

Series enlazadas de los principales agregados nacionales de la EPA, 1964-2014

Ángel de la Fuente Moreno ^(*)

FEDEA e Instituto de Análisis Económico (CSIC)

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 1964-2014 corrigiendo algunas de las rupturas que persisten en las series históricas más recientes del INE.

Palabras clave: Encuesta de Población Activa, series homogéneas.

Clasificación AMS: 62P25, 91C99.

Linked series of the main national aggregates of the Spanish Labor Force survey, 1964-2014

Abstract

This paper collects and extends several historical series of the main national aggregates of the Spanish Labor Force Survey and constructs new homogenous annual series for these variables covering the period 1964-2014 by removing the discontinuities that persist in the most recent historical series of the Spanish National Statistical Institute.

Keywords: Spain, Labor Force Survey, homogeneous series.

AMS classification: 62P25, 91C99.

^(*) 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 ECO2014-59196-P y los comentarios y sugerencias de Luis González Calbet, Alfredo Cristóbal y dos evaluadores anónimos de Estadística Española.

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 1964-2014. 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. Tras discutir brevemente el diseño de la encuesta, en la sección 2 se repasan las principales incidencias que pueden haber generado discontinuidades en las series de la EPA a lo largo de su historia. Entre éstas 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 se parte 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. Seguidamente, se aborda una revisión de la más reciente de estas series. En la sección 5 se estima el tamaño de las distintas rupturas de la misma y en la sección 6 se construye una nueva serie enlazada en la que se corrigen tales discontinuidades. Finalmente, en la sección 7 se recopilan algunas variables auxiliares que pueden resultar de utilidad en futuros trabajos a la hora de pasar de las series de ocupados de la EPA a las series de empleo y horas trabajadas de Contabilidad Nacional. Finalmente, la sección 8 concluye con un breve resumen.¹

2. La Encuesta de Población Activa: consideraciones generales y posibles fuentes de discontinuidades

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

¹ Una versión más larga del presente trabajo (de la Fuente, 2015) está disponible en internet junto con dos archivos Excel que contienen las nuevas series y los datos originales utilizados para construirlas (véase <http://www.fedea.net/datos-mercado-de-trabajo/>). En esa versión del trabajo se incluye una descripción del diseño muestral y de la metodología de estimación de la encuesta, una discusión detallada de las distintas correcciones que aquí se introducen a las series del INE y un listado cronológico detallado de los distintos cambios metodológicos y otras incidencias que ha experimentado la EPA desde su aparición en 1964.

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) trimestral² y proporciona información desagregada territorialmente (por regiones y/o provincias) desde 1972.³ 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) 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 Desde 1987, la metodología de la encuesta sigue las directrices de Eurostat para asegurar que la información se recoge de forma homogénea en todos los países miembros de la UE.

2.1 Diseño muestral y estimación de los agregados poblacionales

Con el fin de identificar las principales fuentes de discontinuidades en las series de la EPA, resulta conveniente repasar brevemente algunos aspectos de su metodología. En la encuesta 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. En la primera etapa del proceso de selección muestral 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 que 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 la población del municipio al que pertenecen y, dentro de cada estrato, en substratos en base a la condición socioeconómica de sus habitantes.

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. Cada

² 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.

³ Ceuta y Melilla se incorporan a la muestra en 1988.

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 en particular INE (2008 y 2012) así como García España (1969), Porras (1999) y Álvarez (2000).

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. Durante el período comprendido entre dos censos también se realizan ocasionalmente revisiones parciales del seccionado 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 se realizan aproximadamente cada tres años en los últimos tiempos.

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). Finalmente, una vez fijadas las secciones censales, se elige aleatoriamente una muestra de hogares a entrevistar dentro de cada una de ellas.

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. Estos factores se construyen a partir de las estimaciones de población intercensal que realiza el propio INE.

2.2 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.⁵ 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 (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 entre 25 y 49 años de edad) están infrarrepresentados.

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.

2.3 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

⁵ Véanse entre otros Toharia (1998), Álvarez (2000) y Pérez Infante (2000 y 2006).

reponderación o calibrado de los factores de elevación, supone sustituir los factores de elevación originales de los hogares por otros 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 extraídas de las proyecciones de población del propio INE.⁶ 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. Desde 2014 se utiliza también el desglose por sexo y grupo de edad de la población provincial y el desglose por tamaños del número de hogares a nivel de comunidades autónomas. 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.

2.4 Posibles fuentes de discontinuidades en las series de la encuesta

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. En primer lugar, resulta evidente que cualquier revisión de las proyecciones de población del INE generará una discontinuidad en las series de la EPA a través de los factores de elevación de la encuesta, 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 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 que, entre otras cosas, permita detectar cada vez mejor situaciones de empleo marginal u ocasional. También han de tenerse en cuenta los sucesivos cambios en las prácticas y

⁶ Para más detalles sobre el procedimiento de calibrado que utiliza el INE y sus efectos, véanse Álvarez (2000), INE (2012) 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 y de los métodos de imputación que pueden utilizarse para paliar tales problemas.

procedimientos de trabajo de campo que se han producido a lo largo de la vida de la encuesta y algunas variaciones en las definiciones de ciertas variables, 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 a de la Fuente (2015) se ofrece un listado cronológico de tales cambios metodológicos.

Cuadro 1

Cambios de criterio en la clasificación de algunos colectivos en la EPA

	<i>Varones que prestan el servicio militar</i>	<i>Activos marginales</i>	<i>Temporeros sin trabajo</i>	<i>Ayudas familiares y trabajadores ocasionales</i>
EPA 1964	no se investigan	inactivos	inactivos	como en 1965
EPA 1965 1965T1 a 1971S2	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 autclasifican como inactivos (estudiantes, amas de casa, jubilados...)
EPA 1972 1972 a 1976T2	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
EPA 1976 1976T3 a 1987T1	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
EPA 1987 1987T2 a 2001	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 2 a de la Fuente (2015)

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 familiares⁷ y los activos marginales, esto es, aquellas personas que, autclasificándose como inactivos (jubilados, estudiantes, amas de casa...), han trabajado o han buscado

⁷ Este colectivo está integrado por aquellas personas que trabajan sin remuneración explícita en el negocio o empresa de un familiar con el que conviven.

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.

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 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 1999⁸ 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. 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.⁹ Finalmente en 2014 se renueva la base poblacional de la Encuesta para adaptarla al Censo de 2011 y, tal como ya se ha dicho, se incorporan variables adicionales de calibrado.

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

⁸ Véase por ejemplo Pérez Infante (2000).

⁹ Sobre las novedades que introduce la EPA 2005 véanse entre otros INE (2005) y Albacete y Laborda (2005).

con los resultados de la encuesta publicados originalmente: están basadas en estimaciones actualizadas de la población y generalmente incorporan aquellas novedades metodológicas que pueden ser aplicadas a posteriori, tales como la reponderación o ciertos cambios en la definición de algunos agregados.

Actualmente (en febrero de 2015) la página web del INE proporciona cuatro 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 económica (en activos, ocupados, parados, inactivos y población contada aparte¹⁰). La más antigua, a la que denominaré “primera serie histórica del INE” o “*serie H1*”,¹¹ cubre el período 1977- 2001. En la segunda (a la que llamaré *serie H2*)¹² el período muestral se extiende hasta 2004 y en la tercera y última por el momento (*serie H3*)¹³ se llega hasta 2013, 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). Finalmente, el INE ha publicado en 2014 una nueva serie retrospectiva desde 2002 (*serie H4*)¹⁴ tras la incorporación de la nueva base poblacional basada en el censo de noviembre de 2011. En lo que sigue llamaré *H3&4* a la serie construida mediante la yuxtaposición de *H3* y *H4*, esto es, utilizando *H4* desde su inicio en 2002 y *H3* en los años anteriores.

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 las 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

¹⁰ 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. Esta categoría desaparece tras 2001 con la supresión del servicio militar.

¹¹ En su página web, el INE se refiere a *H1* como la “serie publicada”, lo que no quiere decir 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 en INE (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.

¹² 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 en INE (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.

¹³ http://www.ine.es/dyngs/INEbase/es/operacion.htm?c=Estadistica_C&cid=1254736032345&menu=resul tados&idp=1254735976595.

¹⁴ http://www.ine.es/dyngs/INEbase/es/operacion.htm?c=Estadistica_C&cid=1254736176918&menu=resul tados&idp=1254735976595 Véase INE (2014)

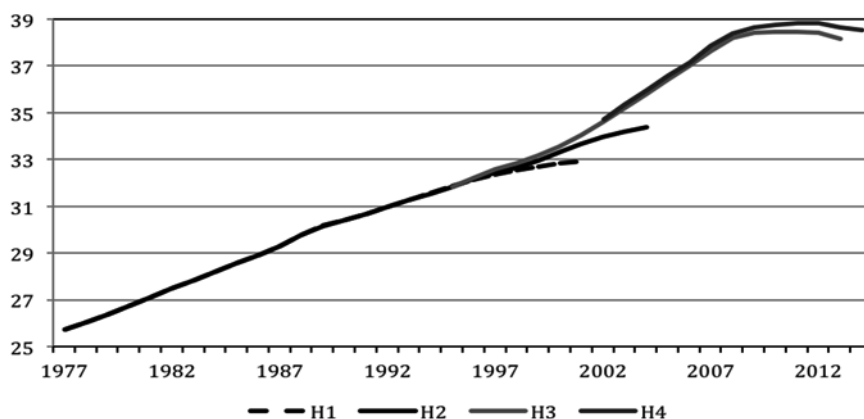
dispuso de los resultados del censo de (noviembre de) 2001, resultó evidente que tal influjo había sido muy superior al estimado inicialmente, lo que hizo necesaria una nueva revisión al alza de las proyecciones de población de 1996 en adelante que se incorpora a la serie *H3*. Lo mismo sucede en 2014 cuando se incorporan las proyecciones de población basadas en el censo de 2011.

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. Finalmente, *H4* también mejora retrospectivamente el calibrado desde 2002 en adelante, introduciendo nuevas variables auxiliares como ya se ha indicado.

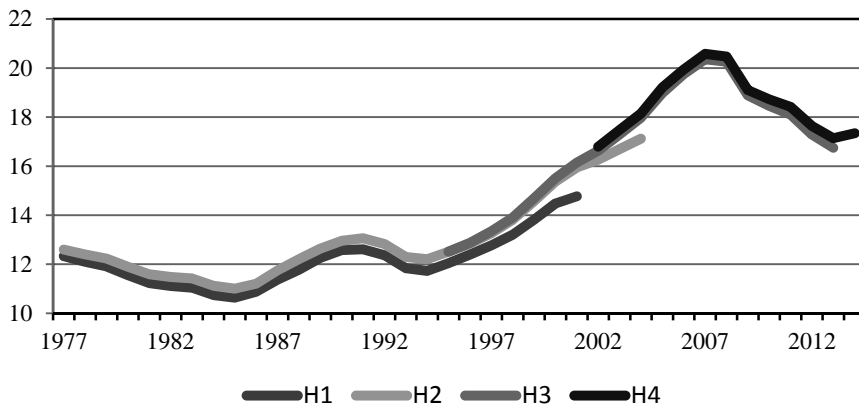
Gráfico 1

Comparación de las sucesivas series históricas del INE, España en su conjunto

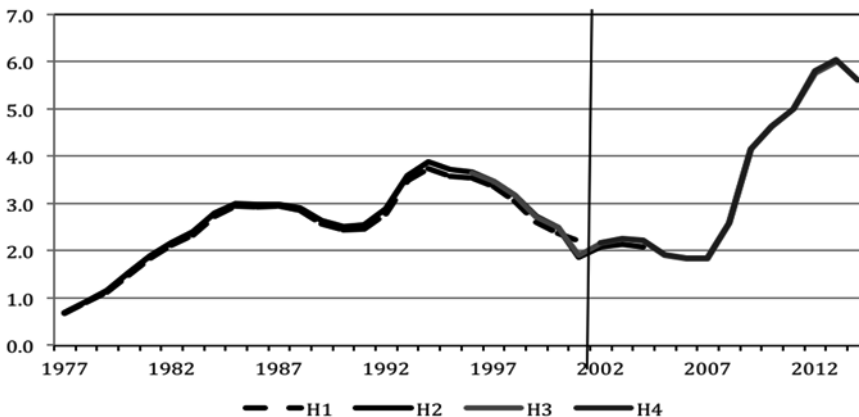
a. Población de 16 o más años de edad en viviendas familiares



b. Población ocupada



c. Población parada



Notas: Millones de personas; datos anuales, contruidos como la media no ponderada de los cuatro trimestres. Panel c: H2 y H3&4 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 INE de la afloración 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: no está claro hasta qué punto el proceso de calibrado corrige el sesgo derivado del envejecimiento del seccionado.

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.¹⁵ 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.

Cuadro 2

Posibles discontinuidades de las series históricas del INE, 1976-2009 (Continúa)

<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	¿?

¹⁵ El cambio de criterio afecta, en particular, a aquellas personas no ocupadas cuya única forma activa 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.

Cuadro 2

Posibles discontinuidades de las series históricas del INE, 1976-2009 (Conclusión)

<i>fecha</i>	<i>cambio metodológico</i>	<i>efecto estimado INE</i>	<i>¿corregido en serie histórica?</i>
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
2014T1	Revisión proyecciones de población		si
	Variables adicionales de calibrado		si

Fuentes: Pérez Infante (2006), EPA, INE (1995, 2002 y 2014) y Anexo 2 a de la Fuente (2015).

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 (serie 2005) y con la antigua definición de desempleo.

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.

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 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. 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 las tasas de crecimiento así corregidas, 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.

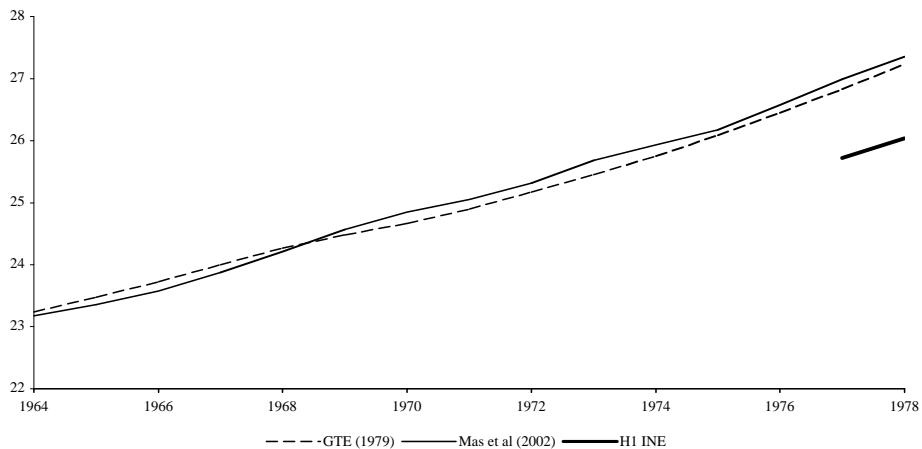
La otra fuente que también cubre el período 1964-76 es el estudio realizado por Mas et al (2002).¹⁶ A diferencia del GTE, Mas et al intentan corregir directamente las discontinuidades de la serie, estimando la evolución de los diversos colectivos cuya clasificación ha variado de una metodología de la 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.

¹⁶ 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. 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, 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.

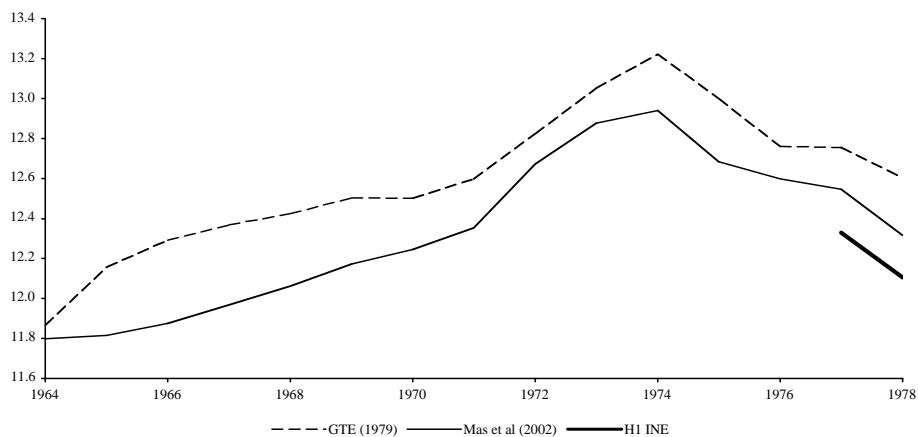
Gráfico 2

Comparación de diversas series, 1964-78

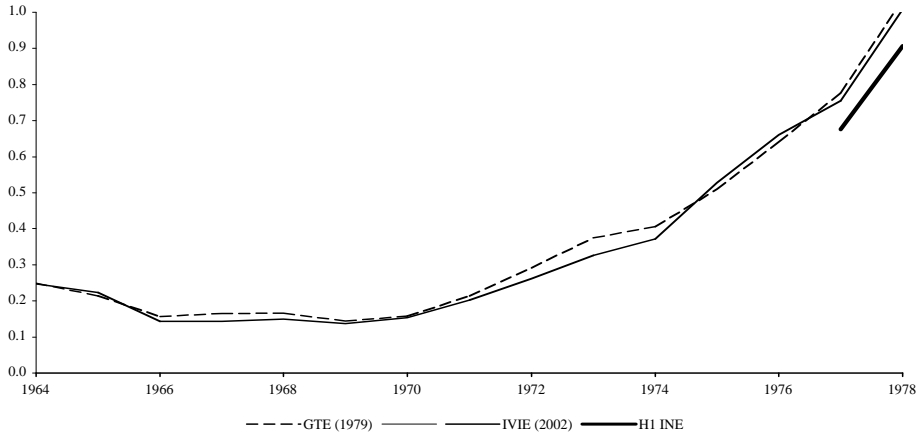
a. Población en edad de trabajar



b. Ocupados



c. Parados



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 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. 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 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&4* (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 conservarla 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) 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

retropolación¹⁷ con la serie *HI* del INE en 1977 para llegar a la “serie *HI* 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 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 por edades simples tomada del censo de 1960 y de las estimaciones intercensales de población del INE (2015c), que están disponibles desde 1971 en adelante.

El peso de la población de 14 y 15 años de edad (*wI415/14+*) 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. 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 retropolación 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.

b) *Población activa*

Una vez completado el paso anterior, disponemos de una estimación de la población total de 14 y 15 años de edad (*POBI415*). Suponiendo que este colectivo es demasiado joven para haber comenzado el servicio militar, queremos estimar su desglose entre ocupados, parados e 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. Dado esto, una posible

¹⁷ Esto es, la serie más reciente se extiende hacia atrás desde su primera observación utilizando las tasas de crecimiento de la serie más antigua.

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.

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 a de la Fuente, 2015). Para cada año entre 1964 y 1977, la fracción no escolarizada de la población relevante 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 retroproyección con la serie *HI* de activos 16+ del INE que comienza en 1977.

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 *HI* 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.

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, *HI*. 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 *HI* 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 la segunda serie histórica del INE y la primera, utilizando datos anuales del período 1977-2000.

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.¹⁸ La variable dependiente es la corrección porcentual de las series originales de ocupados y parados que se introduce como resultado de la reponderación ((serie reponderada – serie *HI*)/serie *HI*). 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 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 a de la Fuente, 2015).

¹⁸ 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 11. Ambas series están disponibles en <http://www.ine.es/epa02/repercusion.htm>.

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]	
constante	0,0127 (2,11)	0,0408 (5,59)		0,0125 (2,08)	0,0406 (5,63)		0,0162 (2,55)	0,0312 (2,82)		
negativas	0,87 (5,19)	-5,31 (3,90)		0,87 (5,28)	-5,20 (3,89)		0,78 (5,22)	-4,08 (2,33)		
negativas*w65+		27,08 (4,41)	6,00 (16,67)		26,49 (4,38)	5,93 (16,23)		23,27 (3,04)	5,28 (10,19)	
distintraprov1	0,07 (1,01)	-1,52 (2,33)								
distintraprov1*w65+		10,36 (2,49)	1,71 (6,74)							
disttot				0,07 (1,06)	-1,29 (2,37)					
disttot*w65+					8,78 (2,54)	1,49 (6,79)				
distintraprov2							0,04 (0,29)	0,83 (0,79)		
distintraprov2*w65+								-4,26 (0,68)	2,59 (5,57)	
Root MSE	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]	
constante	-0,0027 (0,58)	0,0114 (1,55)		-0,0028 (0,60)	0,0112 (1,54)		0,0130 (2,36)	0,0151 (1,26)		
negativas	1,11 (8,42)	-1,83 (1,34)		1,10 (8,53)	-1,74 (1,29)		0,80 (6,18)	-0,31 (0,17)		
negativas*w65+		12,49 (2,03)	5,22 (22,01)		12,00 (1,97)	5,17 (21,49)		5,69 (0,68)	5,17 (11,54)	
distintraprov1	0,15 (2,72)	-1,01 (1,54)								
distintraprov1*w65+		7,50 (1,80)	1,16 (6,95)							
disttot				0,13 (2,76)	-0,84 (1,52)					
disttot*w65+					6,25 (1,79)	1,01 (7,00)				
distintraprov2							-0,11 (1,03)	0,37 (0,33)		
distintraprov2*w65+								-2,79 (0,41)	1,28 (3,20)	
Root MSE	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.

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 capturar (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 de mayor población del resto de los municipios, 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. 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 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

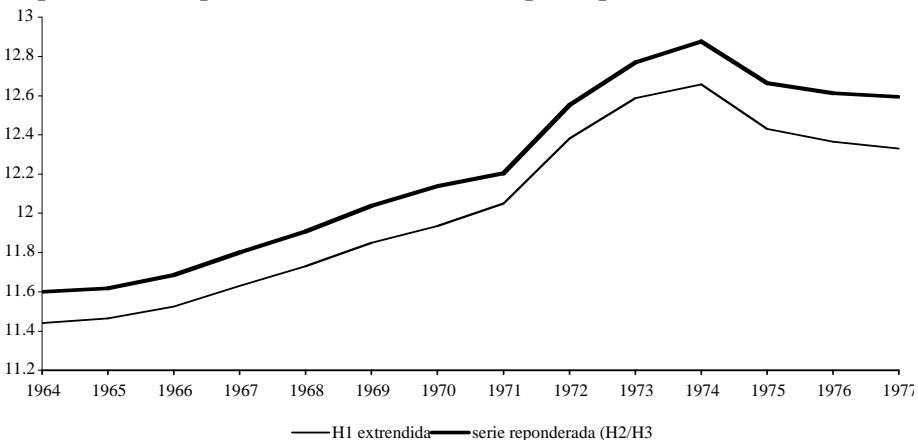
tiene en cuenta las renovaciones menores del seccionado (*distintraprov2*) no resulta significativo en la mayor parte de las especificaciones y suele presentar además el signo “incorrecto” (véanse las ecuaciones [7] y [8]). Este resultado sugiere que las renovaciones menores del seccionado realizadas entre censos podrían no ser demasiado efectivas para combatir el envejecimiento del marco.

4.3 Extensión hacia atrás de la segunda y tercera series históricas del INE (2002 y 2005)

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 3

Reponderación aproximada de la serie de ocupados para 1964-77



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. Tanto en el caso de los ocupados como en el de los parados, la corrección se construye utilizando la ecuación [1], que es la que genera el perfil más suave y plausible (véase de la Fuente, 2015). 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 3 muestra la serie de ocupados obtenida de esta forma y la compara con la serie análoga no reponderada.

Finalmente, la población contada aparte de las series históricas *H2* y *H3* se extiende hacia atrás enlazándola por retroproyección 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 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.

Dentro de las discontinuidades presentes en las series de la EPA conviene distinguir entre las generadas por las actualizaciones del seccionado y aquellas que son resultado 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 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 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.

Tras 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.

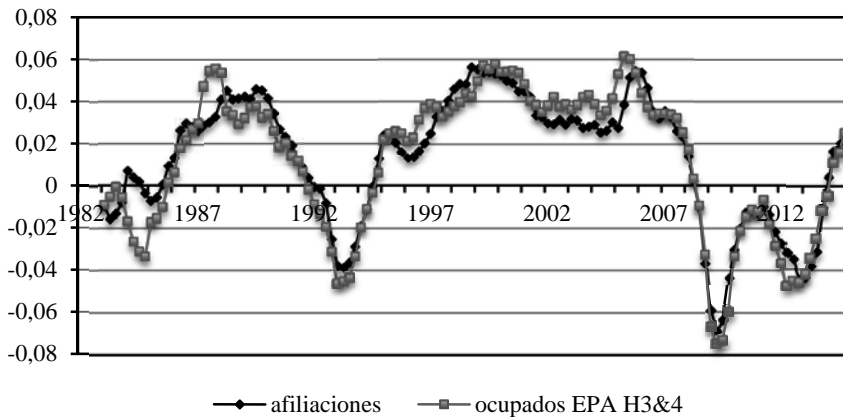
5.1 Fuentes auxiliares

Dos referencias muy útiles en un ejercicio como el que sigue son las que proporcionan la serie de afiliaciones a la Seguridad Social (MESS, 2015a)¹⁹ y la serie de ocupados de la Contabilidad Nacional Trimestral de España (CNTR, en INE, 2015b). 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.

Gráfico 4

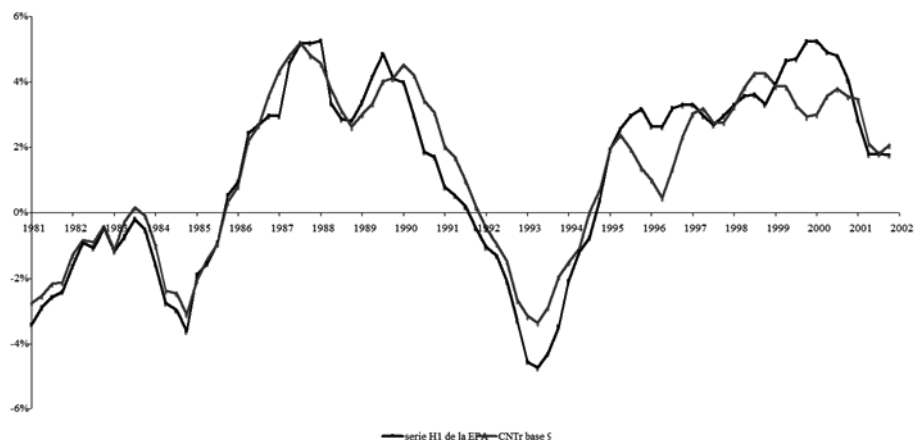
Tasas de crecimiento interanual

a. serie H3&4 de ocupados de la EPA y serie de afiliaciones a la Seguridad Social



¹⁹ 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 (2015a).

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



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 4). 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 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.

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 he empleado consiste en tomar como referencia 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.

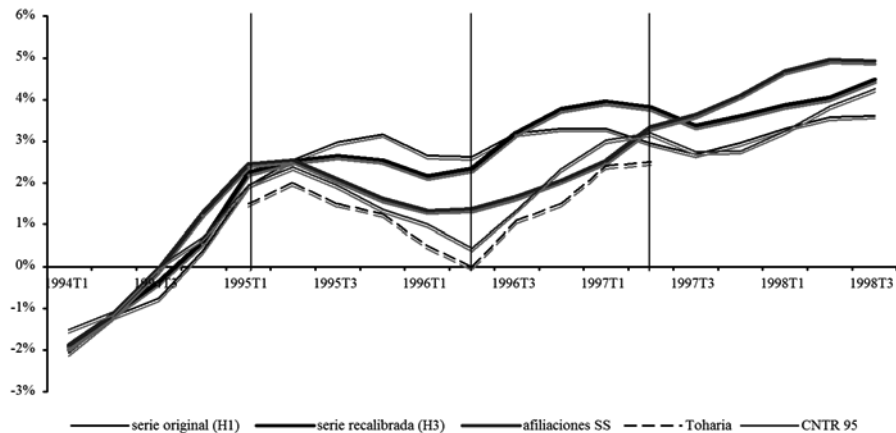
5.2 Un ejemplo: la actualización del seccionado de 1995-96

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 he utilizado para estimar el tamaño de las rupturas en las series históricas del INE 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.

El Gráfico 5 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 (1988) 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 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.

Gráfico 5

Variación interanual de los ocupados cerca de 1995, varias fuentes

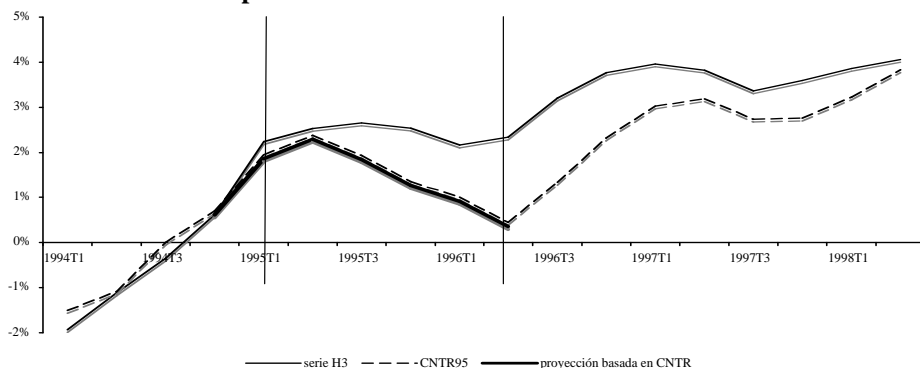


La comparación entre las distintas series que se muestran en el gráfico revela que la renovación del seccionado tuvo el efecto de “camuflar” la clara desaceleración del crecimiento del empleo que se aprecia en las series de Toharia, CNTR y afiliaciones durante 1995 y 1996. Por otra parte, la introducción 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 5, 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.

Tal como se ilustra en el Gráfico 6, 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.²⁰ 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 completada la renovación de la muestra) y la serie vieja depurada en el mismo trimestre (1996-T2). El Gráfico 7 muestra los niveles de las series relevantes y el tamaño estimado de la ruptura debida a la renovación del seccionado.

Gráfico 6

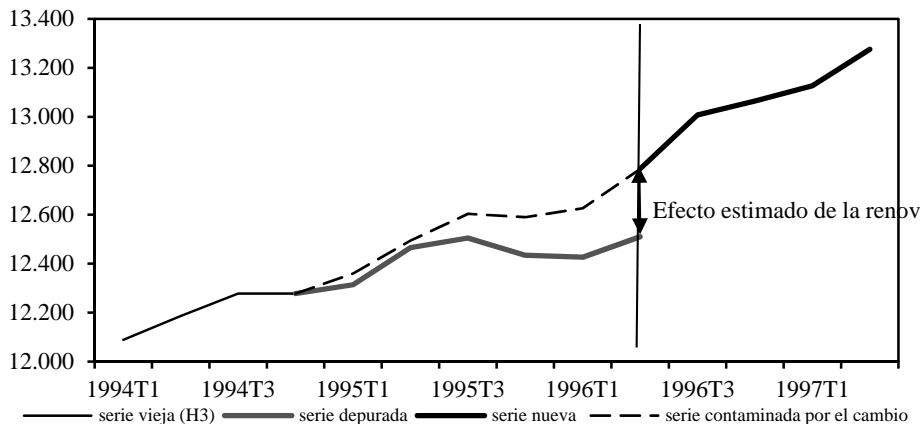
Variaciones interanuales de la serie *H3* de ocupados del INE y de la serie de CNTR95 y proyección de la variación interanual de los ocupados cerca de 1995 en base a la serie de ocupados de la Contabilidad Trimestral base 1995



²⁰ 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:

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

Gráfico 7

Series de empleo en niveles y estimación del impacto de la renovación del seccionado de 1995-96**5.3 Otras incidencias**

El Cuadro 4 identifica los cambios metodológicos que podrían generar discontinuidades en las series históricas extendidas y muestra 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. El cuadro indica también la serie externa que se toma como referencia para estimar el tamaño de la discontinuidad en la serie. El cálculo se realiza por el procedimiento ilustrado en el apartado anterior excepto en el caso de 1972, cuando he optado por interpolar la tasa de crecimiento de la propia EPA entre 1971 y 1973 para eliminar el pico que presenta esta variable en 1972 de una forma que nos deja con un perfil más plausible para la serie que el que sugieren las fuentes alternativas disponibles (véase de la Fuente 2015 para más detalles).

Cuadro 4

Estimación del tamaño de algunas discontinuidades de la serie histórica extendida del INE (serie H3&4 extendida)

<i>fecha</i>	<i>cambio metodológico</i>	<i>efecto estimado ocupados miles/dif log</i>	<i>efecto estimado parados</i>	<i>Serie de referencia</i>	<i>Otros comentarios</i>
1968T1	actualización seccionado	0			
1971S2	actualización seccionado	0			
1972S2	nuevo cuestionario	+204,6/1,64%		CNE86 (M&L, 2005)	Interpolación propia EPA
1978T2	actualización seccionado	0			
1985T1	nueva estratificación	+216,9/2,00%		afiliación	Incertidumbre fecha cambio metodológico
1985T2	actualización seccionado	0			
1987T2	nuevo cuestionario	+176,8/1,52%		afiliación	
1988T2	inclusión de Ceuta y Melilla	+35,6/0,28%	+15,0/0,57%		
1989T1	actualización seccionado				
1995T1	actualización seccionado	+277,4/2,19%		CNTR95	
1999T1	nuevo cuestionario	+113,1/0,79%		afiliación	
2000T1	actualización seccionado	0			
2002T2	nueva definición de paro		-480,1/-22,48% **		
2005T1	actualización seccionado	0			
	nueva EPA 2005	+256,0/1,38%		CNTR00	

* 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.

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, el examen de las series apunta hacia el primer trimestre, que es cuando se detecta un corte claro en la evolución del empleo.

6. Construcción de las series enlazadas de ocupados y parados

Una vez estimada la magnitud de las discontinuidades generadas por distintos cambios metodológicos en la serie más reciente de la EPA (H3&4), 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 enésima incidencia que genera una discontinuidad en algún agregado de la EPA al permitimos 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} – que ha sido estimado en la sección anterior. Así pues, estimaremos y_t^n como

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

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,

$$\hat{D}_t = \sum_n \hat{D}_t^n \quad [2]$$

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

$$\hat{Y}_t = X_t + \hat{D}_t \quad [3]$$

El procedimiento elegido para estimar d_t^n depende de la naturaleza de la incidencia cuyos efectos estemos tratando de corregir. Cuando se trata de una renovación del seccionado, supondremos 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, definiremos CIM_t , como la fracción de los movimientos migratorios totales registrados entre ambas renovaciones del seccionado que se ha acumulado en el año $t \in [T_{n-k}, T_n]$,

$$CIM_t = \frac{\sum_{s=T_n-k}^t IM_s}{\sum_{s=T_n-k}^{T_n} IM_s} \tag{4}$$

y estimamos d_t^n como

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

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 (2009)²¹ parameterizado por la vida media, H , del error de medición aflorado con el cambio metodológico. Fijando el valor de H en 10 años, tenemos

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

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

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.

²¹ 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 (2014).

6.1 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. 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 (población 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 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 de población construidas en de la Fuente (2010) a partir de las estimaciones intercensales de población del INE (2015c).

6.2 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. El INE ha construido una serie que salva la discontinuidad generada por este cambio metodológico 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. 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 8.²² Esto es, el logaritmo de la serie corregida vendrá dado por $\hat{y}_t^n = x_t + \hat{d}_t^n$ con:

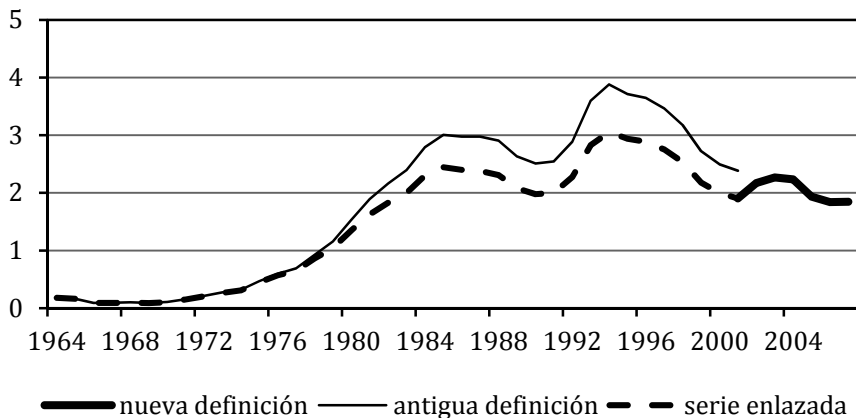
$$\hat{d}_t^n = d_{2001} * \frac{wPOPE_t}{wPOPE_{2001}} \quad \text{para} \quad 1964 \leq t \leq 2000 \quad [7]$$

²² Véase de la Fuente (2015) para más detalles. 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.

donde d_{2001} ha sido estimado por el INE (véase el Cuadro 2 y la discusión relacionada). De 2002 en adelante, se utiliza la nueva serie $H4$ de parados del INE, que ya incorpora las previsiones de población basadas en el censo de 2011.

Gráfico 8

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



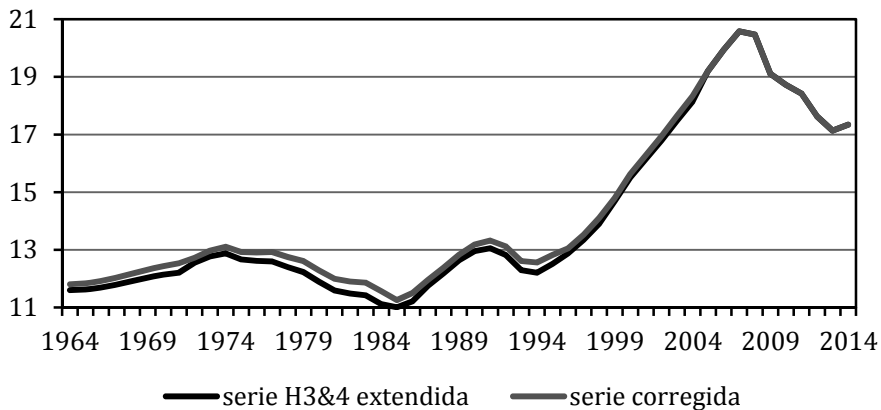
6.3 La serie enlazada

El panel *a* del Gráfico 9 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 9), 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 9

Serie corregida de ocupados vs. serie H3&4 extendida

a. niveles



b. tasas de crecimiento interanual

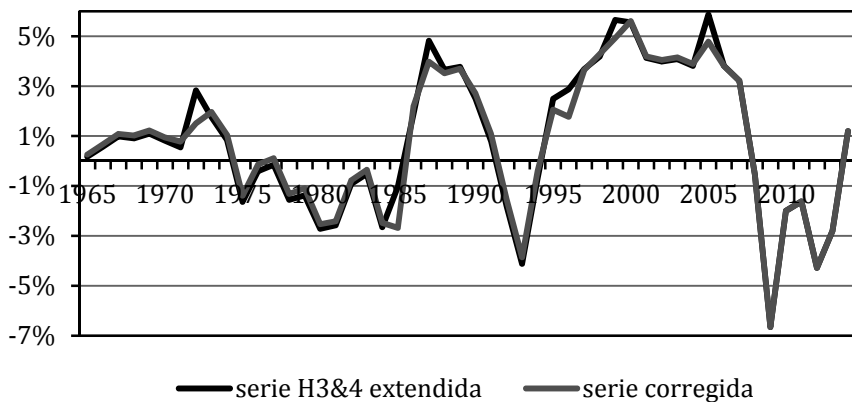
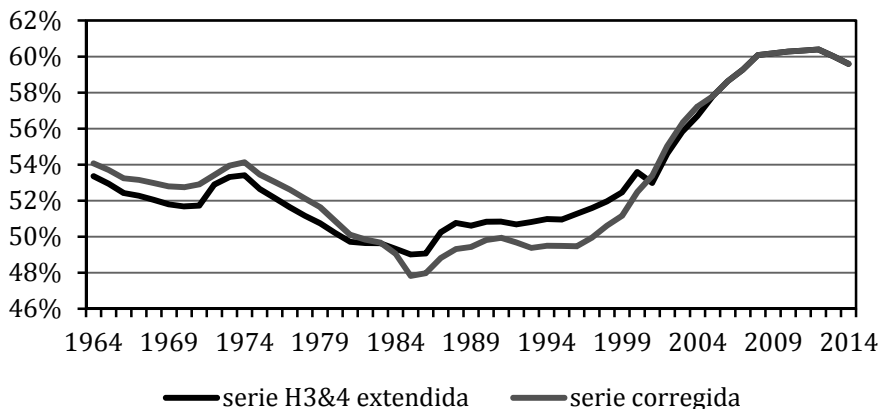


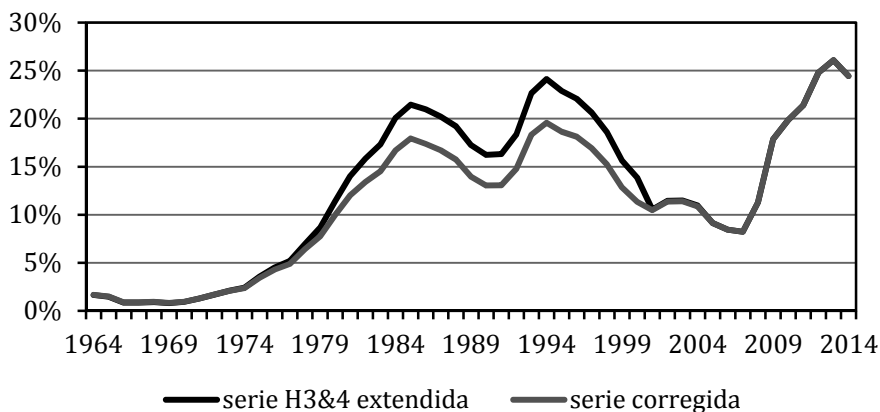
Gráfico 10

Serie corregida vs. serie H3&4 extendida

a. Tasa de actividad de la población 16+



b. Tasa de paro de la población 16+



El Gráfico 10 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. En ambos casos, el 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. También se ofrece una serie consistente con *H4* que comienza en 2008 y coincide en años comunes casi exactamente con la correspondiente serie de *H3*. Recurriendo a las publicaciones en papel de la EPA, finalmente, las dos series citadas pueden extenderse hacia atrás hasta 1964, aunque con ciertas lagunas (véase la sección 4 del Anexo 1 a de la Fuente, 2015) y posibles rupturas en 1972-S2, 1973-T3 y 1987-T2. En el resto de esta sección se construyen series homogéneas de ambas variables mediante el enlace de los distintos segmentos de las series originales.

7.1 Ocupados que declaran una segunda actividad

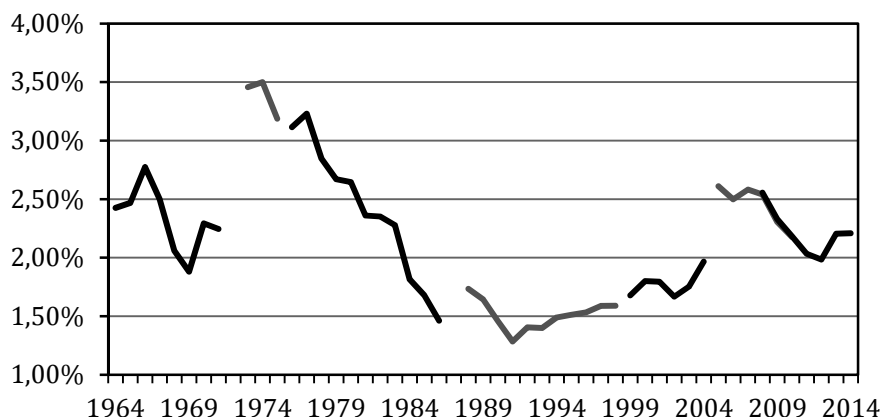
El Gráfico 11 muestra los distintos segmentos de la serie anual del porcentaje de ocupados que declara un segundo empleo (*%SEGEMP*) de acuerdo con la web del INE y con las publicaciones de la EPA en papel.²³ Como ya se ha visto, durante el primer tramo de la encuesta existen lagunas en ciertos trimestres 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).

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 por un procedimiento similar al utilizado en secciones anteriores. A falta de otras referencias, la evolución de *%SEGEMP* en ausencia del cambio metodológico se estima tomando 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 (para más detalles, véase de la Fuente 2015).

²³ 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 11

Porcentaje de los ocupados que declara un segundo empleo (%SEGEMP), series anuales



El Cuadro 5 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).

Cuadro 5

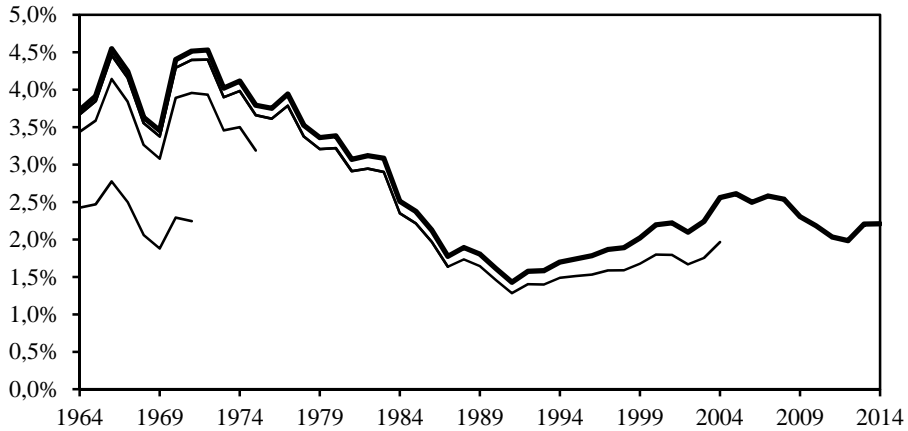
Magnitud estimada de las discontinuidades en la serie del porcentaje de ocupados que declara un segundo empleo

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

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 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 12 ilustra el procedimiento y muestra la serie final corregida, que aparece en trazo grueso.

Gráfico 12

Construcción de la serie corregida de %SEGEMP



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.

7.2 Duración media de la jornada laboral

El Gráfico 13 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.

Gráfico 13

Número medio de horas semanales trabajadas por los ocupados

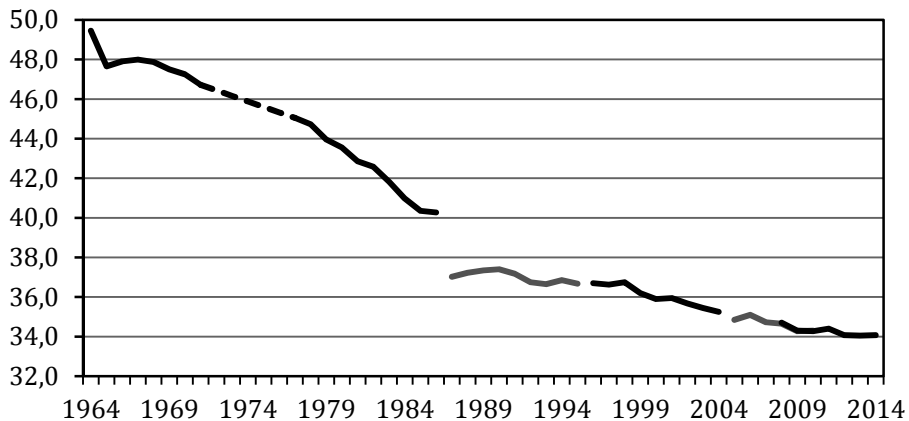
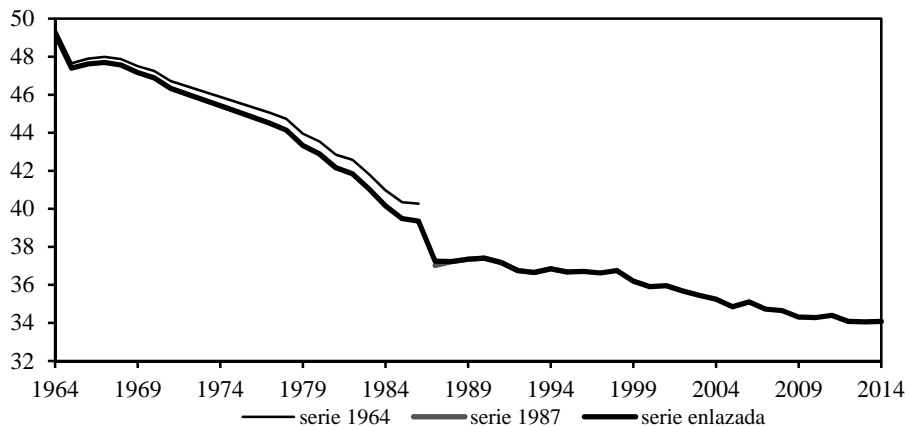


Gráfico 14

Número medio de horas semanales trabajadas por los ocupados

Serie antigua y serie enlazada



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 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 14 (y sigue mostrando un apreciable escalón a la baja en 1987 porque en el tercer trimestre de ese año se registra un número de horas atípicamente reducido pero que ya no puede ser atribuido al cambio metodológico).

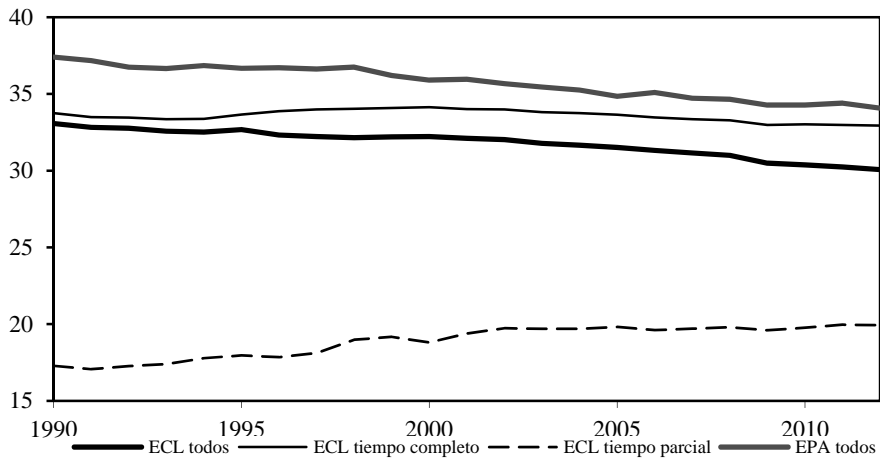
Una fuente adicional de datos sobre la duración media de la jornada laboral es la Encuesta de Coyuntura Laboral (ECL). La ECL (MESS, 2015b) es una encuesta trimestral dirigida a las empresas en la que se recoge información sobre sus efectivos laborales y vacantes, la movilidad del personal, su tiempo de trabajo y otros aspectos del mercado laboral. La encuesta se implantó en el segundo trimestre de 1990, desapareció tras 2012 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.

Gráfico 15

Número medio de horas semanales trabajadas por los ocupados ECL vs EPA



El Gráfico 15 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.

8. Conclusión

En este trabajo se han construido series largas enlazadas de algunos agregados nacionales de la EPA que cubren el período transcurrido desde la aparición de la encuesta hasta nuestros días. Las nuevas series extienden y enlazan entre sí las distintas series históricas que el INE ha construido y eliminan las discontinuidades presentes en las mismas como resultado de los sucesivos cambios en el cuestionario de la encuesta, de las renovaciones de su seccionado y de otros cambios metodológicos. El tamaño de las rupturas en las series de ocupación, que son las más afectadas por estas incidencias, se estima tomando como referencia el comportamiento de otras series relacionadas

(fundamentalmente, las series de empleo de la Contabilidad Nacional Trimestral y de afiliaciones a la Seguridad Social) para aproximar la evolución que habría tenido el empleo de la EPA en ausencia del cambio metodológico.

Las correcciones a las series existentes hasta el momento son significativas pero no cambian cualitativamente su perfil. La corrección al alza del número estimado de ocupados se sitúa por encima de los 200.000 efectivos durante la mayor parte del período muestral, alcanzando un máximo de casi 450.000 personas (un 4% del total) a mediados de los años ochenta. La homogeneización de la serie de parados, por su parte, se traduce en una significativa reducción de la tasa de paro durante los años ochenta y noventa al aplicarse retroactivamente la nueva definición, más restrictiva, de desempleo.

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. (2000). «La Encuesta de Población Activa: cómo mide el mercado laboral.» *Estadística Española* 42, 146, pp. 201-61.
- BANCO DE ESPAÑA (2002). «Cambios metodológicos en la EPA en 2002.» *Boletín Económico*, Abril de 2002, pp. 67-77. Madrid.
- 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.
- DE LA FUENTE, A. (2014). «A mixed splicing procedure for economic time series», *Estadística Española* 56, 183, pp. 107-21
- DE LA FUENTE, A. (2015). «Series enlazadas de los principales agregados nacionales de la EPA, 1964-2014.» FEDEA, *Estudios sobre Economía Española* 2015-07 <http://fedeca.net/documentos-mercado-de-trabajo/>.
- GARCÍA ESPAÑA, E. (1969). «Diseño de la Encuesta General de Población (E.G.P.).» *Estadística Española* 44, pp. 5-42.
- 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, 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, 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, 2012). *Encuesta de Población Activa. Diseño de la encuesta y evaluación de la calidad de los datos. Informe técnico. Edición revisada*. Madrid.
http://www.ine.es/dyngs/INEbase/es/operacion.htm?c=Estadistica_C&cid=1254736176918&menu=metodologia&idp=1254735976595
- INSTITUTO NACIONAL DE ESTADÍSTICA (INE, 2014). *Encuesta de Población Activa. Series revisadas 2002-13. Cambio de base poblacional en las estimaciones de la EPA. Nota metodológica*. Madrid.
http://www.ine.es/dyngs/INEbase/es/operacion.htm?c=Estadistica_C&cid=1254736176918&menu=metodologia&idp=1254735976595
- INSTITUTO NACIONAL DE ESTADÍSTICA (INE, 2015a). *Encuesta de Población Activa. En base de datos INEbase: Mercado Laboral: Actividad, ocupación y paro*. Madrid.
http://www.ine.es/dyngs/INEbase/es/operacion.htm?c=Estadistica_C&cid=1254736176918&menu=ultiDatos&idp=1254735976595
- INSTITUTO NACIONAL DE ESTADÍSTICA (INE, 2015b). *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, 2015c). *Estimaciones intercensales de población. En base de datos INEbase: Demografía y población: Cifras de población y censos demográficos*. Madrid.
http://www.ine.es/inebmenu/mnu_cifraspob.htm
- 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/>
- 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. (2006). *Las estadísticas del mercado de trabajo en España*. Ministerio de Trabajo y Asuntos Sociales. Colección Informes y Estudios Empleo, Madrid.
- PORRAS, J. (1999). «Diseño de muestras en encuestas de población y hogares.» *Qüestió* 23(3), pp. 543-58.
- 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_pdftot.pdf