excluir a la capital y en el resto de la provincia.

Los indicadores de distancia entre censos (o padrones) sucesivos se obtienen sumando los valores absolutos de las variaciones en los pesos de los distintos territorios de un censo a otro. El indicador de *distancia total* entre dos censos sucesivos, realizados en los años t y $t+k$, se calcula como

$$(A.1) D^{tot}(C_t, C_{t+k}) = \frac{1}{2} \sum_p \sum_i |wnac_{ip,t} - wnac_{ip,t+k}|$$

donde $wnac_{ip,t}$ es el peso del territorio i de la provincia p en la población nacional en el año t .

El indicador de *distancia intraprovincial* entre los mismos censos vendría dado por

$$(A.2) D^{\text{intraprov}}(C_t, C_{t+k}) = \frac{1}{2} \sum_p \frac{wnac_{p,t} + wnac_{p,t+k}}{2} \sum_i |wprov_{ip,t} - wprov_{ip,t+k}|$$

donde $wnac_{p,t}$ es el peso de la provincia p en la población nacional y $wprov_{ip,t}$ el peso del territorio i en la población total de la provincia p . Con las modificaciones obvias, las fórmulas (A.1) y (A.2) sirven también para calcular la distancia entre cada censo dado y el padrón que le sigue o le precede.

Tomemos ahora un año $s \in [t, t+k]$ situado entre los dos censos contemplados arriba. Para aproximar la “distancia” entre las distribuciones de la población correspondientes a t y a s utilizaré el indicador anual de intensidad migratoria, IM , construido en de la Fuente (2010) como el número de desplazamientos migratorios (interregionales o internacionales) observados durante el año por cada mil habitantes. En particular, supondré que la distancia en relación al último año censal aumenta con el tiempo siguiendo el mismo patrón que el indicador de intensidad migratoria. Esto es, si definimos $CIM_{t,s}$ como la fracción de los movimientos migratorios totales registrados entre ambos censos que se ha acumulado en el año $s \in [t, t+k]$,

$$(A.3) CIM_{t,s} = \frac{\sum_{n=t}^s IM_n}{\sum_{n=t}^{t+k} IM_n}$$

entonces, la distancia entre el año censal t y el año s se aproximará por

$$(A.4) D^j(C_t, EPA_s) = CIM_{t,s} * D^j(C_t, C_{t+k})$$

donde $j = \text{tot, intraprov}$. Finalmente, si $s > t+k$, entonces definimos

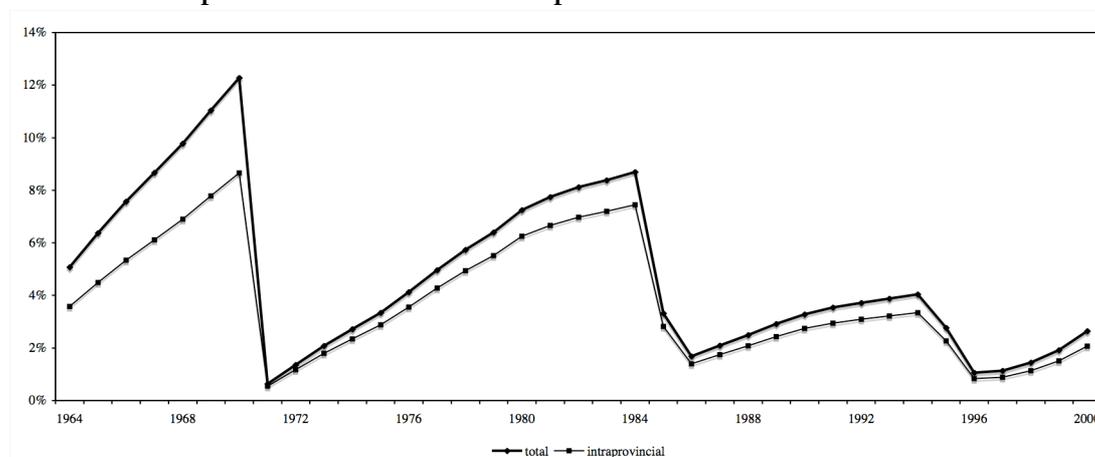
$$(A.5) D^j(C_t, EPA_s) = D^j(C_t, C_{t+k}) + CIM_{t+k,s} * D^j(C_{t+k}, C_{t+2k})$$

Esta última fórmula sería la relevante, por ejemplo, en 1992: aunque hay un censo en 1991, sus resultados todavía no se utilizan para renovar el seccionado de la EPA hasta 1995. Por lo tanto, si ignoramos la revisión parcial basada en el padrón de 1986, el seccionado de la EPA sigue estando basado en el censo de 1981 hasta el año 1995.

Una última complicación afecta a aquellos años en los que se produce una renovación del seccionado y ésta no se realiza de golpe y/o en el primer trimestre. En tales casos, se calcula qué fracción de la muestra total del año se toma del seccionado antiguo y del renovado y la distancia se calcula como una media ponderada de las distancias que corresponderían al mismo año antes y después de la renovación. Así por ejemplo, en 1995 se inició una renovación del seccionado que se introdujo de forma gradual durante seis trimestres, concluyendo en 1996-T2. Durante el primer trimestre de 1995 sólo 1/6 de la muestra trimestral se tomó con el nuevo seccionado. Durante el resto del año, esta fracción fue creciendo hasta alcanzar 4/6 en el cuarto trimestre. Por lo tanto, de las 24 partes en las que

podemos dividir la muestra anual (a razón de 6 partes por trimestre), se tomaron del seccionado renovado un total de $(1+2+3+4)/24 = 10/24$ partes.

Gráfico A.4: Distancia estimada entre la población de cada año y la del año correspondiente al censo utilizado para definir el seccionado de la EPA



El Gráfico A.4 muestra el perfil de las dos variables de distancia intercensal que he construido. Obsérvese que las diferencias entre los dos indicadores son poco importantes, con la excepción de los años 60, que no se incluyen en la muestra utilizada para estimar el modelo.

4. Series anuales de algunas variables de interés tomadas de la EPA en papel

He recogido datos de las publicaciones en papel de la EPA sobre el número de ocupados y ocupados que declaran un segundo empleo o actividad, el número medio de horas trabajadas por semana por los ocupados, el número de parados y el de parados que utilizan oficinas públicas de empleo para buscar trabajo. Las series tienen periodicidad en principio trimestral o semestral, pero presentan algunas lagunas en los períodos en los que no se realizó la EPA o no se recogió alguna de las variables de interés.

Como paso previo para la construcción de las serie de medias anuales con las que se trabaja en ciertas secciones del texto, he rellenado las lagunas en las series originales por un procedimiento similar al utilizado por el GTE. En primer lugar, se completan las series trimestrales o semestrales de tasas de variación interanual, siempre que es posible por interpolación entre las observaciones más próximas disponibles. Seguidamente, se completan las series en niveles aplicando la tasa relevante de crecimiento a la observación más próxima en niveles que esté disponible en cada caso.

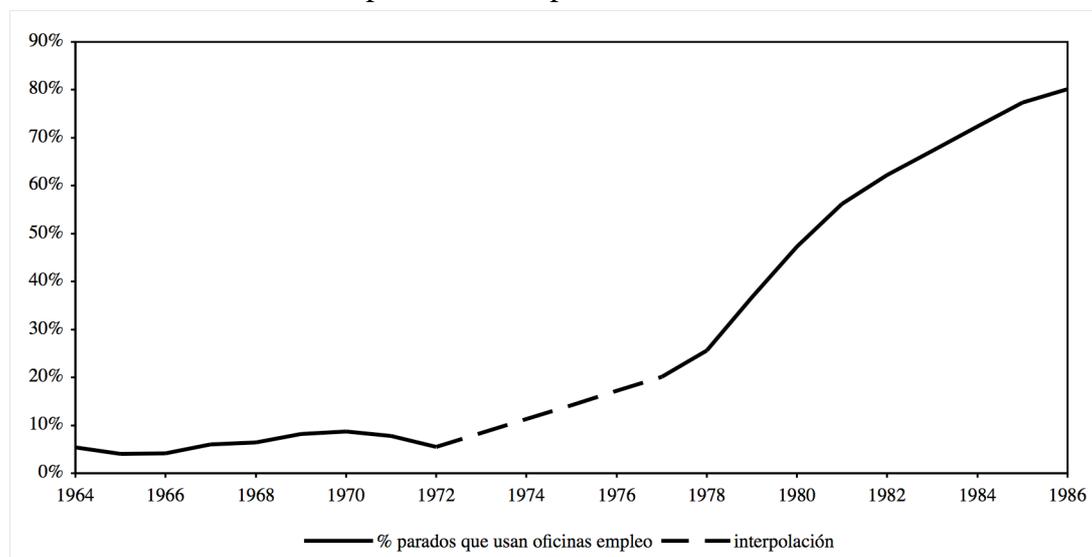
En particular, el nivel de cada variable de interés en 1964-T1 se estima proyectando hacia atrás la observación de 1965-T1 utilizando la variación interanual de la misma serie en 1965-T2 (esto es, la variación observada entre 1964-T2 y 1965-T2). Durante el período 1966-68 la EPA se realiza sólo en los trimestres 2 y 4. Las tasas de variación interanual de los trimestres impares se estiman por interpolación entre los correspondiente trimestres pares y se utilizan para estimar los niveles correspondientes. Por ejemplo, la tasa de variación estimada en 1966-

T1 es la media de las correspondientes a 1965-T4 y 1966-T2 y se utiliza para proyectar hacia delante el nivel observado en 1965-T1 con el fin de estimar el nivel de 1966-T1.

Este procedimiento se aplica a las series de ocupados, ocupados que declaran un segundo empleo, parados, parados que utilizan oficinas públicas de empleo y horas medias semanales trabajadas (*HMSEM*). Los porcentajes de ocupados que declaran un segundo empleo (*%SEGEMP*) y de parados que utilizan las oficinas públicas de empleo (*%PAROFP*) para buscar trabajo se calculan dividiendo las series correspondientes tras haber rellenado sus huecos.

Las series anuales se obtienen promediando las correspondientes observaciones trimestrales o semestrales. En el caso de *%SEGEMP* se descartan las observaciones de 1964 T1 (estimada por el procedimiento indicado arriba) y T2 por ser casi dos veces superior a las que les siguen y la observación anual de 1964 se construye aplicando a la correspondiente a 1965 la razón entre la media de los trimestres III y IV en 1964 y 1965. En 1972 se descarta también la observación estimada para el primer semestre (S1) de la misma serie dado que la ruptura que presenta la misma en 1972-S1 hace que no se disponga de observaciones “no contaminadas” a ambos lados de la que falta. En el caso de las series de *HMSEM* y *%PAROFP* existe un hueco entre 1972 o 1973 en las series originales que se rellena por interpolación lineal.

Gráfico A.5: % de parados que utiliza una oficina pública de ocupación para buscar empleo, serie anual



ANEXO 2: Principales cambios metodológicos y otras incidencias en la EPA

En este anexo se presenta un listado cronológico de los principales cambios metodológicos y de otras incidencias que ha experimentado la EPA a lo largo de su historia. Las principales fuentes utilizadas en la construcción de este listado han sido INE (1995 y 2002), Alvarez (1987), Mas et al (2002), Pérez Infante (2006) y las propias publicaciones de resultados de la EPA.

1964T2

- Puesta en marcha de la encuesta.

1965

- Comienza a investigarse la población que está prestando el servicio militar. Se clasifican como activos o inactivos en función de su situación justo antes de su incorporación a filas.

1966

- T1 y T3: la encuesta no se realiza.

1967

- T1 y T3: la encuesta no se realiza.

1968

- T1 y T3: la encuesta no se realiza.
- Nuevas proyecciones de población basadas en el padrón de 1965.
- Actualización del seccionado a partir del padrón de 1965.
- Revisión de las series de población activa desde 1960.

1969

- S1: La encuesta pasa a tener periodicidad semestral.

1971

- S1: la encuesta no se realiza.
- S2: la encuesta queda totalmente integrada en el diseño de la Encuesta General de Población. Actualización de la estratificación y del seccionado en base al censo de 1970.

1972

- S1: la encuesta no se realiza
- S2: Los trabajadores estacionales que no trabajan ni buscan empleo entre campaña y campaña se contabilizan entre los activos como un grupo aparte de parados y ocupados.
- S2: Introducción de un nuevo cuestionario. Ciertos cambios en la clasificación de las ayudas familiares y trabajadores ocasionales.

1975

- T3: La encuesta vuelve a tener periodicidad trimestral.

1976T3

- Modificación del cuestionario.
- Actualización de las proyecciones de población, tomando como base el censo de 1970 y el padrón de 1975.
- Los varones que cumplen el servicio militar se excluyen de la población activa o inactiva, creándose el concepto de "población contada aparte."
- Los activos marginales pasan a integrarse en la población activa (sacándose de la población inactiva, donde se habían incluido hasta el momento), siempre que hayan trabajado al menos un determinado número de horas durante los tres meses anteriores a la encuesta.
- Desaparece la categoría especial para los trabajadores estacionales que ni trabajan ni buscan trabajo entre campañas. Este colectivo pasa a considerarse inactivo.

1978T2

- Actualización de las probabilidades de selección y del seccionado como consecuencia del Padrón de 1975.

1980T2

- Cambio en la definición de la población en edad de trabajar como consecuencia de la elevación de la edad legal para trabajar de 14 a 16 años.

1984T1

- Los perceptores del subsidio agrario en Andalucía y Extremadura (PER), pasan a ser considerados como parados (los beneficiarios del antecesor inmediato del PER, los planes de empleo comunitario, se incluían en los ocupados estrictos).

1985T2

- Actualización de las probabilidades de selección y del seccionado y nueva estratificación como consecuencia del Censo de 1981. No está claro si el cambio se produce en T1 o en T2.

1986T2

- Actualización de las proyecciones de población que pasan a tomar como base el censo de 1981. Se construye una serie retrospectiva, en principio para 1981T1 a 1986T1, que luego se enlaza de alguna forma con las series anteriores, revisando éstas en alguna medida (INE, 2009, p. 47).

1987 T2

- Con carácter general, a partir de ahora se utiliza sólo la semana de referencia (la anterior a la de la encuesta) para clasificar a los encuestados. Hasta este momento, para algunos grupos (ocupados marginales y ayudas familiares) se tenía en cuenta la situación durante los tres meses anteriores.
- Esto provoca cambios en la definición de ocupado y parado para algunos colectivos específicos. Por ejemplo, para considerar ocupado a los activos marginales ahora sólo se pide que hayan trabajado una hora en la semana de referencia; antes se pedía que hubiesen trabajado 1/3 de la jornada durante 3 meses. Desaparece la distinción entre

ocupados en sentido estricto y ocupados (activos) marginales. Estos últimos pasan a ser considerados ocupados si han trabajado al menos 1 hora durante la semana de referencia.

- Parados: se amplía el período exigido de búsqueda de trabajo de 1 a 4 semanas y se explicita el horizonte de disponibilidad para trabajar, fijándose en no más de 15 días.
- Cambio del criterio de residencia utilizado para identificar a la población encuestable. Se tienen en cuenta los que contribuyen a la producción nacional.⁴²
- Cambios en la estructura del cuestionario y en la formulación de numerosas preguntas.
- Cambios en las clasificaciones de estudios, ocupaciones y actividades.
- Se introduce la distinción entre ocupados a tiempo completo y a tiempo parcial y entre ocupados con contrato indefinido y contrato temporal.
- Cambios en el procedimiento de depuración del cuestionario.

1988 T2

- Incorporación de Ceuta y Melilla a la muestra.

1989 T1

- Actualización de las probabilidades de selección y del seccionado como consecuencia del Padrón de 1986.

1990T4

- Actualización de las proyecciones de población, aunque siguen tomando como base el censo de 1981.

1992T1

- Variaciones menores en el cuestionario
- Nueva clasificación de estudios

1993T1

- Variaciones menores en el cuestionario
- Nueva clasificación de actividades económicas

1994

- T1: revisión de las proyecciones de población.
- T2: Nueva clasificación de ocupaciones

⁴² Comenzando en 1987, la población de referencia para la encuesta es la residente en España, con independencia de su nacionalidad y de que pueda estar trabajando en el extranjero. El período de referencia para fijar la residencia es de un año. Por lo tanto, los españoles que se han ido al extranjero se incluyen si lo han hecho por menos de un año y los extranjeros que viven aquí se incluyen si han permanecido o piensan permanecer en el país durante más de un año. Por convención, se incluyen siempre los diplomáticos y militares españoles destinados en el extranjero y se excluyen los militares y diplomáticos extranjeros destinados en España. En la EPA 1976 se incluía a los extranjeros aunque llevaran menos de un año en el país pero se excluía a parte de los españoles que trabajaban en el extranjero aunque residieran en España.

1995T1

- Actualización de la estratificación y nueva subestratificación más detallada (se pasa de 7 a 15 subestratos).
- Actualización de las probabilidades de selección y del seccionado como consecuencia del censo de 1991. Renovación total de la muestra. El cambio afecta aproximadamente al 85% de las secciones y se produce de forma gradual durante 6 trimestres, concluyendo en 1996T2.

1996T1

- Actualización de las proyecciones de población, que toman como base el censo de 1991.

1996-1998

- Modificaciones en el trabajo de campo: utilización de ordenadores portátiles y llamadas telefónicas.

1999-T1

- Nueva definición de subempleo.
- Modificaciones importantes del cuestionario. Preguntas más claras y precisas. Se busca detectar mejor los empleos de corta duración o jornada muy reducida. Cambios en preguntas sobre nivel de formación.
- Comienzan a realizarse entrevistas también en agosto (hasta entonces período vacacional para los entrevistadores) y en las 13 semanas de cada trimestre (y no en 12 como hasta el momento).

2000

- T1: Actualización de las probabilidades de selección y del seccionado utilizando el padrón continuo (de 1998) y el censo electoral. Se renuevan 143 secciones, un 4,1% del total, de una sola vez, no gradualmente durante 6 trimestres.
- T3: nueva clasificación educativa y pequeños cambios del cuestionario en temas de educación.

2001

- Cambios en el cuestionario motivados por la necesidad de implementar la nueva definición operativa de desempleo a partir de 2002.
- Cambios menores en el tratamiento de personas en principio ocupadas pero que no trabajan en la semana por estar sometidos a un expediente de regulación de empleo con suspensión, por enfermedad, maternidad u otros motivos. Se endurecen algo los requisitos para considerarlos ocupados.
- T2: Actualización de las proyecciones de población, que siguen tomando como base el censo de 1991.

2002T1

- Aplicación de la nueva definición operativa de paro.
- Implantación del procedimiento de reponderación o calibrado de los factores de elevación. Se toma como referencia externa la población de cada comunidad autónoma por sexos y por intervalos de edad de 5 años y la población total provincial de acuerdo con las

proyecciones de población utilizadas en cada momento. Se aplica retroactivamente desde 1976 en la serie corregida que publica el INE.

- Aplicación de las nuevas proyecciones de población y revisión retroactiva de los resultados desde 1996. Elaboración de la serie homogénea 1991-2001.

2005T1

- Actualización de las proyecciones de población, que pasan a tomar como base el censo de 2001 y el nuevo padrón continuo.
- Actualización de las probabilidades de selección y del seccionado como consecuencia del censo de 2001. Se introducen 560 nuevas secciones de un total de 3.588, gradualmente sobre un período de 6 trimestres.
- La variable de nacionalidad pasa a incluirse entre las que se utilizan en el proceso de reponderación o calibrado de los coeficientes de elevación. El nuevo procedimiento de calibrado se aplica retroactivamente en la serie revisada 1996-2004.
- Revisión de las series desde 1996 hasta 2004, incluyendo el recalibrado por nacionalidad (además de las variables ya consideradas previamente).
- Renovación completa del cuestionario. Se simplifican o reescriben muchas preguntas buscando mayor claridad y precisión.
- Cambios en el trabajo de campo, introducción del método CATI de encuestas telefónicas asistidas por ordenador.

Referencias

- Albacete, R. y A. Laborda (2005). "Cambios en la Encuesta de Población Activa y en la Contabilidad Nacional." *Cuadernos de Información Económica* 186, mayo-junio, pp. 44-55.
- Alcaide, J. (2000). "Series históricas españolas 1898 a 1988." En Velarde, J., director, *1900-2000, Historia de un esfuerzo colectivo*. Madrid, Planeta, vol. II, pp. 645-712.
- Alvarez, F. (1987). "La nueva encuesta de población activa (EPA)." *Boletín Trimestral de Coyuntura*, septiembre.
- Alvarez, F. (2000). "La Encuesta de Población Activa: cómo mide el mercado laboral." *Estadística Española* 42(146), pp. 201-61.
- Argentaria (1999). *Economía*. Segunda Época. No. 17, octubre.
- Banco de España (2002). "Cambios metodológicos en la EPA en 2002." *Boletín Económico*, Abril de 2002, pp. 67-77. Madrid.
- Dirección General de Previsión y Coyuntura (DGPC, 1988). "Nota sobre el enlace provisional de las series laborales tras la reforma de la EPA en el segundo semestre de 1987." Ministerio de Economía y Hacienda, Madrid. Reimpreso en *Estudios de Economía del Trabajo en España III. El problema del paro*. Ministerio de Trabajo y Seguridad Social, Madrid, pp. 1185-1207.
- de la Fuente, A. (2009b). "A mixed splicing procedure for economic time series." Mimeo, Instituto de Análisis Económico, CSIC, diciembre de 2009.
- de la Fuente, A. (2010). "Series largas de algunos agregados demográficos regionales, 1950-2009. (RegDat-Dem versión 3.1)." *Estadística Española* 52(175), pp. 501-28.
- Fernández, F., J. Muro y L. Toharia (1991). "Los problemas de enlace de la Encuesta de Población Activa." En *Estudios de Economía del Trabajo en España III. El problema del paro*. Ministerio de Trabajo y Seguridad Social, Madrid, pp. 1167-83.
- García España, E. (1969). "Diseño de la Encuesta General de Población (E.G.P.)." *Estadística Española* 44, pp. 5-42.
- García-Perea, P. (1991). "Elaboración de series homogéneas de ocupados y parados a partir del segundo trimestre de 1964." *Estudios de Economía del Trabajo en España III. El problema del paro*. Ministerio de Trabajo y Seguridad Social, Madrid.
- García-Perea, P. y R. Gómez (G&G, 1994). "Elaboración de series históricas de empleo a partir de la Encuesta de Población Activa (1964-1992)." Documento de Trabajo, núm. 9409. Banco de España, Madrid.
- Grupo de trabajo sobre problemas de empleo (GTE, 1979). *Población, actividad y ocupación en España (reconstrucción de las series históricas 1960-78)*. Madrid, Ministerio de Economía.
- Instituto Nacional de Estadística (INE, 1964). *Encuesta sobre la población activa española, 1964. Plan General*. Madrid.
- Instituto Nacional de Estadística (INE, 1965 a 1973). *Encuesta sobre población activa. Años 1964 a 1971*. Madrid.

- Instituto Nacional de Estadística (INE, 1975). Encuesta de población activa. Años 1972 y 1973. Madrid.
- Instituto Nacional de Estadística (INE, 1976 a 1978). Encuesta de población activa. Años 1974 a 1976. Madrid.
- Instituto Nacional de Estadística (INE, 1990). Encuesta de Población Activa. Series revisadas 1976-87. Madrid.
- Instituto Nacional de Estadística (INE, 1995). *Encuesta de Población Activa. Evolución de las Características Técnicas durante el período 1964-76*. Madrid.
- Instituto Nacional de Estadística (INE, 1996a). Encuesta de Población Activa. Actualización de la muestra de secciones. Adenda. Resultados detallados tercer trimestre de 1995. Madrid.
- Instituto Nacional de Estadística (INE, 1996b). "Incidencia del cambio del seccionado en las estimaciones de la Encuesta de Población Activa." Mimeo, Madrid.
- Instituto Nacional de Estadística (INE, 2002). *Encuesta de Población Activa. Evolución de las Características Técnicas. Período 1976-2002*. Madrid.
- Instituto Nacional de Estadística (INE, 2003). Encuesta de Población Activa. Series históricas revisadas 1976-2001. Madrid. CD-Rom.
- Instituto Nacional de Estadística (INE, 2005). Encuesta de población activa 2005. *Cifras INE 1/2005*.
<http://www.ine.es/revistas/cifraine/0105.pdf>
- Instituto Nacional de Estadística (INE, 2008). *Encuesta de Población Activa. Metodología 2005. Descripción de la Encuesta, definiciones e instrucciones para la cumplimentación del cuestionario*. Madrid.
- Instituto Nacional de Estadística (INE, 2009). *Encuesta de Población Activa. Diseño de la encuesta y evaluación de la calidad de los datos. Informe técnico*. Madrid.
- Instituto Nacional de Estadística (INE, 2012a). *Encuesta de Población Activa*. En base de datos INEbase: Sociedad: Mercado Laboral. Madrid.
http://www.ine.es/inebmenu/mnu_mercalab.htm
- Instituto Nacional de Estadística (INE, 2012b). *Contabilidad Nacional de España*. En base de datos INEbase: Economía: Cuentas Económicas. Madrid.
http://www.ine.es/inebmenu/mnu_cuentas.htm
- Instituto Nacional de Estadística (INE, 2012c). *Contabilidad Nacional Trimestral de España*. En base de datos INEbase: Economía: Cuentas Económicas. Madrid.
http://www.ine.es/inebmenu/mnu_cuentas.htm
- Instituto Nacional de Estadística (INE, 2012d). Cifras de población. En Base de datos electrónica INEbase. Demografía y población. Madrid.
http://www.ine.es/inebmenu/mnu_cifraspob.htm
- Instituto Nacional de Estadística (INE, varios años, a). Evaluación de la calidad de los datos de población activa. Madrid.
- Instituto Nacional de Estadística (INE, varios años, b). Incidencias en los trabajos de campo de la Encuesta de Población Activa. Madrid.
- Instituto Nacional de Estadística (INE, varios años, c). *Anuario Estadístico de España*. En base de datos INEbase Historia. Madrid. <http://www.ine.es/inebaseweb/25687.do>

- Izquierdo, J. F. (1998). "Efectos de la renovación censal de la EPA durante 1995-96." *Situación Trimestral* no. 1, marzo, pp. 75-83. BBV, Madrid.
- Lundström, S. y C. E. Särndal (2001). *Estimation in the presence of non-response and frame imperfections*. Statistics Sweden, Estocolmo.
- Maluquer, J. y M. Llonch (2005). "Trabajo y relaciones laborales." En A. Carreras y X. Tafunell, coordinadores. *Estadísticas históricas de España, siglos XIX-XX*, segunda edición. Fundación BBVA, Bilbao, pp. 1155-1245.
- Mas, M., F. Pérez, E. Uriel, L. Serrano y A. Soler (2002). Capital humano, series 1964-2001. Bancaja.
<http://obrasocial.bancaja.es/publicaciones/publicaciones-ficha.aspx?id=65>
- Ministerio de Empleo y Seguridad Social (MESS, 2012a). Afiliación de trabajadores al sistema de la Seguridad Social. Sitio web del MESS: Estadísticas: Publicaciones de síntesis: Principales series.
<http://www.empleo.gob.es/series/>
- Ministerio de Empleo y Seguridad Social (MESS, 2012b). Encuesta de Coyuntura Laboral. Sitio web del MESS: Estadísticas: Publicaciones de síntesis: Principales series.
<http://www.empleo.gob.es/series/>
- Ministerio de Trabajo y Seguridad Social (MTSS, 1987). *Coyuntura Laboral*, Avance no. 20, septiembre.
- Ministerio de Trabajo y Seguridad Social (MTSS, 1988a). *Anuario de Estadísticas Laborales, 1987*, pp. 201-07. Madrid.
- Ministerio de Trabajo y Seguridad Social (MTSS, 1988b). *Coyuntura Laboral*, Avance no. 22, septiembre
- Pérez Infante, J. I. (1998). "La medición del empleo y el paro en España: situación y problemas." *Cuadernos de Información Económica* 138, septiembre, pp. 11-27.
- Pérez Infante, J. I. (2000). "La medición del desempleo en España: la EPA y el paro registrado." *Revista del Ministerio de Trabajo y Asuntos Sociales. Economía y Sociología* 21, pp. 19-57.
- Pérez Infante, J. I. (2001). "Los problemas de medición del empleo: EPA y CNE." En L. Fina y L. Toharia, coordinadores. *El Empleo en España: situación y perspectivas*. Ministerio de Trabajo y Asuntos Sociales, Madrid, pp. 117-44.
- Pérez Infante, J. I. (2006). *Las estadísticas del mercado de trabajo en España*. Ministerio de Trabajo y Asuntos Sociales. Colección Informes y Estudios Empleo, Madrid.
- Pérez Infante, J. I. y Ll. Fina (1999). "La duración de la jornada de trabajo. Situación y evolución reciente." *Revista del Ministerio de Trabajo y Asuntos Sociales* 16, pp. 13-53.
- Porras, J. (1999). "Diseño de muestras en encuestas de población y hogares." *Qüestió* 23(3), pp. 543-58.
- Serrano, L. y A. Soler (2008). Metodología para la estimación de las series de capital humano, 1964-2007. IVIE y Fundación Bancaja.
<http://www.ivie.es/banco/capital.php>
- Toharia, L., director (1998). *El mercado de trabajo en España*. McGraw-Hill, Madrid.
- Trejo, J. y L. Ortega (2005). "Enlace de las series de paro 1976-2000 según la definición EPA-2002." Subdirección de Estadísticas del Mercado Laboral, INE.
http://www.ine.es/daco/daco42/daco4211/epa_reest_paro_pdf.pdf

