



DemoSoc Working Paper

Paper Number 2010--37

Claves para el trabajo con la Muestra Continua de Vidas Laborales

Irene Lapuerta

E-mail: irene.lapuerta@unavarra.es

Julio de 2010

Department of Political & Social Sciences

Universitat Pompeu Fabra

Ramon Trias Fargas, 25-27

08005 Barcelona

<http://www.upf.edu/dcpis/>



Resumen

Desde el año 2005 la comunidad científica cuenta con una nueva fuente de información anual para el estudio de las dinámicas del mercado de trabajo y del sistema de previsión social de carácter contributivo en España. Sus microdatos, que reciben el nombre de Muestra Continua de Vidas Laborales (MCVL), proceden de tres registros administrativos: la Seguridad Social, el Padrón Continuo Municipal y la Agencia Tributaria. En este trabajo se exponen sus características fundamentales, al tiempo que se plantean algunas pautas básicas para afrontar las dificultades en el manejo de sus datos. Entre ellas destacan las peculiaridades de su estructura panel; el tratamiento del pluriempleo y las situaciones simultáneas; el modo en que se computa una relación laboral; y los problemas para la identificación de la estructura familiar.

Palabras clave

Muestra Continua de Vidas Laborales, Seguridad Social, registro administrativo, dificultades.

Agradecimientos

La autora desea agradecer a la Dirección General de la Seguridad Social y, en especial, a Almudena Durán las aclaraciones y sugerencias recibidas para afrontar el trabajo con la Muestra Continua de Vidas Laborales. Igualmente, agradece la financiación recibida del Institut d'Estadística de Catalunya (IDESCAT), el Ministerio de Trabajo e Inmigración (PREMIOS FIPROS 2009/20) y el Centro de Investigaciones Sociológicas para la elaboración de este trabajo.

1. Introducción.

Desde el año 2005, la comunidad científica cuenta con una nueva fuente de información anual para el estudio de las dinámicas del mercado de trabajo y del sistema de previsión social de carácter contributivo en España. Los microdatos, que reciben el nombre de ‘Muestra Continua de Vidas Laborales’ (en adelante MCVL), provienen de tres registros administrativos: la Seguridad Social, el Padrón Continuo Municipal y la Agencia Tributaria. La MCVL es, por lo tanto, la primera y única fuente de información construida íntegramente a partir de la información en manos de la administración española. Con ella se cubre, en gran medida, nuestro déficit histórico en la explotación estadística de los registros administrativos (Cebrián y Toharia, 2007), que cuenta con una larga tradición en otros países de nuestro entorno europeo y, especialmente, en el caso de los nórdicos (Tonder, 2008).

Pero, más allá del interés que despierta el origen de su información, la MCVL es una fuente tremendamente atractiva por su propio contenido. Entre la información que facilita cabe destacar las fechas exactas de inicio y fin de todos los episodios laborales de los individuos seleccionados desde que la Seguridad Social cuenta con procedimientos informatizados, así como del resto de relaciones administrativas mantenidas con esta institución y a partir de las cuales se reconocen el cobro de algunas prestaciones o el mantenimiento de las cotizaciones. La MCVL también incluye información sobre cada uno de los trabajos desempeñados (régimen y grupo de cotización, tipo de contrato y de jornada, las bases de cotización mensuales, las causas de baja, etc.), los empleadores (tipo de empresa, tamaño de la misma, domicilio social, etc.) en caso de que la relación sea por cuenta ajena, o las características asociadas a la prestación (el tipo, la evolución de su cuantía y la fecha de reconocimiento). A ella se le suma la información sobre los convivientes extraída del Padrón Municipal y los ingresos anuales derivados del trabajo, procedentes de la Agencia Tributaria. La MCVL permite, en definitiva, reconstruir con elevado grado de detalle la vida laboral de los individuos y su historial de percepción de prestaciones de carácter contributivo.

Con la MCVL se abren, así, nuevas líneas de investigación, hasta ahora vetadas por la inexistencia de datos o las limitaciones metodológicas de otras fuentes con información similar. Algunas de ellas son, por ejemplo, el análisis de las trayectorias laborales a lo largo del ciclo de vida; los flujos de entrada y de salida del mercado de trabajo, así como sus dinámicas de cambio y sus repercusiones en distintos colectivos; la intensidad protectora y la evolución de los sistemas de previsión social contributivos (pensiones, desempleo y permisos de cuidado); o la evaluación de impacto de algunas prestaciones y políticas públicas, incluidas las desarrolladas en el ámbito autonómico.

Sin embargo, su utilización no está exenta de dificultades. La mayor parte están relacionadas con el carácter administrativo de la información que facilita y a partir de la cual se diseña. La depuración y el manejo de los datos de la MCVL resultan, además, especialmente arduos en comparación a otras bases. De ahí que, antes de comenzar el trabajo con esta fuente, resulta especialmente conveniente sopesar sus principales ventajas e inconvenientes. Este documento pretende facilitar, en la medida de lo posible, ese proceso. Para ello, se exponen las características fundamentales de la MCVL, al tiempo que se plantean algunas pautas básicas para afrontar el tratamiento de sus datos.

Para ello, su contenido se ha estructurado en seis apartados. En el segundo se expone qué es la MCVL, aludiendo a su representatividad y, por ende, detallando quién forma parte y quién no de la población de referencia; la técnica de muestreo utilizada y su tamaño muestral. El tercer apartado hace referencia a las cuestiones organizativas de la MCVL, entre las que cabe destacar el número y el contenido de los ficheros que la componen y el procedimiento de fusión de los mismos¹. En el cuarto apartado se analizan sus ventajas frente a otras fuentes de información, al tiempo que se señalan sus principales déficits. Posteriormente, en el quinto apartado, se explican las dos formas posibles de trabajo con la MCVL (transversal vs. longitudinal), haciendo especial hincapié en las especificidades de cada una de ellas y sus implicaciones posteriores en el montaje y tratamiento de los datos. El sexto apartado aporta algunas soluciones a las principales dificultades que plantea el manejo de la MCVL y a las que, inevitablemente, el investigador deberá enfrentarse cuando inicie el trabajo con la muestra. Entre ellas cabe destacar las inconsistencias derivadas del proceso de fusión de la información; el tratamiento del pluriempleo y las situaciones simultáneas; el modo en que se computa una relación laboral y sus repercusiones en el cómputo de los periodos de empleo; el cálculo de las duraciones de los episodios de desempleo; y, por último, la identificación de la estructura familiar. El documento cierra con unas breves conclusiones que recogen los principales aspectos abordados en este trabajo.

2. ¿Qué es la MCVL?

Tal y como avanzábamos en el apartado introductorio, la MCVL es una selección de datos individuales y anónimos procedentes de registros administrativos. Así, a diferencia de otras encuestas en las que primero se define el objeto de estudio y, posteriormente, se aborda la recopilación de la información, la MCVL invierte este proceso (MTAS, 2006a). Una característica que tiene consecuencias tanto en su representatividad como en su contenido. En este apartado analizaremos las implicaciones que se derivan del primero de estos aspectos. El siguiente paso es, de este modo, entender quién forma parte y quién no de la **población de referencia** de la MCVL.

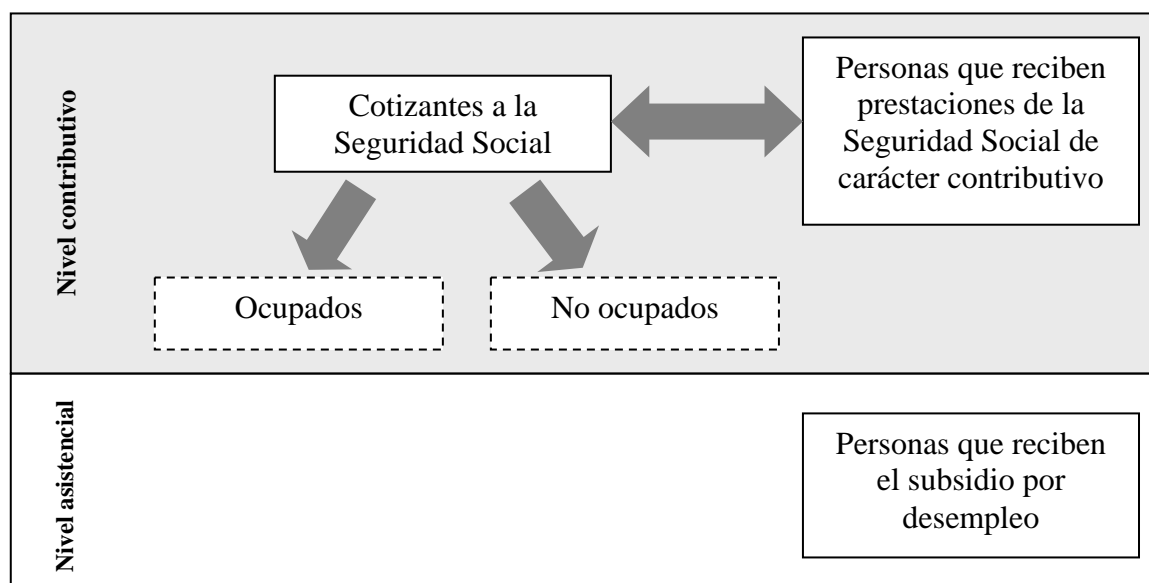
La población de la que se extrae la información está compuesta por “todas las personas que han estado en situación de afiliado en alta, o recibiendo alguna pensión contributiva de la Seguridad Social en algún momento del año de referencia” (MTAS, 2006b: 3). La figura 1 refleja a los tres colectivos que incluye esta definición; los individuos que reciben cualquier tipo de *pensión contributiva* derivada de las contingencias de jubilación, incapacidad permanente o fallecimiento (viudedad y orfandad), e incluidas también las generadas por el Seguro Obligatorio de Vejez e Invalidez (SOVI²); los que trabajan y cotizan en cualquiera de los regímenes de la Seguridad Social (*cotizantes*

¹ Estos dos primeros apartados son, en realidad, un resumen del abundante material metodológico de la MCVL que se ha desarrollado en los últimos años y que se irá citando a lo largo del texto. Su valor añadido no reside en su contenido sino en la organización de una información que se encuentra dispersa en la actualidad.

² El SOVI es un régimen residual de la Seguridad Social que se aplica a aquellas personas (y en su caso familiares), que no tienen derecho a las pensiones del actual sistema de la Seguridad Social y que acreditan tener más de 65 años y haber cotizado con anterioridad al año 1967 (para más información consultar: http://www.seg-social.es/Internet_1/Trabajadores/PrestacionesPension10935/PensionesdelSeguroO10970/index.htm).

ocupados); y los que cotizan para poder percibir una pensión en el futuro, aunque no hayan estado trabajando (*cotizantes no ocupados*). Dentro de este último grupo se incluyen a los beneficiarios de las prestaciones de desempleo y los que tienen un Convenio Especial para continuar aportando cotizaciones a la Seguridad Social³. Existían algo más de 1,6 millones de relaciones de este tipo en la población de referencia del año 2004 (MTAS, 2006a y 2006b).

Imagen 1: Población de referencia de la MCVL.



Fuente: Elaboración propia.

En cambio, se excluyen de la población de referencia a aquellas personas que, aún estando registradas en la Seguridad Social, reciben únicamente asistencia sanitaria o son perceptores de prestaciones no contributivas (es decir, aquellas que no están ligadas a una cotización previa), a excepción de aquellos que perciben el *subsidio de desempleo*⁴ que sí forman parte de ella. Además, al referirse exclusivamente al ámbito de la Seguridad Social no están incluidos los que tienen sistemas de previsión distintos como, por ejemplo, los funcionarios de Clases Pasivas o los profesionales liberales con protección social a través de sus Colegios Profesionales. No obstante, se estima que el sistema de la Seguridad Social cubre al 95% de los ocupados en España (Durán, 2007).

La población de referencia de la MCVL recoge, de este modo, relaciones administrativas con la Seguridad Social que reflejan situaciones tanto de actividad (ya sean trabajadores o preceptores de la prestación o el subsidio de desempleo), como de no-actividad (cotizantes que no tienen trabajo o pensionistas). Sin embargo, la

³ Bajo la figura del Convenio Especial se recoge a aquellos individuos que han trabajado en el pasado y continúan figurando en la Seguridad Social como si siguieran haciéndolo. Un ejemplo de ello son los individuos que viven de las indemnizaciones por despido y mantienen su cotización a la Seguridad Social hasta que pueden acceder a una pensión.

⁴ El subsidio por desempleo tiene carácter asistencial y se reconoce a aquellos trabajadores que, cumpliendo determinados requisitos, han agotado la prestación por desempleo de carácter contributivo o no cumplen el periodo mínimo de cotización para acceder a ésta.

población de referencia de la MCVL no es asimilable a la totalidad de la población activa, debido a que excluye a las personas cubiertas por otros sistemas de previsión social distintos al de la Seguridad Social -como en el caso de los funcionarios de clases pasivas u otras mutualidades- y a los demandantes de empleo que no reciben ninguna ayuda económica por parte de la Seguridad Social. Tampoco resulta asimilable al conjunto de la población inactiva, a pesar de incluir a algunos colectivos que la integran (MTAS, 2006a y 2007b).

A partir de la población de referencia así definida, cada versión de la MCVL se construye a través de un **muestro aleatorio simple, no estratificado**, que equivale al 4% de la misma. En consecuencia, el factor elevación de la MCVL para obtener resultados poblacionales es 25. La selección de los individuos se realiza a través de la localización de unas cifras escogidas aleatoriamente dentro de su código de identificación personal.

Este tipo de muestreo presenta, tal y como señala Durán (2007), dos ventajas fundamentales. En primer lugar, asegura la selección de los mismos individuos en las sucesivas versiones de la MCVL, siempre y cuando “continúen manteniendo relación con la Seguridad Social y no cambien de código identificador” (MTAS, 2006a: 6). Así se consigue una estructura tipo panel, lo que facilita el seguimiento de los mismos individuos y limita, hasta cierto punto, el problema de sesgo muestral que supone la definición de la población de referencia (mantener una relación administrativa con la Seguridad Social, bien como cotizante, bien como perceptor de prestaciones contributivas). En segundo lugar, “proporciona un procedimiento automático para seleccionar las nuevas incorporaciones, garantizando además que éstas son representativas de las altas en la población de referencia” (MTAS, 2006a: 6). De este modo, se mantiene la representatividad de la muestra en los sucesivos años de referencia.

Cuadro 1: Tamaño de la población de referencia y tamaño muestral de la MCVL (olas 2004-2008).

<i>Olas:</i>	Población de referencia	Tamaño muestral España⁽¹⁾
2004	27.435.159	1.089.016
2005	28.571.142	1.140.929
2006	29.257.294	1.169.870
2007	30.003.000	1.200.076
2008	30.310.000	1.212.884

⁽¹⁾ El tamaño muestral refleja el número de individuos que componen la MCVL CDF una vez depurados los identificadores repetidos (ver la dificultad 1 del apartado 4 para obtener una visión más detallada de esta problemática).

Fuente: Elaboración propia.

El cuadro 1 refleja la evolución de estos parámetros -tamaño de la población de referencia y de la muestra- en cinco de las seis olas disponibles en la actualidad. Cabe señalar que la Seguridad Social ofrece dos versiones de la MCVL cada año: una de ellas

sin datos fiscales (MCVL SDF) y la otra con datos fiscales (MCVL CDF)⁵. La primera incluye la información extraída de la Seguridad Social y el Padrón Continuo Municipal⁶, mientras que la segunda añade a la información anterior la procedente de la Agencia Tributaria. Debido a razones de confidencialidad, los identificadores de ambas versiones cambian, por lo que resulta conveniente trabajar directamente con la MCVL CDF, ya que es la que incluye el mayor volumen de información⁷.

De esta forma, la primera ola, correspondiente al año 2004, está compuesta por 1.089.016 individuos, cifra que ha ido incrementándose en las siguientes hasta superar 1.212.000 en el año 2008. Este crecimiento se debe, lógicamente, al aumento de la población de referencia, que ha pasado de los veintisiete millones y medio de personas a más de treinta millones trescientos mil en esos mismos años, como consecuencia del incremento en el número de cotizantes y perceptores de prestaciones. El crecimiento de la población de referencia ha alcanzado, por lo tanto, el 10,5%, cifra ligeramente superior en el caso de la muestra, que se sitúa en el 11,4%.

3. Organización y contenido de la MCVL.

El origen administrativo de los datos también tiene consecuencias en la información que provee la MCVL y en la forma en que ésta se organiza. El cuadro 2 refleja las principales características de los distintos tipos de ficheros que integran la MCVL, detallando su nombre y contenido, la fuente de la que proviene la información, el ámbito temporal a la que hace referencia y el modo en que se organizan los datos.

En este sentido, cabe señalar que la versión SDF de la MCVL incluye siete tipos distintos de ficheros, mientras que la MCVL CDF añade un fichero adicional que contiene los datos de la Agencia Tributaria (MTAS, 2006c). Dado el gran volumen de dos de ellos -el relativo a la historia laboral de los individuos y el de bases de cotización por cuenta ajena-, se proporcionan divididos en varias partes con la intención de facilitar su manejo, pero manteniendo la misma estructura. Todos los ficheros se distribuyen, además, comprimidos y en formato de texto. Sin embargo, la organización de los ficheros en las sucesivas olas y en ambas versiones -con y sin datos fiscales- es la misma, aunque su estructura ha sufrido algunas modificaciones que serán destacadas a lo largo de este apartado.

Además, la Seguridad Social facilita con cada una de las versiones de la MCVL varios documentos que recogen, por un lado, la estructura de los campos de cada archivo⁸ y,

⁵ La versión MCVL SDF suele estar disponible antes del verano y la MCVL CDF en el otoño. Las instrucciones para tramitar su solicitud están disponibles en la web de la Seguridad Social http://www.seg-social.es/Internet_1/Estadistica/Est/Muestra_Continua_de_Vidas_Laborales/index.htm.

⁶ Excepto en la primera ola, correspondiente al año 2004, que sólo incluye la información de la Seguridad Social. La información del Padrón Continuo fue incluida, por primera vez, en la ola del 2005.

⁷ Por este motivo, la información que se presenta en este documento hace referencia a la versión CDF. Para el año 2009 solo está disponible la versión sin datos fiscales.

⁸ Para la versión MCVL 2008 SDF la estructura de los ficheros está disponible en <http://www.seg-social.es/stpri00/groups/public/documents/binario/120865.pdf>. La estructura del fichero que contiene la información fiscal esta disponible en: <http://www.seg-social.es/stpri00/groups/public/documents/binario/123574.pdf>.

por el otro, las fichas metodológicas de cada una de las variables⁹, en las que, junto a este aspecto, se detallan las fuentes de información de las que se obtienen los datos, sus posibles valores y el grado de cobertura de la información. La revisión y utilización de esta documentación resulta, en consecuencia, imprescindible cuando se inicia el trabajo con la MCVL. No obstante, en aras a facilitar esta tarea, el cuadro 2 resume esta información.

De este modo, el **fichero ‘PERSANON’** contiene información relativa a la persona. Las variables que incluye son el sexo, la fecha de nacimiento (mes y año), la nacionalidad, la provincia de nacimiento, la provincia en la que se afilió por primera vez en la Seguridad Social, el municipio de residencia y, en su caso, la fecha de fallecimiento. Todas ellas son extraídas de la información obrante en los registros de la Seguridad Social. A ellas se añadieron 2 variables más en la ola del 2005 -nivel de estudios y el país de nacimiento- a partir de la información procedente del Padrón Municipal. Las variables municipio de residencia y provincia de nacimiento pasaron a adoptar también la información disponible en el Padrón Municipal, a no ser que éste no dispusiese de ella. En estos casos, se ha dejado la información procedente de la Seguridad Social (MTAS, 2006c).

La información de este fichero figura, en cada una de las olas, con el valor que tenían en el momento en que se extrajo la información (MTAS, 2006c). La fecha de extracción es, de este modo, distinta en cada una de las olas y posterior al año al que hace referencia la ola. Así, para la ola del 2004, la información se obtuvo en junio de 2005, mientras que en la ola del 2005 en mayo de 2006, para la información procedente de los registros de la Seguridad Social, y en septiembre de ese mismo año, para la información del Padrón Municipal. En las posteriores, correspondientes a los años 2006, 2007 y 2008, la información se extrajo en el mes de marzo de los años 2007 y 2008 y en abril del 2009 respectivamente. La información se organiza en los ficheros ‘PERSANON’, en un único registro por individuo, por lo que en cada fichero existen tantos registros como individuos incluya la MCVL en la ola correspondiente.

El **fichero ‘CONVIVI’**, incorporado en la ola del 2005 y extraído del Padrón Continuo Municipal, también mantiene estas dos últimas características relativas al plazo temporal al que hace referencia la información y la organización de la misma. Las variables que ofrece son la fecha de nacimiento (mes y año) y el sexo de la persona que aparece en la muestra y de las personas que figuran en su misma hoja padronal, hasta un máximo de 9. Este archivo repite, por lo tanto, dos variables facilitadas en el archivo ‘PERSANON’, pero añade nueva información sobre los convivientes. No obstante, cuando en la misma hoja padronal figuran inscritos más de 20 individuos, no se incluye esta información en el fichero, al considerarse que en esos casos se trata probablemente de una institución (MTAS, 2006c).

⁹ La Seguridad Social ofrece la información de las variables en 5 documentos distintos. Los relativos a la versión MCVL 2008 SDF pueden descargarse en: http://www.seg-social.es/Internet_1/Estadistica/Est/Muestra_Continua_de_Vidas_Laborales/Descripci_n_de_ficheros_y_variables/MCVLSINDATOS/index.htm, mientras que el documento que describe las variables fiscales está disponible en: <http://www.seg-social.es/stpri00/groups/public/documents/binario/120874.pdf>

Cuadro 2: Estructura de los ficheros de la MCVL.

Nombre del fichero	Nº de ficheros	Contenido	Ámbito temporal	Fuente	Organización de los registros
PERSANON	1	Datos personales	Fecha extracción	Seguridad Social ⁽¹⁾	Uno por cada individuo
CONVIVI	1	Convivientes inscritos en la misma hoja padronal	Fecha extracción	Padrón Municipal	Uno por cada individuo
AFILANON	3	Historia Laboral	Retrospectiva (desde 1967) ⁽²⁾	Seguridad Social	Uno por cada relación laboral que el individuo haya mantenido, así como los trabajos por cuenta propia o las situaciones en las que haya cotizado sin estar ocupado
COTIANON	12	Bases de cotización por cuenta ajena	Retrospectiva (desde 1980)	Seguridad Social	Uno por cada año natural de cada relación laboral por cuenta ajena ⁽⁴⁾
CPROANON ⁽³⁾	1	Bases de cotización por cuenta propia	Retrospectiva (desde 1980)	Seguridad Social	Uno por cada año natural de cada relación laboral por cuenta propia ⁽⁴⁾
PREANON	1	Pensiones	Retrospectiva (desde 1996)	Seguridad Social	Uno por cada año natural de pensión que haya recibido el individuo
BLOQUE5 ⁽⁵⁾	1	Datos Fiscales	Año ola	Agencia Tributaria	Uno por cada retenedor
DIVISION	1	Mapa ⁽⁶⁾	-	-	Uno por cada individuo

⁽¹⁾ A partir de la ola del 2005, ésta incluida, en el fichero PERSANON también se incluye información extraída del Padrón Municipal.

⁽²⁾ Excepto la información relativa al empleador, que contiene los valores de la fecha de extracción de cada una de las olas de la MCVL.

⁽³⁾ A partir de la ola del 2006 este fichero pasa a llamarse 'COTIANON13'. Por lo tanto,

⁽⁴⁾ Disponibles a partir de junio de 1980. El archivo COTIANON no incluye las relaciones laborales mantenidas bajo el Régimen Especial Agrario y el de Empleados del Hogar, ni las amparadas bajo Convenios Especiales, facilitadas todas ellas en el archivo CPROANON.

⁽⁵⁾ Este fichero solo esta disponible en la versión CDF. A partir de la ola del 2006 adopta el nombre de 'DATOS_FISCALES'.

⁽⁶⁾ Indica en cuál de los subficheros se encuentran los datos de cada individuo. Por lo tanto, el único objetivo de este fichero es facilitar la localización de los casos.

Fuente: Elaboración propia.

En los tres ficheros 'AFILANON' se incluyen todas las relaciones administrativas mantenidas por el individuo con la Seguridad Social desde 1967. En consecuencia, a partir de estos archivos se puede reconstruir, con carácter retrospectivo, la historia laboral de los individuos incluidos en cada ola de la MCVL¹⁰. Entre su información se

¹⁰ Cabe advertir que la historia laboral que refleja la MCVL puede estar incompleta para los individuos de mayor edad. Según señala el informe del MTAS (2006), en la MCVL apenas podemos encontrar periodos

incluyen variables relativas a los trabajos desempeñados (fechas exactas de alta y baja, régimen de cotización, grupo de cotización, tipo de contrato, coeficiente de tiempo parcial, causas de baja, minusvalía si consta en el contrato, colectivo al que pertenece el trabajador) y el empleador o responsable de la cotización (actividad económica, tamaño de la empresa, fecha de alta del primer trabajador, tipo de empleador, provincia donde tiene su sede el domicilio social de la empresa, etc.). La misma información se incluye, si procede, para otras situaciones de no ocupación pero que conllevan el mantenimiento de la cotización a la Seguridad Social, al tratarse de relaciones administrativas -tal y como explicamos en el apartado anterior- sobre las que se construye la población de referencia de la MCVL.

No obstante, cabe señalar que estos ficheros sólo ofrecen información retrospectiva para las variables que describen el trabajo desempeñado o la relación administrativa mantenida con la Seguridad Social. La información relativa al empleador hace referencia a la fecha de extracción de cada una de las olas de la MCVL, por lo que su valor puede no coincidir con el valor real del momento en que se desarrolló la relación laboral (MTAS, 2006c). La información de este grupo de variables, especialmente las relativas al tamaño y la sede del domicilio social de la empresa, deben utilizarse con precaución con anterioridad al año 2004 y tiene consecuencias, tal y como explicaremos más adelante, a la hora de construir nuestra base de datos.

El contenido de estos los ficheros ‘AFILANON’ ha sufrido, además, algunas modificaciones en las sucesivas olas de la MCVL que es necesario tener en cuenta cuando se procede a convertir los archivos. En el año 2005, se incorporaron 8 variables que recogían información sobre las modificaciones que pueden haberse producido durante una relación laboral sin que ésta se interrumpa. Las variables que se facilitan son la fecha de la modificación, el tipo de contrato y el coeficiente de tiempo parcial, para las dos primeras modificaciones, y el grupo de cotización y la fecha del cambio del grupo de cotización para la primera modificación. A partir de ese mismo año, pero solo en la versión CDF, se incluyó una nueva variable para identificar al empleador o pagador que permite la fusión de esta información con la facilitada por la Agencia Tributaria (‘BLOQUE5’). Posteriormente, en la ola del 2006, se ampliaron los posibles valores de la variable ‘colectivo al que pertenece el trabajador’, permitiendo la identificación no solo de las empresas de trabajo temporal, sino también de los trabajadores del sector público u otros colectivos.

De este modo, los datos de los ficheros ‘AFILANON’ reflejan, en término promedio, más de doce relaciones administrativas por individuo, pero pueden alcanzar varios centenares, especialmente en aquellos colectivos con una alta temporalidad (MTAS, 2006a). De ahí que la organización de la información en estos ficheros sea distinta a los anteriores, que reflejaban en una única fila la información de un individuo. En este caso, los registros hacen referencia a cada una de las relaciones administrativas que el individuo ha mantenido con la Seguridad Social. Los registros se facilitan, además, ordenados de forma ascendente por la fecha de inicio de cada relación.

de afiliación anteriores a 1967, a pesar de que la mayoría de los actuales jubilados ya trabajan en ese año. La explicación radica en que la informatización de los registros de la Seguridad Social ha sido muy desigual y relativamente reciente. Por lo tanto, cuando se trabaje con datos antiguos de este fichero “hay que ser consciente de que la ausencia de información no significa necesariamente que no hubiera relación con la Seguridad Social en ese periodo” (MTAS, 2006: 124).

Los **ficheros ‘COTIANON’ y ‘CPROANON’** contienen información sobre las bases de cotización para los trabajadores por cuenta ajena y por cuenta propia respectivamente. La excepción la constituyen las relaciones laborales mantenidas bajo el Régimen Especial Agrario y el de Empleados del Hogar y las amparadas por Convenios Especiales, facilitadas todas ellas en el archivo ‘CPROANON’ independientemente de la situación profesional del individuo. La información de ambos ficheros proviene, al igual que los ficheros ‘AFILANON’, de la Seguridad Social y ofrecen información desde el año 1980 de las bases de cotización mensuales por contingencias comunes y las fechas de inicio y fin de cada relación laboral. El fichero ‘COTIANON’ incluye, además, las bases de cotización mensuales por accidentes de trabajo (MTAS, 2006c).

No obstante, la información se organiza en ambos ficheros en tantos registros como años naturales haya durado cada relación laboral. Esto significa que si un individuo ha tenido dos relaciones laborales en un mismo año con dos empresarios distintos, en el fichero aparecerán dos registros para ese año; uno para cada relación laboral. Los campos de las bases de cotización correspondientes a cada uno de los meses reflejarán o no información en función de la duración de cada contrato. Sin embargo, si una persona ha mantenido con la misma empresa dos o más contratos dentro de un mismo mes, las bases de cotización aparecerán sumadas como si fueran una sola cotización, aunque en el fichero ‘AFILANON’ aparecerán separadas. Por último, cabe advertir que, dado el gran volumen de la información contenida en el fichero ‘CONTIANON’, éste se presenta dividido en 12 partes en todas las olas, mientras que el ‘CPROANON’ representa un único fichero que, a partir de la ola del 2006, se denomina ‘COTIANON13’.

El **fichero ‘PREANON’** proporciona información sobre las pensiones de carácter contributivo percibidas por los individuos desde 1996, año en que la Seguridad Social creó la base de datos sobre prestaciones de la que se extrae la información (MTAS, 2006c). Entre las 31 variables que componen el fichero se incluyen la cuantía de cada pensión, la fecha de reconocimiento, la clase y el régimen de la pensión, así como una serie de características que influyen en el cálculo de su cuantía (base reguladora, grado de minusvalía, número de años cotizados y bonificados, complementos a mínimos recibidos, revalorizaciones, etc.). La información se ofrece en tantos registros como años naturales y tipos de pensión haya recibido el individuo, ordenados también de manera ascendente. Pero, a diferencia de los ficheros que contienen las bases de cotización, la información no se presenta mensualmente sino para los totales anuales de cada una de las pensiones que perciba el individuo. De este modo, el fichero facilita en el tercer campo un identificador para cada pensión, que nos permite distinguir las distintas prestaciones percibidas por la misma persona desde el momento en que éstas le son reconocidas.

El **fichero ‘BLOQUE5’** -denominado ‘DATOS_FISCALES’ a partir de la ola del 2006-, es facilitado por la Agencia Tributaria y contiene la información que los empleadores o pagadores de pensiones o prestaciones por desempleo transmiten a este organismo a efectos del Impuesto de la Renta de las Personas Físicas (IRPF). La información se extrae, por lo tanto, a partir de la explotación del modelo 190¹¹ e

¹¹ Aunque la información del modelo 190 tiene efectos sobre la declaración del IRPF, no se puede confundir con ésta. El modelo 190 es el “resumen anual de retenciones e ingresos a cuenta del IRPF sobre

identifica, entre otras variables el salario total, sin los topes mínimos y máximos que afectan a las bases de cotización, así como la cuantía de las prestaciones públicas o privadas (MTAS, 2006c y 2006d). Igualmente, también incluye la información que los trabajadores o perceptores de prestaciones facilitan a la entidad que les paga para calcular las retenciones de las que son susceptibles y, por ende, establecer la cuantía de los pagos y el borrador del IRPF. Así pues, el fichero incluye 26 variables que describen distintos aspectos de la situación familiar del individuo (estado civil, número de descendientes y ascendientes, edad de los mismos y si están afectados por alguna discapacidad), variables que afectan todas ellas al tipo de retención aplicado a cargo del IRPF (MTAS, 2006c y 2006d). Sin embargo, cabe señalar que el trabajador no está obligado a notificar estas situaciones y, en caso de no hacerlo, puede recuperar posteriormente las retenciones aplicadas al realizar la declaración del IRPF.

A ello hay que añadir que la información de este fichero solo se facilita para los individuos que estaban en el campo de aplicación del modelo 190 en la ola correspondiente de la MCVL. En consecuencia, quedan excluidos los que tienen su residencia en una Comunidad con régimen fiscal propio (este es el caso de Navarra y el País Vasco); los trabajadores por cuenta propia, con independencia del régimen al que estén adscritos¹²; y los trabajadores pertenecientes al Régimen Especial de Empleados del Hogar. Tampoco se incluye información sobre los trabajadores cubiertos por otros sistemas de previsión social distintos al de la Seguridad Social (funcionarios de Clases Pasivas u otras mutualidades) que, a pesar de que tienen retenciones para el IRPF, no forman parte de la población de referencia de la MCVL (MTAS, 2006d).

La información del ‘BLOQUE5’ se presenta en tantos registros como retenedores haya tenido el individuo en el año de referencia de la MCVL, bien sea el retenedor un empresario, bien una entidad pública o privada en el caso del cobro de una pensión o prestación por desempleo. Los datos ofrecidos por la Agencia Tributaria aluden, de este modo, a los ingresos percibidos, tanto en dinero como en especie, y a las retenciones aplicadas en cada una de las olas de la MCVL y que están sujetos al IRPF a través del modelo 190 (MTAS, 2006d). La información puede adquirirse, en consecuencia, carácter longitudinal si se enlazan las 5 ficheros distribuidos hasta el momento -uno por cada ola disponible de la MCVL CDF-, pero no se dispone de información retrospectiva con anterioridad al año 2004.

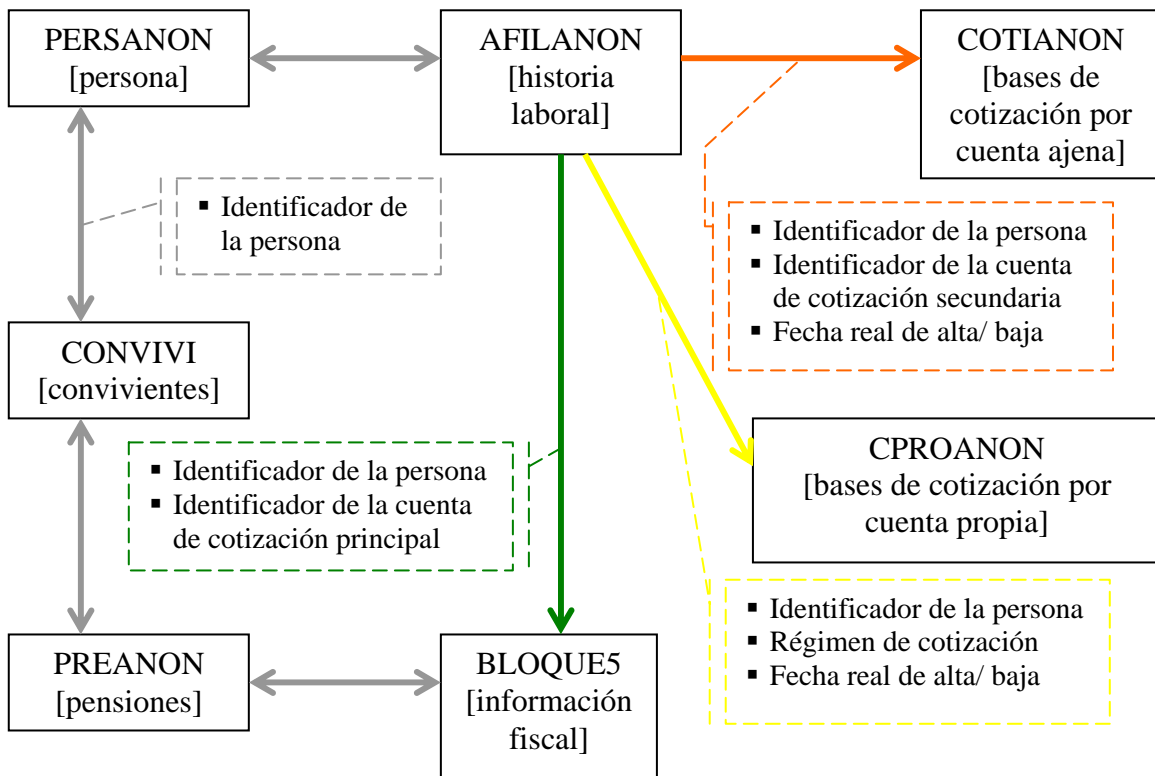
A los archivos hasta aquí descritos se añade otro, denominado ‘**DIVISION**’, cuya finalidad es meramente operativa y dirigida a localizar la información de cada persona en los ficheros ‘AFILANON’ y ‘COTIANON’ que, tal y como hemos explicado anteriormente, están divididos en varios subficheros para facilitar su uso. Por lo tanto, el fichero ‘DIVISION’ solo contiene tres variables: el identificador de la persona, el número del fichero en que se localizan las relaciones administrativas mantenidas con la Seguridad Social y, por último, el número del fichero en el que se encuentran sus bases de cotización por cuenta ajena (MTAS, 2006d). De este modo, el fichero ‘DIVISION’ contiene el identificador de todos los individuos que forman parte de la MCVL en cada ola, con independencia de si tienen asignado o no un valor en los dos últimos campos.

rendimientos de trabajo, de determinadas actividades económicas, premios y determinadas imputaciones de renta”. Está disponible en: <http://www.boe.es/boe/dias/2004/10/29/pdfs/A35759-35794.pdf>.

¹² Sí están incluidos en la MCVL los trabajadores autónomos cuando desempeñan su actividad en una empresa por cuenta ajena y el tipo de retribución que perciben está sujeta a retención (MTAS, 2006d).

Una vez descrita la organización general de la MCVL y el contenido de cada uno de los ficheros, es necesario avanzar en otra cuestión de carácter operativo: la fusión de la información. La figura 2 refleja el procedimiento a seguir, indicando en los rectángulos de línea discontinua las variables a través de las cuáles podemos unir la información de los distintos ficheros (representados por los rectángulos de línea continua y de color negro).

Figura 2: Fusión de los ficheros de la MCVL.



Fuente: Elaboración propia.

La variable ‘identificador de la persona’ (en adelante ‘ip’) funciona como nexo de unión de todos los ficheros y constituye el primer campo de todos ellos. No obstante, la fusión de algunos ficheros opera a través de variables complementarias. Así, el fichero ‘AFILANON’, que recoge la historia laboral de los individuos, y los ficheros ‘COTIANON’ y ‘CPROANON’, que reflejan las bases de cotización mensuales, tienen como nexo de unión, además del ‘ip’, otras tres variables; el ‘identificador de la cuenta de cotización secundaria’, en el caso del primer fichero, el ‘régimen de cotización’, en el caso del segundo, y la ‘fecha real de alta’ y la ‘fecha real de baja’ para ambos. Estas variables permiten, en consecuencia, unir cada relación laboral con su base de cotización correspondiente. La fusión del fichero ‘AFILANON’ con la información fiscal del ‘BLOQUE5’ también opera a través de una variable complementaria: el identificador de la cuenta de cotización principal, que en el ‘BLOQUE5’, incluido en la versión CDF, adopta el nombre de identificador de la entidad pagadora. En este sentido, cabe señalar, que se mantienen los mismos códigos anónimos para las entidades pagadoras en las sucesivas olas de la MCVL CDF.

4. Principales ventajas y limitaciones de la MCVL.

La descripción realizada en los dos apartados anteriores evidencia, ya de antemano, el enorme potencial analítico de la MCVL para el estudio del mercado de trabajo y de la previsión social en España. Entre sus ventajas cabe señalarse, en primer lugar, las derivadas del modo en que se ha definido su población de referencia y que, tal y como adelantábamos en el primer apartado, está compuesta por todos los individuos que han tenido una relación administrativa con la Seguridad Social en ‘algún momento del año de referencia’, con independencia de su duración (MTAS 2006b: 3). Esta definición temporal evita, por un lado, la influencia que el calendario tiene sobre el proceso de altas y bajas laborales (especialmente las de carácter estacional) (MTAS 2006a, 2006b). Por otro lado, permite una mejor representación de las personas que solo trabajan ocasionalmente. Un colectivo bastante numeroso en España que añade a la población de referencia de la MCVL alrededor de tres millones de personas y que no son captadas tan fidedignamente en otras fuentes de información que definen la población a partir de su situación en una fecha concreta (MTAS, 2006a). Este es el caso, por ejemplo, de la Encuesta de Población Activa (EPA), la otra fuente de información imprescindible para el estudio del mercado de trabajo en España. Una característica que resulta todavía más relevante si tenemos en cuenta que las personas con mayor número de entradas y salidas en el mercado de trabajo son, fundamentalmente, mujeres y jóvenes (García Pérez, 2008).

En segundo lugar, frente a los análisis transversales que permiten establecer asociaciones entre variables, su estructura de tipo panel¹³ posibilita el desarrollo de estudios longitudinales y la aplicación de técnicas de análisis de historia de acontecimientos basadas en el concepto clave de ‘curso de vida’. Desde esta perspectiva, la trayectoria de los individuos es interpretada como una serie de transiciones que, a su vez, son el resultado de la situación del individuo en un momento determinado pero, también, de las decisiones tomadas en el pasado y que, en cierta manera, condicionaron transiciones posteriores (Mortimer y Shanahan, 2003). Con la MCVL podemos reconstruir las trayectorias laborales y, por ende, investigar cuáles son los mecanismos causales que generan procesos de cambio en la relación de los individuos con el mercado de trabajo. Además, al proveer las fechas exactas de inicio y fin de las relaciones laborales, la MCVL posibilita la aplicación de técnicas de análisis en tiempo continuo¹⁴. El periodo de análisis es, además, más amplio que el de la EPA, limitado a un año y medio de duración. Un periodo relativamente corto si lo comparamos con los cinco años que la MCVL provee en la actualidad y que irán aumentando con las sucesivas olas.

El tercer valor añadido radica en su gran tamaño muestral¹⁵. La MCVL permite estudiar colectivos relativamente pequeños -que en otro tipo de encuestas no suelen ser significativos- o incluso profundizar en el análisis de las diferencias territoriales y el impacto de algunas políticas desarrolladas por las Comunidades Autónomas. De hecho, la MCVL abre la posibilidad de desarrollar diseños de investigación cuasi-experimentales, gracias a los cuales podemos contrastar los resultados obtenidos tras la implementación de una política autonómica concreta (grupo de tratamiento) con los de

¹³ Su estructura panel presenta algunas peculiaridades que serán explicadas en el apartado siguiente.

¹⁴ Véase Bernardi (2006) para una explicación más detallada de estas técnicas.

¹⁵ Ver el cuadro 1 del primer apartado.

otras comunidades que no la han aplicado (grupo de control). Ello es factible debido a que la MCVL identifica la localidad de residencia cuando ésta es superior a los 40.000 habitantes.

Sin embargo, la MCVL dista mucho de ser una fuente de información perfecta. Entre sus limitaciones cabe destacar, en primer lugar, que sus datos se organizan en torno a la persona física. Por lo tanto y a diferencia de la EPA, no es posible conectar las trayectorias laborales de los individuos que integran un hogar sino, simplemente, incluir en el análisis algunas de sus características¹⁶. En segundo lugar, al tratarse de una fuente de información que recoge las relaciones administrativas mantenidas con la Seguridad Social, no refleja a los colectivos que quedan al margen de su acción protectora: los trabajadores de la economía sumergida, inactivos o desempleados que no reciben prestaciones de carácter contributivo (a excepción de los perceptores del subsidio por desempleo) y las personas cubiertas por otros sistemas de previsión social (funcionarios de Clases Pasivas o algunos profesionales liberales). Sin embargo, tampoco recoge información de todos los actos administrativos asociados a derechos de carácter contributivo. Por ejemplo, la MCVL no dispone de información de los periodos de suspensión de contrato por adopción o acogimiento, riesgo durante el embarazo o la lactancia, o de cobro de las prestaciones por maternidad y paternidad (pero sí sobre la excedencia por cuidado de hijos menores de tres años o la reducción de jornada por guarda legal). Por estos motivos, es importante destacar que la MCVL no es una fuente de información alternativa a la EPA, sino complementaria.

Además, los investigadores deberán hacer frente a otras limitaciones que hunden sus raíces en el origen administrativo de los datos y que, en cualquier caso, son rasgos comunes a todas las fuentes de información que se nutren de registros (Marcos, 2006; Durán, 2007; Wallgren y Wallgren, 2007; García-Segovia y Durán, 2008). Entre ellas cabe señalar el elevado porcentaje de registros incompletos. En la mayor parte de administraciones el proceso de recogida de datos solo se realiza de forma exhaustiva en aquellos campos cuya cumplimentación es estrictamente obligatoria. En el caso de la Seguridad Social están relacionados con la gestión de las prestaciones o las cotizaciones de los ciudadanos. El resto de la información demandada puede no ser facilitada por los propios ciudadanos o incluso no ser introducida con el mismo nivel de exigencia y calidad por el propio personal técnico. De ahí el elevado porcentaje de casos perdidos en algunas variables de la MCVL. Ejemplo de ello es la variable 'provincia de nacimiento', ya que solo contiene información para una cuarta parte de los individuos seleccionados. Además, en caso de que se recoja la información, no significa que su contenido esté actualizado. De hecho, algunos individuos tienen escaso contacto con la Administración y no siempre notifican los cambios que tienen pocas o nulas repercusiones en la gestión de sus prestaciones. En caso de hacerlo, tampoco suelen comunicarlos de forma inmediata. La variable 'nivel educativo', extraída del Padrón Municipal, ejemplifica esta problemática¹⁷.

En ocasiones resulta también muy complicado traducir el lenguaje jurídico que caracteriza a la MCVL en otro más sociológico. Los cambios legislativos acontecidos en las últimas décadas plantean dificultades añadidas en el manejo e interpretación de

¹⁶ Ver la dificultad 5 del próximo apartado.

¹⁷ García Pérez (2008) compara la distribución de esta variable en la ola del 2007 con los valores que se obtienen en el segundo cuatrimestre de la EPA en ese mismo año y llega a la conclusión de que el nivel de educación en la MCVL está muy sesgado a la baja (García Pérez, 2008: 9).

sus datos¹⁸. Tal y como señalan Argimón y González (2006), al introducir nuevas excepciones y casos particulares incrementan el detalle de la información disponible en la MCVL, pero simultáneamente aumentan las dificultades a la hora de realizar el análisis. Ejemplo paradigmático de ello es la variable ‘tipo de contrato’ que recoge más de 500 modalidades distintas, algunas de ellas abolidas en la actualidad y otras prácticamente marginales, pero que, en cualquier caso, han estado presentes en la legislación laboral española en el transcurso de los años. El esfuerzo que el investigador deberá dedicar a la comprensión, depuración y recodificación de las variables es, en consecuencia, más elevado con la MCVL que cuando se trabaja con fuentes de información primaria. A las limitaciones planteadas hasta aquí hay que sumar algunas dificultades concretas en el manejo de sus datos y que, por sus implicaciones y características específicas, pasamos a detallar en los dos siguientes apartados.

5. Las dos formas de abordar el trabajo con la MCVL: la perspectiva estática vs. la dinámica.

Al igual que la Encuesta de Población Activa y otras fuentes de información de características similares, la MCVL permite la explotación de su información trasversal o longitudinalmente. Sin embargo, ambas perspectivas cuentan en la MCVL con ciertas peculiaridades de las que tenemos que ser conscientes cuando iniciemos el trabajo con esta base de datos.

La primera de ellas, que denominamos **transversal o estática**, consistente en la utilización de una única ola cuya elección depende del interés del investigador. Esta forma de explotación es la más factible en el corto y medio plazo, pero conlleva una cautela: ser rigurosos en el tratamiento del ámbito temporal al que hace referencia la MCVL. La información retrospectiva que nos ofrecen algunos de sus ficheros puede hacernos perder de vista que la MCVL es únicamente representativa de aquellos individuos que, en el año al que alude cada versión de la MCVL, tuvieron una relación administrativa con la Seguridad Social, pero no de las cohortes anteriores. Por lo tanto, cuanto más atrás nos remontemos en la reconstrucción de la vida laboral de los individuos, más probabilidades habrá de que ésta no refleje la realidad de su época.

El motivo es doble; en primer lugar, porque cuanto más antigua es la información, más factible es también que existan lagunas o incorrecciones en los registros de la Seguridad Social (MTAS, 2006a: 123); y, en segundo lugar, porque estaremos analizando a personas que han ‘sobrevivido’ en el mercado de trabajo. Es decir, estas personas se caracterizan primordialmente porque, en el año al que hace referencia la MCVL, continuaban manteniendo una relación administrativa con la Seguridad Social. En consecuencia, no son representativas de los individuos que tenían relación con esta institución en épocas pasadas, ya que muchos de ellos -posiblemente los que coinciden con perfiles de mayor inestabilidad- habrán abandonado el mercado de trabajo. En definitiva, cuando utilicemos una de las olas de la MCVL no podemos caer en la tentación de realizar análisis dinámicos con la información retrospectiva que nos ofrece.

No obstante, como cada año se facilita una nueva ola de la MCVL manteniéndose la misma definición de la población de referencia y el mismo criterio de selección de los

¹⁸ Recordamos que la MCVL incluye información de las bases de cotización desde la década de los 80 y que la reconstrucción de la vida laboral de los individuos es posible desde principios de los 60.

individuos¹⁹, sí es posible realizar este tipo de explotación. El análisis **longitudinal o dinámico** requiere el montaje de una base de datos tipo panel que, en este caso, presentará algunas peculiaridades derivadas de los procesos de entrada y de salida de la MCVL y que la alejan de una estructura panel convencional. Por lo tanto, como paso preliminar, resulta imprescindible entender como se producen estos procesos. El cuadro 3, que adjuntamos a continuación, refleja las distintas casuísticas que, en este sentido, podemos encontrar en la MCVL.

Cuadro 3: Ejemplo de entradas y salidas de los individuos en las olas de la MCVL.

<i>Identificador personal</i> (‘ip’):	Olas de la MCVL:				
	2004	2005	2006	2007	2008
0000001	x	x	x	x	x
0000022	x	-	-	-	-
0000333	-	-	-	-	x
0004444	x	-	-	-	x

Fuente: Elaboración propia

El ‘ip 0000001’ es el caso más fácil, ya que representa a una persona que se mantiene en las 5 olas de la MCVL disponibles en la actualidad. Bajo esta pauta podríamos ubicar, por ejemplo, a aquellos individuos que han tenido un contrato de trabajo estable en el periodo 2004-2008, a estudiantes que trabajaron durante varios veranos consecutivos y posteriormente consiguieron un contrato en prácticas o a los individuos que han pasado a cobrar una pensión de jubilación tras su paso por un periodo de percepción de la prestación por desempleo. A pesar de representar historias de relaciones administrativas con la Seguridad Social muy distintas, todos ellos se caracterizan por formar parte de la población de referencia de la MCVL en cada una de las olas. De ahí que su ‘ip’ aparecerá incluido en todas ellas. Según García Pérez (2008) el 75% de los individuos incluidos en la ola del 2004 han sido observados en las tres olas posteriores.

Los tres ‘ip’ siguientes representan a individuos que no aparecen en alguna de las olas de la MCVL. El ‘ip 0000022’ identifica a una persona incluida en la ola del 2004, pero no así en las posteriores. Los motivos de esta salida pueden ser diversos y entre ellos se encuentran, por ejemplo, el fallecimiento del individuo, su paso a una situación de inactividad o de desempleo -pero, en este último caso, sin cobro de prestaciones-, o a cualquiera de las situaciones que originan la exclusión de la población de referencia de la MCVL. Los individuos que desaparecen son, en mayor proporción, mujeres, inmigrantes y personas mayores de 65 años (García Pérez, 2008).

Por el contrario, el ‘ip 0000333’ representa a un individuo que se incorpora en la ola del 2008, como consecuencia de los procesos de renovación de la muestra. Tal y como avanzábamos en el primer apartado, cada año permanecen en la MCVL los individuos que ya estaban y continúan teniendo relación con la Seguridad Social²⁰ y se incorporan a ella nuevas personas que, por primera vez, son cotizantes o pensionistas y que

¹⁹ Ver la explicación del epígrafe 2 sobre el tipo de muestreo utilizado en la MCVL.

²⁰ En la medida en que el individuo no cambie de código de identificador personal. Este problema se explica con detalle en el marco de la ‘dificultad 1’ del apartado siguiente.

cumplen los mismos criterios de selección que los anteriores. Por ambos motivos, las sucesivas olas de la MCVL son representativas de las altas de la población de referencia y funcionan como un panel.

Mención específica requiere el 'ip 0004444' que aparece en la primera y quinta ola, correspondientes respectivamente a los años 2004 y 2008, mientras que desaparece en las restantes. En esta pauta se enmarcan personas con una mayor inestabilidad en el mercado de trabajo. Así, por ejemplo, podríamos pensar en un joven que ha tenido uno o varios contratos temporales de corta duración a lo largo del 2004, permaneciendo en situación de desempleo y sin cobrar ninguna prestación de carácter contributivo entre el 2005 y 2007 y comenzando a trabajar nuevamente en el 2008. En ella también encaja una mujer con cargas familiares que, por ejemplo, abandona transitoriamente el mercado de trabajo tras el nacimiento de un hijo y, después de un tiempo, se incorpora nuevamente al mercado de trabajo.

La peculiaridad de un panel con la MCVL es que, para aquellas personas con pautas de entrada y salida semejantes al 'ip 0004444', tendremos información sobre la evolución de su historia laboral, pero desconoceremos los cambios que se puedan producir en otras variables cuando esa persona no formaba parte de la muestra, siempre que éstos no tengan efectos perdurables en el tiempo. Por ejemplo, puede que el joven identificado con el 'ip 0004444' aprovechase el periodo de desempleo para formarse y obtuviese un grado de formación profesional en el 2006. En la MCVL ese cambio aparecerá reflejado en la ola del 2008 en la variable nivel educativo²¹. A partir de la información de la MCVL desconocemos, por lo tanto, cuándo se produce esa variación, pero podemos detectarla posteriormente. No ocurriría lo mismo en el caso, por ejemplo, de que ese joven cambiase temporalmente de municipio de residencia con el objetivo de buscar trabajo. La modificación de esta variable no quedará registrada si no se mantiene en la fecha de extracción de la MCVL del 2008. La estructura longitudinal de la MCVL no se ajusta, por lo tanto, a la de un panel convencional en el que se observan a todos los individuos durante un periodo concreto.

A la hora de abordar el montaje de la base de datos y vistos estos cuatro casos podríamos pensar que, si trabajamos con los ficheros de la ola del 2008 que contienen información retrospectiva -'AFILANON', 'COTIANON', 'CPROANON' y 'PREANON'-, prácticamente tenemos montado un panel para el periodo 2004-2008. Solo faltaría añadir la información del 'ip 0000022', incluida en la ola del 2004. Sin embargo, esta forma de construcción de la base de datos de adelante (o desde la última ola disponible) hacia atrás (recuperando los individuos que han salido de la muestra) tiene un inconveniente. En la MCVL, junto con la información de carácter retrospectivo, coexiste información relativa a la fecha de extracción. De este modo, el fichero 'PERSANON' contiene variables cuyos valores pueden sufrir modificaciones con el transcurso del tiempo como, por ejemplo, el nivel educativo o el municipio de residencia. Igualmente, el fichero 'CONVIVI' puede reflejar cambios en el número de personas empadronadas en la misma vivienda. A ellos hay que añadir, la información relativa al empleador o responsable de la cotización del archivo 'AFILIANON' que, tal y como adelantamos en el segundo apartado, también hacen referencia a la fecha de extracción.

²¹ Suponiendo que ese individuo haya notificado el cambio en el padrón municipal (ver lo comentado en este sentido en el apartado anterior).

Por este motivo, el panel con la MCVL se construye, al igual que en otras fuentes de información de estas características, de atrás hacia adelante. La única precaución que deberemos adoptar es borrar aquellos registros del fichero 'AFILANON' cuya fecha de baja sea 31 de diciembre del año siguiente al que hace referencia la ola. Ello es debido a que las relaciones administrativas censuradas por la derecha, es decir, aquellas que todavía siguen en vigor en la fecha de extracción de la MCVL aparecen, por defecto, con este valor. Por lo tanto, si un individuo dispone de un contrato indefinido desde el 2006 hasta la actualidad, en la ola del 2007 presentará como 'fecha real de baja' (nombre que adopta la variable) el 31 de diciembre del 2008. Para añadir la información de la ola del 2008 tendremos, en consecuencia, que eliminar dicho registro de la ola del 2007.

6. Dificultades a tener en cuenta en la explotación estadística de la MCVL.

La MCVL es una fuente de información excepcional. Sin embargo, a la hora de iniciar su explotación estadística es necesario, además de una buena dosis de paciencia, no perder de vista varias consideraciones de carácter general. Algunas de ellas se traducen en dificultades más o menos salvables, mientras que otras constituyen rasgos de la MCVL con los que el investigador deberá aprender a convivir. En cualquier caso, todas constituyen peculiaridades de la MCVL a las que, inevitablemente, el investigador deberá enfrentarse y sobre las que tendrá que tomar decisiones. En la exposición de cada una de ellas se apuntan posibles líneas de solución con el ánimo de facilitar, en la medida de lo posible, este proceso.

Dificultad 1. Inconsistencias derivadas del proceso de fusión ('*matching*') de las distintas fuentes de información de la MCVL.

Todos los ficheros que componen la MCVL se unen entre sí, tal y como adelantábamos en el segundo apartado, por el 'identificador de la persona' ('ip'). El 'ip' está basado en el Documento Nacional de Identidad (DNI) -o NIE en el caso de los extranjeros-, aunque se presenta transformado para preservar el anonimato de los individuos incluidos en la muestra. De este modo, el DNI es una de las variables utilizadas para relacionar los datos de un mismo individuo en las tres fuentes de las que se extrae la información. Sin embargo, su uso presenta algunos problemas que se visualizan de forma directa en los ficheros de la MCVL.

El primero de ellos está relacionado con la asignación del número identificativo a cada individuo. A pesar de que, en la actualidad, cada persona cuenta con un número único e intransferible que no es adjudicado a otra tras su fallecimiento, no ocurría así en épocas pasadas. Por este motivo en la MCVL encontramos 'ip' duplicados. El cuadro 4 refleja el alcance de este problema, mostrando el número de 'ip' repetidos en los ficheros 'PERSANON' de las sucesivas olas de la MCVL en su versión CDF. Así, en el año 2004 encontramos un total de 508 'ip' repetidos y 595 en el 2005, cifras que descienden hasta situarse en 503, 450 y 397 en las olas del 2006, 2007 y 2008 respectivamente.

La última columna del cuadro 4 recoge el número total de registros afectados por las duplicaciones. En concreto, la cifra se sitúa en 1.017 en la ola del 2004 y se reduce de forma progresiva hasta alcanzar los 822 registros en la del 2008. En ningún caso, el

número total de registros afectados por este error supera el 0,1% de los incluidos en cada ola. Su eliminación no afecta, en consecuencia, a la representatividad de la MCVL y resulta recomendable a la hora de abordar el proceso de fusión de los distintos ficheros.

Cuadro 4: Identificadores personales ('ip') repetidos en el fichero 'PERSANON' de la MCVL CDF, según las sucesivas olas.

<i>Olas:</i>	'ip' repetidos			Registros a eliminar
	2 veces	3 veces	4 veces	
2004	507	1	-	1.017
2005	594	1	-	1.189
2006	485	17	1	1.025
2007	432	14	4	922
2008	372	22	3	822

Fuente: Elaboración propia.

El segundo problema que plantea la utilización del DNI radica en que no todas las personas están obligadas a su obtención y utilización. Así pues, algunos individuos mantienen una relación con la Seguridad Social pero carecen de DNI como, por ejemplo, los menores con pensión de orfandad, los españoles residentes en el exterior o algunos extranjeros. En estos casos se utiliza otro código identificativo a la hora de extraer la información obrante en la Seguridad Social: el 'número de afiliación' (NAF). El NAF es el código en torno al que se organizan los datos de la Seguridad Social y que es asignado de forma automática por esta institución a las personas que entran en su ámbito de aplicación. Por lo tanto, a partir de este identificador, es posible recuperar toda la historia de relaciones administrativas con la Seguridad Social de aquellos individuos que no tienen DNI, pero no se puede fusionar la información procedente de los registros de la Agencia Tributaria o el Padrón Municipal. Además, el NAF tampoco está exento de problemas, ya que puede darse el caso -a pesar de ser poco habitual- de que una misma persona tenga asignado más de uno. En consecuencia, un porcentaje pequeño de las incorporaciones o desapariciones de individuos entre una ola y otra de la MCVL no serán debidas al fallecimiento o a variaciones en la actividad sino, simplemente, a un cambio del NAF (MTAS, 2006a).

Por su parte, la fusión de la información procedente de la Seguridad Social con la del Padrón Municipal se efectúa, además de con el DNI, a través de una segunda variable: la fecha completa de nacimiento del individuo. De este modo, la fusión de ambas fuentes solo se considera válida cuando ambos campos coinciden. Con ello se ha evitado el problema de la duplicidad de algunos DNI, pero no así el de aquellas personas que carecen de él. Para éstas la muestra no incorpora información padronal, tal y como adelantábamos previamente. El cuadro 5 refleja el número de individuos para los que se dispone de información del Padrón Municipal y de la Agencia Tributaria, según las distintas olas.

En la ola del 2005, la primera en la que se incorporó el Padrón Municipal, no se disponía de información procedente de esta fuente para el 17,1% de los individuos incluidos en la MCVL, debido a fallos en el proceso de fusión o porque no existía información sobre ellos. Igualmente, cabe recordar que tampoco se incluye la

información procedente del Padrón Municipal en aquellos casos en que figuran inscritos más de 20 individuos en la misma hoja padronal. No obstante, el porcentaje de casos perdidos se redujo hasta el 4,8% en la siguiente ola, cifra que se mantiene estable en la actualidad. La excepción la constituye la ola del 2007, en la que el 17,3% de los individuos no tienen información padronal, debido a causas desconocidas.

Cuadro 5: Individuos incluidos en la MCVL CDF con información del Padrón Municipal y de la Agencia Tributaria, según las sucesivas olas.

<i>Olas:</i>	Padrón Municipal		Agencia Tributaria	
	N.	% ⁽¹⁾	N.	% ⁽¹⁾
2004	-	-	913.875	83,9
2005	946.350	82,9	976.726	85,6
2006	1.113.920	95,2	1.006.499	86
2007	992.678	82,7	1.048.692	87,4
2008	1.160.934	95,7	1.068.805	88,1

⁽¹⁾ Ambas columnas representan el porcentaje de individuos que disponen de la información procedente de cada una de las fuentes respecto al total de individuos incluidos en la MCVL en la correspondiente ola. Fuente: Elaboración propia.

Para la fusión de la información de la Agencia Tributaria se utiliza, además del DNI, el código de identificación del empleador o la entidad pagadora. Así, en la ola del 2005 se dispone de información fiscal para el 83,9% de los individuos, mientras que en la del 2008 la cifra se sitúa en el 88,1%. Sin embargo, la mayor parte de las lagunas de información en los datos fiscales no se justifica por los fallos del proceso de fusión, sino en el ámbito de aplicación del modelo 190 del que se extrae esta información. De hecho, tal y como se ha señalado en el segundo apartado, la Agencia Tributaria no ofrece información de los ingresos de los trabajadores autónomos ni de los residentes en Navarra o el País Vasco, al tener ambas comunidades un régimen fiscal propio²².

Por lo tanto, cuando se proceda al montaje de la base de datos conviene identificar a los individuos que no disponen de información padronal o fiscal. En el caso de la el caso de la primera fuente resulta sencillo, ya que el fichero ‘CONVIVI’ solo incluye los ‘ip’ para los que se ha localizado información padronal, aunque puede darse el hecho de que no haya más personas empadronadas con el individuo incluido en la muestra. En el caso de la información procedente de la Agencia Tributaria habrá que distinguir entre los casos perdidos (aquellos de los que no tenemos información por errores en el proceso de fusión) de los casos en que no procede dicha información, al estar fuera del ámbito de aplicación del modelo 190. No obstante, su identificación es posible a partir de las variables domicilio de residencia y régimen de cotización.

²² Ver la explicación referida a la información del fichero ‘BLOQUE5’ incluida en el segundo apartado para obtener una descripción más detallada sobre los colectivos de los que no se dispone de información fiscal en la MCVL.

Dificultad 2. El tratamiento del pluriempleo y las situaciones simultáneas.

La unidad de análisis de la MCVL es el individuo, aunque gran parte de la información que se incluye en los distintos ficheros se refiere a situaciones (por ejemplo; trabajar, cobrar una prestación por desempleo, una pensión, etc.). Por lo tanto, una misma persona puede haber disfrutado de varias de estas situaciones a lo largo de su vida. Algo que a priori puede resultar obvio y de escasa complejidad a la hora de abordar el trabajo con la MCVL. Lo que no lo es tanto es el hecho de que estas situaciones también pueden desarrollarse de forma simultánea. Por ejemplo, una persona puede presentar fases de pluriempleo -imaginemos un médico que trabaja en un hospital, pero paralelamente lo hace como autónomo en una consulta privada- o también puede trabajar y recibir una pensión al mismo tiempo. Este último es el caso de las jubilaciones a tiempo parcial o el de aquellas pensiones cuya percepción es compatible con la actividad remunerada. En definitiva, cada persona puede pertenecer simultáneamente a varias categorías distintas, tanto a lo largo de toda su vida como en momentos puntuales.

Esta posibilidad plantea una dificultad extra a la hora de trabajar con la MCVL ya que, a diferencia de lo que ocurre con la EPA en dónde el entrevistado indica cuál es su situación o empleo principal, en la MCVL es el propio investigador quien tiene que realizar, de forma deductiva, esta apreciación²³. Sin embargo, este problema resulta más gravoso a la hora de realizar análisis multivariados que en la construcción de indicadores descriptivos. En este último caso, la solución pasa por clarificar la unidad de observación (individuos vs. situaciones). Por poner un ejemplo, no es lo mismo hablar de la distribución según sexo y edad de los afiliados dados de alta en la Seguridad Social en el 2008 que de las altas en la Seguridad Social en ese mismo año. El resultado de este último indicador siempre será más elevado, debido a que un mismo individuo puede presentar varias altas en un mismo año e incluso puede tener dos altas simultáneas en regímenes distintos.

En cambio, si se abordan análisis multivariados, el investigador deberá definir previamente los criterios de selección de la situación (por ejemplo, empleado vs. pensionista) o el empleo principal. Así, en el caso de la primera la decisión dependerá fundamentalmente del objeto de estudio del investigador, mientras que en el caso de la segunda los criterios deberán ser explícitos y construirse a partir de la información que nos ofrece la MCVL. Por ejemplo, a la hora de determinar el contrato principal en las situaciones de pluriempleo, Ramos (2007) escoge; en primer lugar, el que tiene un carácter indefinido; en segundo lugar y en caso de igualdad, el contrato de mayor duración; y como tercera opción en caso de que se mantenga la coincidencia, el contrato con primera fecha de alta (Ramos, 2007: 3). La duración es, por ende, el criterio de selección más idóneo en las situaciones simultáneas de trabajo por cuenta ajena y propia.

²³ El documento del MTAS que describe la MCVL reconoce que este es uno de los aspectos que “más dolores de cabeza dará a los usuarios” (MTAS, 2006b: 4).

Dificultad 3: Cómo distinguir una relación contractual en los ficheros ‘AFILANON’ de la MCVL y calcular la duración de los episodios de empleo.

Una de las informaciones más interesantes que podemos obtener a partir de la MCVL es la duración media de los episodios laborales de los individuos. Para ello, hay que utilizar las variables ‘fecha real de alta’ y ‘fecha real de baja’ del fichero ‘AFILANON’. De este modo, su duración equivaldría al tiempo transcurrido entre ambas fechas.

Sin embargo, a la hora de realizar este cálculo deben tenerse en cuenta dos precauciones. La primera de ellas es que no todos los episodios que recoge la MCVL coinciden con altas laborales, es decir, periodos en los que el individuo está efectivamente ocupado. También incluye relaciones en las que una persona mantiene su cotización a la Seguridad Social, pero no está ocupado. Este es el caso de los episodios amparados bajo algún Convenio Especial o cuando se percibe una prestación o subsidio por desempleo que reconoce este derecho. Por este motivo es muy importante identificar a qué situación se corresponde cada periodo de cotización y discernir los episodios de empleo, ya sea en su modalidad por cuenta ajena o propia, del resto²⁴.

La segunda precaución está relacionada con la forma en que se organiza la información en el fichero ‘AFILANON’. Tal y como explicábamos en el segundo apartado, cada relación administrativa mantenida con la Seguridad Social viene recogida, como regla general, en un registro, pero existen excepciones. De hecho, algunos contratos laborales, originan más de una relación administrativa con la Seguridad Social y, en consecuencia, vienen reflejados en más de un registro. Esto sucede, por ejemplo, cuando cambia algún aspecto fundamental del contrato, especialmente el grupo de cotización o el coeficiente de tiempo parcial²⁵ o cuando se produce alguna de las circunstancias que dan lugar a una suspensión del mismo, pero se mantiene la reserva del puesto de trabajo o el derecho a la reincorporación laboral²⁶. De ahí la necesidad de ser cuidadosos en la identificación de la fecha de inicio y fin de una relación contractual, ya que lo que puede estar pareciendo un cambio de contrato puede tratarse, en realidad, de un mero ajuste de carácter administrativo.

A ello hay que añadir que la Seguridad Social también ha ido modificando el procedimiento de recogida de datos y, desde hace algunos años, no siempre se generan nuevos registros cuando varían el grupo de cotización o el coeficiente de tiempo parcial e incluso el tipo de contrato. Por esta razón, a partir de la ola del 2005, se incorporaron en la MCVL ocho nuevas variables que identifican los cambios sufridos dentro de una misma relación laboral²⁷ en los tres aspectos mencionados. De este modo, en cada

²⁴ Ello es posible gracias a la variable ‘régimen de cotización’ que identifica a los individuos que cotizan a través de un Convenio Especial con los códigos 140, 540, 640, 740, 840, 940, y 1240 y a la variable ‘tipo de relación laboral’ que señala a los individuos que perciben la prestación o el subsidio por desempleo con el código 400, del 751 al 756 y el 980.

²⁵ Recordamos que esta variable indica la duración de la jornada de trabajo.

²⁶ El artículo 45 del Estatuto de los Trabajadores recoge las distintas causas y los efectos de las suspensiones de los contratos. Sin embargo, solo algunas son directamente identificables en la variable ‘causas de baja’ de la MCVL. Entre ellas cabe señalar la baja por servicio militar (código 61), por incapacidad transitoria (códigos 65 y 70), por excedencia voluntaria o forzosa (código 63), por huelga legal (código 64), de excedencia por cuidado de hijos (código 68) y de excedencia por cuidado de familiares (código 73).

²⁷ Por este término entendemos la existencia de un vínculo contractual entre un empresario (identificado en la MCVL por el ‘código de la cuenta de cotización secundaria’) y un trabajador, pero que puede estar

registro se recoge el último valor que han adoptado en cada relación laboral y, de manera adicional, también se reflejan sus dos valores iniciales.

Por lo tanto, las relaciones contractuales están reflejadas en la MCVL de distintas formas. La imagen 3 ilustra, a través de un extracto del fichero 'AFILANON', las tres casuísticas que podemos encontrarnos en este sentido; desde la más sencilla (un contrato por cada registro) hasta las más complejas (bien porque un mismo contrato ha dado lugar a distintos registros, o bien porque en un mismo registro se refleja una relación laboral compuesta por varios contratos). Las columnas muestran una selección de variables que nos ayudarán a entender cómo se refleja la información sobre los contratos y, lo que es todavía más importante, sus implicaciones posteriores en el manejo de los datos. Las filas representan los registros del fichero.

El 'ip 0000018' ejemplifica el caso más fácil: cada uno de los tres registros que presenta este individuo muestra una relación contractual diferente. En concreto, podemos observar que esta persona tuvo un contrato temporal por obra y servicio con una duración total de tres meses en el año 2004, relación que queda recogida en el registro número 1. Posteriormente, con fecha 15 de septiembre de 2005 inició un contrato de carácter temporal y formativo (registro número 2) que, un año después, dio lugar a un contrato indefinido ordinario para la misma empresa²⁸ que a fecha de extracción de la ola del año 2007 seguía en vigor (registro número 3).

No ocurre lo mismo para el 'ip 0000182', en el que la información de un mismo contrato queda recogida en dos registros distintos (el número 4 y el 5). De hecho, esta persona fue contratada indefinidamente el 7 de octubre de 2002, situación que se mantuvo hasta el 30 de noviembre de 2007. Sin embargo, este contrato fue suspendido el 16 de marzo del 2005, debido al disfrute de una situación de excedencia por cuidado de hijos (identificada en la variable 'causa de baja' con el código 68 en el registro número 4). Por lo tanto, es la reincorporación al trabajo, acontecida el 15 de septiembre del 2005, lo que da lugar a la aparición de un nuevo registro (el número 5) que, en realidad, refleja la continuidad de una relación contractual iniciada con anterioridad.

El 'ip 0001563' ejemplifica la última casuística y muestra una relación laboral compuesta por varios contratos, que quedan recogidos en un único registro (el número 7). Dicha relación se inició el 1 de febrero de 1999 y continúa en vigor en la fecha de extracción de la ola del 2007. No obstante, las variables 'fecha_1', 'contrato_1' y 'jornada_1' indican que el primer contrato de ese individuo de carácter temporal y formativo y con jornada completa finalizó el 31 de enero del 2001, dando paso a un segundo contrato temporal de dos tercios de la jornada que terminó el 1 de junio de 2003, tal y como revelan las variables 'fecha_2', 'contrato_2' y 'jornada_2'. A partir de esta última fecha desconocemos las modificaciones que haya podido sufrir esa relación laboral, ya que la MCVL solo incorpora a partir de la ola del 2005 sus dos valores iniciales. No obstante, los primeros campos, etiquetados como 'contrato' y 'jornada' reflejan la última situación laboral, siendo ésta la de fijo indefinido a tiempo completo.

formado por un solo contrato o una sucesión de ellos, siempre y cuando éstos se desarrollen de forma consecutiva.

²⁸ Tal y como podemos observar, el código identificador de la cuenta de cotización es el mismo en el segundo y el tercer registro del 'ip 0000018'.

Imagen 3: Ejemplos de relaciones contractuales a partir de un extracto del fichero ‘AFILANON’.

	IP	Contrato ⁽¹⁾	Jornada ⁽²⁾	Fecha alta	Fecha baja	Causa Baja ⁽³⁾	Identificador cuenta cotización	Fecha_1	Contrato_1 ⁽¹⁾	Jornada_1 ⁽²⁾	Fecha_2	Contrato_2 ⁽¹⁾	Jornada_2 ⁽²⁾
1	0000018	401	0	01/07/2004	31/09/2004	54	0111180R00099						
2	0000018	421	0	15/09/2005	14/09/2006	99	0613040047115						
3	0000018	100	0	15/09/2006	31/12/2008	0	0613040047115						
4	0000182	100	0	07/10/2002	16/03/2005	68	0111012043010						
5	0000182	100	0	15/09/2005	30/11/2007	0	0111012043010						
7	0001563	100	0	01/02/1999	31/12/2008	0	0111017022180	31/01/2002	421	0	01/06/2003	501	663

⁽¹⁾ Los códigos de la variable tipo de contrato son los siguientes; ‘100’ contrato indefinido ordinario a tiempo completo; ‘421’ contrato de duración determinada de formación a tiempo completo; ‘501’ contrato de duración determinada por obra y servicio a tiempo parcial.

⁽²⁾ En estas columnas se recoge el coeficiente de trabajo a tiempo parcial que puede oscilar entre el 0 (que equivale a una jornada a tiempo completo) y el 999 (que representaría la reducción total de la jornada).

⁽³⁾ Los códigos de la variable causa de baja son los siguientes; ‘0’ contrato en vigor (a fecha de extracción de la MCVL todavía no se ha producido la baja de ese contrato); ‘54’ baja no voluntaria; ‘68’ baja de excedencia por cuidado de hijos; ‘99’ otras causas de baja.

Fuente: Elaboración propia a partir de casos reales extraídos de la MCVL del año 2007.

Así pues, los casos equiparables al 'ip 0000018' no plantean ningún problema a la hora de proceder al cálculo de la duración de un contrato, ya que cada uno de ellos está recogido en un registro distinto. La duración será, de este modo, la diferencia entre la 'fecha de baja' y la 'fecha de alta'. En cambio, la forma de tratar las dos casuísticas más complejas dependerá del tipo de análisis que se quiera realizar y ambas cuentan con algunos inconvenientes.

En las situaciones en la que una relación contractual está recogida en varios registros, puede optarse -dependiendo del objeto de estudio- por su unificación, previa confirmación de que la relación contractual se ha desarrollado bajo la misma empresa, igual tipo de contrato y de forma consecutiva. Algunos autores recomiendan seguir esta pauta incluso cuando medie una interrupción no superior a 7 ó 15 días (García Pérez, 2008: 26). Para ello es necesario ordenar previamente los registros del fichero 'AFILANON' por las variables 'ip', 'fecha real de alta' 'fecha real de baja' y el código 'cuenta de cotización secundaria' que identifica a la empresa. Sin embargo, en los casos en que se produzca una suspensión del contrato pero se mantenga el derecho a la reincorporación laboral, como en el 'ip 0000182', el periodo de interrupción laboral puede ser superior a los 15 días. En consecuencia, también habrá que identificar los registros que presenten dentro de la variable 'causa de baja laboral' alguno de los códigos que identifican estas situaciones²⁹ y fusionarlos con el registro siguiente, siempre y cuando también se correspondan con periodos de empleo para la misma empresa y bajo la misma forma contractual.

Cabe advertir que al proceder de este modo podemos correr el riesgo de estar sumando duraciones de contratos que, en realidad, son diferentes. Pensemos, por ejemplo, en un trabajador que ha encadenado tres contratos por obra y servicio para la misma empresa. Para discernir estos casos, conviene señalar que la mayoría de ellos presentan el código 54 dentro de la variable 'causa de baja', que indica la baja no voluntaria por cese o finalización del contrato. En cambio, resulta imposible detectarlos cuando no consta el motivo de la baja (codificado con el valor 0 en esta variable).

Por el contrario, en los casos semejantes al 'ip 0001563', en vez de fusionar los registros, habrá que dividirlos, generando tantos registros como modificaciones haya tenido esa relación laboral en el tipo de contrato. Dado que la MCVL solo recoge información sobre los dos primeros valores y el último, tendremos que crear un máximo de tres. De este modo, el registro número 7 daría lugar a tres nuevos, tal y como muestra la imagen 4.

Así, el nuevo registro 7 adquiriría los valores del primer contrato, reflejados en las siguientes variables de la imagen 3; 'fecha de alta', que indica el inicio de la relación contractual; 'fecha_1', que indica la fecha de finalización; 'contrato_1', que recoge el tipo de contrato; y, por último 'jornada_1', que señala el coeficiente de trabajo a tiempo parcial. El registro número 8 adoptaría los valores de la segunda modificación, manteniendo las pautas antes indicadas. No obstante, en este caso, la fecha de alta sería el día siguiente a la finalización del primer contrato. Para terminar, el registro número 9 tomaría los valores de las variables 'contrato' y 'jornada'. Las variables 'identificador de la cuenta de cotización secundaria' y 'causa de baja' adoptarían los mismos valores,

²⁹ Ver la nota 26.

aunque en esta última variable también puede señalarse la finalización de los dos primeros contratos con el código 54.

Imagen 4: Ejemplo de conversión de una relación laboral a partir de un extracto del fichero ‘AFILANON’.

	IP	Contrato (1)	Jornada (2)	Fecha alta	Fecha baja	Causa Baja (3)	Identificador cuenta cotización
7	0001563	421	0	01/02/1999	31/01/2002	0	0111017022180
8	0001563	501	663	01/02/2003	01/06/2003	0	0111017022180
9	0001563	100	0	02/06/2003	31/12/2008	0	0111017022180

Fuente. Elaboración propia.

El inconveniente de esta operacionalización radica en que desconocemos qué ha pasado entre la segunda modificación del contrato y el valor final que refleja la MCVL. Por lo tanto, puede que la duración de la última relación contractual (computada dentro de la imagen 4 en el registro 9) no se ajuste a la realidad. No ocurre así, en las relaciones laborales que solo sufren una modificación.

Dificultad 4: El cálculo de las duraciones de los periodos de desempleo.

Al igual que sucede con los periodos de ocupación, también es posible calcular los periodos de desempleo³⁰ a partir de la información de la MCVL. En este caso su duración equivaldría al tiempo transcurrido entre la fecha de baja de una relación laboral y la primera fecha de alta de la siguiente. Pero, también aquí, hay que tener en cuenta varias consideraciones que, aunque han sido mencionadas en puntos anteriores, en este caso tienen implicaciones específicas.

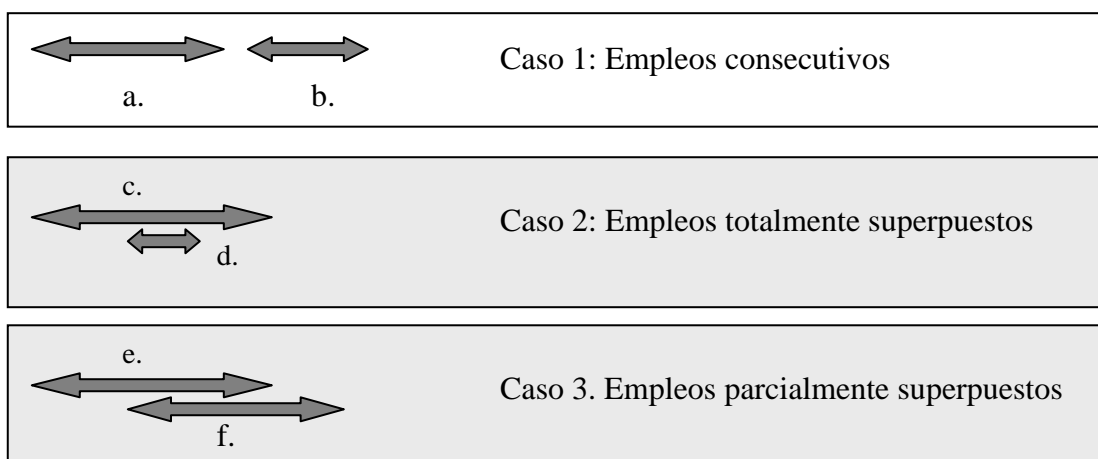
La primera de ellas está relacionada con lo que hemos comentado en la dificultad anterior: la MCVL registra los periodos de cobro de las ayudas por desempleo³¹. Por lo tanto, a la hora de computar la duración total del episodio de desempleo habrá que sumar el tiempo de percepción de la prestación y el tiempo transcurrido hasta la próxima fecha real de alta. La segunda está relacionada con lo que apuntábamos al describir la dificultad 2: las situaciones de pluriempleo³². La imagen 5 refleja en los rectángulos sombreados de color gris las dos casuísticas que, en este sentido, podemos encontrar en la MCVL.

³⁰ Cuando utilizamos este término también nos referimos a los periodos de inactividad, ya que en la MCVL no es posible distinguir ambas situaciones.

³¹ Las personas amparadas por un Convenio Especial se caracterizan por no tener un empleo, pero no tienen la consideración de personas desempleadas.

³² Para ver una explicación exhaustiva de las implicaciones de esta problemática en el cálculo de las duraciones de desempleo véase el artículo de García Pérez (2008). Lo que aquí se presenta es un resumen del mismo.

Imagen 5: Tipología de las situaciones a considerar en el tratamiento de la MCVL.



Fuente: García-Pérez, J.I. (2007: 26).

Por un lado, tenemos empleos o periodos de cotización de corta duración que están totalmente insertados dentro de otro. Es decir, su fecha de alta es posterior a la de un empleo de duración más larga y su fecha de baja anterior a la finalización del mismo. Esta situación se corresponde en la imagen con el caso 2, etiquetado como “empleos totalmente superpuestos”. En la MCVL del 2007 el 9,9% de los registros en el caso de los hombres y el 7,4% en el de las mujeres se corresponden con esta pauta (García Pérez, 2008: 25). Por el otro lado, también podemos observar empleos o periodos de cotización parcialmente superpuestos. Éstos comparten la característica de que un empleo empieza antes de que termine otro. El tercer caso de la imagen 5 refleja dicha situación, bajo la que pueden englobarse el 3% del total de los registros de los hombres y el 2,7% de las mujeres en la MCVL 2007 (García Pérez, 2008: 26).

El problema que plantean ambos casos es común; al calcular el periodo de desempleo intermedio obtendremos valores negativos. En consecuencia, esas duraciones sesgarán nuestros resultados, al no corresponderse con la realidad. Su solución, en cambio, es distinta. Así, en situaciones como las del caso 3, en donde la superposición es parcial, lo lógico es “considerar los dos empleos por separado como dos experiencias distintas y, no considerar el periodo de desempleo intermedio al ser inexistente” (García Pérez, 2008: 26).

El caso 2, en donde la superposición es total, resulta algo más complejo de resolver. Algunos autores apuntan la posibilidad de eliminar empleos como el etiquetado bajo la letra d., siempre y cuando se correspondan con periodos de cotización muy cortos (García Pérez, 2008: 25). Esto permite tener computada correctamente la historia laboral del individuo, evitando las distorsiones de empleos con escasa relevancia debido a su exigua duración. Sin embargo, en la MCVL podemos encontrarnos con situaciones de pluriempleo o pluriactividad que se mantienen durante periodos muy prolongados de tiempo. En ese caso es conveniente seguir las recomendaciones expuestas en la dificultad 2 para determinar cuál es la situación principal del individuo y controlar el análisis por el hecho de que ese individuo es pluriempleado.

Dificultad 5: La identificación de la estructura familiar a partir de la información del Padrón Municipal y de la Agencia Tributaria.

Lamentablemente, la MCVL no incluye datos sobre la situación familiar procedentes de los registros administrativos de la Seguridad Social debido a su escasa actualización. En cambio, las otras dos fuentes sí ofrecen información en este sentido, aunque en ningún caso podemos reconstruir la composición del hogar de forma exacta a partir de ellas. De hecho, la información extraída del Padrón Municipal se refiere a las personas inscritas en la misma hoja padronal. De este modo, las variables incluidas son el sexo y la fecha de nacimiento de hasta un máximo de 9 personas, además del individuo seleccionado en la MCVL. No existe, por lo tanto, información sobre la relación de parentesco de estas personas con el individuo de referencia.

Por su parte, la información procedente del modelo 190 de la Agencia Tributaria versa sobre los familiares que tienen una relación de dependencia económica con la persona de referencia, en la medida en que dicha dependencia tenga algún efecto sobre las retenciones que son aplicables al trabajador y siempre y cuando éste haya notificado previamente su situación familiar al empleador. Así, podemos saber si el individuo está ‘casado’ o ‘soltero, viudo o divorciado con hijos a cargo’³³. Igualmente, en algunos casos se conoce el número de descendientes menores de 18 años o de ascendientes que están a cargo del perceptor y si éstos son minusválidos o no. Sin embargo, cabe recordarse nuevamente que esta información no es similar a la que el declarante facilita al realizar el IRPF ya que está sujeta a las características y limitaciones que impone el modelo 190 explicadas en el segundo apartado. Las fichas metodológicas correspondientes al fichero de la Agencia Tributaria tampoco contienen información sobre la validez y fiabilidad de estas variables.

A partir de la información que nos facilitan ambas fuentes solo podemos obtener algunos indicadores que, en cualquier caso, quedan lejos del concepto de estructura familiar. Por lo tanto, se desaconseja la utilización de esta información para la elaboración de indicadores de carácter general. Sin embargo, cuando se afrontan estudios concretos sí es posible inferir relaciones de parentesco a través del cálculo de las edades de los individuos empadronados en la misma hoja. No obstante, el riesgo de error es elevado y el ejercicio resulta imposible en aquellos casos en los que se detectan dos o más adultos del mismo sexo y tramo de edad conviviendo en el mismo hogar. En consecuencia, en los análisis que utilizan la información padronal de la MCVL siempre quedan excluidos los hogares plurinucleares, que tienen una incidencia especial entre determinados colectivos, como por ejemplo, el de los inmigrantes. La información que ofrece la Agencia Tributaria cubre -aunque solo parcialmente- estas carencias, al permitir introducir en el análisis indicadores más aproximados sobre la composición familiar. En este último caso, no debemos perder de vista las limitaciones que impone el modelo 190 del que se extrae la información.

7. Conclusiones:

La MCVL es la primera fuente de información en España que aprovecha, con fines investigadores, la información de la Seguridad Social, el Padrón Continuo Municipal y

³³ La variable ‘situación familiar’ que ofrece la información de la Agencia Tributaria no distingue cada uno de los últimos estados civiles mencionados.

la Agencia Tributaria. Sus microdatos permiten la reconstrucción de las trayectorias laborales de los individuos, así como su historial de percepción de prestaciones de carácter contributivo. La MCVL es, de este modo, una fuente de información de referencia para el estudio de las dinámicas del mercado de trabajo o la evolución de nuestro sistema de previsión social.

Entre sus ventajas destaca, en primer lugar, el modo en que se ha definido la población de referencia, que permite una mejor representatividad de personas con trayectorias discontinuas -mayoritariamente mujeres y jóvenes- y que no son captados tan fidedignamente por otras fuentes de información (García Pérez, 2008). En segundo lugar, su gran tamaño muestral posibilita el estudio de colectivos relativamente pequeños e incluso incorporar en el análisis la dimensión territorial, manteniendo simultáneamente altos niveles de significatividad estadística. En tercer lugar, su estructura tipo panel permite el desarrollo de estudios longitudinales y la aplicación de técnicas de análisis de historia de acontecimientos basadas en el concepto clave de 'curso de vida' (Mortimer y Shanahan, 2003; Bernardi, 2006).

Sin embargo, la depuración y el manejo de sus datos plantean numerosas dificultades, especialmente en comparación con otras fuentes de información primaria. En este trabajo se han desgranado algunas de ellas, aportando claves fundamentales para su tratamiento y resolución. Con este objetivo como trasfondo, los dos primeros apartados describen las características fundamentales de la MCVL, al tiempo que detallan la información que en ella podemos encontrar. La comprensión del ámbito poblacional al que hace referencia la MCVL, así como los colectivos que quedan excluidos de la misma es una tarea ineludible que el investigador deberá acometer como paso preliminar al montaje de la base de datos. El cuarto apartado explica cómo abordar este proceso, profundizando en las peculiaridades de la estructura panel de la MCVL. Así pues, para acometer con garantías de éxito esta labor resulta fundamental comprender, por un lado, los flujos de entrada y de salida de la muestra y, por el otro, el ámbito temporal al que hace referencia la información de cada uno de estos ficheros. La descripción realizada en el segundo apartado también pretende facilitar esta tarea.

Dificultades añadidas que el investigador deberá tener en cuenta son; la duplicación de los identificadores de los individuos seleccionados como consecuencia del procedimiento utilizado para la extracción de la información de los registros; la detección del empleo o actividad principal en las situaciones de pluriempleo; o los problemas en la identificación de la estructura familiar. Mención específica requiere la explotación de una de las informaciones más interesantes de la MCVL: las duraciones de los episodios de empleo y desempleo. Su cálculo resulta complicado por la evolución de las distintas formas en que la MCVL registra las relaciones laborales. El planteamiento de estas casuísticas y de sus posibles soluciones es una de las contribuciones fundamentales de este trabajo. Su vocación es, en cualquier caso, aproximar a los investigadores a las ventajas y limitaciones de una fuente de información con enormes potencialidades.

8. Referencias:

Argimón, M.I. y González, C.I. (2006) 'La Muestra Continua de Vidas Laborales de la Seguridad Social', Boletín Económico, Madrid: Banco de España.

Bernardi, F. (2006) *Análisis de la Historia de Acontecimientos*, Madrid: Cuadernos Metodológicos, CIS.

Durán, A. (2007) 'La Muestra Continua de Vidas Laborales de la Seguridad Social', *Revista del Ministerio de Trabajo y Asuntos Sociales*, 67(E): 231-240. [Disponible en: <http://www.mtas.es/es/publica/revista/numeros/ExtraSS07/default.htm>. Consultado el 22 de septiembre de 2009].

Cebrián, I. y Toharia, I. (2007) 'La mejora del sistema de información estadística de los registros de la Seguridad Social. Informa final'.

García Pérez, I. (2008) 'La Muestra Continua de Vidas Laborales: una guía de uso para el análisis de transiciones', *Revista de Economía Aplicada*, 1(E): 5-28. [Disponible en: <http://www.revecap.com/revista/numeros/e1/pdf/garcia.pdf> . Consultado el 18 de noviembre de 2009].

García-Segovia, F. y Durán, A. (2008) 'Nuevos Avances en la información laboral: la Muestra Continua de Vidas Laborales', *Economistas*, 116(E): 228-231.

Marcos, C. y Durán, A. (2005) 'Proyecto de microdatos de vidas laborales', *Revista de Estadística y Sociedad*, 11: 18-19, Madrid: Instituto Nacional de Estadística [Disponible en: <http://www.revistaindice.com/numero11/p18.pdf> Consultado: 16 de noviembre de 2009].

Marcos, C. (2006) 'Ventajas e inconvenientes, oportunidades y algunas propuestas para el desarrollo de la estadística de base administrativa', Ponencia presentada en el 'Seminario Internacional de Estadística: El papel de los registros administrativos en relación con el análisis social y económico y el desarrollo del sistema estadístico nacional', 30-31 de enero de 2006, Madrid: Instituto de Estudios Fiscales [Disponible en: <http://www.ief.es/investigacion/recursos/Seminarios/Estadistica/RegistrosAdministrativos.htm> Consultado: 15 de diciembre de 2009].

Mortimer, J., y Shanahan, M. (ed.) (2003): *Handbook of the Life Course*, Kluwer Academic/ Plenum Publishers.

MTAS (2006a) *La Muestra Continua de Vidas Laborales*, Colección de Informes y estudios de la Seguridad Social, nº24, Madrid: Ministerio de Trabajo y Asuntos Sociales.

MTAS (2006b) *La Muestra Continua de Vidas laborales. Descripción general*. Madrid: Dirección General de la Seguridad Social. [Disponible en [http://www.seg-social.es/Internet_1/Estadistica/Est/Muestra Continua de Vidas Laborales/CONTUTIL2009/index.htm](http://www.seg-social.es/Internet_1/Estadistica/Est/Muestra%20Continua%20de%20Vidas%20Laborales/CONTUTIL2009/index.htm) Consultado: 12/7/2010].

MTAS (2006c) *La Muestra Continua de Vidas laborales. Cómo se organiza la información*. Madrid: Dirección General de la Seguridad Social. [Disponible en [http://www.seg-social.es/Internet_1/Estadistica/Est/Muestra Continua de Vidas Laborales/CONTUTIL2009/index.htm](http://www.seg-social.es/Internet_1/Estadistica/Est/Muestra%20Continua%20de%20Vidas%20Laborales/CONTUTIL2009/index.htm) Consultado: 12/7/2010].

MTAS (2006d) *La Muestra Continua de Vidas laborales. Módulo fiscal*. Madrid: Dirección General de la Seguridad Social. [Disponible en <http://www.seg-social.es/stpri00/groups/public/documents/binario/120874.pdf> Consultado: 12/7/2010].

Ramos, B. (2007) ‘Comparación de las magnitudes estadísticas del empleo según la Encuesta de Población Activa y la Muestra Continua de Vidas Laborales’, documento de trabajo presentado en las ‘*Jornadas de Usuarios de la MCVL*’, 4-5 de octubre de 2007, Madrid: Fedea.

Tonder, J.K. (2008) ‘The Register-based Statistical System: Preconditions and Processes’, Documento presentado en la ‘International Association for Official Statistics Conference’, Shanghai. 14-18 de Octubre del 2009. [Disponible en: http://www.stats.gov.cn/english/specialtopics/iaos/Papers/CS_1_2_Tonder.doc Consultado: 30 de junio de 2010].

Wallgren, A. y Wallgren, B (2007) *Register-based Statistics - Administrative Data for Statistical Purposes*, Chichester: John Wiley & Sons.