



# La MCVL como valor añadido a las estadísticas laborales. Explotación transversal, retrospectiva y longitudinal

Instituto de Estadística y Cartografía de Andalucía

Isabel Padilla Sánchez  
Joaquín Planelles Romero

## Resumen:

El mercado laboral está inmerso en un rápido proceso de transformación hacia nuevas formas de organización del trabajo (economía colaborativa, automatización de funciones, digitalización) más flexibles y alejadas del concepto tradicional de empleo. Están apareciendo nuevos oficios y profesiones, y también nuevos modelos organizativos. La producción estadística debe adaptarse a esta nueva realidad.

En esta ponencia se presenta el trabajo que se está realizando desde el Instituto de Estadística y Cartografía de Andalucía con la Muestra Continua de Vidas Laborales (MCVL), una fuente estadística que permite informar sobre la estabilidad del empleo y las trayectorias laborales/salariales (algunas de las nuevas necesidades de información).

Se presentan en primer lugar las explotaciones transversal y retrospectiva que ya están disponibles en la web del Instituto de Estadística y Cartografía de Andalucía. En la actualidad se está trabajando en la ampliación de la tabulación, con el fin de profundizar en el análisis longitudinal de las trayectorias profesionales, atendiendo a los procesos de incorporación, supervivencia y salida del mercado laboral de los andaluces.

Palabras clave: MCVL, Seguridad Social, Trayectorias Profesionales

## **1. Introducción**

Los datos de la Muestra Continua de Vidas Laborales proceden de una selección aleatoria de los afiliados y/o pensionistas de la Seguridad Social en el año de referencia. Para ellos, la Seguridad Social incorpora información relativa a cada episodio de cotización y de cobro de prestaciones. A estos datos se les añaden otros procedentes del Padrón Continuo Municipal (INE) y del resumen anual de retenciones e ingresos a cuenta del IRPF de la Agencia Tributaria.

El proceso de selección da lugar a una muestra de 1,2 millones de personas para España y de unas 200 mil para Andalucía, con la particularidad de que para cada una de ellas se incorpora todo su historial de relaciones con la Seguridad Social, no sólo las del periodo de referencia, lo que permite analizar la situación del mercado de trabajo en un momento del tiempo, así como las trayectorias laborales previas.

Además del gran número de individuos que componen la muestra, el rango temático que abarca es también muy amplio. En las distintas tablas en las que se estructura la información se recogen casi 200 campos que permiten realizar un análisis minucioso sobre las características del mercado laboral español. Ahora bien, debido precisamente al gran volumen de información (la edición de cada año contiene más de mil millones de datos) así como a la procedencia administrativa de la información, los requerimientos informáticos y los procesos de verificación y control son complejos.

El análisis de las trayectorias laborales permite profundizar en aspectos tales como el número de episodios que ha tenido cada persona y la duración temporal de los mismos, la permanencia en la empresa o en la situación de desempleo, la movilidad sectorial y ocupacional, así como la trayectoria salarial asociada.

La ponencia se estructura en 7 apartados. Para empezar, en el apartado 2, se realiza una descripción más detallada de la estructura de la MCVL y de cómo tratar la dimensión temporal para los distintos tipos de análisis (transversal, retrospectivo y longitudinal). En los apartados 3 y 4 se abordarán respectivamente las explotaciones transversal y retrospectiva de la MCVL que ya están disponibles en la web del IECA. En el apartado 5 las posibles extensiones a la información actualmente disponible en la web: explotación longitudinal y monografías dedicadas a determinados colectivos. El documento termina con las conclusiones y con un anexo metodológico en el que se tratan algunas vicisitudes en la selección de registros y tratamiento de los datos así como las soluciones que hemos adoptado.

## **2. La estructura de la MCVL y la dimensión temporal**

Anualmente la Seguridad Social elabora dos versiones de la MCVL: la primera, que difunde en torno al mes de julio del año siguiente al de referencia, se denomina versión "sin datos fiscales". La elabora la Seguridad Social, incluyendo para las personas seleccionadas información procedente de los registros de la Seguridad Social y del Padrón Continuo.

La segunda versión o "versión con datos fiscales" se difunde unos meses más tarde. La elabora la Agencia Tributaria a partir de las tablas sin datos fiscales. La AEAT cambia los identificadores anonimizados de los individuos seleccionados y añade información fiscal que procede del Modelo 190 (resumen anual de retenciones e ingresos a cuenta). Esta última versión, que está disponible desde el año 2006, es la que se explota en el IECA.

La información se organiza en 6 tablas, que se pueden cruzar entre ellas utilizando el identificador anonimizado de la persona (que aparece en todas ellas) y, de ser necesario, otros campos comunes.

Tablas con un único registro por persona<sup>1</sup>:

- **1 - Tabla Personas:** contiene información relativa a las personas seleccionadas que procede de la Seguridad Social y del Padrón Municipal, como fecha y lugar de nacimiento, lugar de residencia, sexo, nivel de estudios...
- **2 - Tabla Convivientes:** contiene para cada persona seleccionada el sexo y la fecha de nacimiento de las personas registradas con ella en la hoja padronal.

Tablas con uno o más registros por persona:

- **3 - Tabla Afiliación:** muestra los datos esenciales de cada episodio de cotización a la Seguridad Social que el individuo ha mantenido a lo largo de su vida. Para cada episodio se dispone de la fecha de alta y la fecha de baja e información sobre características del mismo: tipo de contrato, grupo de cotización, coeficiente de parcialidad, etc.
- **4 - Tabla Pensiones:** muestra las características fundamentales de las pensiones contributivas que perciben, o han percibido, en el pasado los individuos seleccionados.
- **5 - Tabla Bases de Cotización:** refleja el importe mensual de las bases de cotización de las personas en la MCVL.
- **6 - Tabla Datos Fiscales:** Los datos proceden del Modelo 190 y contiene información individualizada sobre las retribuciones satisfechas y las retenciones practicadas por IRPF a las personas incluidas en la MCVL durante el año de referencia, así como algunos datos relativos a su situación familiar cuando son necesarios para la aplicación de reducciones o beneficios fiscales.

El proceso de selección de la muestra y la información que se incorpora en las tablas anteriores para cada individuo seleccionado, presenta ciertas particularidades que proporcionan a la MCVL una gran ductilidad en cuanto a los distintos tipos de análisis que se pueden realizar con ella, en concreto en lo relativo a la perspectiva temporal.

Una primera particularidad de la MCVL se refiere a la propia definición de la población de referencia, pues la información que contiene no está referida a un momento del tiempo, sino a todas las personas que han tenido relación con la seguridad social a lo largo del año, por breve que haya sido

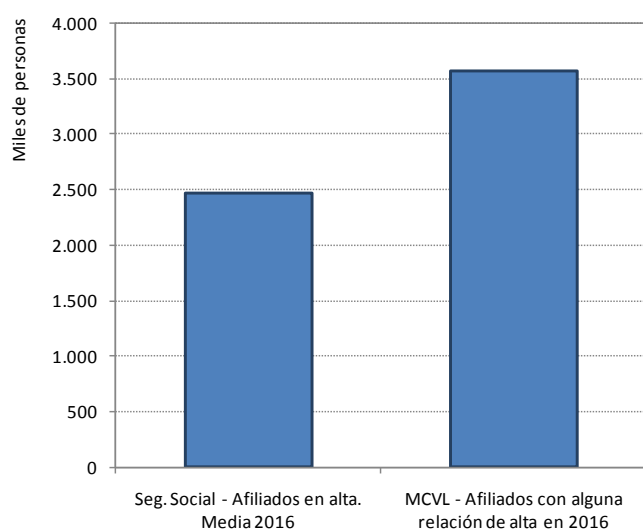
---

<sup>1</sup> En el anexo metodológico se realizan algunos comentarios respecto a la aparición de duplicados en el fichero de personas y también en los ficheros de afiliados y pensiones.

el periodo. Esto permite conocer mejor la relación con el mercado de trabajo de determinados colectivos, que entran y salen del mismo y cuyo volumen queda sólo parcialmente recogido en las estadísticas laborales habituales, referidas a una fecha concreta o a una semana de referencia.

Un ejemplo claro de un colectivo que entra y sale del mercado de trabajo son los jóvenes (figura 1), tanto más cuanto más jóvenes sean. En el año 2016 el número medio de jóvenes menores de 30 años que se encontraban en situación de alta laboral en España fue ligeramente inferior a 2,5 millones de personas. Una información complementaria a la anterior y que podemos elaborar con la MCVL es el número de jóvenes que en algún momento del año han estado en alta, 1,1 millones más.

Figura 1. Jóvenes (<30 años) afiliados en alta laboral. España. Año 2016



Una segunda particularidad tiene que ver con la disponibilidad de información referente a cada episodio de cotización y cobro de prestaciones contributivas, así como de los episodios de retención fiscal, incluyendo las fechas de alta y de baja de dichos episodios. Esta característica permite por ejemplo estudiar la duración de los episodios dentro del año (número de días cotizados en el año o número de días cobrando una prestación/subsidio por desempleo).

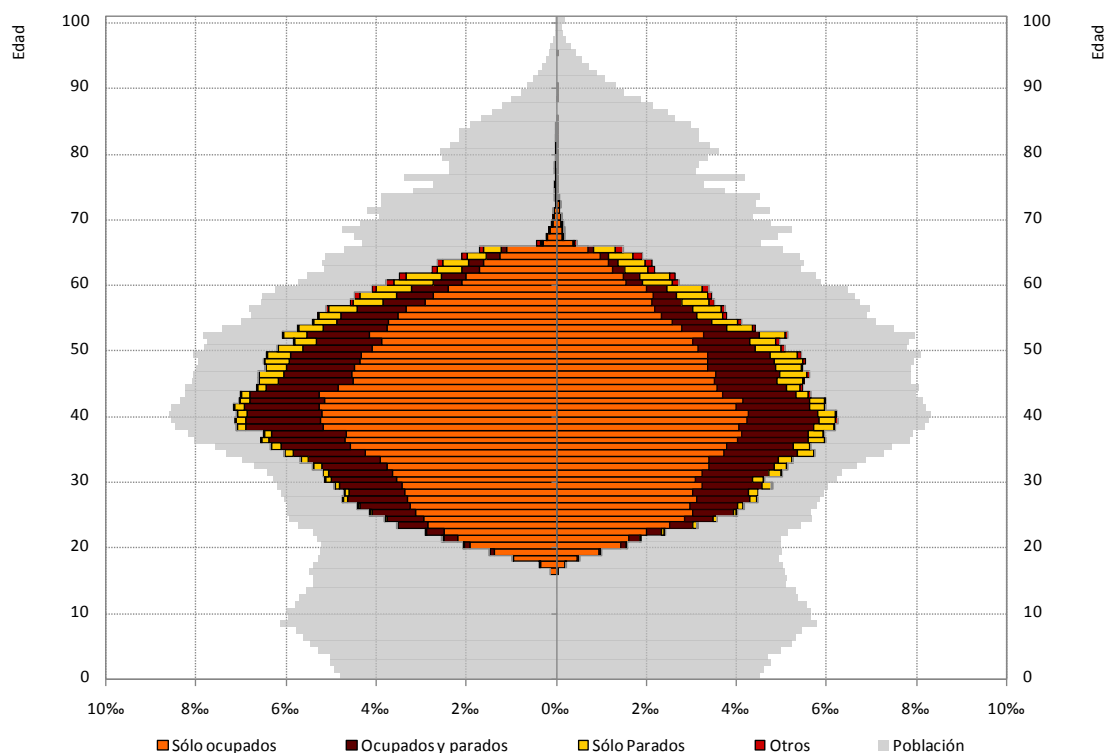
Disponer de las fechas de alta y baja (y con ellas de las duraciones) de cada episodio también posibilita estimar qué parte del año pasan los individuos en cada situación. Por ejemplo, atendiendo al tipo de cotización (figura 2), encontramos que en el año 2016 en Andalucía hubo cotizantes que sólo tuvieron relación laboral de ocupación (66%) o sólo relación de desempleo (7%), mientras que el 23% de cotizantes andaluces tuvo relaciones de ocupación y de desempleo a lo largo del año<sup>2</sup>.

De algún modo, la presencia de las fechas de alta y baja de cada episodio, permite superar las habituales representaciones dicotómicas/binarias de la realidad y atender a la proporción de tiempo

<sup>2</sup> La diferencia hasta 100% se debe a situaciones en las que la cotización no se considera ni ocupación ni desempleo, que se han denominado en la publicación "otra situación". Se trata principalmente de Convenios Especiales con la Seguridad Social.

que pasa un individuo en cada situación. Frente al cero y el uno, podemos interponer ahora las innumerables coordenadas posibles de números racionales dentro del intervalo unidad<sup>3</sup>.

Figura 2. Pirámide de población. Cotizantes según tipo de cotización. Andalucía. Año 2016<sup>4</sup>



Fuentes: Instituto de Estadística y Cartografía de Andalucía. Explotación de la MCVL, Cifras de Población, MNP y EVR.

Una tercera particularidad de la dimensión temporal es que se incluye todo el historial de relaciones con la Seguridad Social para las personas seleccionadas y no sólo las del periodo de referencia. Esto permite realizar estudios retrospectivos, como el que se ha elaborado en el IECA, que se detallará en el apartado 4.

Para realizar este tipo de análisis hay que tener en cuenta que el esfuerzo por informatizar los registros disponibles en la Seguridad Social empezó en los años 60 y para la mayoría de individuos se conservan todos los datos de afiliación desde 1967. No obstante, para algunas personas faltan las relaciones laborales más antiguas. A partir de mediados de los años 70 las omisiones son muy infrecuentes y se puede tener cierta seguridad de que los registros de la MCVL recogen todo el historial de cotizaciones de la persona. Por ello, los datos de la MCVL ofrecen información relativa al

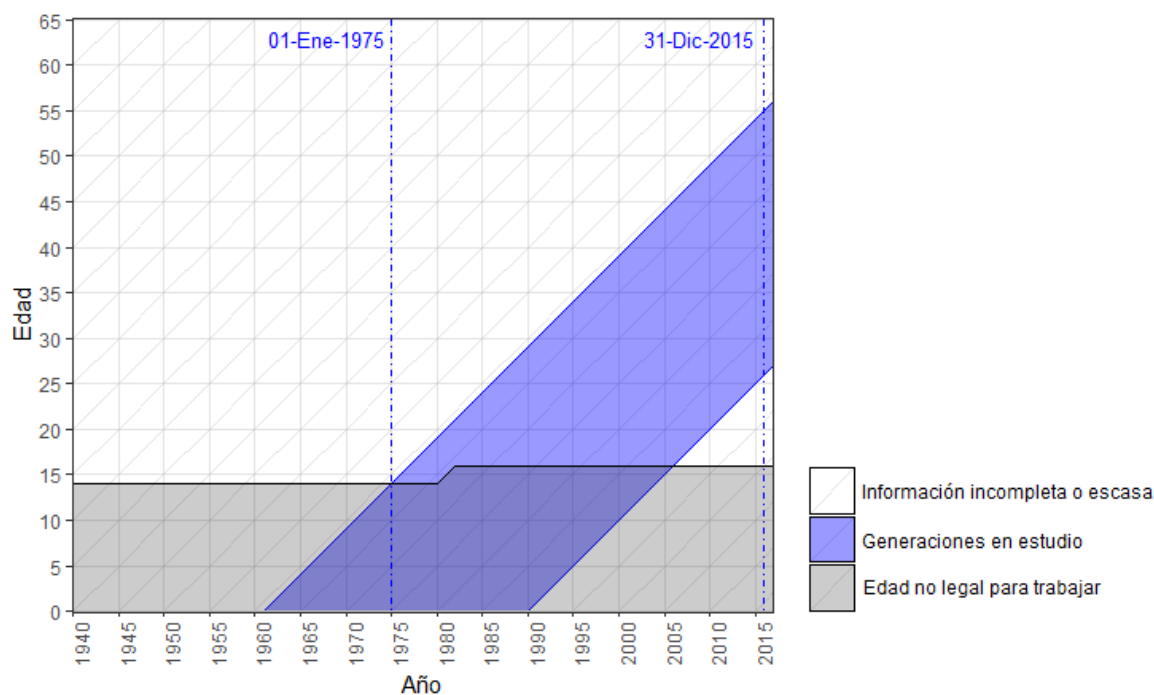
<sup>3</sup> Esta misma versatilidad de la MCVL implica cierta complejidad. Hay que definir bien a qué colectivo nos referimos en cada caso y cómo se definen las distintas categorías de las variables que se vayan a explotar.

<sup>4</sup> Se ha añadido la población para tener una idea del grado de cobertura de los cotizantes, por edad. Los datos de población se han elaborado a partir de las cifras de población a 1/1/2017 del INE, a las que se han añadido las defunciones (MNP) y emigraciones (EVR) de 2016, para obtener el conjunto de personas que en algún momento del año fueron residentes en Andalucía.

conjunto de la vida laboral (desde su inicio) para las generaciones posteriores a 1960, que cumplen la edad legal de cotización a mediados de los años 70, figura 3.

En nuestro caso se ha optado por proporcionar indicadores relativos a la vida laboral de las generaciones de 1960 a 1990, pues las generaciones posteriores a esa fecha presentan problemas de representatividad derivados de un reducido tamaño muestral (un número relativamente pequeño de ellos se ha incorporado en al mercado laboral).

Figura 3. Diagrama de Lexis de la vida laboral para las generaciones 1960-1990



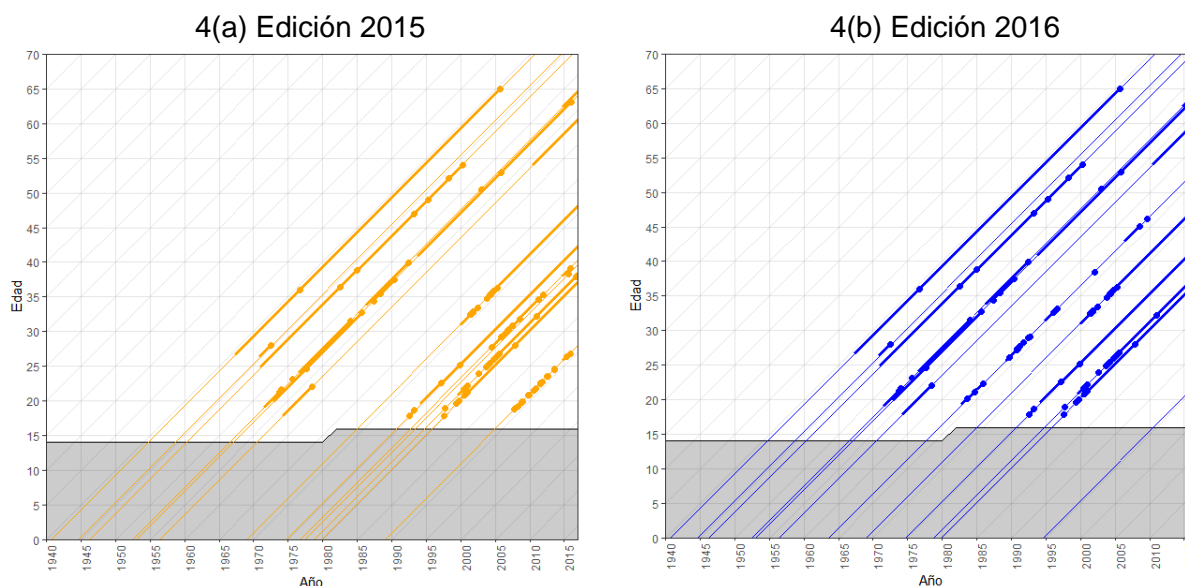
La cuarta y última particularidad es la que nos permite realizar análisis longitudinal con la MCVL (supervivencias y duraciones más allá de los límites del año natural). El diseño muestral se realiza de modo que una persona que es seleccionada para la muestra un determinado año, seguirá formando parte de la misma en el futuro, lo que permite tratar la MCVL como un panel.

Esta característica deriva del método de selección, que se describirá a continuación. En la primera edición de la MCVL se seleccionó un 4% de la población que ese año (2004) tuvo relación con la Seguridad Social. Para ello, se acudió al identificador real de la persona, para el 90% de la población objetivo el DNI, para el 9% el NIE y para el 1% restante otros documentos. Se seleccionaron 2 de las posiciones numéricas de estos identificadores (estudiando que dieran lugar a una selección aleatoria y representativa, equivalente a si se hiciera un muestreo aleatorio simple) y se escogieron 4 de las 100 combinaciones posibles con esas 2 posiciones. En las siguientes ediciones (2005 a 2017) se mantuvieron las mismas 4 combinaciones.

De este modo, una persona cuyo DNI contenga una de las 4 combinaciones originalmente seleccionadas, formará parte de la muestra todos los años en los que haya tenido alguna relación administrativa con la Seguridad Social, sea como cotizante y/o pensionista, e independientemente de lo breve que fuera la relación.

Finalmente, para realizar análisis longitudinal a partir de los datos de las distintas ediciones de la MCVL es necesario crear un fichero único a partir de la información contenida en las distintas ediciones. En la figura 4 se representa cómo se generaría un fichero único de historiales de cotización a partir de las ediciones de la MCVL de 2015 y 2016<sup>5</sup>. La mayor parte de la información que contienen las distintas ediciones es redundante (por ejemplo, porque hace referencia a episodios de cotización ocurridos muchos años antes). Por ello, el fichero único debe partir de la información de la última edición disponible (2016 en el ejemplo, en color azul). A esta selección habría que incorporarle la información de personas que formaron parte de la edición anterior (personas que fueron cotizantes y/o pensionistas en 2015), pero no de la edición de 2016. Este proceso seguiría sucesivamente, añadiendo información de ediciones anteriores, hasta llegar al año que se haya marcado como año de inicio del análisis longitudinal. Como la MCVL se empezó a realizar en 2004 (2006 en el caso de la versión con datos fiscales), el trabajo longitudinal que se puede realizar con la MCVL no puede comenzar antes de ese año.

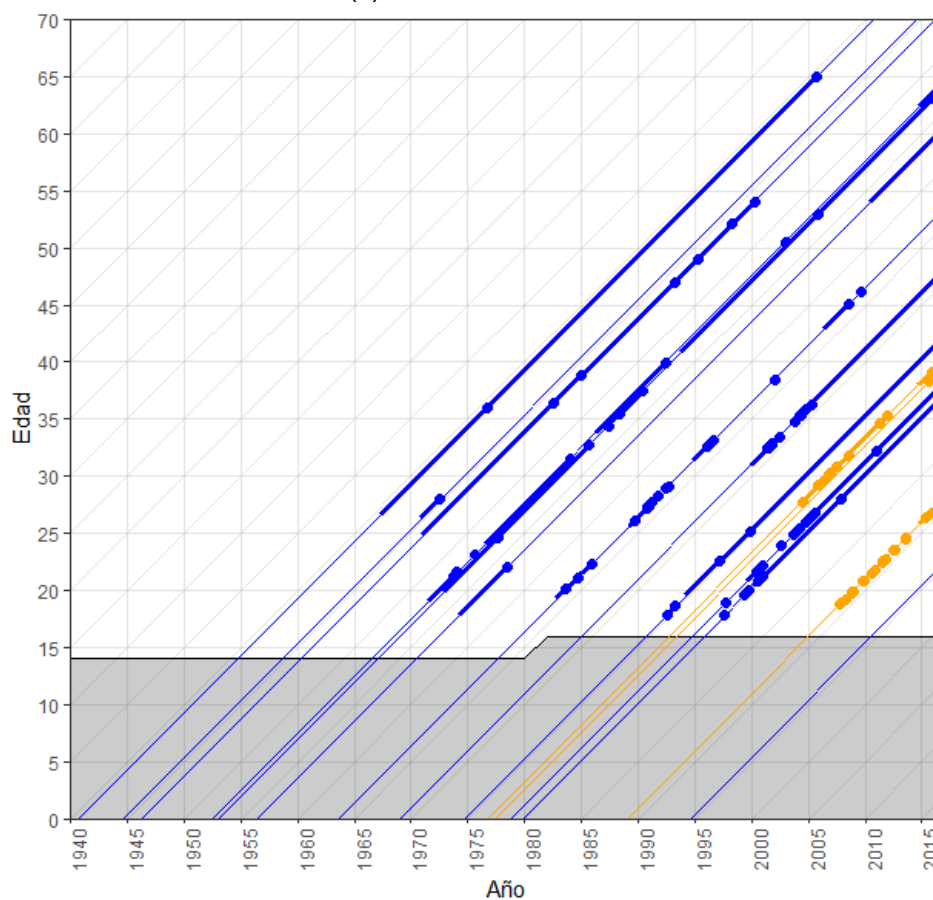
Figura 4. Creación de un fichero único para los años 2015 y 2016 a partir de las ediciones anuales



<sup>5</sup> Las líneas gruesas representan los periodos cotizados y los puntos el final de cada episodio de cotización.

Cada edición de la MCVL contiene información de más de 1 millón de individuos. Para que el gráfico resulte legible, se ha seleccionado uno de cada cien mil individuos de la MCVL. Para ello, se han escogido los individuos cuyo identificador anonimizado termina en una combinación concreta de los últimos 5 dígitos numéricos. La combinación escogida se ha realizado de acuerdo al resultado del último sorteo de Navidad.

4(c) Fichero único 2015-2016



### 3. Explotación transversal

El primer paso consiste en definir la población de referencia. En la tabulación elaborada en el IECA se han utilizado dos poblaciones de referencia distintas:

- La población residente y/o cotizante en Andalucía, para la que se han elaborado tres tablas únicamente.
- La población residente en Andalucía, para la que se ha hecho una tabulación mucho más detallada. Para sistematizar la tabulación, se han considerado cuatro variables principales: sexo, edad, ámbito territorial y tipo de relación con la seguridad social. Estas cuatro variables se han considerado transversales y se han cruzado en cada tabla con una quinta variable, específica para cada tabla. De este modo, se ha buscado que toda la tabulación tuviera un aspecto homogéneo, en lugar de realizar un diseño específico para cada tabla.



La edad se obtiene por medio de la variable fecha de nacimiento, como resultado del cálculo a 31 de diciembre del año de referencia<sup>6</sup>. En cuanto al territorio, se ofrecen datos para los siguientes ámbitos: el conjunto de Andalucía, cada una de las 8 provincias andaluzas, los municipios mayores de 100 mil habitantes (las 8 capitales de provincia más los municipios de Algeciras, Dos Hermanas, Marbella y Jerez) y finalmente, se realiza una tabulación mucho más agregada para los 35 municipios andaluces que superan los 40 mil habitantes.

Las relaciones administrativas con la Seguridad Social pueden tener una naturaleza muy distinta: asalariado que cotiza a la Seguridad Social, receptor de una prestación por desempleo ó de una pensión de jubilación, etc. Para cada una de estas relaciones la información disponible y la información de interés será distinta. Al incorporar como variable transversal el "tipo de relación con la seguridad social" se pretende precisamente distinguir colectivos para los que la información disponible o la información de interés es homogénea. En concreto, en la explotación realizada se distinguen los siguientes colectivos: (1) total (cotizantes y/o pensionistas), (2) cotizantes, (3) cotizantes con relaciones de ocupación, (4) cotizantes con relaciones de ocupación por cuenta ajena, (5) cotizantes con relaciones de desempleo, (6) pensionistas, (7) pensionistas con pensiones de jubilación, (8) pensionistas con pensiones de viudedad, (9) pensionistas con pensiones de incapacidad.

En cuanto a la variable "adicional" que se incorpora a cada tabla, se trata de un conjunto de 17 variables, que se cruzan con las anteriores.

- ❖ Algunas son variables demográficas: nacionalidad, lugar de nacimiento, número de convivientes. En futuras ediciones se pretende incorporar tipologías de hogar, a partir de los sexos y edades de los convivientes.
- ❖ Otras ayudan a caracterizar los episodios de cotización tenidos en el año: número de episodios de cotización, número de días cotizados, tipo de cotización (sólo relaciones de ocupación, sólo desempleo, ocupación y desempleo, otros).
- ❖ Otras extraen características del último episodio de ocupación ó del último episodio de ocupación por cuenta ajena que haya tenido la persona: sector de actividad, duración, grupo de cotización.
- ❖ Otras son propias de los episodios de desempleo: número de días cotizados como desempleado, tipo de prestación (sólo prestación por desempleo, sólo subsidio por desempleo, ambos).
- ❖ Finalmente, hay variables que son propias de los episodios de cobro de pensión: tipo de pensión, número de años considerados cotizados.

---

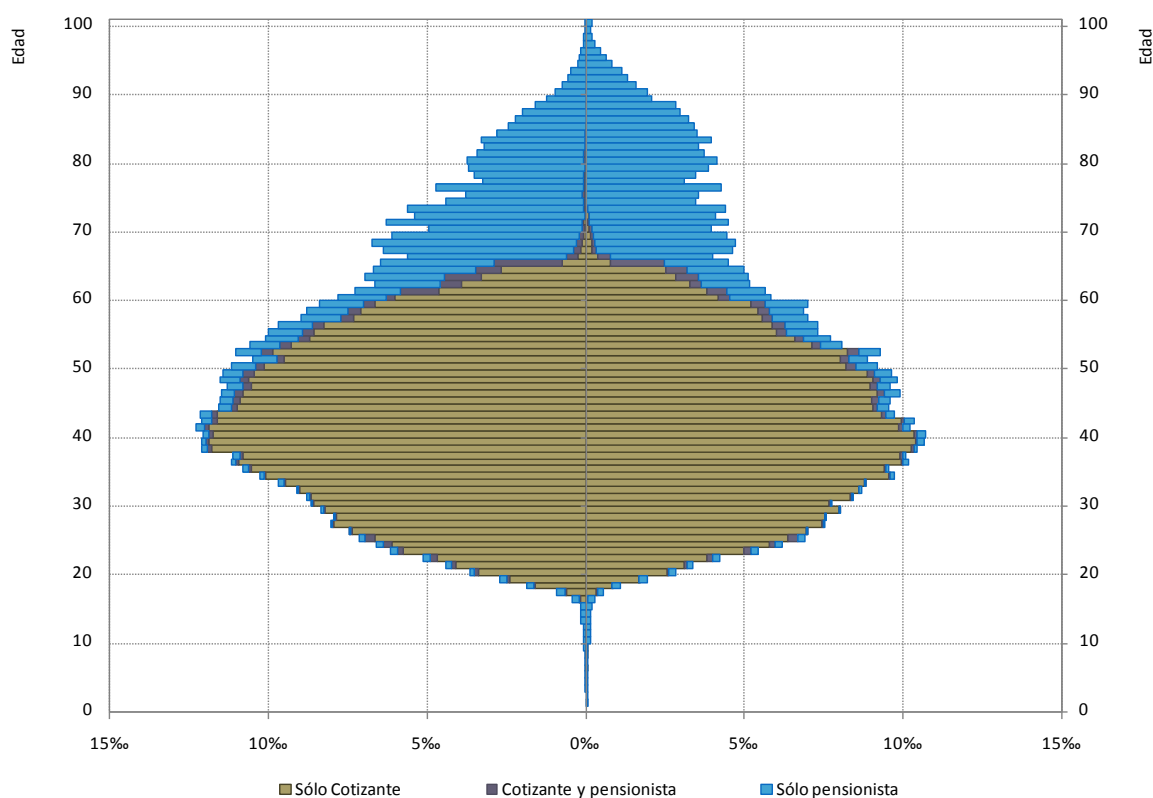
<sup>6</sup> En general se ha utilizado la edad que consta en los propios ficheros de la Seguridad Social. No obstante, cuando el dato no existía ó resultaba poco creíble (por ejemplo, al dar lugar a cotizaciones por debajo de la edad legal), se ha acudido al fichero de convivientes, donde consta la fecha de nacimiento según el Padrón Municipal.

### 3.1. Resumen de resultados

A modo de resumen de la información que se elabora y difunde con la explotación transversal y que no es habitual encontrar en otras fuentes estadísticas, se incorporan a continuación algunos comentarios, extraídos de la última edición disponible (2016), y referidos a la población residente en Andalucía.

A lo largo del año 2016 un total de 5,1 millones de andaluces mantuvieron algún tipo de relación administrativa con la Seguridad Social. En la mayor parte de los casos, la relación fue únicamente como cotizantes a la seguridad social (72% del total). Otro 25% fueron sólo pensionistas. El 3% restante fueron las dos cosas, cotizantes y pensionistas, a lo largo del año, figura 5.

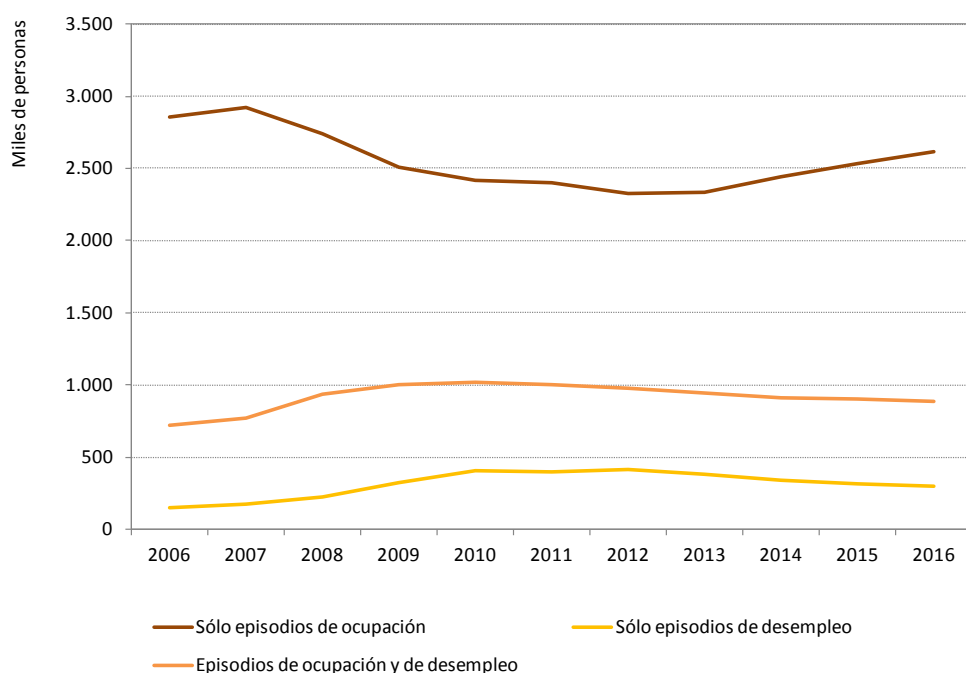
Figura 5. Pirámide de cotizantes y pensionistas en Andalucía. Año 2016



Fuente: Instituto de Estadística y Cartografía de Andalucía. Explotación de la MCVL

Del mismo modo, hay cotizantes que sólo tuvieron relación laboral de ocupación (66%) o sólo relación de desempleo (7%), mientras que 883 mil andaluces tuvieron relaciones de ocupación y de desempleo a lo largo del año (23%).

Figura 6. Evolución de los cotizantes, según tipo de cotización. Andalucía 2006-2016



Fuente: Instituto de Estadística y Cartografía de Andalucía. Explotación de la MCVL.

Este último colectivo (personas con episodios de cotización y desempleo en el año) creció con intensidad entre los años 2006 y 2010, alcanzando un máximo ligeramente superior a 1 millón de andaluces, el 26% de los cotizantes ese año y viene disminuyendo desde entonces, figura 6.

Tabla 1. Cotizantes, según características de sus episodios de cotización. Andalucía, 2006 y 2016

	Nº de personas		Peso	
	2006	2016	2006	2016
<b>Cotizantes, según nº de episodios tenidos en el año</b>				
1 único episodio	1.960.550	2.082.925	52%	54%
2-5 episodios	1.485.225	1.331.150	39%	35%
6+ episodios	314.350	417.100	8%	11%
<b>Cotizantes, según nº de días cotizados en el año</b>				
Menos de 1 mes	102.075	145.425	3%	4%
Entre 1 y 11 meses	1.195.400	1.200.375	32%	31%
12 meses	2.462.650	2.485.375	65%	65%
<b>Cotizantes según nº de días percibiendo prestación ó subsidio de desempleo</b>				
0 días	2.890.400	2.650.175	77%	69%
Menos de 1 mes (1 a 30 días)	82.475	93.400	2%	2%
Entre 1 y 11 meses	687.750	949.150	18%	25%
12 meses	99.500	138.450	3%	4%

Fuente: Instituto de Estadística y Cartografía de Andalucía. Explotación de la MCVL.

Más de la mitad de los cotizantes tuvieron un único episodio de cotización en el año, con un 11% de casos de 6 o más episodios de cotización (tabla 1), un porcentaje ligeramente superior al de una década atrás (8% en 2006).

En cuanto al periodo de cotización, el 65% de los cotizantes lo hicieron durante todo el año 2016, mientras que el 4% lo hizo por un periodo inferior a un mes.

Como ya se ha comentado, la MCVL también contiene información sobre las pensiones y sobre la población pensionista. A modo de resumen, se ha elaborado la tabla 2, que incide en otro de los aspectos que se pueden tratar con la MCVL, la brecha de género en pensiones. A lo largo del año 2016 más de 1,4 millones de andaluces percibieron alguna pensión. El 60% de los pensionistas tuvieron una pensión de jubilación (75% de los hombres y 44% de las mujeres). Por su parte, el 28% de los pensionistas percibieron pensiones de viudedad en 2016 (4% de los hombres y 54% de las mujeres). El 46% de los pensionistas residentes en Andalucía tuvieron en 2016 unos ingresos inferiores a 10 mil euros anuales (33% en los hombres y 61% en el caso de las mujeres).

Tabla 2. Pensionistas según características de sus pensiones y sexo. Andalucía, año 2016

	Nº de personas			Peso		
	Hombres	Mujeres	Total	Hombres	Mujeres	Total
<b>Pensionistas, según tipo de pensión</b>						
Sólo incapacidad	134.100	70.875	204.975	18%	10%	14%
Sólo jubilación	544.975	206.200	751.175	72%	30%	52%
Sólo viudedad	9.725	267.275	277.000	1%	39%	19%
Sólo orfandad y a favor de familiares	42.150	42.200	84.350	6%	6%	6%
Jubilación y viudedad	20.675	99.150	119.825	3%	14%	8%
Otras combinaciones	2.850	6.100	8.950	0%	1%	1%
<b>Pensionistas, según intervalo de ingresos</b>						
<10 mil euros	243.825	416.275	660.100	33%	61%	46%
10-30 mil euros	415.875	235.275	651.150	56%	34%	46%
>30 mil euros	87.075	31.700	118.775	12%	5%	8%

Fuente: Instituto de Estadística y Cartografía de Andalucía. Explotación de la MCVL.

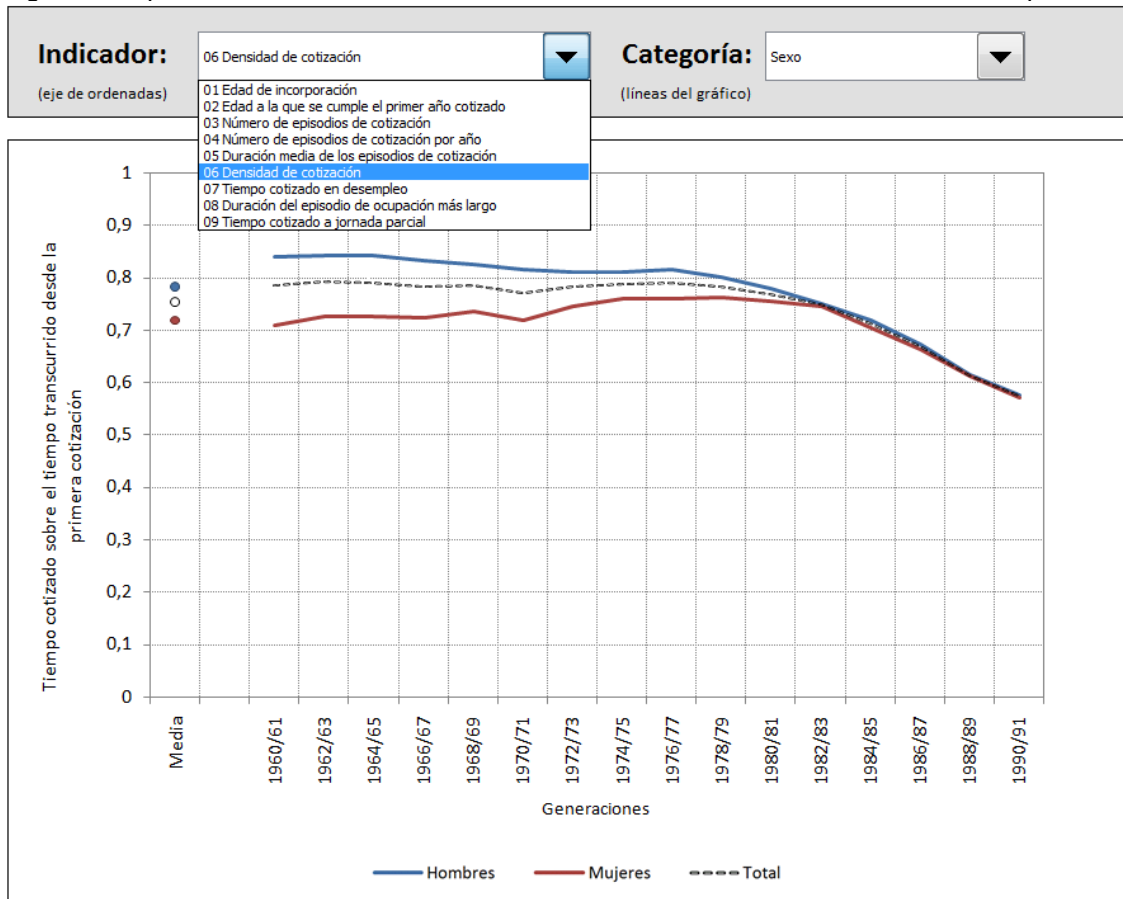
#### 4. Explotación retrospectiva

El segundo producto que se ha elaborado en el IECA a partir de la MCVL es un estudio retrospectivo que parte de la población que tuvo relación con la Seguridad Social en el último año disponible (2016 actualmente) y ofrece información relativa no solo a ese año sino a toda la vida laboral de esas personas, así como a etapas de la misma.

La publicación genera un gran volumen de datos con el fin de describir cómo han sido los itinerarios laborales de los andaluces. Se pone a disposición de los usuarios una batería de indicadores novedosos y un conjunto de variables categóricas que identifican subgrupos de población que comparten ciertas características (como el sexo o el año de nacimiento). Para visualizar los resultados, se han elaborado unos ficheros interactivos en formato Excel, que permiten seleccionar

una variable categórica y un indicador, lo que permite al usuario visualizar gráficamente los resultados, de manera ágil e interactiva, figura 7.

Figura 7. Aspecto del fichero interactivo elaborado. Indicador "densidad de cotización" por sexo



Fuente: Instituto de Estadística y Cartografía de Andalucía. Explotación de la MCVL.

También existe la posibilidad de distinguir etapas de la vida laboral. En ese caso, el grupo de generaciones para los que se puede ofrecer el dato es distinto para cada etapa. Por ejemplo, para las generaciones de los años 40 no se puede saber con los registros informáticos de la Seguridad Social cómo fue el principio de su vida laboral, pero sí cómo fue el final. Para ello, se ha dividido la vida laboral en tres etapas definidas por sendos aniversarios significativos (30, 50 y 65 años respectivamente):

Etapas 1: [16; 30)

Etapas 2: [30; 50)

Etapas 3: [50; 65)

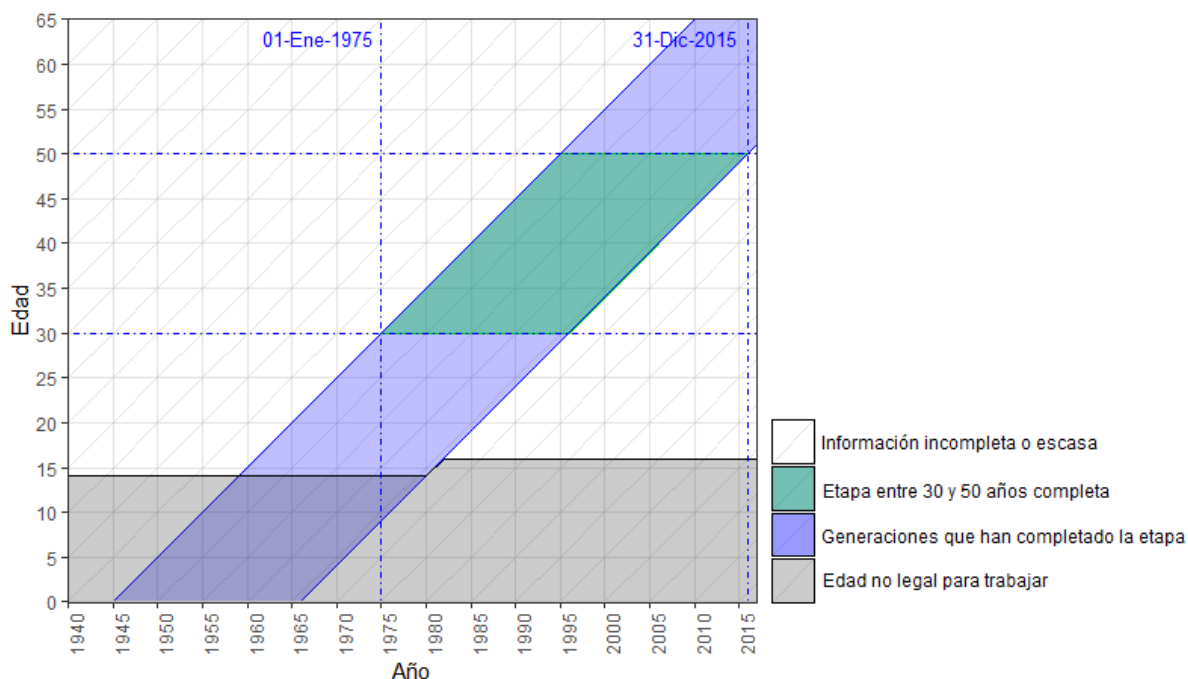
Se ha decidido ofrecer el dato para las generaciones que han completado cada etapa de observación. Para ello, la última generación para la que podemos dar el dato vendrá determinada por dos factores: la edad fijada como final de cada etapa y el último año disponible de la MCVL.

Edades [16; 30]  $\Rightarrow$  Generaciones [1975-147; 2015-30] = [1961; 1985)

Edades [30; 50]  $\Rightarrow$  Generaciones [1975-30; 2015-50] = [1945; 1965)

Edades [50; 65]  $\Rightarrow$  Generaciones [1975-50; 2015-65] = [1925; 1950)

Figura 8. Diagrama de Lexis para la segunda etapa de la vida laboral (30 a 50 años)



#### 4.1. Resumen de resultados

Las vidas laborales de las mujeres contienen un menor número de episodios de cotización que los hombres (25,8 y 30,7 episodios en media respectivamente). Además, las mujeres han pasado, en media, una proporción menor de tiempo cotizando a la Seguridad Social (72% del tiempo transcurrido desde la primera cotización ha sido tiempo cotizado, frente al 78% de los hombres). En cambio, la proporción de tiempo que se ha estado cotizando desde la situación de desempleo es mayor en mujeres (15% frente al 12% en hombres), diferencia que se va reduciendo para las generaciones más jóvenes, tabla 3.

<sup>7</sup> Desde la aprobación del Estatuto de los Trabajadores, en 1980, la edad mínima legal para acceder al mercado laboral queda fijada en 16 años (hasta entonces era 14). Este cambio afectó a las generaciones posteriores a 1966.

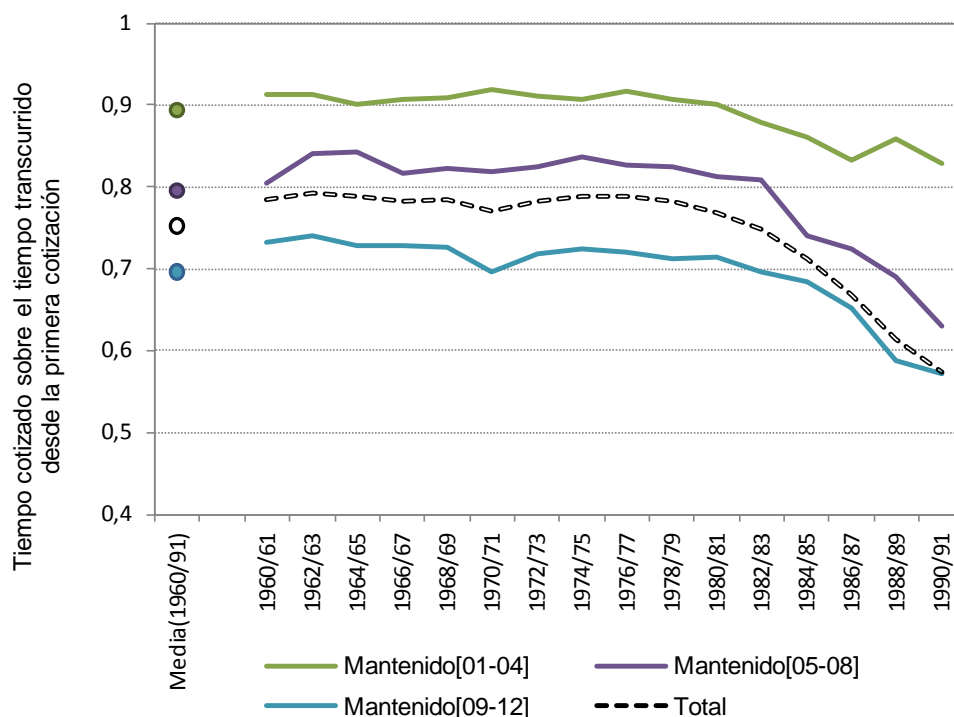
Tabla 3. Ejemplo de tabla con diversos indicadores por generaciones y variables categóricas

Indicador	Nº total episodios de cotización			Densidad de cotización <sup>a</sup>			Tiempo cotizado en desempleo <sup>b</sup>			
	1960/61	1990/91	1960/91	1960/61	1990/91	1960/91	1960/61	1990/91	1960/91	
<b>Total población</b>	<b>33,4</b>	<b>13,9</b>	<b>28,4</b>	<b>0,78</b>	<b>0,57</b>	<b>0,75</b>	<b>0,14</b>	<b>0,09</b>	<b>0,13</b>	
<b>Sexo</b>	Hombres	35,8	15,6	<b>30,7</b>	0,84	0,58	<b>0,78</b>	0,13	0,09	<b>0,12</b>
	Mujeres	30,2	12,1	<b>25,8</b>	0,71	0,57	<b>0,72</b>	0,16	0,09	<b>0,15</b>
<b>Movilidad laboral (evolución grupo de cotización)</b>	Ha empeorado	37,1	15,5	<b>30,6</b>	0,81	0,54	<b>0,77</b>	0,14	0,09	<b>0,12</b>
	Ha mejorado	31,9	12,3	<b>26,7</b>	0,70	0,50	<b>0,68</b>	0,20	0,11	<b>0,18</b>
	Mantenido [01-04]	20,2	5,5	<b>20,1</b>	0,91	0,83	<b>0,89</b>	0,02	0,02	<b>0,03</b>
	Mantenido [05-08]	18,7	9,5	<b>18,4</b>	0,80	0,63	<b>0,80</b>	0,09	0,04	<b>0,09</b>
	Mantenido [09-12]	37,0	16,1	<b>32,3</b>	0,73	0,57	<b>0,70</b>	0,18	0,10	<b>0,17</b>
<b>Intervalo ingresos</b>	<10 mil euros	41,8	14,2	<b>32,2</b>	0,71	0,50	<b>0,66</b>	0,23	0,10	<b>0,20</b>
	10-20 mil euros	39,0	14,5	<b>30,1</b>	0,78	0,71	<b>0,80</b>	0,14	0,07	<b>0,12</b>
	20-30 mil euros	29,1	11,7	<b>25,4</b>	0,87	0,74	<b>0,86</b>	0,07	0,03	<b>0,07</b>
	>30 mil euros	20,4	16,0	<b>23,0</b>	0,91	0,74	<b>0,89</b>	0,03	0,03	<b>0,04</b>

Fuente: Instituto de Estadística y Cartografía de Andalucía. Explotación de la Muestra Continua de Vidas Laborales con Datos Fiscales de la Seguridad Social  
 1 Proporción de tiempo cotizado sobre tiempo transcurrido desde la primera cotización, en tanto por 1  
 2 Proporción de tiempo cotizado en desempleo sobre el total de tiempo cotizado, en tanto por 1

También se ha estudiado la movilidad laboral a partir de la evolución de los grupos de cotización. Las personas que se han mantenido en los grupos de cotización más altos (01-04) han cotizado durante un 89% de su vida laboral, cifra superior a la de las personas que se han mantenido en los grupos bajos (09-12) que es del 70%, figura 9.

Figura 9. Densidad de cotización según evolución del grupo de cotización.

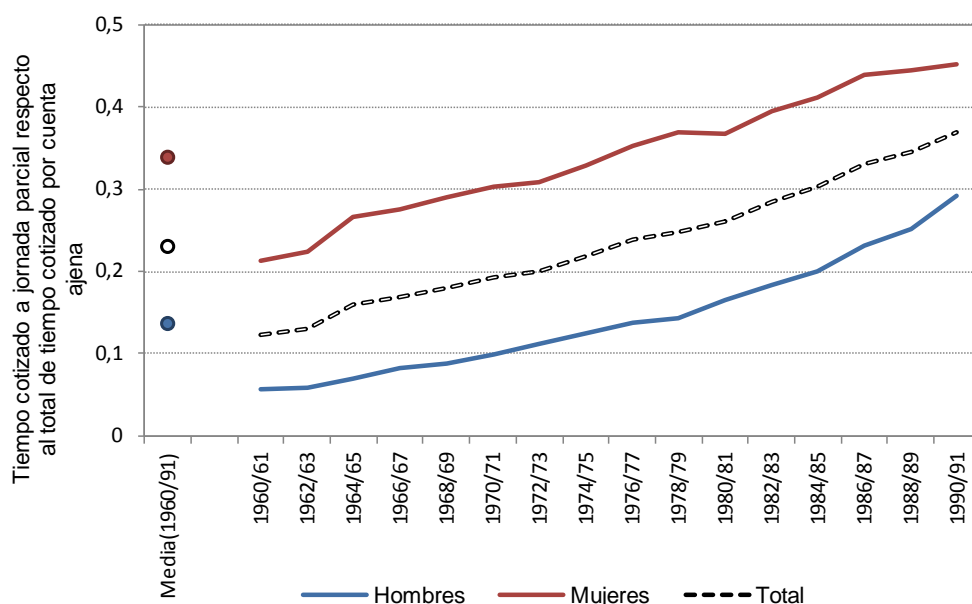


Fuente: Instituto de Estadística y Cartografía de Andalucía. Explotación de la MCVL.

Las personas que se han mantenido en grupos de cotización altos han tenido también vidas laborales con escasa presencia de desempleo (3% de media), mientras que ese porcentaje llega hasta el 17% para los que se han mantenido en grupos de cotización más bajos.

También se observa un incremento de las jornadas a tiempo parcial en las generaciones más jóvenes. Los episodios de cotización a tiempo parcial suponen el 12% de la vida laboral de la personas nacidas en 1960. En cambio, representan el 37% en la generación de 1990, figura 10.

Figura 10. Proporción de tiempo cotizado a jornada parcial según sexo. Media por generación y total.

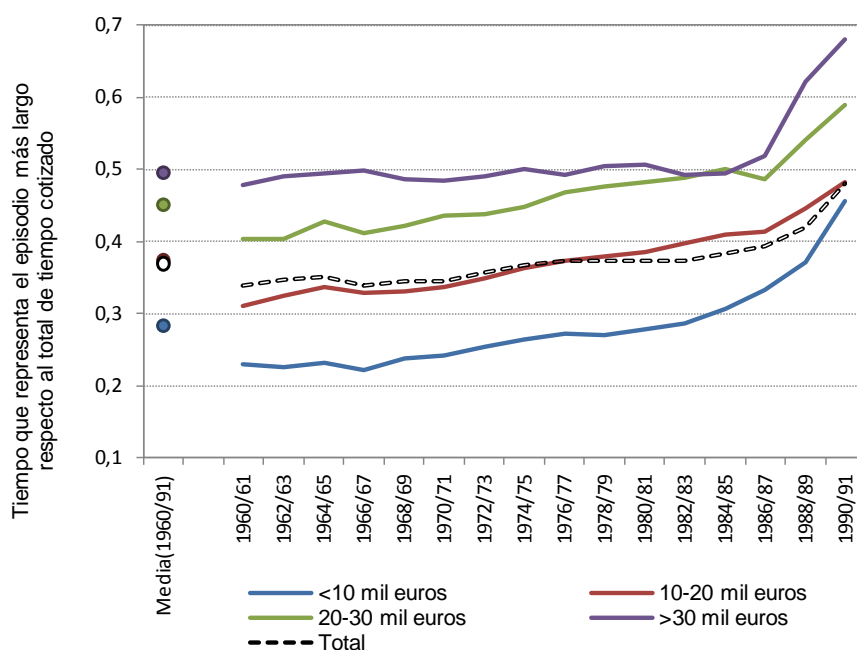


Fuente: Instituto de Estadística y Cartografía de Andalucía. Explotación de la MCVL.

La estabilidad de la trayectoria laboral también incide (o como mínimo muestra correlación positiva) en el nivel de ingresos que la persona acaba teniendo. El episodio de ocupación más largo de la vida laboral supone el 29% del total de tiempo cotizado para las personas con unos ingresos anuales inferiores a 10 mil euros, mientras que para las personas con ingresos superiores a 30 mil euros, el episodio más largo dura la mitad de toda la vida laboral, figura 11.



Figura 11. Duración del episodio de ocupación más largo según intervalo de ingresos en 2016. Media por generación y total.



Fuente: Instituto de Estadística y Cartografía de Andalucía. Explotación de la MCVL.

## 5. Posibles extensiones a la información actualmente disponible

En los apartados inmediatamente anteriores se han mostrado los dos productos actualmente disponibles en la página web del IECA, las explotaciones transversal y retrospectiva de la MCVL. En este apartado se tratarán las posibles extensiones a la información actualmente disponible, cosa que se está planteando en tres sentidos: (1) ampliar la tabulación existente (2) incorporar la perspectiva longitudinal y (3) realizar monografías específicas para determinados colectivos de interés.

Respecto a la ampliación de la tabulación existente, en las próximas ediciones de la MCVL se pretende incorporar como nuevo ámbito territorial de referencia el conjunto de España, lo que nos permitirá comparar los datos de Andalucía con los de su principal ámbito de referencia. Además, aunque no están previstos grandes cambios en la estructura de los productos ya existentes, se ha previsto la inclusión de alguna variable adicional<sup>8</sup>.

La explotación longitudinal permite hacer algo que no era posible con los enfoques anteriores: el estudio de supervivencias. A modo de ejemplo, la tabla 4 presenta las altas de autónomos en los años 2006 a 2016 en España. Se ofrece también una matriz triangular con las supervivencias (el número de emprendimientos que van cumpliendo determinados aniversarios significativos). Dividiendo supervivencias sobre altas, se obtienen las probabilidades de supervivencia.

<sup>8</sup> Por ejemplo, se está valorando la posibilidad de incorporar una variable relativa a la tipología de hogar a partir de la información de sexos y edades del fichero de Convivientes. Hasta ahora la única información relativa a la composición del hogar es el número de miembros.

Por ejemplo, se puede ver cómo la supervivencia a 12 meses de los emprendimientos autónomos se redujo en los primeros años de la crisis, pasando del 70% en 2006 a 54% en 2011. Desde entonces se ha mantenido en un nivel bastante estable, ligeramente por encima del 50%.

Tabla 4. Supervivencia de autónomos<sup>(a)</sup>. Residentes en España

	Año de alta										
	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016
Personas que se dan de alta en autónomos, [1]	488.325	492.000	416.425	420.500	455.875	494.825	533.400	601.825	648.775	606.500	593.475
Nº de altas de autónomos, [2]	502.550	507.325	430.700	439.850	478.475	522.300	568.300	643.900	689.850	642.875	629.500
Supervivencia, [3]											
Un año (12 meses ininterrumpidos)	353.800	350.100	279.925	252.675	273.450	283.475	295.375	329.075	367.100	342.275	
Dos años (24 meses ininterrumpidos)	285.500	253.725	212.300	193.250	207.600	215.975	226.625	252.275	275.700		
Tres años (36 meses ininterrumpidos)	225.300	203.950	174.250	157.650	170.825	179.550	191.500	209.075			
Cuatro años (48 meses ininterrumpidos)	187.875	170.800	143.550	131.375	144.350	155.500	167.225				
Cinco años (60 meses ininterrumpidos)	162.425	146.675	125.075	114.750	127.225	138.600					
Seis años (72 meses ininterrumpidos)	142.575	126.450	111.250	102.625	114.875						
Siete años (84 meses ininterrumpidos)	125.550	113.900	102.550	94.075							
Ocho años (96 meses ininterrumpidos)	114.375	105.200	94.525								
Nueve años (108 meses ininterrumpidos)	106.325	97.650									
Diez años (120 meses ininterrumpidos)	99.400										
Probabilidades de supervivencia de autónomos (%), [3]/[1]											
Un año	72%	71%	67%	60%	60%	57%	55%	55%	57%	56%	
Dos años	58%	52%	51%	46%	46%	44%	42%	42%	42%		
Tres años	46%	41%	42%	37%	37%	36%	36%	35%			
Cuatro años	38%	35%	34%	31%	32%	31%	31%				
Cinco años	33%	30%	30%	27%	28%	28%					
Seis años	29%	26%	27%	24%	25%						
Siete años	26%	23%	25%	22%							
Ocho años	23%	21%	23%								
Nueve años	22%	20%									
Diez años	20%										
Probabilidades de supervivencia de episodios por cuenta propia (%), [3]/[2]											
Un año	70%	69%	65%	57%	57%	54%	52%	51%	53%	53%	
Dos años	57%	50%	49%	44%	43%	41%	40%	39%	40%		
Tres años	45%	40%	40%	36%	36%	34%	34%	32%			
Cuatro años	37%	34%	33%	30%	30%	30%	29%				
Cinco años	32%	29%	29%	26%	27%	27%					
Seis años	28%	25%	26%	23%	24%						
Siete años	25%	22%	24%	21%							
Ocho años	23%	21%	22%								
Nueve años	21%	19%									
Diez años	20%										

Fuente: Instituto de Estadística y Cartografía de Andalucía. Explotación de la Muestra Continua de Vidas Laborales con Datos Fiscales de la Seguridad Social (a) RETA (Régimen especial de trabajadores autónomos) y antiguo Régimen especial agrario cuenta propia (integrado en RETA en el año 2008).

Nótese que el ejemplo presentado se refiere tan sólo a un régimen de cotización. El estudio de supervivencias se puede ampliar y sistematizar para informar sobre múltiples aspectos de la situación laboral. Por ejemplo, se podría estudiar la supervivencia de episodios de ocupación según tipo de contrato o la duración de los episodios de desempleo, etc.

Por último, la idea de realizar monografías específicas para determinados colectivos de interés. Se ha planteado la posibilidad de realizar tres monografías: "Incorporación de los jóvenes al mercado de trabajo", "Older workers y transiciones a la jubilación" y "Brecha de género en salarios y pensiones".

Esto no agota el tipo de estudios que se puede realizar con la MCVL. Por ejemplo, uno de los campos disponibles en el fichero de Personas es la fecha de defunción, lo que permitiría realizar un estudio de la mortalidad según características laborales.

## 6. Conclusiones

La MCVL es un conjunto de datos representativo de los afiliados y/o pensionistas de la Seguridad Social. Dada la naturaleza de la información, permite ver cómo fue la vida laboral de los individuos en su inicio, en su desarrollo y también como fueron las transiciones hacia la jubilación.

Por tanto, la MCVL conforma un conjunto completo de datos para el análisis del mercado laboral dada la versatilidad de análisis que se pueden aplicar a los registros representativos de la Seguridad Social. Este es el caso del análisis transversal, retrospectivo y longitudinal que permiten describir las trayectorias profesionales de los individuos.

El Instituto de Estadística y Cartografía tiene ya a disposición de los usuarios información transversal de cotizantes y pensionistas andaluces así como resultados de un estudio retrospectivo con indicadores novedosos sobre las trayectorias profesionales de los mismos. Se tiene la intención de incluir la perspectiva longitudinal, la elaboración de monografías para determinados colectivos de interés así como también ampliar toda esta información a nivel nacional.

Por todos estos motivos, la MCVL conforma un valor añadido a las estadísticas socio-laborales existentes en el panorama nacional.

## 7. Anexo metodológico

En el presente anexo se abordan algunas vicisitudes en la selección de registros y tratamiento de los datos así como las soluciones que se han adoptado en el IECA. Las restricciones de tiempo disponible y espacio permitido nos han llevado a tratar únicamente siete aspectos. Hemos priorizado los que nos han parecido más importantes.

### 7.1. Duplicados

En teoría, en cada fichero de la MCVL existe un campo o una combinación de campos única cuyos valores permitirían identificar un único registro de la tabla. No obstante la realidad resulta más prosaica y nos encontramos con duplicados en los ficheros de Personas, Afiliación y Pensiones.

#### Fichero de Personas

La propia Seguridad Social advierte sobre la existencia de duplicados en el fichero de personas de la versión con datos fiscales que elabora la Agencia Tributaria, es decir, puede existir más de una fila o registro para un mismo identificador (0,2% de los registros) . En la mayoría de los casos, es debido al proceso de identificación propio de dicha Agencia ya que existen pensionistas de orfandad sin DNI propio a los que se les asigna el mismo identificador que al progenitor que sigue vivo. En otros casos, aún más infrecuentes, la Agencia Tributaria identifica como una única persona lo que la Seguridad Social considera que son individuos distintos.

Para los registros duplicados en los que existen pensiones de orfandad, hemos asumido que éstas corresponden al individuo más joven y que el resto de relaciones con la Seguridad Social (cotizaciones y pensiones de viudedad en su mayoría) corresponden al progenitor que sigue vivo. Para diferenciar unos individuos de otros, hemos generado unos identificadores ficticios.

En última instancia, de persistir los duplicados, se han excluido del análisis.

#### Fichero de Afiliación

La propia guía de la Seguridad Social indica que un mismo episodio de alta puede quedar registrado varias veces en las bases de datos de afiliación. Es decir, que puede haber más de un registro con la combinación identificador de persona física, código de cuenta de cotización y fecha de alta y distintos valores en las demás columnas, cuando en realidad, debería ser única. Se trata de episodios que se duplican cuando se producen cambios en las características que los describían originalmente. Se trata de muy pocos casos (0,05%).

Para paliar este problema se ha optado en primera instancia por seleccionar el episodio que tenga una fecha de baja posterior de entre los que estén duplicados. De persistir el empate, el criterio ha sido simplemente escoger el primer episodio que aparezca en la base de datos.

#### Fichero de Pensiones

En algunos casos un mismo episodio con idéntico identificador de la persona, identificador de la prestación y año de la prestación, da lugar a dos registros duplicados (aproximadamente un 0,2% de los registros del fichero). Se trata de prestaciones que han cambiado de situación dos o más veces a lo largo del año, pero que por lo demás, presentan idéntica información. Para salvar estas duplicidades se ha optado por escoger siempre el primer registro que aparece en la base de datos.

## **7.2. Recuperaciones de provincia y fechas de nacimiento**

En el fichero de personas aparecen ciertos valores perdidos para algunos individuos como son el “lugar de residencia” y la “fecha de nacimiento”. En más de la mitad de los casos el motivo de la desaparición de la información es que la persona ha muerto (al gestionar la Seguridad Social las bajas por fallecimiento, se pierde información de la fecha y lugar de nacimiento del fallecido). En los demás casos serán otras vicisitudes las que provocan la pérdida de información. Para unos y otros se ha hecho un esfuerzo por recuperar parte de esta información perdida.

Para el “lugar de residencia” en primer lugar se busca la información faltante en el fichero de personas del año anterior al año de referencia. Si no está aquí, optamos por examinar el fichero de prestaciones ya que muchos de los fallecidos para los que carecemos de información estaban cobrando una pensión. Hay que destacar que en las recuperaciones hechas en el fichero de prestaciones, lo único que podemos rescatar es la provincia que gestionaba el pago de la correspondiente pensión.

La “fecha de nacimiento” del fichero de personas viene del documento TA.1 que rellena la persona cuando hace su primera inscripción en la Seguridad Social. En el caso de que no esté el dato, lo buscamos en el fichero de convivientes donde aparece la fecha de nacimiento que procede del Padrón.

### **7.3. Depuración de fechas**

Aparecen ciertas inconsistencias entorno a las fechas de alta y de baja de los episodios, propias de la naturaleza administrativa de la MCVL. En primer lugar vamos a ver cómo se trata el caso en el que la fecha de baja es anterior a la de alta y seguido, el problema de las fechas imposibles, es decir, fechas no válidas en el calendario gregoriano.

#### *Fechas de baja anteriores a las de alta*

En algunos episodios antiguos (todos ellos anteriores a la primera edición de la MCVL) encontramos que el número de días cotizados es cero ó negativo. Ésta circunstancia implica que este problema afecta al estudio retrospectivo, pero no afecta ni al transversal ni al longitudinal. Son un 0.04% de los episodios de afiliación.

Entendemos que se trata de errores y que por tanto, una de las dos fechas, o las dos, se introdujo de forma incorrecta.

Para solventar el problema seleccionamos los individuos con episodios problemáticos y sobre el episodio crítico vemos si “engancha” con el inicio o el final de otro episodio de cotización de la persona. Si se da esta situación es indicativo de que la fecha de alta o la fecha de baja debe estar bien y que el problema hay que atribuírselo a la otra fecha. En base a estos “enganches” pueden ocurrir cuatro cosas, que optemos por modificar la fecha de alta, o la fecha de baja o las dos (porque consideremos que ambas fechas están permutadas) o ninguna.

Tras realizar este procedimiento se modifican el 80% de los episodios cuya diferencia entre la fecha de baja y de alta resulta negativa o cero. El 20% de episodios restantes se excluyen del estudio.

#### *Fechas imposibles*

Hemos observado que existen alrededor de 230 episodios cuyas fechas de alta son 19640000. La mayoría de ellos pertenecen a personas del antiguo régimen especial agrario cuenta propia (0721). Se trata de episodios largos y significativos en la vida de esas personas y por tanto, se ha decidido mantenerlos.

El criterio a seguir en este caso es poner como fecha de alta 19640701 para todos ellos (a mitad de año).

Otro inconveniente que hemos encontrado en las fechas de alta es la aparición de fechas imposibles (por ejemplo 30 de febrero). En este caso se ha solucionado reemplazando por el último día del mes según corresponda.

#### **7.4. Cotizaciones por debajo de edad legal para trabajar**

Desde la aprobación del Estatuto de los Trabajadores, en 1980, la edad mínima legal para acceder al mercado laboral queda fijada en 16 años (hasta entonces era 14). Este cambio afectó a las generaciones posteriores a 1966. Por tanto nos quedamos sólo con las personas que tuvieron su primera cotización con 14 años o más (si tienen primera cotización anterior a 1980) o con 16 años o más (si tienen primera cotización en 1980 o año posterior).

#### **7.5. Sector de actividad de las cuentas de cotización que ya no existen**

La MCVL se realiza desde el año 2004. Desde entonces el marco normativo y las clasificaciones de aplicación no se han mantenido estáticas. Por ejemplo han aparecido nuevas formas de contratación (y han desaparecido otras). Un cambio que revistió gran importancia fue la introducción de una nueva Clasificación Nacional de Actividades Económicas (CNAE). En el año 2009 entró en vigor una nueva clasificación (CNAE-2009) que sustituyó a la antigua CNAE-93.

En las primeras ediciones de MCVL (2004-2008) se incluía en el fichero de afiliados un campo relativo a la actividad económica de la cuenta de cotización conforme a la clasificación vigente en ese momento (CNAE-93). En la edición de 2009 se incorpora un nuevo campo con la CNAE-2009 y desaparece el campo CNAE-93. No obstante, esto generó un problema: dado que la información sobre la actividad económica va asociada a la cuenta de cotización, todas aquellas cuentas que no habían estado activas a partir del 01-01-2009 no tenían relleno el campo CNAE 2009. Esto afectaba a gran parte de los episodios de cotización de la persona, especialmente a los más antiguos.

En la edición del año siguiente (2010) se introdujo un nuevo cambio para solventar este problema que se ha mantenido hasta la actualidad. Además del nuevo campo CNAE 2009 se reintrodujo el campo con el código CNAE-93 que tuviera la cuenta de cotización. Esto permite (no sin problemas como ahora explicaremos) reconstruir la información sobre la actividad para las cuentas de cotización que no la tenían.

La limitación al proceder de este modo deriva de que aunque existe una relación biunívoca entre CNAE-93 y CNAE-2009 establecida por el INE, esta correspondencia sólo es válida al máximo nivel de desagregación (5 dígitos) cuando en las CNAEs de la MCVL solo disponemos de los tres primeros.

Para construir una tabla de correspondencias a 3 dígitos se ha acudido a la propia información empírica que contiene la MCVL, es decir, para las cuentas de cotización en las que constan ambos códigos se establece la relación en base a cuál de las relaciones existentes es más frecuente. Con esta información se reconstruye la CNAE-2009 para las cuentas de cotización que no las tenían y finalmente se agrupan los códigos CNAE a 3 dígitos en 6 grandes sectores (Agricultura, Industria, Construcción, Comercio y hostelería, Servicios públicos, Otros). Al agrupar en grandes sectores disminuyen enormemente las asignaciones ambiguas.

Algunas cuentas de cotización (en su mayoría fuera del Régimen General y del Régimen de Autónomos), carecen de información de actividad. Para ellas asignamos el código de actividad en función del régimen de cotización que aparezca en el episodio, tabla 5.

Tabla 5. Asignación de la actividad a partir del régimen de cotización

Régimen de Cotización	CNAE09	CNAE93	Sector de Actividad
<b>0138-1211-1221</b>	970	950	Hogar
<b>0163-0613</b>	01	01	Agricultura
<b>0721-0722-0723</b>	01	01	Agricultura
<b>0811-0825</b>	031	050	Pesca
<b>0911</b>	051	101	Industria (Extracción)

*Fuente: Instituto de Estadística y Cartografía de Andalucía. Elaboración propia*

Por último, el año 2009 presenta ciertas particularidades, al no incluir el campo CNAE-93. Por ello, ha habido que hacer una adaptación para recuperar la información faltante a partir de las ediciones posteriores de la MCVL:

- Paso 1 - Identificamos en el fichero de 2009 los registros que carecían de información de actividad. Los registros se corresponden con el identificador único de episodio (tripla Identificador de persona - Código de cuenta de cotización - Fecha de alta).
- Paso 2 - Agrupamos a todas las personas que han formado parte de la MCVL en las ediciones posteriores. Con ellas construimos un fichero único de Afiliación, siguiendo el esquema comentado para el análisis longitudinal (ver figura 4).
- Paso 3 - Finalmente, cruzamos los ficheros generados en los Pasos 1 y 2, utilizando el identificador único presente en ambos.

## 7.6. Recuperación de registros omitidos en la edición de 2011

El Régimen Especial Agrario Cuenta Ajena se integró el 01-01-2012 en el Régimen General. Como comenta la propia Seguridad Social en la "Guía de Contenido" de la MCVL, en el proceso de adaptar la extracción de información a este hecho, en la edición de 2011 se omitieron algunos episodios. Concretamente, se omitieron los episodios de paro (tipo de relación laboral 751 a 756) que habían tenido las personas en el antiguo régimen especial agrario cuenta ajena.

Para poder realizar una tabulación transversal correcta relativa al año 2011, en el IECA hemos realizado una adaptación específica para el código de ese año. El procedimiento para recuperar los episodios faltantes es similar al que se ha comentado en el apartado anterior para el rescatar los códigos de actividad del año 2009, obteniendo el dato de las ediciones posteriores, que sí lo incluyen. La diferencia radica en que ahora no hay que traer de las ediciones posteriores el valor de un campo concreto faltante en un registro (fila) existente, sino que hay que traer el registro completo.

## 7.7. Tratamiento de los datos de la MCVL

Este apartado se ocupará de las herramientas utilizadas en la elaboración de la información. Se ha considerado oportuno incluirlo por la complejidad que presenta el manejo de los datos de la MCVL dada la naturaleza de la información y el gran volumen de la misma.

Los datos han sido tratados principalmente con R, software libre y versátil que cuenta con librerías adecuadas para el manejo de este tipo de información. Los primeros resultados obtenidos fueron los de la explotación transversal y después, la retrospectiva. Ambas se elaboraban extrayendo la información de los ficheros de la MCVL en texto plano donde previamente se establecía un filtro para la comunidad autónoma de Andalucía con el fin de reducir el volumen de información a tratar. Pese a que R es un programa muy potente presenta ciertas limitaciones respecto a la memoria RAM ya que es un devorador de la misma cuando los dataframe son muy grandes, de ahí que se extrajera solo la información de la comunidad andaluza de los ficheros para ser tratada.

Para solventar este inconveniente, con miras a publicar resultados para España y realizar análisis longitudinal de los datos (lo que implica tratar simultáneamente varias ediciones de la MCVL) se ha procedido a volcar la información en Oracle. Desde esta ubicación se extrae la información mediante una conexión a R y se opera mediante una librería novedosa, **dbplyr**, para obtener los resultados. Esta librería, muy potente, se encuentra aún en desarrollo y traduce las sentencias **dplyr** (otra librería, de uso muy extendido en el tratamiento de datos con R) a consultas sql sin consumir memoria volátil.



## Bibliografía

Arranz, J., García Serrano, C., & Herranz, V. (2013). How do we pursue "labormetrics"? An application using the MCVL. *Estadística Española*.

Duran, A. (2007). La Muestra Continua de Vidas Laborales de la Seguridad Social. *Revista del Ministerio de Trabajo y Asuntos Sociales*.

Lapuerta, I. (Julio de 2010). Claves para el trabajo con la Muestra Continua de Vidas Laborales. *DemoSoc Working Paper*.

Veiguela Fernández, N., & Romero Martínez, P. (2014). Procesamiento y depuración de la Muestra Continua de Vidas Laborales para el estudio del mercado laboral gallego. *Jornadas Estadísticas de las Comunidades Autónomas*.

Willekens, F. (2014). *Multistate Analysis of Life Histories with R*. Rostock, Germany: Springer.